MATERIALS AND METHODS FOR ATOMISTIC CHARACTERIZATION OF EMERGENT NANOPOROUS ADSORBENTS

A Dissertation
Presented to
The Academic Faculty

by

Jeffrey S. Camp

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
School of Chemical and Biomolecular Engineering

Georgia Institute of Technology
August 2016

MATERIALS AND METHODS FOR ATOMISTIC CHARACTERIZATION OF EMERGENT NANOPOROUS ADSORBENTS

Approved by:

Dr. David Sholl, Advisor
School of Chemical and Biomolecular Engineering
Georgia Institute of Technology

Dr. Sankar Nair
School of Chemical and Biomolecular Engineering
Georgia Institute of Technology

Dr. Rigoberto Hernandez
School of Chemistry and Biochemistry
Georgia Institute of Technology

Dr. Krista Walton
School of Chemical and Biomolecular Engineering
Georgia Institute of Technology

Dr. Ryan Lively
School of Chemical and Biomolecular Engineering
Georgia Institute of Technology

Date Approved: June 21st, 2016

To Christine

ACKNOWLEDGEMENTS

I wish to express gratitude to all of my teachers throughout my education. Many thanks to the faculty and administration of the Charter School of Wilmington for building an outstanding school from limited resources. CSW is a model for what secondary education needs to be in our country. Thanks to the Honors Seminar professors at the University of Maryland, especially Imad Ahmad, John Franklin, Howard Smead, and Shannen Hill. It was a great privilege to take excellent liberal arts courses while pursuing an engineering degree. Above all, thanks to my mentor, teacher, and advisor Dr. David Sholl for his guidance and support throughout my PhD work. David is an inspiration both as a scientist and as a man of the highest degree of character.

Thanks to the Department of Energy and everyone at the Nanoporous Materials Genome Center for funding my graduate studies. In particular, thanks to Professor Gregory Chung for his friendly and effective collaboration on our MOF database.

Thanks to all of my friends in the Sholl Group and throughout Georgia Tech who have made graduate school an enjoyable experience both inside and outside the office. In particular, thanks to Salah Boulfelfel for teaching me the ropes as a new graduate student. I have enjoyed our conversations about everything from simulation methods to politics. Thank you, Dalar Nazarian, for being a great friend and research collaborator. Dalar is co-author on a large portion of this work. Also, thanks to Grace Chen, Jason Gee, Ross Verploegh, Emmanuel Haldoupis, Jack Evans, and the whole PACE staff for many productive technical discussions.

Finally, I would like to thank my family for all of their love and encouragement. My dad never suggested that I follow in his footsteps as a chemical engineer, but inspired me to do so through his enthusiasm for alternative energy and problem solving. Thanks to my mom – her cheerfulness, wisdom, and strength through adversity is always admirable. Most importantly, I would like to express my love and appreciation for my wife Christine He. She has been my better half, best friend, and study buddy since we were both nineteen years old. This thesis is dedicated to Christine.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

SUMMARY

Metal organic frameworks (MOFs) are an emerging class of nanoporous materials that have shown promise in applications including gas storage, separations, and catalysis. The complexity and diversity of MOF chemical space frustrates experimental efforts to examine even a representative subset of the thousands of known MOFs. High-throughput computational screening can guide experimental efforts by identifying top candidate structures for applications of interest. These screening efforts require a database of crystallographic structural information that has been prepared for molecular simulations by removal of solvent molecules, partially occupied atoms, and disordered atoms. In this thesis, we describe algorithms to automatically prepare MOF structures for molecular simulations. The outcome of this work was a publicly available database of over 5,000 computation-ready MOF structures. As an example of using our database, we perform simulations of CH_4 adsorption in each material and identify key thermodynamic parameters influencing adsorbed natural gas storage. In additional, we assign framework charges to nearly 3,000 structures using periodic density functional theory calculations. These DFT derived point charges were used to identify materials potentially useful in the adsorptive removal of a corrosive sulfurous odorant (*tert*-butyl mercaptan) from methane.

Detailed atomistic simulation can be used to understand the properties of the most promising materials identified by computational screening. Of particular interest are adsorbate diffusivities, which are an important predictor of material performance in both equilibrium and kinetic separation applications. We describe methods to measure adsorbate diffusion in flexible nanoporous materials at timescales inaccessible to conventional molecular dynamics. These methods are applied to a novel class of porous molecular cage compounds that crystallize in the solid state without intermolecular covalent or coordination bonds. Our results show that cage crystal 3 has promise in the diffusive separation of rare gases and aromatics.

1

INTRODUCTION

1.1 Emergent Nanoporous Adsorbents

Metal–organic frameworks (MOFs) are a class of nanoporous crystalline polymers synthesized by bonding ionic metal nodes to organic bridging ligands. Thousands of unique MOFs structures have been synthesized in the past two decades owing to the combinatorial possibilities afforded by selection of different building blocks, synthesis conditions, and activation procedures.[1] For many reasons, MOFs can be fascinating from a pure science perspective. For example, MOFs have been synthesized that exhibit properties including ferroelectricity[2], ferromagnetism[3], proton conduction[4], ion conduction[5], and luminesce[6].[7] Study of the complexities of MOF framework geometries has led to advances in the mathematical classification of periodic reticular ("net like") structures.[8] Several hundred different net topologies have been observed among experimentally synthesized MOFs, some of which are unknown in any other type of crystal.[9] MOFs often undergo phase transitions between crystalline phases upon initial desolvation or guest adsorption into activated structures.[10] This can lead to interesting gas adsorption behavior including isotherms with substantial hysteresis.[11]

MOFs have been investigated for a wide range of potential applications including catalysis[12], molecular sensing[13], drug delivery[14], and optoelectronics[7]. Much applied academic and commercial research into uses of MOFs has focused on the possibility of using MOFs as nanoporous adsorbents in gas storage[15] and separations[16] applications. Among MOFs are example of materials with the highest surface areas (7140 m^2/g) and lowest densities (0.126 g/cm^3) among all known adsorbents.[17] MOFs have a wide range of internal pore and aperture dimensions, which can lead to high diffusion selectivities in membrane separations. Chemical properties of some MOFs are also uniquely suited to sorption applications. Many MOFs have coordinatively unsaturated "open" metal nodes that interact favorably with adsorbate multipoles, increasing heats of adsorption.[18] MOFs have the potential to be modified after synthesis, which can increase adsorbate saturation capacities and equilibrium adsorption selectivities.[19]

MOFs are candidate adsorbent materials for the storage of natural gas and hydrogen fuel. As nanoporous adsorbents in vehicle fuel tanks, MOFs have the potential to dramatically lower pressures needed to achieve energy densities comparable to compressed natural gas (CNG) at 250 bar in tanks without adsorbent fillers. This could increase safety, reduce fuel compression costs, and allow for conformable tank geometries in place of current bulky spherical or cylindrical CNG tanks.[20] Hydrogen gas is an attractive energy carrier because of its high gravimetric energy density, but suffers from low volumetric density and standard temperature and pressure. Physisorption of hydrogen gas into MOFs has been extensively investigated as an alternative to chemical storage in metal hydrides, cryogenic compression, or ultra-high pressure storage for increasing the volumetric density of H_2.[21] Early studies of H_2 storage in MOF-5 showed considerable improvement (66 g/L) over the volumetric density of pure hydrogen (31 g/L) at 100 bar and 77 K.[22] MOFs also have some promise for use in the storage of acetylene[23] and oxygen gases[24].

Adsorption-based separation processes such as simulated moving bed chromatography are used extensively used to purify chemical mixtures. Examples include the UOP Parex® process for separation of xylene isomers and the Molex® process for separation of *n*-paraffins from branched and cyclic hydrocarbons.[25] These technologies use nanoporous aluminosilicate zeolites as adsorbents.[26] MOF adsorbent materials with greater equilibrium selectivities, better adsorption capacities, or more favorable mass transfer properties than currently employed zeolites could save huge amounts of energy in these processes. More importantly, MOFs may play a role in replacing thermally driven separations such as distillation of alkenes from alkanes and ethanol from water with much more energy efficient membrane separation processes. Experimental and theoretical reports have confirmed the molecular sieving potential of MOFs such as ZIF-8 for short-chain hydrocarbons[27] and $Zn_2(BDC)_2(TED)$ for alcohol in water[28]. MOF based membrane processes could be revolutionary, as distillation processes account for nearly half of the total energy spent on all chemical separations, which together comprise 10-15% of global energy consumption.[29]

Unfortunately, MOFs have a number of drawbacks that have hampered their widespread industrial application in the 20+ years since they were first synthesized. Foremost, MOFs are currently extremely expensive relative to commodity nanoporous

sorbents such as activated carbons and zeolites as vehicles for raw surface area. Although progress has been made in scaling up MOF synthesis, the organic ligands present in MOFs are inherently costly.[30] The economic viability of MOF depends on finding applications that exploit chemical and physical properties not found in other types of nanoporous materials. Furthermore, some MOF structures are structurally unstable when exposed to water vapor[31] or used in repeated cycles of adsorption and desorption, which precludes many practical applications. Finally, the sheer number of extant MOF structures is a significant obstacle to identifying the best material for a given application of interest.[32]

Atomistic simulation methods may assist in current efforts to commercialize MOFs. First, detailed simulations of individual materials contribute to fundamental understanding of the chemistry that governs structural stability in MOFs and emergent material properties. Second, simulations of adsorbate-framework interactions within MOFs can reveal details responsible for materials performance that are challenging or impossible to capture with experimental techniques such as x-ray diffraction. Third, high-throughput computational screening of the thousands of synthesized MOF structures can identify materials promising for a given application of interest prior to time consuming and expensive experimental testing. Finally, analysis of simulations of large numbers of MOF structures can reveal structure-property relationships common to high performing materials.

1.2 Atomistic Simulation Methods

Prior to any atomistic simulation, the atomic positions of experimentally synthesized structures must be crystallographically resolved. Most nanoporous crystals are resolved through either single crystal x-ray diffraction or powder x-ray diffraction. The Cambridge Structural Database (CSD)[33] hosts the majority of organometallic crystal structures, while zeolites are primarily stored in the International Zeolite Association (IZA) database[34]. MOF structures reported in the CSD often include solvent molecules and partially occupied or disordered atoms. These features are crystallographically meaningful but must be removed prior to computer simulations of fully activated, solvent-free structures. This is of little consequence for simulations of individual materials, but is a significant impediment to applying high-throughput computations to thousands of materials. In Chapter 2, I describe algorithms to automatically prepare MOF structures for

molecular simulations. We applied these algorithms to over 5,000 nanoporous MOF structures from the CSD, and then made these computation-ready structures publicly available.[35]

Using only the experimentally resolved atomic coordinates of a MOF structure, a number of important textural properties can be mathematically determined that have a direct analog with an experimental measurement.[36] The experimentally determined BET surface area is usually in reasonably good agreement with the accessible surface area created by rolling a spherical probe nitrogen molecule of each atom of the computer model structure.[37] Significant disagreement between experimentally derived BET or Langmuir surface areas and the geometric value may indicate MOF framework collapse upon desolvation or an incompletely activated experimental sample. Similarly, the pore volume of an *in-silico* (computer model) MOF structure can be determined with trial insertions of a helium atom. This directly corresponds to experimental helium porosimetry.[36] The *in-silico* model also yields the pore limiting diameter (PLD), which is the size of the largest spherical probe accessible to continuous nanopores within the material. The PLD is a reasonable upper limit on the size of adsorbate molecules that can be accommodated by a MOF.[32] In Chapter 2, I discuss calculation of these textural properties for each MOF in our Computation-Ready Experimental MOF (CoRE MOF) database.

Characterization of electrical and magnetic properties, as well as rigorous description of interactions between adsorbates and framework atoms, requires consideration of the distribution of electrons within MOFs. All computational electronic structure methods applicable to MOFs use the Born-Oppenheimer approximation, which decouples total electronic energy from the nuclear repulsion in approximate solutions to the Schrodinger equation.[38] These methods are broadly divided into two different approaches for considering electron correlation. Density functional theory (DFT) approximates electron correlation in terms of the three dimensional electron density.[39] DFT is routinely applied to periodic MOF structures with unit cells consisting of hundreds of atoms. Wave function theory (WFT) methods, such as Møller-Plesset perturbation theory and coupled cluster theory, are usually more accurate but too computationally expensive to use on fully periodic models of MOF structures.[38] WFT methods have been

used independently[40] and in concert with DFT[41] to parameterize first-principles derived classical force fields for describing interactions between adsorbates and MOF frameworks.

Density functional theory can be used directly to investigate adsorbate-framework interaction energies in MOFs. For example, the energetics of water[41] and CO_2 adsorption[42] in the well-known MOF copper benzene-1,3,5-tricarboxylate (CuBTC) have been calculated using periodic DFT. It is computationally infeasible, however, to use electronic structure methods to simulate phenomena such as adsorbate diffusion occurring over time scales of nanoseconds or longer. Similarly, grand canonical Monte Carlo (GCMC) simulations of adsorption within nanopores require thousands or millions of computational iterations to converge and thus cannot be directly simulated with electronic structure methods. For these reasons, simulations of MOFs often rely on classical force fields parameterized using DFT. Examples include DFT-derived force fields for modeling hydrocarbon adsorption in MIL-47(V)[43] and CO_2 adsorption in MOF-74 variants[44].

Classical force-fields for modeling adsorbate framework interactions typically decompose electronic interactions into a Lennard-Jones potential to model dispersion and a pairwise Coulombic term to model long range electrostatic forces.[45] The pairwise Coulombic energy is almost always calculated by assigning point charges to the nuclei of each framework atom. These point charges are most accurately assigned by partitioning the electron density from a fully periodic DFT calculation. This represents a significant computational investment for large numbers of materials. In Chapter 2, I describe our works towards creating a nearly comprehensive, publicly available database of atomic point charges for a large number of experimental MOF structures.[46]

In high-throughput screening studies, a generic force field such as the Universal Force Field[47] or DREIDING[48] is typically used to model Lennard-Jones dispersion. A recent study by McDaniel et al. compared these generic force fields to more sophisticated polarizable models and found that while generic force fields are often quantitatively inaccurate in regimes of low adsorbate loading, they were typically able to reproduce correct rankings of MOFs in terms of adsorbate saturation capacities.[49] In Chapter 3, I describe using generic force fields to predict the deliverable capacity of methane in each CoRE MOF structure for adsorbed natural gas applications using GCMC simulations.[35] We also used generic force fields in concert with our DFT derived atomic point charges to

identify materials suitable for the adsorptive removal of a potentially corrosive sulfurous odorant from natural gas.[46]

Almost all GCMC simulations of adsorption treat the nanoporous framework as rigid. However, in simulations of adsorbate diffusion, experiments and classical molecular dynamics (MD) simulations have demonstrated that local framework flexibility associated with rotation of organic ligands, oscillation of coordination bonds, and variation in pore aperture dimensions is quite significant.[50] This is particularly true when adsorbate dimensions are comparable to the size of pore apertures. Members of the Sholl group have found that considering framework flexibility in MOFs[51] and zeolites[52] can result in simulated adsorbate diffusivities orders of magnitude higher than in rigid experimentally derived crystal structures. Despite this, studies of adsorbate diffusion in large numbers of MOF materials typically still use the rigid crystal structures because of the paucity of classical force fields that model MOF framework flexibility.[53]

For some highly studied MOFs such as CuBTC[54], MOF-5[55], and ZIF-8[56], force fields have been parameterized using DFT calculations to describe internal flexibility with classical bond, angle, and torsional parameters. However, even with the benefit of a custom flexible force field, classical MD simulations of slow adsorbate diffusion can be computationally prohibitive. In Chapter 4, I demonstrate classical transition state theory methods to resolve this "MD timescale problem" that complicates measuring slow adsorbate diffusion in MOFs and other nanoporous materials.[57] These methods were applied to light gas diffusion in porous organic cage crystal 3 (CC3)[58], a highly flexible molecular crystal that has shown promise in gas separation applications.[59] In chapter 5, these methods are applied to measure the diffusivities of aromatics in additional diamondoid topology porous organic cage compounds.

In Chapter 6, I conclude this thesis by speculating on potential future directions in simulations of nanoporous materials41. In particular, I discuss how the methods I have used for high-throughput screening of MOFs can be used in conjunction with strategies to incorporate MOF framework flexibility into simulations of adsorption.

1.3 References

[1] H. Furukawa, K. E. Cordova, M. O'Keeffe, O. M. Yaghi, *Science* **2013**, *341*, 1230444.

[2] D. Di Sante, A. Stroppa, P. Jain, S. Picozzi, *Journal of the American Chemical Society* **2013**, *135*, 18126-18130.

[3] P. Canepa, Y. J. Chabal, T. Thonhauser, *Physical Review B* **2013**, *87*, 094407.

[4] J. M. Taylor, K. W. Dawson, G. K. Shimizu, *Journal of the American Chemical Society* **2013**, *135*, 1193-1196.

[5] S. Horike, D. Umeyama, S. Kitagawa, *Accounts of Chemical Research* **2013**, *46*, 2376-2384.

[6] M. Allendorf, C. Bauer, R. Bhakta, R. Houk, *Chemical Society Reviews* **2009**, *38*, 1330-1352.

[7] M. D. Allendorf, M. E. Foster, F. Léonard, V. Stavila, P. L. Feng, F. P. Doty, K. Leong, E. Y. Ma, S. R. Johnston, A. A. Talin, *The Journal of Physical Chemistry Letters* **2015**, *6*, 1182-1195.

[8] N. W. Ockwig, O. Delgado-Friedrichs, M. O'Keeffe, O. M. Yaghi, *Accounts of Chemical Research* **2005**, *38*, 176-182.

[9] M. O'Keeffe, M. A. Peskov, S. J. Ramsden, O. M. Yaghi, *Accounts of Chemical Research* **2008**, *41*, 1782-1789.

[10] A. J. Fletcher, K. M. Thomas, M. J. Rosseinsky, *Journal of Solid State Chemistry* **2005**, *178*, 2491-2510.

[11] J. A. Mason, J. Oktawiec, M. K. Taylor, M. R. Hudson, J. Rodriguez, J. E. Bachman, M. I. Gonzalez, A. Cervellino, A. Guagliardi, C. M. Brown, *Nature* **2015**, *527*, 357-361.

[12] J. Liu, L. Chen, H. Cui, J. Zhang, L. Zhang, C.-Y. Su, *Chemical Society Reviews* **2014**, *43*, 6011-6061.

[13] B. Chen, Y. Yang, F. Zapata, G. Lin, G. Qian, E. B. Lobkovsky, *Advanced Materials* **2007**, *19*, 1693-1696.

[14] P. Horcajada, T. Chalati, C. Serre, B. Gillet, C. Sebrie, T. Baati, J. F. Eubank, D. Heurtaux, P. Clayette, C. Kreuz, *Nature Materials* **2010**, *9*, 172-178.

[15] S. Ma, H.-C. Zhou, *Chemical Communications* **2010**, *46*, 44-53.

[16] J.-R. Li, R. J. Kuppler, H.-C. Zhou, *Chemical Society Reviews* **2009**, *38*, 1477-1504.

[17] O. K. Farha, I. Eryazici, N. C. Jeong, B. G. Hauser, C. E. Wilmer, A. A. Sarjeant, R. Q. Snurr, S. T. Nguyen, A. O. Yazaydin, J. T. Hupp, *Journal of the American Chemical Society* **2012**, *134*, 15016-15021.

[18] B. Chen, N. W. Ockwig, A. R. Millward, D. S. Contreras, O. M. Yaghi, *Angewandte Chemie* **2005**, *117*, 4823-4827.

[19] Z. Wang, S. M. Cohen, *Chemical Society Reviews* **2009**, *38*, 1315-1329.

[20] C. M. Simon, J. Kim, D. A. Gomez-Gualdron, J. S. Camp, Y. G. Chung, R. L. Martin, R. Mercado, M. W. Deem, D. Gunter, M. Haranczyk, D. S. Sholl, R. Q. Snurr, B. Smit, *Energy & Environmental Science* **2015**, *8*, 1190-1199.

[21] M. P. Suh, H. J. Park, T. K. Prasad, D.-W. Lim, *Chemical Reviews* **2011**, *112*, 782-835.

[22] N. L. Rosi, J. Eckert, M. Eddaoudi, D. T. Vodak, J. Kim, M. O'Keeffe, O. M. Yaghi, *Science* **2003**, *300*, 1127-1129.

[23] J.-P. Zhang, X.-M. Chen, *Journal of the American Chemical Society* **2009**, *131*, 5516-5521.

[24] J. B. DeCoste, M. H. Weston, P. E. Fuller, T. M. Tovar, G. W. Peterson, M. D. LeVan, O. K. Farha, *Angewandte Chemie International Edition* **2014**, *53*, 14092-14095.

[25] P. S. Gomes, M. Minceva, A. E. Rodrigues, *Adsorption* **2006**, *12*, 375-392.

[26] S. Kulprathipanja, *Zeolites in Industrial Separation and Catalysis*, Wiley, **2010**.

[27] U. Böhme, B. Barth, C. Paula, A. Kuhnt, W. Schwieger, A. Mundstock, J. r. Caro, M. Hartmann, *Langmuir* **2013**, *29*, 8592-8600.

[28] G. F. de Lima, A. Mavrandonakis, H. A. de Abreu, H. l. A. Duarte, T. Heine, *The Journal of Physical Chemistry C* **2013**, *117*, 4124-4130.

[29] D. S. Sholl, R. P. Lively, *Nature* **2016**, *532*, 435-437.

[30] M. Rubio-Martinez, M. P. Batten, A. Polyzos, K.-C. Carey, J. I. Mardel, K.-S. Lim, M. R. Hill, *Scientific Reports* **2014**, *4*.

[31] N. C. Burtch, H. Jasuja, K. S. Walton, *Chemical Reviews* **2014**, *114*, 10575-10612.

[32] E. Haldoupis, S. Nair, D. S. Sholl, *Journal of the American Chemical Society* **2010**, *132*, 7528-7539.

[33] F. H. Allen, *Acta Crystallographica Section B: Structural Science* **2002**, *58*, 380-388.

[34] C. Baerlocher, L. B. McCusker, D. Olson, W. M. Meier, I. Z. A. S. Commission, *Atlas of Zeolite Framework Types*, Elsevier, **2007**.

[35] Y. G. Chung, J. Camp, M. Haranczyk, B. J. Sikora, W. Bury, V. Krungleviciute, T. Yildirim, O. K. Farha, D. S. Sholl, R. Q. Snurr, *Chemistry of Materials* **2014**, *26*, 6185-6192.

[36] L. Sarkisov, J. Kim, *Chemical Engineering Science* **2015**, *121*, 322-330.

[37] K. S. Walton, R. Q. Snurr, *Journal of the American Chemical Society* **2007**, *129*, 8552-8556.

[38] S. O. Odoh, C. J. Cramer, D. G. Truhlar, L. Gagliardi, *Chemical Reviews* **2015**, *115*, 6051-6111.

[39] D. Sholl, J. A. Steckel, *Density Functional Theory: A Practical Introduction*, Wiley, **2011**.

[40] A. L. Dzubak, L.-C. Lin, J. Kim, J. A. Swisher, R. Poloni, S. N. Maximoff, B. Smit, L. Gagliardi, *Nature Chemistry* **2012**, *4*, 810-816.

[41] L. Grajciar, O. Bludsky, P. Nachtigall, *Journal of Physical Chemistry Letters* **2010**, *1*, 3354-3359.

[42] L. s. Grajciar, A. D. Wiersum, P. L. Llewellyn, J.-S. Chang, P. Nachtigall, *Journal of Physical Chemistry C* **2011**, *115*, 17925-17933.

[43] A. R. Kulkarni, D. S. Sholl, *Langmuir* **2015**, *31*, 8453-8468.

[44] E. Haldoupis, J. Borycz, H. Shi, K. D. Vogiatzis, P. Bai, W. L. Queen, L. Gagliardi, J. I. Siepmann, *Journal of Physical Chemistry C* **2015**, *119*, 16058-16071.

[45] D. Frenkel, B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, Elsevier Science, **2001**.

[46] D. Nazarian, J. S. Camp, D. S. Sholl, *Chemistry of Materials* **2016**, *28*, 785-793.

[47] A. K. Rappé, C. J. Casewit, K. Colwell, W. Goddard Iii, W. Skiff, *Journal of the American Chemical Society* **1992**, *114*, 10024-10035.

[48] S. L. Mayo, B. D. Olafson, W. A. Goddard, *Journal of Physical Chemistry* **1990**, *94*, 8897-8909.

[49] J. G. McDaniel, S. Li, E. Tylianakis, R. Q. Snurr, J. R. Schmidt, *Journal of Physical Chemistry C* **2015**, *119*, 3143-3152.

[50] R. J. Verploegh, S. Nair, D. S. Sholl, *Journal of the American Chemical Society* **2015**, *137*, 15760-15771.

[51] E. Haldoupis, T. Watanabe, S. Nair, D. S. Sholl, *ChemPhysChem* **2012**, *13*, 3449-3452.

[52] R. V. Awati, P. I. Ravikovitch, D. S. Sholl, *The Journal of Physical Chemistry C* **2013**, *117*, 13462-13473.

[53] B. Borah, H. Zhang, R. Q. Snurr, *Chemical Engineering Science* **2015**, *124*, 135-143.

[54] L. Zhao, Q. Yang, Q. Ma, C. Zhong, J. Mi, D. Liu, *Journal of Molecular Modeling* **2011**, *17*, 227-234.

[55] M. Tafipolsky, S. Amirjalayer, R. Schmid, *Journal of Computational Chemistry* **2007**, *28*, 1169-1176.

[56] B. Zheng, M. Sant, P. Demontis, G. B. Suffritti, *The Journal of Physical Chemistry C* **2012**, *116*, 933-938.

[57] J. Camp, D. S. Sholl, *Journal of Physical Chemistry C* **2016**, *120*, 1110-1120.

[58] T. Tozawa, J. T. A. Jones, S. I. Swamy, S. Jiang, D. J. Adams, S. Shakespeare, R. Clowes, D. Bradshaw, T. Hasell, S. Y. Chong, C. Tang, S. Thompson, J. Parker, A. Trewin, J. Bacsa, A. M. Z. Slawin, A. Steiner, A. I. Cooper, *Nature Materials* **2009**, *8*, 973-978.

[59] L. Chen, P. S. Reiss, S. Y. Chong, D. Holden, K. E. Jelfs, T. Hasell, M. A. Little, A. Kewley, M. E. Briggs, A. Stephenson, K. M. Thomas, J. A. Armstrong, J. Bell, J. Busto, R. Noel, J. Liu, D. M. Strachan, P. K. Thallapally, A. I. Cooper, *Nature Materials* **2014**, *13*, 954-960.

2

DEVELOPMENT OF THE COMPUTATION-READY EXPERIMENTAL METAL-ORGANIC FRAMEWORK DATABASE*

2.1 Introduction

In principle, the metallic nodes and organic linkers that comprise MOFs can be rationally selected to tailor structures for given applications. In practice, it is difficult to predict the complex relationship among the building blocks, the resulting framework structures, and the emergent physical properties prior to synthesis.[1] Before attempting de novo design of a new structure for a particular application, it is wise to consider whether any existing MOFs might be suitable.[2] However, synthesis, characterization, and experimental testing of thousands of MOFs to find the best material for a given application would be prohibitively time-consuming and expensive.

High-throughput computational screening can guide experimental efforts by identifying top candidate structures for applications of interest.[3] For example, experimental MOF crystal structures from the Cambridge Structural Database (CSD)[4] have been computationally screened to identify top-performing MOFs for applications in light gas[5], noble gas[6], and CO_2/N_2[7] separations. These large-scale computational screening efforts require a computation-ready database with the crystallographic information for each structure. However, experimentally refined crystal structures reported in the CSD often include solvent molecules and partially occupied or disordered atoms. Such features are crystallographically meaningful but must be removed prior to computer simulations of fully activated, solvent-free structures. Prior to this work, the lack of a publicly available database of computation-ready crystal structures was a major impediment to applying high-throughput computations to MOFs, because each research

*Portions of this chapter have been published previously:

1. Yongchul G. Chung, Jeffrey S. Camp, et al. Computation-ready, experimental metal-organic frameworks: a tool to enable high-throughput screening of nanoporous crystals. *Chemistry of Materials* **2014**, *26*, 6185-6192.
2. Dalar Nazarian‡, Jeffrey S. Camp‡, and David S. Sholl. A comprehensive set of high-quality point charges for simulations of metal–organic frameworks. *Chemistry of Materials* **2016**, *28*, 785-793.
‡These authors contributed equally to this work.

group interested in such an approach had to first construct an appropriate database of materials.

Prior to our work detailed here, Goldsmith et al.[8] developed an in-house database of computation-ready MOF structures derived from the CSD. In their database, solvents were removed, mislabeled structures were identified, and some notable MOF structures (e.g., PCN-610 and NU-100) were manually repaired. These structures were used to predict the theoretical upper bound for hydrogen storage in MOFs. In their database, structures with interpenetrated frameworks and charge-balancing ions were not included. Furthermore, neither the software they used to prepare the MOF structures for simulations nor their final database was made available to the public.

The central result of this work is the availability of a nearly comprehensive set of porous MOF structures that are derived directly from experimental data but are immediately suitable for molecular simulations or visualization.[9] Our computation-ready, experimental (CoRE) MOFs have several characteristics that make them broadly useful. First, each structure in the database was desolvated by the efficient graph-labeling algorithm described by Goldsmith et al.[8] Additional solvent molecules were removed that were bound to unsaturated metal atoms. Second, charge-balancing ions were retained where necessary, so each CoRE MOF structure is charge neutral overall. Third, interpenetrated structures were also retained on the basis of a graph-labeling algorithm. Finally, important geometric properties including the helium void fraction, pore-limiting diameter, and largest cavity diameter were reported alongside each computation-ready structure.

Not all MOFs that have been reported experimentally are included in the CoRE MOF database. In some instances, incomplete data or a high degree of disorder makes it challenging to generate a plausible computation-ready crystal structure in an automated procedure. Despite not including materials of this kind, the first version of the CoRE MOF database includes over 5,000 porous structures with a pore-limiting diameter of at least 2.4 Å. A subset of these structures (2,932 total) were published with accompanying atomic point charges derived from periodic DFT calculations.[10] These point charges are highly useful for modeling Coulombic interactions between MOF framework atoms and adsorbate molecules – necessary, for instance, in any classical simulation of the adsorption or

diffusion of polar or quadrupolar adsorbates within MOFs. The CoRE MOF structures, including the subset with DFT derived atomic point charges, are available at http://dx.doi.org/10.11578/1118280.

2.2 CoRE MOF Database Construction

2.2.1 Collection of 3-D MOF Structures

The procedure for generating the CoRE MOF database is shown schematically in Figure 2.1. First, potential MOF structures were collected from the CSD, version 5.35, which includes submissions through February 2014.[4] The CCDC Conquest program was used to search for structures with more than one bond between metals and the elements O, N, B, P, S, and C. Additionally, we required the structures to form any kind of bond from these six elements to C, N, P, or S atoms. This search yielded over 60,000 candidate MOF structures.

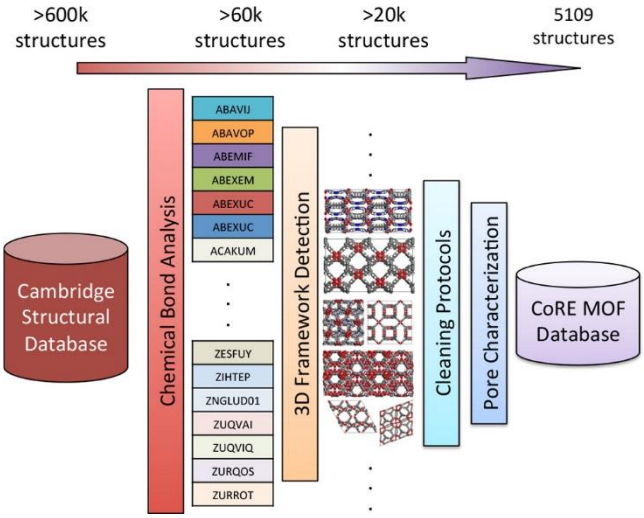

Figure 2.1 Schematic illustration of the CoRE MOF database construction. Chemical bond analysis was performed using the CCDC Conquest program, 3D framework detection and pore characterization were performed using the Zeo++ Open Source Software, and cleaning protocols were implemented in Python using the ASE and SciPy libraries. All structures in the CoRE MOF database have pore-limiting diameters (PLDs) >2.4 Å.

This set of structures includes 1-D coordination polymers and 2-D hydrogen-bonded "planar MOFs" in addition to 3-D MOF structures. We identified the 3-D MOF

structures in this set by applying an algorithm previously used to determine the dimensionality of void channels.[11] Here, bonded components in a molecular graph of the structure are analyzed to determine the dimensionality of the basis vectors that describe the connection between the bonded component and its images in neighboring simulation cells. A bonded component of a molecular graph refers to a connected set of atoms based on an internal bond criterion from the Zeo++ software.[11] Two atoms are considered bonded if the distance between them is less than the sum of their covalent radii plus a skin distance. This procedure yielded over 20,000 3-D MOF structures that were further considered for inclusion in the CoRE MOF database.

2.2.2 Structure Preparation

An automated text editor was then used to remove all atoms marked by special characters ("*" and "?") in the crystallographic information file for each structure. The atoms marked with "*" are symmetry-related copies of atoms already present in the structure. The atoms marked with "?" are atoms with partial occupancy. In most cases, removing one copy of these partially occupied atoms leaves a single representation of chemical moieties such as aromatic rings. Note that this procedure introduces a degree of order which may not exist in the experimental structure. Each structure was then converted to its primitive unit cell, and the symmetry was set to P1 using a Perl script that interfaces with the Materials Studio software.

2.2.3 Categorizing Structures

The chemical formulas associated with each structure were searched for "+" and "-" symbols to find MOF frameworks that have associated charge-balancing ions. About half of these charged structures were flagged in the CSD as "disordered" and were discarded after visual inspection showed that most of these structures contain major disorder in the framework atoms or lack resolved ionic coordinates.

Some of the MOF frameworks without associated charge-balancing ions that are flagged as disordered only contain disorder in the solvent molecules and were retained in the database. These structures were identified by searching the CSD entries for phrases such as "a N, N-dimethylformamide solvent molecule is disordered". Each of these

structures was visually checked for disorder in the framework atoms, and manual editing was done where appropriate by referring to the literature. For example, the disordered benzene rings in PCN-68 (CSD: HABRAF)[12] were manually corrected. The structures without associated charge-balancing ionic species were passed directly to the solvent removal step.

2.2.4 Retention of Charge-Balancing Ions

Many MOF structures with associated charge-balancing ions also contain undesirable neutral solvent molecules. To discriminate between ionic species and neutral solvent molecules, the elemental compositions of the bonded components in a molecular graph of each structure were compared to the chemical formulas reported by the CSD. The bonded components are the independent "molecules" within each structure; these include the MOF framework, the ionic species, and any neutral solvent molecules. First, the Atomic Simulation Environment[13] NeighborList module was used to construct the periodic adjacency matrix for each structure. Two atoms are considered bonded if the distance between them is less than the sum of their CSD covalent radii[14] plus a skin distance of 0.3 Å. The skin distance is chosen to be slightly smaller than the CSD definition (0.4 Å), so that the terminal atom connected to the metal atom does not form another bond with other nearby atoms. The adjacency matrix was then passed to the SciPy connected components module to identify the bonded components in each structure. The bonded components with elemental compositions matching the composition of the ions reported by the CSD were exempted from deletion in the solvent removal step.

2.2.5 Solvent Removal

In the solvent removal step, all bonded components in the molecular graph of each structure other than the MOF framework and charge-balancing ions were removed (Figure 2.2). The MOF framework was defined as the highest molecular weight bonded component of the graph. Interpenetrated MOF frameworks were retained by identifying the number of atoms, N, in the largest bonded component in the structure and retaining all additional components having at least 0.5 N atoms.

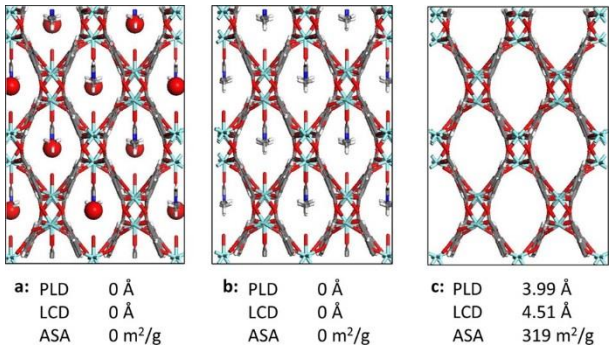

Figure 2.2 Illustration of solvent removal from a candidate MOF structure (CSD REFCODE NADZEZ): (a) original structure from CSD. Atoms shown with the CPK model are free solvents; (b) structure with free solvent removed; (c) structure with both free and bound solvent removed. The pore-limiting diameter (PLD), largest cavity diameter (LCD), and gravimetric accessible surface area (ASA) are shown for each structure.

The bonded component corresponding to the MOF framework often includes undesirable solvent molecules bound to unsaturated metal centers. To remove these coordinated solvent molecules, we performed a trial "cut" on all bonds between metal centers and oxygen atoms. If the number of bonded clusters detected by the connected component algorithm remained constant, the bond was restored. If the number of bonded components increased, the entire new component was considered a solvent molecule and removed. For example, the dimethylformamide solvent molecules in the MOF JUC-64 (CSD: OFODET)[15] that are bonded to manganese metal centers were automatically removed in this step. An exception was built into the algorithm to retain hydroxyl groups bonded to metal centers. In some cases, these solvent molecules are necessary to stabilize the MOF frameworks. Our desolvation procedure simply removes these molecules without considering whether the structural integrity of the framework would be retained upon removal of all solvent molecules.

2.2.6 Manual Structure Editing

Structures flagged as disordered were retained if the CSD comment field indicated that only the uncharged solvent atoms have disordered coordinates or if the comment field did not explicitly state that the structure is disordered. Each of these structures was visually examined to confirm that the framework atoms are not disordered. This process identified

106 MOF structures with minor framework disorder that were fixed manually by referring to the original literature for the MOF crystal structures. Missing hydrogen atoms were added automatically to an additional 63 structures using Materials Studio (Accelrys, San Diego, CA). Following manual editing, a geometric optimization was run using the Materials Studio Forcite module. The positions of each atom were allowed to relax to a tolerance of 0.002 kcal/mol, while the simulation cell was fixed at the experimental lattice constants. Universal force field parameters[16] were assigned to each atom in the simulation cell to model all bonded and nonbonded interactions.

2.2.7 Geometric Characterization

Each CoRE MOF structure was characterized with Zeo++[11] using the high-accuracy setting[17], which uses Voronoi decomposition to identify probe-accessible regions of void space and calculate the accessible surface area, accessible volume, largest cavity diameter (LCD), and pore-limiting diameter (PLD).[5] All Zeo++ calculations used a probe of radius 1.86 Å (corresponding to N_2)[18] and covalent radii from the Cambridge Crystallographic Data Centre for all framework atoms. The reported accessible surface areas only include pore regions accessible through windows large enough to admit N_2. Helium void fractions were calculated via random Widom particle insertions with a Lennard-Jones helium atom at 298 K.

2.3 Atomic Point Charge Calculation

2.3.1 Background

Electronic structure calculations such as density functional theory (DFT) and other quantum chemistry methods have been shown to reliably match experimental measurements of properties of MOFs such as adsorbate interaction energies.[19] It is computationally infeasible, however, to use electronic structure methods to simulate phenomena such as adsorbate diffusion occurring over time scales of nanoseconds or longer. Similarly, grand canonical Monte Carlo (GCMC) simulations of adsorption within nanopores require thousands or millions of computational iterations to converge and thus cannot be directly simulated with electronic structure methods. For these reasons,

simulations of MOFs often rely on classical force fields, especially in efforts to examine large numbers of materials.[6, 20-23]

In essentially all force field calculations with MOFs, Coulombic interactions between atoms in a MOF and also between those atoms and adsorbate molecules are modeled by assigning point charges to each atom of the framework. Because there is no unique solution to the task of assigning point charges to represent the full three-dimensional distribution of charge in a material[24], multiple methods have been explored for assigning charges in MOFs. Semiempirical methods such as charge equilibration have been used because they can be applied without performing an electronic structure calculation.[25-26] When possible, it is preferable to use atomic charges derived from the electron density calculated from an electronic structure calculation for either discrete clusters cleaved from MOF structures or fully periodic representations of MOF crystals.[27] Methods for assigning charges based on partitioning the electron density of MOF clusters include ChelpG[28] and more recent charge model techniques[29]. Cluster techniques have been used to screen small numbers (~20 MOFs) of experimentally synthesized MOFs for CO_2 storage.[30-31] Fully periodic methods for partitioning the electron density such as DDEC[24], fitting the local electrostatic field around atoms such as REPEAT[32], or period population analysis such as CM5[29] avoids the problem of ambiguous bond termination inherent to cluster-based methods.[27, 33] Both the DDEC and REPEAT methods were designed in part to accurately reproduce the electrostatic potential energy surface for locations outside the van der Waals radius of atoms in the material, a property that is desirable in modeling adsorption in MOFs. Other methods that have been widely used to assign point charges to periodic materials such as Bader charges do not have this property.[24, 34] Unlike the Bader method, DDEC incorporates spherical averaging and uses reference ion densities to enhance the transferability and chemical meaning of the charges.

2.3.2 Computational details

A single self-consistent ionic step was attempted in the VASP 5.3.5 plane-wave DFT package for each CoRE MOF structure without charge compensating ionic species to generate the electron and spin density distributions used as inputs for point charge assignment.[35] Nazarian et al. have previously shown that there is a negligible difference

in the DDEC-derived atomic point charges from electronic densities generated with the PBE, PW91, M06L, or vdw-DF2 functionals for a diverse test set of MOF structures.[36] This is consistent with earlier results of Manz and Sholl for a broad range of materials.[24] The PBE functional was used throughout this work to minimize computational expense. Calculations on the same test set of MOFs mentioned above also indicated negligible differences in point charges between calculations using the experimental structures reported in the CoRE MOF database and structures that were fully geometrically optimized with DFT.[36] As a result, geometric relaxation of the MOFs was not employed for any of the calculations described below. For most structures, the Brillouin zone was sampled with a 1000 points per atom density Monkhorst–Pack grid. For ~200 structures, calculations with a Gamma grid were necessary for proper convergence. Spin polarization was included for all calculations, adopting ferromagnetic high-spin states for magnetic elements.[37-38] While there are many structures in the database that exhibit antiferromagnetic spin ordering, it is difficult to efficiently identify these structures a priori. We have found that for Cu-BTC, which includes a copper dimer with a ground state antiferromagnetic spin state[39-40], a ferromagnetic calculation results in a Cu charge that is different by <0.02 electron from that of the antiferromagnetic structure. This observation suggests that using ferromagnetic states is sufficient for assigning point charges.

The electron density was successfully computed for ~75% of the structures in the starting data set. Of the calculations that did not converge in VASP, approximately half exceeded the maximal virtual memory imposed by our computing resources, and most of these calculations were for the largest CoRE MOF structures, with primitive cells of several hundred atoms or more. The remaining calculations failed because of other VASP errors, including issues with k-point grid density requirements and unresolved segmentation faults.

The converged electron densities from VASP were used as inputs to the January 2014 version of the Density Derived Electrostatic and Chemical (DDEC) charge assignment code distributed by Manz et al.[24, 34, 41] Atomic point charges were successfully calculated for 2,932 structures. All charges below are reported in units of electron charge. A small number of these structures (14 MOFs, including nine with silver atoms) were assigned unphysical negative charges to cationic metal centers. These MOFs were found

to be missing bound solvent atoms in the proximity of metal centers that were removed in the construction of the CoRE MOF database. In these cases, restoring the bound solvent molecules to their crystallographically refined positions produces realistic positive charges for the cationic metal atoms. To remain consistent with the structures in the CoRE MOF database, we excluded these 14 structures from further analysis.

2.4 Results and Discussion

2.4.1 Geometric and Physical Characteristics of CoRE MOFs

The CoRE MOF database contains 5,109 3-D MOF structures with pore-limiting diameters greater than 2.4 Å, which corresponds to approximately the diameter of a hydrogen molecule. We note that some MOF structures are duplicated in our database because multiple representations of the same MOF are reported in the CSD. Notably, there are at least 13 entries for IRMOF-1 and 50 entries for HKUST-1 in the CSD. The authors of new MOF structures often report more than one crystal structure at different activation conditions and different temperatures or with different guest molecules inside. Additional crystal structures may also be reported by investigators interested in different applications of the same MOF. The lattice constants of these "duplicate" representations often vary, and this may have an effect on simulated adsorption properties. Since the different structures may be of interest to different researchers, we have not removed duplicates from the CoRE MOF database.

The number of structures in the CoRE MOF database does not represent the number of unique porous MOFs synthesized so far because there are duplicate MOF structures in our database and highly disordered structures are not included. Nevertheless, the CoRE MOF database captures a great deal of the chemical and structural diversity in experimentally synthesized MOFs. To assess the structural diversity of the structures, we determined the underlying topology of more than 2,000 CoRE MOF structures with the TOPOS program (Appendix A.1)[42]. Over 350 unique topologies were observed in this subset of the database. Among the structures to which we assigned nets, the most common are pcu (16%), dia (12%), ths (3%), sql (3%), rtl (3%), srs (3%), and bcu (3%). This distribution of underlying nets qualitatively agrees with the topological analysis performed

by Proserpio and co-workers, who found that the most common topologies in 4,709 noninterpenetrating MOFs are pcu (9%) and dia (6%).[43] In contrast, only six topological nets are represented in a recent database of hypothetical MOFs[44], where over 90% of structures have the 6-coordinated pcu topology.[45]

We determined the 3-D space group of each structure using the pymatgen library's Symmetry Finder module after the structure went through the cleaning procedure.[46] The CoRE MOF database contains structures with 190 of the 230 possible 3-D space groups. This represents a remarkable degree of structural diversity. The most common space groups are P21/c (17%), C2/c (14%), P1 (12%), R3 (3%), and Fm3m (3%). We also find that 297 out of 5109 structures have space groups that are different from what was reported in the CSD following our cleaning procedure. We have also tabulated the metals associated with each structure in the CoRE MOF database. The database includes over 50 types of metal clusters, including lanthanides. The most common metal species in the database are Zn and Cu, which is not surprising given the widespread use of Zn_4O and copper and zinc paddlewheel metal centers for the synthesis of metal–organic frameworks.

Figure 2.3 shows the distribution of calculated volumetric and gravimetric accessible surface areas in the CoRE MOF database (a, b) and in the database of hypothetical MOFs from Wilmer et al.[44] (c, d). The plots show that the structures from the CoRE MOF database have a flatter distribution for the volumetric surface area but a more peaked distribution for the gravimetric surface area. Synthesis of large surface area MOFs has been an active area of research for the past decade[47-50], and the skewed distribution of gravimetric surface areas in the CoRE MOF database might be attributed to the lower stability of MOFs with large surface areas and the difficulty in synthesizing such structures.

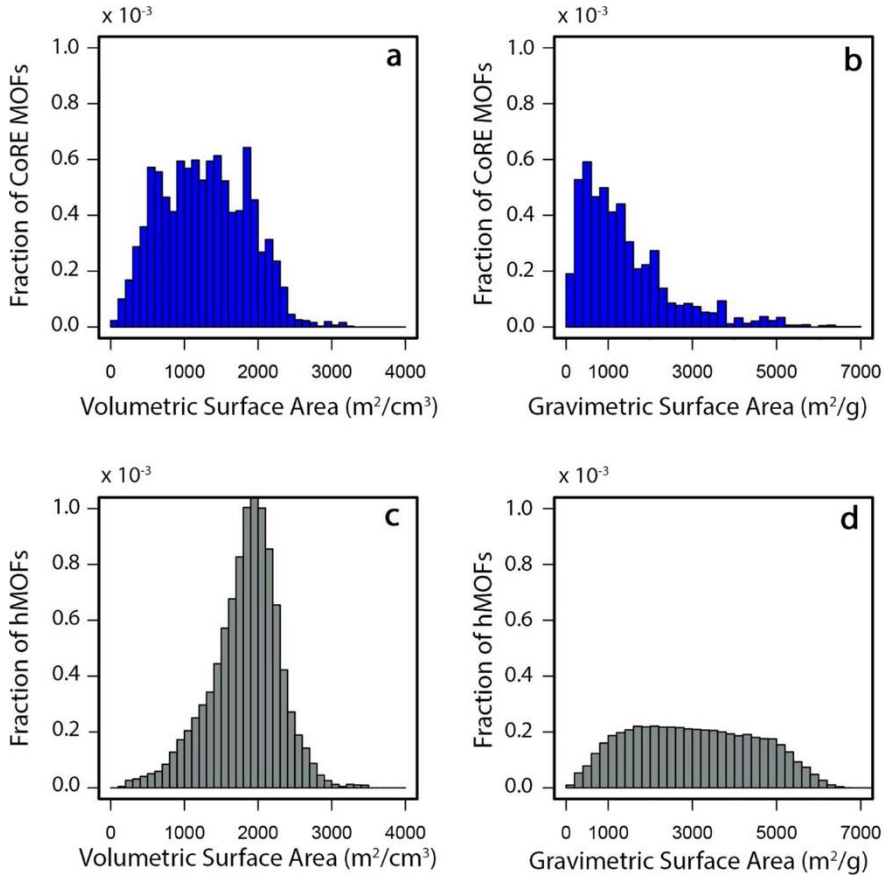

Figure 2.3 Probability distribution of calculated volumetric and gravimetric accessible surface areas: (a) volumetric accessible surface area from the CoRE MOF database; (b) gravimetric accessible surface area from the CoRE MOF database; (c) volumetric accessible surface area from the hypothetical MOF database of Wilmer et al.; (d) gravimetric accessible surface area from the hypothetical MOF database of Wilmer et al. All properties were calculated using Zeo++ with a probe radius of 1.86 Å (corresponding to N_2).

2.4.2 Comparison of DFT/DDEC to Charge Equilibration

Electronic structure methods for calculating atomic point charges require significant computational investment. Generating the electron and spin density distribution for a single MOF using DFT typically requires between tens and hundreds of hours of CPU time. Semiempirical methods such as extended charge equilibration (EQeq)[26] and periodic charge equilibration (PQeq)[51] that use tabulated elemental properties such as electronegativity and ionic affinity to estimate point charges are orders of magnitude more computationally efficient. This enables point charges to be assigned to very large numbers

of structures quickly. For example, the EQeq method has been used to assign point charges to a database of more than 137,000 hypothetical MOFs.[52] To evaluate the accuracy of these methods, we compared the EQeq charges to DDEC charges for each MOF in our data set. EQeq charges were computed with the stand-alone code distributed by NuMat Technologies using default metallic oxidation states. This stand-alone code was found to produce charges consistent with the EQeq equilibration method implemented in RASPA 1.0.[53]

Figure 2.4 shows the comparison between DDEC and EQeq for each of the more than 10000 distinct metal atoms in our data set. A small number of neodymium and uranium metals with unrealistically large (> +4) EQeq charges are excluded from this plot.

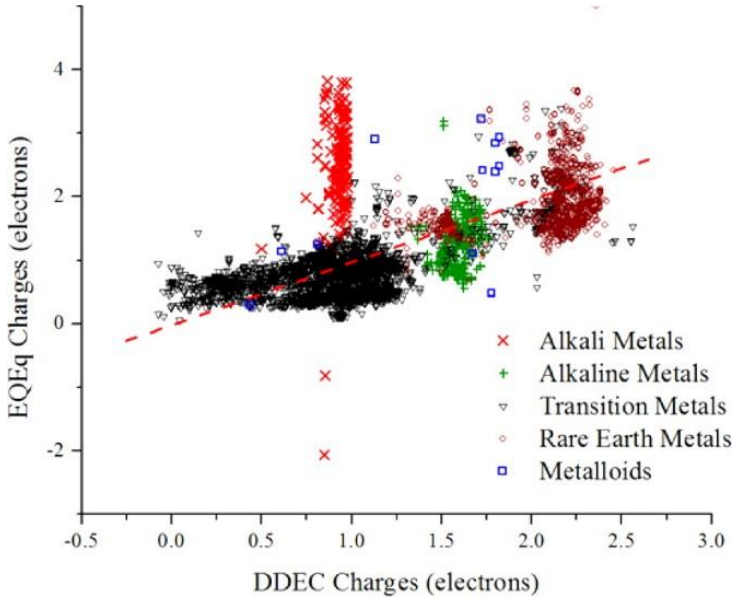

Figure 2.4 Charges from EQeq charge equilibration compared to DDEC-derived charges for more than 10000 distinct metal atoms in MOFs. Rare earth metals include lanthanides and actinides.

Figure 2 shows that EQeq predicts charges for metals higher than those predicted by DDEC on average. This observation has previously been made by Haldoupis et al.[25] EQeq predicts highly unrealistic charges for many structures containing alkali metals. While DDEC charges for alkali metals cluster around the +1 oxidation state expected from their position on the periodic table, EQeq predicts a range of alkali charges from −2 to +4 electrons.

2.5 Conclusions

We have constructed a database of over 5,000 computation-ready porous MOF structures that were derived directly from experimental crystal data. Efficient algorithms were used to retain charge-balancing ionic species and remove solvent molecules bound to unsaturated metal centers. Physical and chemical properties of each structure including surface areas and pore characteristics are tabulated in the online supplemental information. In addition, we have produced a set of high-quality point charges for nearly 3,000 of these experimentally synthesized MOF structures using plane-wave DFT calculations and the DDEC charge partitioning method. By using a periodic representation of each MOF, these charges avoid problem of ambiguous bond termination inherent to cluster-based methods. Because the DDEC method was designed in part to accurately reproduce the electrostatic potential energy surface for locations outside the van der Waals radius of atoms in the material, our data set of charges is well suited for modeling adsorption in MOFs. These data are publically available online (http://dx.doi.org/10.11578/1118280).

2.6 References

[1] M. O'Keeffe, M. Eddaoudi, H. Li, T. Reineke, O. M. Yaghi, *Journal of Solid State Chemistry* **2000**, *152*, 3-20.

[2] H. Furukawa, K. E. Cordova, M. O'Keeffe, O. M. Yaghi, *Science* **2013**, *341*, 1230444.

[3] Y. J. Colon, R. Q. Snurr, *Chemical Society Reviews* **2014**, *43*, 5735-5749.

[4] F. H. Allen, *Acta Crystallographica Section B: Structural Science* **2002**, *58*, 380-388.

[5] E. Haldoupis, S. Nair, D. S. Sholl, *Journal of the American Chemical Society* **2010**, *132*, 7528-7539.

[6] T. Van Heest, S. L. Teich-McGoldrick, J. A. Greathouse, M. D. Allendorf, D. S. Sholl, *Journal of Physical Chemistry C* **2012**, *116*, 13183-13195.

[7] T. Watanabe, D. S. Sholl, *Langmuir* **2012**, *28*, 14114-14128.

[8] J. Goldsmith, A. G. Wong-Foy, M. J. Cafarella, D. J. Siegel, *Chemistry of Materials* **2013**, *25*, 3373-3382.

[9] Y. G. Chung, J. Camp, M. Haranczyk, B. J. Sikora, W. Bury, V. Krungleviciute, T. Yildirim, O. K. Farha, D. S. Sholl, R. Q. Snurr, *Chemistry of Materials* **2014**, *26*, 6185-6192.

[10] D. Nazarian, J. S. Camp, D. S. Sholl, *Chemistry of Materials* **2016**, *28*, 785-793.

[11] T. F. Willems, C. H. Rycroft, M. Kazi, J. C. Meza, M. Haranczyk, *Microporous and Mesoporous Materials* **2012**, *149*, 134-141.

[12] D. Yuan, D. Zhao, D. Sun, H. C. Zhou, *Angewandte Chemie International Edition* **2010**, *49*, 5357-5361.

[13] S. R. Bahn, K. W. Jacobsen, *Computing in Science & Engineering* **2002**, *4*, 56-66.

[14] B. Cordero, V. Gomez, A. E. Platero-Prats, M. Reves, J. Echeverria, E. Cremades, F. Barragan, S. Alvarez, *Dalton Transactions* **2008**, 2832-2838.

[15] M. Xue, G. Zhu, Y. Li, X. Zhao, Z. Jin, E. Kang, S. Qiu, *Crystal Growth & Design* **2008**, *8*, 2478-2483.

[16] A. K. Rappé, C. J. Casewit, K. Colwell, W. Goddard Iii, W. Skiff, *Journal of the American Chemical Society* **1992**, *114*, 10024-10035.

[17] M. Pinheiro, R. L. Martin, C. H. Rycroft, M. Haranczyk, *CrystEngComm* **2013**, *15*, 7531-7538.

[18] Y.-S. Bae, A. Ö. Yazaydın, R. Q. Snurr, *Langmuir* **2010**, *26*, 5475-5483.

[19] L. s. Grajciar, A. D. Wiersum, P. L. Llewellyn, J.-S. Chang, P. Nachtigall, *Journal of Physical Chemistry C* **2011**, *115*, 17925-17933.

[20] R. Krishna, J. M. van Baten, *Physical Chemistry Chemical Physics* **2011**, *13*, 10593-10616.

[21] L.-C. Lin, A. H. Berger, R. L. Martin, J. Kim, J. A. Swisher, K. Jariwala, C. H. Rycroft, A. S. Bhown, M. W. Deem, M. Haranczyk, *Nature Materials* **2012**, *11*, 633-641.

[22] A. O. z. r. Yazaydın, R. Q. Snurr, T.-H. Park, K. Koh, J. Liu, M. D. LeVan, A. I. Benin, P. Jakubczak, M. Lanuza, D. B. Galloway, *Journal of the American Chemical Society* **2009**, *131*, 18198-18199.

[23] J. G. McDaniel, S. Li, E. Tylianakis, R. Q. Snurr, J. R. Schmidt, *Journal of Physical Chemistry C* **2015**, *119*, 3143-3152.

[24] T. A. Manz, D. S. Sholl, *Journal of Chemical Theory and Computation* **2010**, *6*, 2455-2468.

[25] E. Haldoupis, S. Nair, D. S. Sholl, *Journal of the American Chemical Society* **2012**, *134*, 4313-4323.

[26] C. E. Wilmer, K. C. Kim, R. Q. Snurr, *Journal of Physical Chemistry Letters* **2012**, *3*, 2506-2511.

[27] S. Hamad, S. R. Balestra, R. Bueno-Perez, S. Calero, A. R. Ruiz-Salvador, *Journal of Solid State Chemistry* **2015**, *223*, 144-151.

[28] C. M. Breneman, K. B. Wiberg, *Journal of Computational Chemistry* **1990**, *11*, 361-373.

[29] A. V. Marenich, S. V. Jerome, C. J. Cramer, D. G. Truhlar, *Journal of Chemical Theory and Computation* **2012**, *8*, 527-541.

[30] R. Babarao, J. Jiang, *Langmuir* **2008**, *24*, 6270-6278.

[31] C. Zheng, D. Liu, Q. Yang, C. Zhong, J. Mi, *Industrial & Engineering Chemistry Research* **2009**, *48*, 10479-10484.

[32] C. Campana, B. Mussard, T. K. Woo, *Journal of Chemical Theory and Computation* **2009**, *5*, 2866-2878.

[33] S. Keskin, J. Liu, R. B. Rankin, J. K. Johnson, D. S. Sholl, *Industrial & Engineering Chemistry Research* **2008**, *48*, 2355-2371.

[34] T. A. Manz, D. S. Sholl, *Journal of Chemical Theory and Computation* **2012**, *8*, 2844-2867.

[35] G. Kresse, J. Furthmüller, *Physical Review B: Condensed Matter* **1996**, *54*, 11169.

[36] D. Nazarian, P. Ganesh, D. S. Sholl, *Journal of Materials Chemistry A* **2015**, *3*, 22432-22440.

[37] A. Jain, G. Hautier, C. J. Moore, S. P. Ong, C. C. Fischer, T. Mueller, K. A. Persson, G. Ceder, *Computational Materials Science* **2011**, *50*, 2295-2310.

[38] Y. Cai, A. R. Kulkarni, Y. G. Huang, D. S. Sholl, K. S. Walton, *Crystal Growth and Design* **2014**, *14*, 6122-6128.

[39] T. Watanabe, S. Keskin, S. Nair, D. S. Sholl, *Physical Chemistry Chemical Physics* **2009**, *11*, 11389-11394.

[40] L. Grajciar, O. Bludsky, P. Nachtigall, *Journal of Physical Chemistry Letters* **2010**, *1*, 3354-3359.

[41] T. Watanabe, T. A. Manz, D. S. Sholl, *The Journal of Physical Chemistry C* **2011**, *115*, 4824-4836.

[42] V. Blatov, *Structural Chemistry* **2012**, *23*, 955-963.

[43] E. V. Alexandrov, V. A. Blatov, A. V. Kochetkov, D. M. Proserpio, *CrystEngComm* **2011**, *13*, 3947-3958.

[44] C. E. Wilmer, M. Leaf, C. Y. Lee, O. K. Farha, B. G. Hauser, J. T. Hupp, R. Q. Snurr, *Nature Chemistry* **2012**, *4*, 83-89.

[45] B. J. Sikora, R. Winnegar, D. M. Proserpio, R. Q. Snurr, *Microporous and Mesoporous Materials* **2014**, *186*, 207-213.

[46] S. P. Ong, W. D. Richards, A. Jain, G. Hautier, M. Kocher, S. Cholia, D. Gunter, V. L. Chevrier, K. A. Persson, G. Ceder, *Computational Materials Science* **2013**, *68*, 314-319.

[47] K. Koh, A. G. Wong-Foy, A. J. Matzger, *Journal of the American Chemical Society* **2010**, *132*, 15005-15010.

[48] H. Furukawa, N. Ko, Y. B. Go, N. Aratani, S. B. Choi, E. Choi, A. O. Yazaydin, R. Q. Snurr, M. O'Keeffe, J. Kim, O. M. Yaghi, *Science* **2010**, *329*, 424-428.

[49] O. K. Farha, A. Ö. Yazaydın, I. Eryazici, C. D. Malliakas, B. G. Hauser, M. G. Kanatzidis, S. T. Nguyen, R. Q. Snurr, J. T. Hupp, *Nature Chemistry* **2010**, *2*, 944-948.

[50] O. K. Farha, I. Eryazici, N. C. Jeong, B. G. Hauser, C. E. Wilmer, A. A. Sarjeant, R. Q. Snurr, S. T. Nguyen, A. O. Yazaydin, J. T. Hupp, *Journal of the American Chemical Society* **2012**, *134*, 15016-15021.

[51] S. Ramachandran, T. Lenz, W. Skiff, A. Rappe, *Journal of Physical Chemistry* **1996**, *100*, 5898-5907.

[52] C. E. Wilmer, O. K. Farha, Y.-S. Bae, J. T. Hupp, R. Q. Snurr, *Energy & Environmental Science* **2012**, *5*, 9849-9856.

[53] D. Dubbeldam, S. Calero, D. E. Ellis, R. Q. Snurr, *Molecular Simulation* **2015**, *42*, 81-101.

3

HIGH-THROUGHPUT SCREENING OF MOFS: NATURAL GAS STORAGE AND PURIFICATION*

3.1 Introduction

Natural gas (primarily CH_4) is both more abundant worldwide and cleaner burning than gasoline derived from crude oil.[1] This makes natural gas an attractive potential fuel for passenger vehicles. Compressed natural gas (CNG) is already widely used to power buses for public transit. However, conventional CNG requires compression of natural gas to pressure of up to 250 bar to achieve required energy densities. This necessitates the use of heavy, thick walled tanks of a cylindrical or spherical geometry.[2] MOFs adsorbents have the potential to dramatically lower pressures needed to achieve energy densities comparable to conventional CNG and allow for conformable tank geometries.

As an illustration of the utility of the CoRE MOF database, we performed grand canonical Monte Carlo (GCMC) simulations of methane adsorption in each structure to determine the storage capacity at 65 bar and the deliverable capacity from 65 to 5.8 bar. These pressures correspond to the DOE ARPA-E targets for methane storage in adsorbed natural gas fuel tanks within passenger vehicles.[2] Similar GCMC simulations were previously conducted on a database of over 137,000 hypothetical MOFs (hMOFs) to find structural properties that govern methane storage[1, 3] and deliverable capacity[1]. The best performing hMOF structures were found to have helium void fractions of around 0.8 and methane heats of adsorption between 10 and 15 kJ/mol. Our GCMC simulations of methane storage and delivery in the CoRE MOF database demonstrate that the structure–property relationships identified using the hMOF database are also found in real structures.

*Portions of this chapter have been published previously:

1. Yongchul G. Chung, Jeffrey S. Camp, et al. Computation-ready, experimental metal-organic frameworks: a tool to enable high-throughput screening of nanoporous crystals. *Chemistry of Materials* **2014**, *26*, 6185-6192.
2. Dalar Nazarian‡, Jeffrey S. Camp‡, and David S. Sholl. A comprehensive set of high-quality point charges for simulations of metal–organic frameworks. *Chemistry of Materials* **2016**, *28*, 785-793.
‡These authors contributed equally to this work.

As an example of applying our data set of atomic point charges described in Chapter 2, we screened each MOF in our data set for potential use in the adsorptive removal of *tert*-butyl mercaptan from methane. TBM is one of the principal gaseous sulfur odorants used in pipeline natural gas. Combustion of TBM in natural gas produces undesirable SO_x compounds that can corrode turbines. Removal of TBM from natural gas fuel streams is traditionally accomplished by a two-step catalytic hydrodesulfurization process.[4] Recently, three common MOFs and the zeolite NaY were experimentally tested for use in the selective adsorption of TBM from natural gas.[5] UiO-66(Zr) was found to have promising properties that make it a good candidate material for this application. Below, we use a hierarchical high-throughput screening approach based on GCMC simulations to identify additional MOF materials with high selectivity for TBM over CH_4 and a high saturation capacity for TBM.

3.2 Methods and Computational Details

3.2.1 GCMC Simulations of MOF Adsorption Capacity: CH_4 Storage for Adsorbed Natural Gas Applications

Classical grand canonical Monte Carlo (GCMC) simulations of methane adsorption were conducted on all CoRE MOF structures in the RASPA 1.0 molecular simulation package. Methane adsorption was simulated at 0.01, 5.8, and 65 bar at T = 298 K. Fugacity values necessary to impose equilibrium between the system and the external gas reservoir at each pressure were calculated with the Peng- Robinson equation of state.[6] Methane-methane and methane-framework interactions were modeled with the Lennard-Jones (LJ) 12-6 potential:

$$V_{ij} = 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \qquad (1)$$

Here, i and j index the interacting atoms, r_{ij} is the distance between atoms i and j, and ε and σ are the LJ parameters. LJ interaction parameters between atoms of different types were calculated using the Lorentz-Berthelot mixing rules:

$$\sigma_{ij} = \frac{\sigma_i + \sigma_j}{2} \qquad (2)$$

$$\varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j} \qquad (3)$$

Framework atoms were held fixed in the positions reported in the CoRE MOF database. LJ parameters for all framework atoms were obtained from the Universal Force Field (UFF).[7] LJ parameters for methane (ε/kB = 148.0 K; σ = 3.73 Å) were obtained from the TraPPE force-field.[8] In the TraPPE model, methane is modeled as a single sphere with one LJ interaction site. LJ parameters for helium (ε/kB = 10.9 K; σ = 2.64 Å) were taken from Talu and Myers. All LJ interaction potentials were truncated at 12.8 Å. Accordingly, the simulation cells were replicated to at least 25.6 Å along each axis to satisfy the minimum image convention.

All GCMC simulations included a 1,000-cycle equilibration period followed by a 5,000- cycle production period. A cycle consists of N Monte Carlo steps, where N is the number of atoms in the system. GCMC simulations included random insertion, deletion, translation, and re-insertion moves with equal probabilities.

The isosteric heats of adsorption were calculated based on the fluctuation method[9]:

$$Q_{st} = RT - \frac{\langle VN \rangle - \langle V \rangle \langle N \rangle}{\langle N^2 \rangle - \langle N \rangle^2} \qquad (4)$$

Here, N is the number of methane molecules in the system, V is the potential energy per adsorbed methane, R is the ideal gas constant, and T is the temperature. The angled brackets indicate an ensemble average over the Monte Carlo steps. The heat of adsorption is calculated at low loading (P = 0.01 bar) to exclude contributions from methane-methane interactions.

The helium void fraction of each structure was calculated using the Widom particle insertion method by probing the structure with a Lennard-Jones helium atom at 5000 random points at T = 298 K.[10] The energy difference with and without the helium atom was calculated, and the average Boltzmann weight resulting from the energy difference corresponds directly to the helium void fraction.

3.2.2 GCMC Simulations of MOF Adsorption Selectivity: *tert*-Butyl Mercaptan Removal from CH_4

A hierarchical high-throughput screening approach using grand canonical Monte Carlo simulations was used to identify candidate MOF structures for selective adsorption of trace *tert*-butyl mercaptan (TBM) from methane (CH_4). MOFs were evaluated on the

basis of selectivity for TBM in Henry's regime, the saturation capacity of TBM, the binary selectivity for trace TBM over CH_4, and synthetic and structural properties.

All GCMC simulations of adsorption were performed in the RASPA 1.0 molecular simulation package.[11] The configurational bias Monte Carlo (CBMC) method was used to model the internal flexibility of TBM molecules as described by the TraPPE force field[12], while CH_4 was modeled a single TraPPE united atom. Dispersion forces between adsorbates and the MOF were described by combining Lennard-Jones parameters from the Universal Force Field[7] (MOF atoms) and TraPPE (adsorbates) with the Lorentz–Berthelot mixing rule. Electrostatic interactions were modeled by using our DDEC-derived atomic point charges for MOF atoms and TraPPE charges for TBM and CH_4 molecules. All Lennard-Jones interactions were truncated at 16 Å; all electrostatic interactions were computed pairwise to 16 Å, and a long-range Ewald summation scheme was used thereafter.

Henry's constants of TBM and CH_4 were calculated from 5×10^5 Widom particle insertions. The adsorption selectivity in Henry's regime was defined as the ratio of the single-component TBM and CH_4 Henry's constants.[13] The single-component saturation capacity of each MOF structure for TBM was calculated by performing GCMC at a very high fugacity (5×10^4 bar) using 3×10^4 initialization and 4×10^5 production Monte Carlo cycles. The binary selectivity for TBM over CH_4 was calculated at a composition representative of a natural gas pipeline composition (10 ppm TBM in CH_4) and pressure (18.1 atm) using 2×10^5 initialization cycles and 5×10^5 production cycles. The binary selectivity was defined as

$$\frac{x_{TBM}}{x_{CH_4}}\frac{y_{CH_4}}{y_{TBM}} \qquad (5)$$

where x and y are the concentrations in the adsorbed phase and vapor phase, respectively.[5] Pipeline natural gas can contain a range of other species at low concentrations[14], including low-molecular weight hydrocarbons and CO_2; these components have not been considered in our screening calculations.

3.3 Results and Discussion

3.3.1 Methane Uptake Capacity of CoRE MOFs

Figure 3.1 shows the simulation results for the absolute methane uptake at 65 bar and the deliverable capacity from 65 to 5.8 bar. We found over 800 structures with methane capacity predicted on the basis of our model of greater than 200 vol_{STP} vol^{-1} at 65 bar.

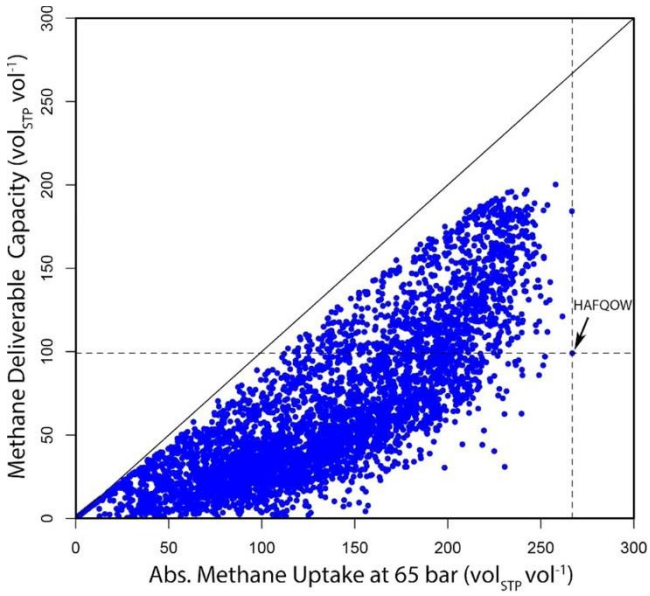

Figure 3.1 Methane deliverable capacity from 65 to 5.8 bar as a function of absolute methane uptake at 65 bar calculated from GCMC simulations for the structures in the CoRE MOF database at 298 K. The data point at the intersection between the vertical and horizontal dotted lines is for CSD: HAFQOW (MIL-53(Al)). HAFQOW is the best MOF in terms of methane storage (267 vol_{STP} vol^{-1}) but not in terms of methane deliverable capacity (~100 vol_{STP} vol^{-1}).

The GCMC simulations predict that the top-performing structure in the CoRE MOF database for absolute methane uptake is MIL-53(Al) (CSD: HAFQOW)[15], with an uptake at 65 bar of 267 vol_{STP} vol^{-1}, which exceeds the current methane storage target set by the ARPA-E MOVE program (263 vol_{STP} vol^{-1}), if the packing efficiency loss is ignored.[2] This was a surprising result, since this is a well-known MOF and only moderate methane uptake has previously been reported for this material up to 30 bar.[16] We speculated that if special attention were paid to activating the MOF and maximizing its surface area, we might see very high methane uptake in the laboratory as predicted by simulation. To test this, our experimental collaborators synthesized MIL-53(Al) and activated it with a

procedure designed to remove solvent and unreacted organic ligands to produce a high-quality sample (S_{BET} = 1530 m^2/g).[17] Details of the synthesis are given in Appendix A.2. However, as shown in Appendix A.3, subsequent experiments with this high-quality MIL-53(Al) sample failed to confirm the simulation prediction, with the methane uptake measured at 65 bar and 298 K being only 190 vol_{STP} vol^{-1}. To assess the origin of this discrepancy, we identified all of the MIL-53(Al) structures in the CSD and computed full methane isotherms from 0 to 65 bar. There are 13 different crystal structures for MIL-53(Al) with somewhat different experimentally resolved lattice parameters and atomic coordinates. We found that there are large variations in the simulated methane isotherms in the 13 MIL-53(Al) structures. At 65 bar and 298 K, the methane storage capacity of these structures varies from 180 to 267 vol_{STP} vol^{-1}. Further discussion about methane adsorption in MIL-53(Al) is provided in Appendix A.3. This case study serves as a cautionary example in using the CoRE MOF structures.

Figure 3.2 shows a parity plot between experimental (Brunauer–Emmett–Teller (BET) or Langmuir) surface areas from the literature and our calculated accessible surface areas for the 53 MOFs ranked highest for methane storage and delivery in our GCMC simulations. As found by Goldsmith et al., the correlation between calculated surface areas and experimentally measured surface areas is poor.[18] Only 16 out of the 53 MOFs examined have experimental surface areas larger than 85% of the calculated value. The discrepancy may be a result of defects in the experimentally synthesized MOFs or incomplete removal of solvents. Our computational analysis assumes that every material can be completely desolvated without resulting in framework collapse. For some materials, it may not be possible to access the porosity at all due to framework collapse (experimental surface areas near zero), while in others the experimental synthesis and activation conditions may not have been optimized. Activation of MOFs is an active area of research, and the activation conditions of MOFs often need to be varied to yield high-quality MOF samples.[19] For example, the correlation between the experimental and theoretical surface areas of IRMOF-1 has improved as better activation protocols have been introduced.[20]

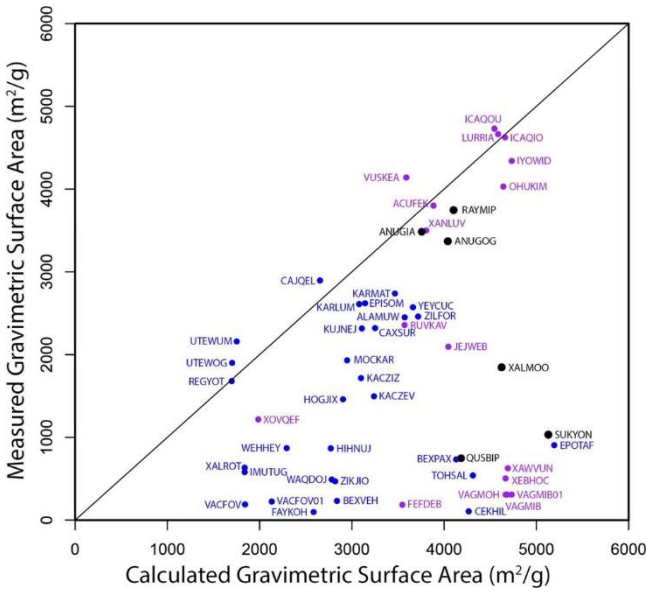

Figure 3.2 Comparison between gravimetric surface areas from the experimental literature and those calculated geometrically from the crystal structures for structures with methane deliverable capacity greater than 180 vol_{STP} vol^{-1} or methane uptake greater than 240 vol_{STP} vol^{-1}. CSD reference codes are shown next to the data points: purple for top methane deliverable capacities, blue for top methane uptake values, and black for both. If the experimental BET surface area is not reported, the Langmuir surface area was used.

3.3.2 Comparison between CoRE MOFs and Hypothetical MOFs for Methane Uptake

In Figure 3.3, we compare the simulated methane uptake values at 65 bar with those from the hypothetical MOF (hMOF) database of Wilmer et al.[3] Both the hMOFs and the CoRE MOFs span a wide range of textural properties (surface area, LCD, etc.), but the CoRE MOFs are much more topologically diverse. Nevertheless, as shown in Figure 3.3, the trends for how methane uptake correlates with different textural properties are remarkably similar for the two sets of MOFs. For example, Figure 3.3c shows that the simulated methane adsorption capacity at 65 bar shows a pronounced maximum at a helium void fraction of around 0.8 for both databases.

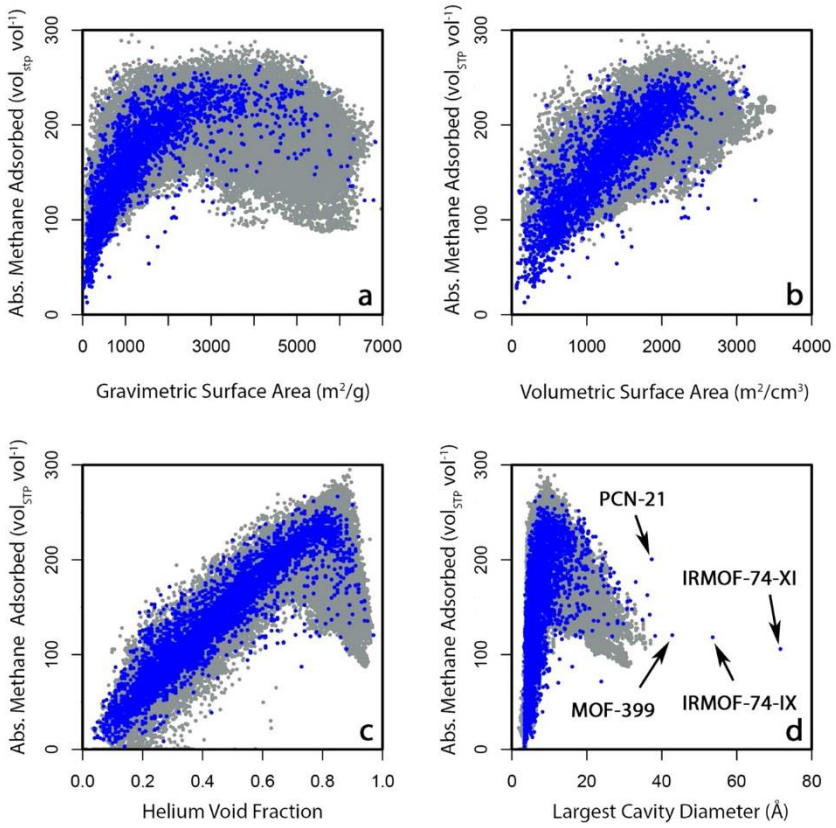

Figure 3.3 Absolute methane storage capacity at 65 bar for the CoRE MOF structures (blue) and the hMOF structures (gray) plotted as a function of (a) gravimetric accessible surface area, (b) volumetric accessible surface area, (c) helium void fraction, and (d) largest cavity diameter.

Figure 3.3d shows that IRMOF-74-XI (CSD: RAVXOD)[21], IRMOF-74-IX (CSD: RAVXIX)[21], MOF-399 (CSD: BAZGAM)[22], and PCN-21 (CSD: YUSWEP)[23] have LCDs greater than those of all of the structures in the hMOF database because large organic linkers in these MOFs were not included by Wilmer et al. in the library of building blocks used to construct the hMOF database.[3]

Figure 3.4 shows that the two databases also produce similar trends for methane deliverable capacity versus heat of adsorption, with the highest methane deliverable capacities occurring at heats of adsorption between 10 and 15 kJ/mol (vertical lines in Figure 3.4). Figures 3.3 and 3.4 suggest that a wide range of textural properties may be more important for developing structure–property relationships for methane storage in MOFs than a wide range of topologies. This is an unexpected result and suggests future work to see if this tentative conclusion holds for other performance properties of MOFs.

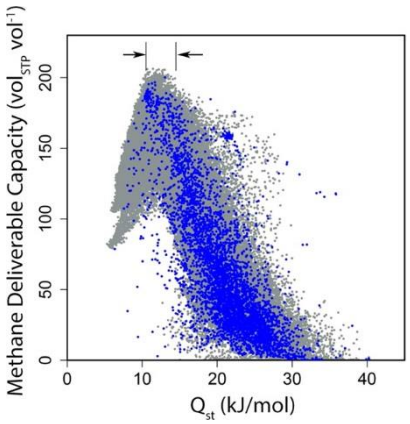

Figure 3.4 Methane deliverable capacity from 65 to 5.8 bar for the CoRE MOF database (blue) and the hMOF database of Wilmer et al. (gray) as a function of the heat of adsorption (Q_{st}) at 0.01 bar.

3.3.3 High-throughput Screening for TBM/ CH_4 Selectivity

In Chapter 2, we discussed comparing point charges assigned with periodic DFT calculations and the DDEC charge partitioning method with results from the semiempirical EQeq method. While this comparison is useful, it is equally useful to remember that point charges are not a direct experimental observable, so the implications of the charge density in a material for its physical properties are typically more important than the numerical value of a given charge. The availability of the charges we have reported above for the CoRE MOF database immediately opens the possibility of screening the materials in this database for physical properties that involve nonpolar adsorbate species. As an initial illustration, we have examined the selective adsorption of *tert*-butyl mercaptan (TBM) from methane. As mentioned above, TBM is ubiquitous in pipeline natural gas at parts per million levels as an odorant. We know of no previous simulations of TBM adsorption in MOFs. Below, we report on the predicted adsorption properties of TBM and methane in the MOFs in our data set. Although there are 2234 unique MOFs in our data set, we performed adsorption calculations on all 2932 MOFs, so these calculations include multiple experimental structures of several MOFs.

First, we calculated Henry's constant and isosteric heats of adsorption of CH_4 and TBM in each structure in the data set. These quantities are computationally inexpensive to compute for large numbers of materials using Widom insertions in RASPA. The Henry's

regime selectivity in each MOF was defined as the ratio of the TBM and CH_4 Henry constants, where values above 1 indicate preferential adsorption of TBM in the Henry regime. Approximately one-quarter of structures exhibited very low Henry regime selectivities ($< 10^{-12}$) and were excluded from further analysis. Most of these structures have a largest cavity diameter (LCD) of < 4.5 Å, indicating nanopores too small to accommodate a TBM molecule. Even if these structures are excluded, our results have MOFs that span an enormous range of selectivities. Figure 3.5 shows the Henry regime selectivities for the remaining structures plotted as a function of the largest cavity diameter and TBM isosteric heat of adsorption.

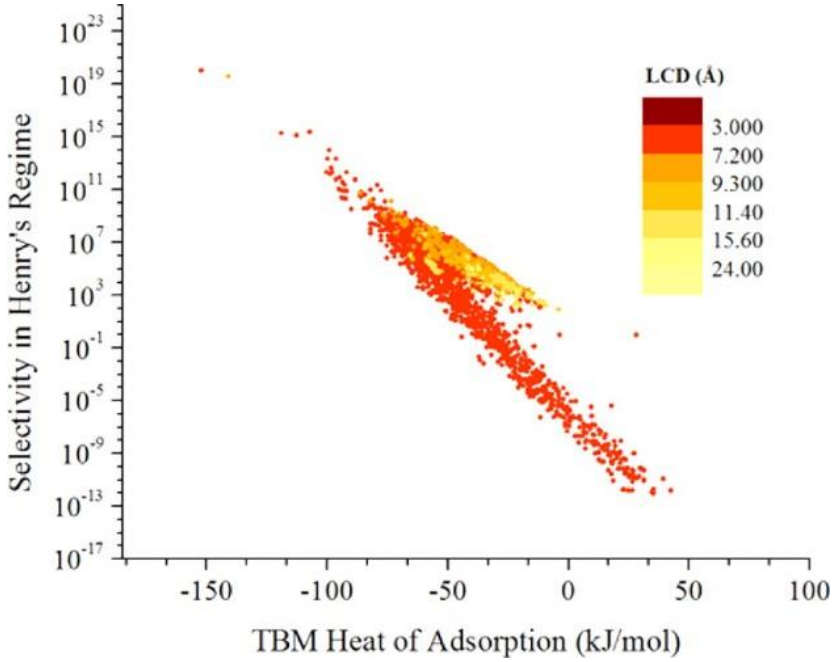

Figure 3.5 Henry's regime selectivity vs TBM heat of adsorption, where negative heats indicate energetically favorable adsorption. More lightly colored data points are associated with structures with larger LCDs.

The results in Figure 3.5 allow a relatively simple description of the competing effects that control selective adsorption of TBM relative to CH_4. Structures with very small pores cannot easily accommodate the TBM molecule but readily adsorb CH_4, leading to Henry's regime selectivities of < 1 and positive TBM heats of adsorption. For materials with slightly larger pores (LCD ~ 5–6 Å), steric repulsive forces become less important than energetically favorable electrostatic and dispersive forces. This regime is associated

with negative heats of adsorption for TBM and high Henry's regime selectivities for TBM over CH_4. The lowest TBM heats of adsorption (less than −70 kJ/mol) are associated with LCDs of 6–6.5 Å. In structures with LCDs in this range, the TBM molecules fit optimally into the largest pore within each MOF. Although these MOFs exhibit very high Henry's regime selectivities (> 10^8), the adsorption of TBM is likely irreversible within these materials and access of TBM into the pores may be subject to severe kinetic limitations. At LCDs of > 8 Å, TBM molecules experience somewhat weaker energetic interactions with nearby MOF atoms, although these interactions are still typically considerably stronger than for CH_4. The structures with the largest LCDs (> 20 Å) have relatively modest Henry's regime selectivities (< 10^4).

Next, we evaluated the correlation between these Henry's regime selectivities and selectivities computed from binary GCMC calculations at a representative pipeline composition of natural gas (10 ppm TBM in CH_4, 18.1 atm total pressure).[24] Binary GCMC simulations simulate competitive adsorption effects at finite loadings that single-component Henry's regime calculations cannot capture. However, these simulations are considerably more computationally expensive than calculations in Henry's regime. We first performed these binary calculations for a subset of around 100 MOFs with Henry's regime selectivities ranging from 1 to 10^{10}. The 100 MOFs were chosen to represent a range of selectivities for TBM and produced well-converged results within 5×10^5 Monte Carlo cycles. Figure 3.6 shows the binary selectivity (eq 5) as a function of Henry's regime selectivity for these 100 materials. With the bulk phase condition we considered a binary selectivity of 105 corresponds to an equimolar adsorbed mixture of TBM and CH_4.

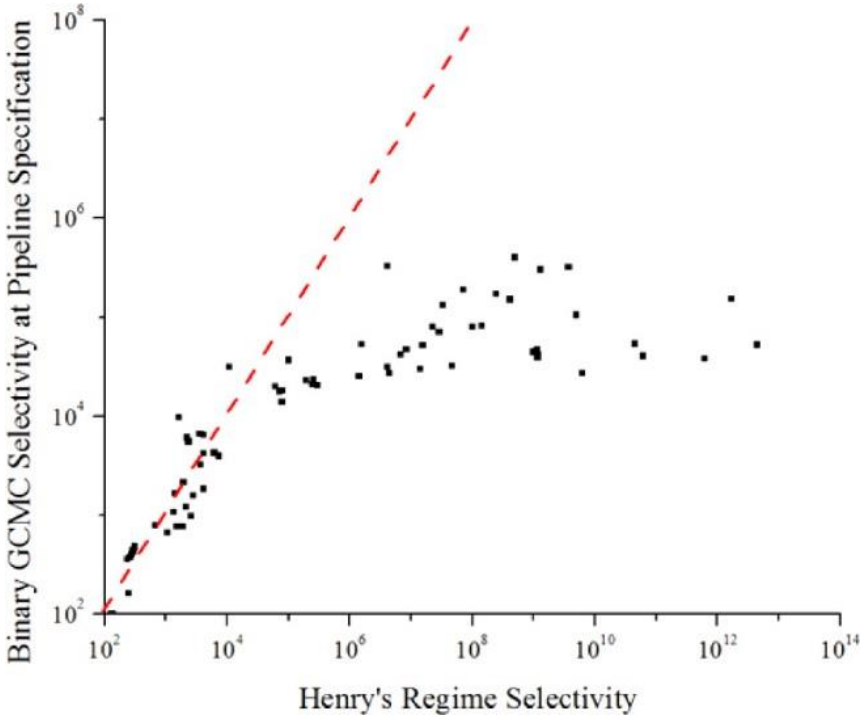

Figure 3.6 Binary GCMC selectivity at a representative pipeline composition of natural gas (10 ppm TBM in CH_4, 18.1 atm) compared to Henry's regime selectivity for 100 MOFs. At values above 10^4, selectivities from binary GCMC deviate significantly from Henry's regime.

Figure 3.6 shows that when the Henry's regime selectivity is less than $\sim 10^4$, Henry's regime selectivity is strongly correlated with the binary selectivity. For Henry's regime selectivities above $\sim 10^4$, the Henry's regime prediction tends to strongly overestimate the binary selectivity. It is challenging to achieve complete numerical convergence in binary GCMC for the most selective materials, and we observed relatively large fluctuations in the observed GCMC selectivity for some structures because of the very small amounts of CH_4 observed. Nevertheless, the uncertainties associated with this effect are small enough to allow us to conclude that the trend shown in Figure 3.6 for high-selectivity materials is a physical effect. This effect arises because the TBM adsorption is not accurately described by Henry's law in these highly selective materials under the bulk phase conditions we examined, meaning that the adsorbed amount of TBM is overestimated by using Henry's regime results. Although this means that using binary GCMC calculations is necessary to quantitatively describe adsorption of the TBM/CH_4 mixture in the most selective materials we have considered, Figure 3.6 shows that using

the Henry's regime selectivity is a useful way to order materials using computationally efficient methods.

On the basis of the results described above, we narrowed our attention to materials with Henry's regime selectivity of $>10^3$. While MOFs with high selectivity for TBM are desired, MOFs with a large TBM heat of adsorption are likely to irreversibly adsorb TBM. To include the feature in our calculations in a simple way, we also eliminated MOFs with a Henry's regime heat of adsorption for TBM more favorable than −70 kJ/mol. We also removed all MOFs containing lanthanide metals from further consideration. For the remaining 1497 distinct MOFs, we performed calculations to evaluate each material's capacity for TBM and the binary selectivity under the natural gas pipeline conditions defined above. The TBM saturation capacity was calculated using single-component GCMC at a fugacity (50 kPa) above the vapor pressure of TBM at 25 °C.[25] As expected, these results correlate strongly with the MOF pore volume. Figure 3.7 shows the saturation TBM loading of each MOF as a function of selectivity at the natural gas pipeline composition calculated using binary GCMC.

We anticipate that top-performing MOFs for TBM removal will have a saturation loading for TBM of > 200 mg/g. We found 354 MOFs from the CoRE MOF database meeting these criteria. Among these promising candidates are multiple representations of the commonly studied MOFs reported in different experimental reports. Among these common MOFs are MIL-53 and Cu-BTC, which were identified by Chen et al. as being highly selective but structurally unstable during TBM adsorption.[5] ZIF-8, a commonly studied and readily available MOF that is stable under humid conditions, is predicted to have a binary selectivity of 1.17×10^4 and saturation loading of approximately 270 mg of TBM/g of adsorbent. It has been shown in both experiments and using molecular modeling that ZIF-8 can adsorb molecules that are considerably larger than its nominal pore diameter because of flexibility in the small windows that control molecular diffusion in this material.[26-28] These observations mitigate concerns that TBM adsorption in ZIF-8 would be limited by kinetic considerations. Other less studied but water stable and promising candidates include BIBXUH, a nickel-based MOF with a 691 mg/g capacity for TBM, and MFU-4, a zinc- and chlorine-based MOF with a 6.85×10^5 selectivity for TBM over methane.

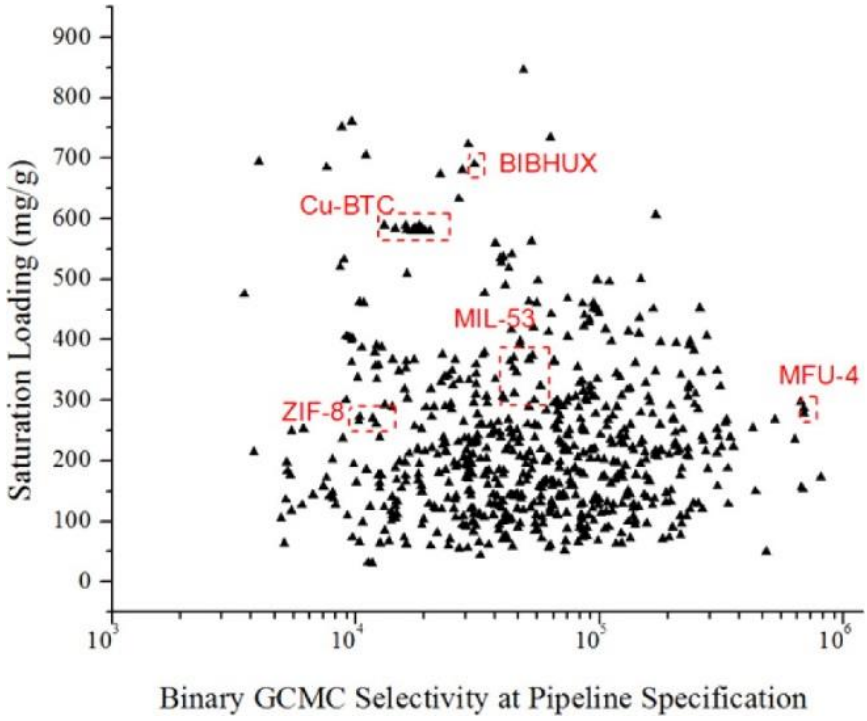

Figure 3.7 TBM/CH_4 selectivity at the pipeline composition as a function of saturation loading of TBM in 1497 MOFs. Cu-BTC and MIL-53, two MOFs studied experimentally for TBM adsorption by Chen et al., are highlighted. Three other promising and water stable MOFs are also highlighted.

We have not attempted in these screening calculations to pick a single "winning" material for the challenge of selectively adsorbing TBM from CH_4. Finding appropriate materials for practical use must involve considerations that are beyond the scope of our current calculations, including the long-term stability of materials, the cost and ease of synthesis of materials, etc. Nevertheless, the observation that our calculations have identified a large number of materials with appealing adsorption selectivities and adsorption capacities for TBM provides a strong basis for continued development of high-performance materials for this application.

3.4 Conclusions

Simulations of methane adsorption in the CoRE MOF database show that the predicted structure–property relationships agree well with those predicted in a database of hypothetical MOFs, even though the hypothetical MOFs are much less topologically diverse. Notably, both the CoRE MOF and hypothetical MOF databases predict that

methane storage capacity is optimized at a helium void fraction of around 0.8 and that methane deliverable capacity is maximized at heats of adsorption between 10 and 15 kJ/mol. Future work is needed to better understand the role of topology on structure–property relationships in MOFs. For example, are certain combinations of textural properties only accessible with certain topologies? We anticipate that the CoRE MOF database will be useful in answering these questions, in facilitating high-throughput identification of candidate MOFs for future applications, and in revealing structure–property relationships that could suggest design principles for optimum materials.

As an example of applying our data set of atomic point charges, we have screened each MOF in our data set for potential use in the adsorptive removal of tert-butyl mercaptan from methane. Our efficient screening procedure has identified hundreds of MOFs with high selectivity and capacity for TBM. These results suggest multiple directions for future experimental efforts, including the identification of some well-known materials as potential candidates for this separation.

The high selectivities of MOFs in our study reveal a potential challenge with the application of MOFs for methane storage. In the original CoRE MOF report, MIL-53 was found to have among the highest capacities for methane storage.[17] In our study, we find that MIL-53 is highly selective for TBM and most likely for other polar components of natural gas. This selectivity may drastically reduce methane capacity during cyclic adsorption, especially if TBM accumulates over the many cycles in the lifetime of the material. Zhang et al. have studied the adsorption and shown the accumulation of ethane, propane, and butane in some common MOFs,[14] but there has yet to be a study of the impact of adsorption of trace components such as TBM and other polar species such as H_2O and CO_2 in natural gas.

3.5 References

[1] D. A. Gómez-Gualdrón, C. E. Wilmer, O. K. Farha, J. T. Hupp, R. Q. Snurr, *The Journal of Physical Chemistry C* **2014**, *118*, 6941-6951.

[2] C. M. Simon, J. Kim, D. A. Gomez-Gualdron, J. S. Camp, Y. G. Chung, R. L. Martin, R. Mercado, M. W. Deem, D. Gunter, M. Haranczyk, D. S. Sholl, R. Q. Snurr, B. Smit, *Energy & Environmental Science* **2015**, *8*, 1190-1199.

[3] C. E. Wilmer, M. Leaf, C. Y. Lee, O. K. Farha, B. G. Hauser, J. T. Hupp, R. Q. Snurr, *Nature Chemistry* **2012**, *4*, 83-89.

[4] I. Babich, J. Moulijn, *Fuel* **2003**, *82*, 607-631.

[5] G. Chen, S. Tan, W. J. Koros, C. W. Jones, *Energy & Fuels* **2015**, *29*, 3312-3321.

[6] D.-Y. Peng, D. B. Robinson, *Industrial & Engineering Chemistry Fundamentals* **1976**, *15*, 59-64.

[7] A. K. Rappé, C. J. Casewit, K. Colwell, W. Goddard Iii, W. Skiff, *Journal of the American Chemical Society* **1992**, *114*, 10024-10035.

[8] M. G. Martin, J. I. Siepmann, *The Journal of Physical Chemistry B* **1998**, *102*, 2569-2577.

[9] T. J. H. Vlugt, E. García-Pérez, D. Dubbeldam, S. Ban, S. Calero, *Journal of Chemical Theory and Computation* **2008**, *4*, 1107-1118.

[10] B. Widom, *The Journal of Chemical Physics* **1963**, *39*, 2808.

[11] D. Dubbeldam, S. Calero, D. E. Ellis, R. Q. Snurr, *Molecular Simulation* **2015**, *42*, 81-101.

[12] N. Lubna, G. Kamath, J. J. Potoff, N. Rai, J. I. Siepmann, *Journal of Physical Chemistry B* **2005**, *109*, 24100-24107.

[13] T. Van Heest, S. L. Teich-McGoldrick, J. A. Greathouse, M. D. Allendorf, D. S. Sholl, *Journal of Physical Chemistry C* **2012**, *116*, 13183-13195.

[14] H. D. Zhang, P. Deria, O. K. Farha, J. T. Hupp, R. Q. Snurr, *Energy & Environmental Science* **2015**, *8*, 1501-1510.

[15] S. V. Potts, L. J. Barbour, D. A. Haynes, J. M. Rawson, G. O. Lloyd, *Journal of the American Chemical Society* **2011**, *133*, 12948-12951.

[16] G. Férey, C. Mellot-Draznieks, C. Serre, F. Millange, J. Dutour, S. Surblé, I. Margiolaki, *Science* **2005**, *309*, 2040-2042.

[17] Y. G. Chung, J. Camp, M. Haranczyk, B. J. Sikora, W. Bury, V. Krungleviciute, T. Yildirim, O. K. Farha, D. S. Sholl, R. Q. Snurr, *Chemistry of Materials* **2014**, *26*, 6185-6192.

[18] J. Goldsmith, A. G. Wong-Foy, M. J. Cafarella, D. J. Siegel, *Chemistry of Materials* **2013**, *25*, 3373-3382.

[19] J. E. Mondloch, O. Karagiaridi, O. K. Farha, J. T. Hupp, *CrystEngComm* **2013**, *15*, 9258-9264.

[20] S. S. Kaye, A. Dailly, O. M. Yaghi, J. R. Long, *Journal of the American Chemical Society* **2007**, *129*, 14176-14177.

[21] H. Deng, S. Grunder, K. E. Cordova, C. Valente, H. Furukawa, M. Hmadeh, F. Gándara, A. C. Whalley, Z. Liu, S. Asahina, *Science* **2012**, *336*, 1018-1023.

[22] H. Furukawa, Y. B. Go, N. Ko, Y. K. Park, F. J. Uribe-Romo, J. Kim, M. O'Keeffe, O. M. Yaghi, *Inorganic Chemistry* **2011**, *50*, 9147-9152.

[23] W. Zhuang, S. Ma, X.-S. Wang, D. Yuan, J.-R. Li, D. Zhao, H.-C. Zhou, *Chemical Communications* **2010**, *46*, 5223-5225.

[24] G. Chen, R. P. Lively, C. W. Jones, W. J. Koros, *Industrial & Engineering Chemistry Research* **2014**, *53*, 7113-7120.

[25] SigmaAldrich, *Safety Data Sheet: 2-Methyl-2-propanethiol* **2015**.

[26] R. J. Verploegh, S. Nair, D. S. Sholl, *Journal of the American Chemical Society* **2015**, *137*, 15760-15771.

[27] C. Zhang, R. P. Lively, K. Zhang, J. R. Johnson, O. Karvan, W. J. Koros, *Journal of Physical Chemistry Letters* **2012**, *3*, 2130-2134.

[28] K. Zhang, R. P. Lively, C. Zhang, R. R. Chance, W. J. Koros, D. S. Sholl, S. Nair, *Journal of Physical Chemistry Letters* **2013**, *4*, 3618-3622.

4

TRANSITION STATE THEORY METHODS TO MEASURE DIFFUSION IN FLEXIBLE NANOPOROUS MATERIALS: APPLICATION TO A POROUS ORGANIC CAGE CRYSTAL*

4.1 Introduction

Adsorbates within nanoporous materials such as zeolites and MOFs experience strong confinement effects that can result in transport properties dramatically different than bulk phases. Atomistic simulations of these adsorbed phases can provide insight into the kinetics of transport in nanopores where experimental characterization would be challenging.[1] For example, molecular dynamics (MD) simulations can give adsorbate diffusivities, which are an important predictor of material performance in both equilibrium and kinetic separation applications.[2] These MD simulations frequently treat the crystalline framework as a rigid body.[3-4] This approach greatly reduces the computational cost of MD because adsorbate–framework potential energies can be mapped to a grid and do not need to be recomputed over the course of a simulation.

The rigid framework assumption may be reasonable for stiff frameworks or where adsorbate dimensions are significantly smaller than pore apertures.[5] In highly flexible materials or where adsorbates are comparable in size to pore apertures, framework flexibility has a significant influence on diffusion.[6] This presents a technical challenge, as including flexibility dramatically increases the cost of MD simulations of nanoporous materials because many framework degrees of freedom must be updated with each MD step. It is computationally expensive to capture enough rare diffusive hops in straightforward molecular dynamics simulations to make accurate measurements of a diffusion coefficient. In rigid systems, a practical lower bound for diffusivities that can be captured using straightforward molecular dynamics is 10^{-7} cm^2/s, corresponding to about 1 hop between nanopores per 50 nanoseconds of simulation time.[6] In flexible simulations,

*Portions of this chapter have been published previously:

Jeffrey S. Camp and David S. Sholl. Transition state theory methods to measure diffusion in flexible nanoporous materials: application to a porous organic cage crystal. *Journal of Physical Chemistry C* **2016**, *120*, 1110-1120.

this value is perhaps an order of magnitude higher due to the additional expense associated with computing framework motions.

Adsorbate diffusion by activated hopping between nanoporous cavities through small apertures (windows) is an example of a "rare event" in molecular simulations. Systems with rare events are characterized by regions of phase space separated by energetic barriers such that mean residence times in each stable region are much longer than the timescale of transitions between regions.[7] The prototypical method for studying rare events is transition state theory (TST), developed in the 1930s by Eyring and Wigner.[8] Central to TST is the concept of the dividing surface, which partitions phase space into two regions along some reaction coordinate. TST reduces the dynamical problem of rate constant calculations to the equilibrium problem of finding the reversible work to reach the dividing surface from a stable state.[9]

In the context of diffusion in nanoporous materials, the hopping rate between adjacent cavities is calculated by finding the reversible work necessary to move an adsorbate from its energy minimized location within a cavity to the window. This free energy profile can be computed by sampling adsorbate positions with MD or Monte Carlo and then histogramming these positions over the reaction coordinate.[10] This method is computationally expensive because the adsorbate rarely visits the high energy regions of phase space. Various methods have been developed to overcome this problem including Widom particle insertions[11], umbrella sampling[12], and thermodynamic integration[13]. If entropic contributions to the energetic barrier are neglected, the problem reduces to calculation of the minimum potential energy path through the window. The method of Haldoupis et al.[14] calculates this minimum energy path on a spatially discretized energy surface, while nudged elastic band calculations[15] use the continuous energy surface.

Transition state theory methods have been applied extensively to the diffusion in rigid zeolites[16] and metal-organic frameworks[17]. However, few reports have used TST in the study of flexible nanoporous materials.[18] The primary objective of this work is to demonstrate two numerically efficient TST methods that are broadly applicable to measuring slow diffusion in highly flexible nanoporous materials that is computationally inaccessible to direct MD. These methods are an alternative to more complicated transition path sampling methods used to simulate slow diffusion in flexible nanoporous materials.[19]

We also describe techniques to analyze the relationship between pore aperture (window) sizes, adsorbate hopping activation energies, and adsorbate hopping rates.

4.2 Porous Organic Cages

Porous organic cages (POCs) crystallize in the solid state without forming intermolecular covalent or coordination bonds.[20-21] The weak nature of these intermolecular forces makes POCs inherently flexible.[22] One of the first POCs synthesized, known as "cage crystal 3" (CC3),[23] has a number of desirable properties that make the material potentially useful in kinetic separation applications. CC3 forms a 3-D diamondoid pore network with a BET surface area[24] of 409 $m^2\ g^{-1}$ which is thermally stable to 398 C,[23] stable in boiling water,[25] and reversibly adsorbs over 20 wt % water[25]. MD simulations of adsorbates within isolated cage molecules were used to predict that mesitylene molecules cannot overcome a reorientational barrier to diffusion in CC3 while smaller aromatics such as 4-ethyltoluene can. These predictions were consistent with single phase gas uptake experiments which demonstrated that CC3 can adsorb a significant amount of 4-ethyltoluene, but no mesitylene.[26] Subsequent MD studies of diffusion in crystalline CC3 at 300 K showed that 4-ethyltoluene molecules hop between cages and intercage voids several times over the course of a 20 ns simulation, but mesitylene molecules are completely immobile. Recently, CC3 was reported to have attractive performance in the adsorptive separation of Kr, Xe, and Rn from air.[27]

Although qualitative insights from MD can be valuable, typically hundreds of adsorbate hops must be observed in MD to yield a well-converged diffusion coefficient. This "MD time scale problem" is a well-known challenge in the simulation of diffusion in nanoporous materials.[10] In CC3, this challenge is also relevant for smaller spherical adsorbate molecules. Evans et al. performed MD simulations of light gas diffusion to evaluate the performance of mixed-matrix membranes with CC3 and other porous organic cage additives.[28] The diffusion coefficient of CH_4 in CC3 measured in that study at 298 K and 10 bar adsorbate loading pressure was $2.16 \times 10^{-6}\ cm^2\ s^{-1}$, which corresponds to only a few CH_4 hops per nanosecond.[28] Similarly, Holden et al. found that the diffusion coefficient of Xe in CC3 at 298 K is $1.83 \times 10^{-6}\ cm^2\ s^{-1}$. It is computationally expensive

to measure diffusion coefficients on this order of magnitude using MD, particularly at low adsorbate loadings. Larger adsorbates of interest in CC3 such as 1-phenylethanol[29] are expected to diffuse orders of magnitude slower, making direct MD simulation of diffusion infeasible. Here, we apply two efficient TST methods to compute the self-diffusivities of light gases in CC3 in the regimes of low temperature and infinite dilution where diffusion is inaccessible to direct MD simulation.

4.3 Modeling Diffusion in CC3

4.3.1 Structure and Force Fields

A 2 × 2 × 2 supercell of the homochiral CC3-R structure (Cambridge Structural Database: PUDXES)[30] was used throughout this work. All window sizes reported here were calculated by the arene carbon method described by Chen et al.[27] for consistency with window sizes previously reported in CC3.[27, 31] This approach is convenient but less rigorous than a definition of window size based on percolating sphere diameters[14] or Voronoi decomposition.[32]

Cage Adsorption Force Field (CAFF)

The cage-adsorption force field (CAFF)[27] was used to model adsorbate–cage interactions for CH_4 and the noble gases. CAFF is based upon DREIDING[33] with rescaled epsilon parameters to fit experimental adsorption data. United atom models of carbon disulfide (CS_2) and sulfur hexafluoride (SF_6) were represented using experimentally derived 12–6 Lennard-Jones parameters from the literature.[34] Parameters from EPM2[35] and UFF[36] for CO_2 and the framework, respectively, were combined with the Lorentz–Berthelot mixing rules to model carbon dioxide–framework dispersion interactions. CSFF partial charges were used to model electrostatic interactions with CO_2. CAFF parameters were also used to model adsorbate–adsorbate interactions. All adsorbate–cage and adsorbate–adsorbate interactions were truncated at 10 Å.

Cage Specific Force Field (CSFF)

CSFF was parametrized by Holden et al.[37] to describe the framework dynamics of CC3 and other porous organic cages. In CSFF, intramolecular nonbonded interactions are

completely excluded, meaning that the 168 atoms within individual cage molecules do not interact by Lennard-Jones or Coulombic forces. To describe intermolecular forces between individual cage molecules, CSFF uses partial charges and scaled 9–6 Lennard-Jones parameters from the polymer consistent force field (PCFF).[38] Following the original CSFF report, the dispersive portion of Lennard-Jones potential was scaled by a factor of 1.20. All CSFF intermolecular Lennard-Jones interactions were truncated at 10 Å. Coulombic interactions were computed pairwise to 10 Å, and a long-range particle–particle mesh Ewald correction was used thereafter. Bonded forces were applied as described in the original CSFF report with the following exceptions: two angle and three torsion parameters absent from the original CSFF report[37] were adapted directly from PCFF by Holden and co-workers and used throughout this work. These parameters are necessary to reproduce the geometry of the cyclohexyl groups on the CC3 molecules. The coefficient leading the class2 trigonometric dihedral potential was also modified, as described in Appendix B.

4.3.2 Energy Minimization of the CC3 Structure

The atomic positions of the CC3 structure were minimized with the CSFF force field using a damped dynamics algorithm[39] to a tolerance of 10^{-11} kcal mol^{-1}. Independent minimization runs were performed at variable lattice constants created by isotropically scaling the CC3 atomic positions by factors ranging 0.97 to 1.03 (corresponding to changes in unit cell volumes between −9% and +9%).

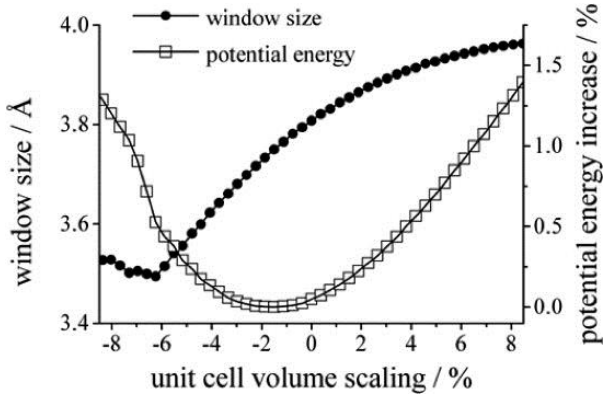

Figure 4.1 Energy minimized CC3 window sizes and potential energies as a function of unit cell volume scaling.

Figure 4.1 shows the energy minimized window sizes and total potential energies for unit cell volumes ranging from 9% smaller to 9% larger than the experimental CC3 structure. At the experimentally refined lattice constants (0% volume scaling), the CC3 windows relax to 3.80 Å, about 5% larger than the window size of the experimental structure (3.62 Å). The potential energy of the relaxed CC3 structure at the experimental unit cell volume is slightly higher (0.05%) than the relaxed CC3 structure found after applying a 1.5% volumetric contraction in the unit cell. Over the range of unit cell volumes considered, small differences in potential energy (less than 1.5%) are associated with significant changes in the relaxed CC3 window size (3.49–3.97 Å). In molecular dynamics simulations using the flexible CSFF force field frameworks, the CC3 windows assume many different conformations as the simulation progresses.

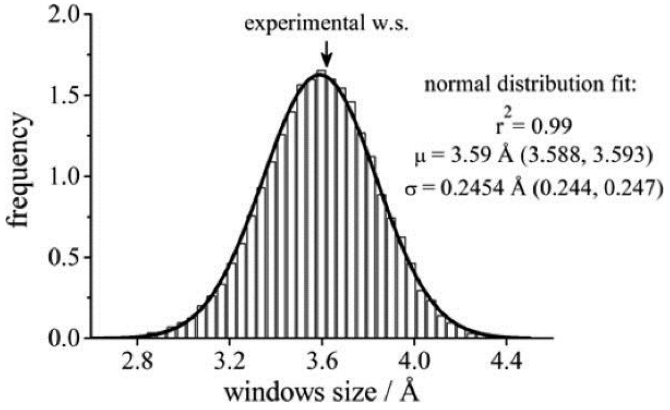

Figure 4.2 Normalized histogram of CC3 window sizes in fully flexible NVT MD at 300 K. The experimental window size (3.62 Å) is indicated with an arrow.

After 10 ps of NPT equilibration at 300 K, the CC3 structure contracts by 4.2%. Figure 4.2 shows the normalized window size distribution at 300 K in the empty CC3 structure over a 100 ps NVT production period at these NPT equilibrated lattice constants. The window size distribution observed in NPT simulations is essentially identical to the NVT results shown in Figure 4.2 (data not shown). The window size distribution is very well fit by a normal distribution, which is a common feature of many nanoporous materials.[6, 28] The window size of the energy minimized structure at the corresponding lattice parameters (3.61 Å at 4.2% contraction) is close to the mean of the distribution (μ = 3.59 Å). The fit parameters are in reasonably good agreement with the CC3 window size distribution reported in the original CSFF publication, which is centered near μ = 3.75 Å.[37]

4.3.3 Diffusion Coefficients by MD

The diffusion coefficients of Kr, CH_4, and CO_2 were measured with NVT molecular dynamics at 300 K using a Nosé–Hoover thermostat and a 0.5 fs time step. For the other adsorbates and temperatures considered below, diffusion was too slow to be readily simulated with straightforward MD. Each of the adsorbates simulated with MD was loaded to a concentration of 2 adsorbate molecules per unit cell, which each contain 8 CC3 molecules. This corresponds to a total of 16 adsorbate molecules in a 2 × 2 × 2 CC3 supercell, or a ratio of 1 adsorbate atom for every 4 porous organic cage molecules. The experimentally measured saturation loading of CC3 for Kr and Xe is 2.1 and 2.69 adsorbate atoms per cage, respectively, at room temperature.[27] We expect that our simulated loading of 0.25 adsorbates per cage results in negligible adsorbate–adsorbate interaction effects on self-diffusivities measured by MD. This facilitates comparison to TST calculations performed at infinite dilution.

The loaded structure was subject to a 20 ps NPT equilibration period at a pressure of 1 bar followed by a 20 ps NVT equilibration at the equilibrated cell volume. Twenty-eight (28) independent 1 ns NVT simulations were averaged and then fit to the Einstein relation to give the self-diffusivity, Ds. In each case, the MSD vs time curve was renormalized to zero by subtracting 4 $Å^2$ from the MSD, which is associated with the ballistic movement of adsorbate atoms within individual cage molecules. The mean-squared displacements plots and the associated fits are shown in Appendix B.

4.3.4 Calculating Diffusion Coefficients by TST Hopping Rates

Diffusion in CC3 proceeds by activated adsorbate hopping of molecules between a cage and the void space separating adjacent cages, also known as "window cavities".[27] The self-diffusion coefficient is given by weighting the hopping rates out of the 4-coordinated cage sites ($k_{C\rightarrow V}$) and the 2-coordinated void sites ($k_{V\rightarrow C}$) by the equilibrium probability of occupying cages or voids (P_C and P_V, respectively):[40]

$$D_s = \frac{1}{6}\lambda^2(4k_{C\rightarrow V}P_C + 2k_{V\rightarrow C}P_V) \quad (1)$$

Here, λ is the hopping distance between cages and voids, which is 5.32 Å in the NPT equilibrated structure at 300 K. Since the likelihood of occupying a cage or void is

directly proportional to the residence time in each site (the inverse of the hopping rate), the occupancy probabilities can be eliminated from eq 1, giving

$$D_s = \frac{1}{6}\lambda^2 \left(4\left(\frac{k_{V\to C}}{2k_{C\to V} + k_{V\to C}}\right)k_{C\to V} + 2\left(\frac{2k_{C\to V}}{2k_{C\to V} + k_{V\to C}}\right)k_{V\to C}\right) \quad (2)$$

The hopping rates $k_{C\to V}$ and $k_{V\to C}$ were calculated using the 1-dimensional TST methods described below.

4.4 Computing Hopping Rates Using TST

4.4.1 TST in the Rigid Experimental Structure

The local framework flexibility associated with window size deformation is understood to strongly influence diffusion in CC3.[28, 41-42] We first performed simulations of spherical adsorbate diffusion in the rigid experimental structure to construct an appropriate reaction coordinate and to establish a basis of comparison for our fully flexible simulations. The 1-dimensional reaction coordinate (denoted q) used in all of our simulations is coincident with the vector that connects the centers of mass of two adjacent cage molecules. The coordinate system along this line is centered (q = 0) at the midpoint between these positions in the void space that spans adjacent cages. Along this line, the centers of mass are located at q = ± 5.4 Å in the experimental structure and each window is centered at q = ± 2.5 Å. For convenience, we constructed a reaction coordinate between two cage molecules that do not span periodic boundaries in the 2×2×2 CC3 supercell.

To determine the free energy profile, *F(q)*, for adsorbate hopping along this reaction coordinate, we calculated the mean energy of insertion of spherical adsorbate molecules using CAFF parameters in planes orthogonal to the reaction coordinate:

$$F(q) = -k_B T \ln\langle e^{-\beta\Delta U}\rangle_q \quad (3)$$

Here, the brackets denote averaging the Boltzmann factor over square grids of 0.2 Å resolution perpendicular to the reaction coordinate at positions along q. Each of these squares is 11.4 × 11.4 Å across (57 × 57 insertions) to fully enclose the cage cavities. These square planes were positioned at 0.2 Å intervals to form a contiguous 3-dimensional rectangular cuboid grid oriented parallel to q. Figure 4.3 illustrates the construction of the square grid "slices" orthogonal to the reaction coordinate.

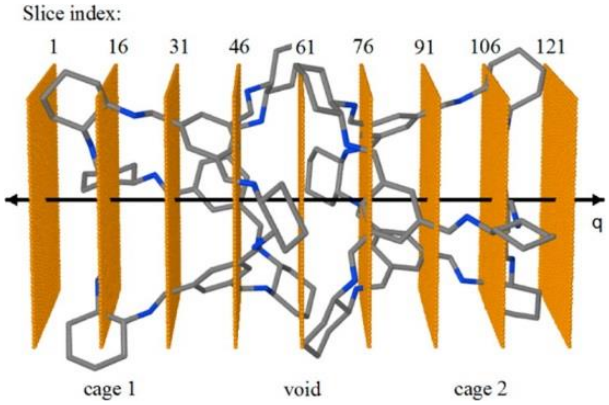

Figure 4.3 TST reaction coordinate superimposed on the experimental CC3 structure. Every 15th slice is shown by an orange square.

The cage to void hopping rate $k_{C\rightarrow V}$ was calculated by applying 1-dimensional TST[10]:

$$k_{C\rightarrow V} = \kappa \sqrt{\frac{k_B T}{2\pi m}} \frac{e^{-\beta F(q^*)}}{\int_{cage} e^{-\beta F(q)} dq} \qquad (4)$$

Here, m is the mass of the adsorbate molecule and κ is the Bennett–Chandler dynamic correction factor. We have assumed a dynamic correction factor of 1 throughout this work, which was shown to be a good approximation for spherical adsorbates in LTL and LTA-type zeolites.[10] $F(q^*)$ is the free energy at the transition states, which correspond to the two local maxima in $F(q)$. The denominator of eq 4 was evaluated by integrating over the points on the reaction coordinate associated with the respective cage microstates. $F(q)$ is symmetric in the rigid experimental structure (Figure 4.4b), meaning that $k_{\text{cage 1}\rightarrow\text{void}}$ and $k_{\text{cage 2}\rightarrow\text{void}}$ are equal. We evaluated the void to cage hopping rate $k_{V\rightarrow C}$ by integrating over the void microstate, corresponding to the points on $F(q)$ between q_1^* and q_2^*.

To visualize the cage and void microstates in CC3, we applied the Fernand–Meyer watershed segmentation algorithm[43] to the 3-dimensional adsorbate potential energy grid described above. In 2D image processing, this algorithm is used to segment images into regions based on pixel intensity. For example, a 2D grayscale image can be considered a 3D topographic map where altitude at each point is proportional to pixel intensity. The Fernand–Meyer algorithm identifies the 2D "catchment basins" and 1D "ridge lines" that separate adjacent regions in this topographic projection. In our 3-dimensional application, the adsorbate potential energy grid in each voxel is mapped to a grayscale pixel intensity.

The Fernand–Meyer algorithm then identifies the 3D microstates and 2D dividing surfaces in potential energy space. This procedure is similar to the algorithm of Henkelman et al. for grid-based Bader decomposition of charge density.[44]

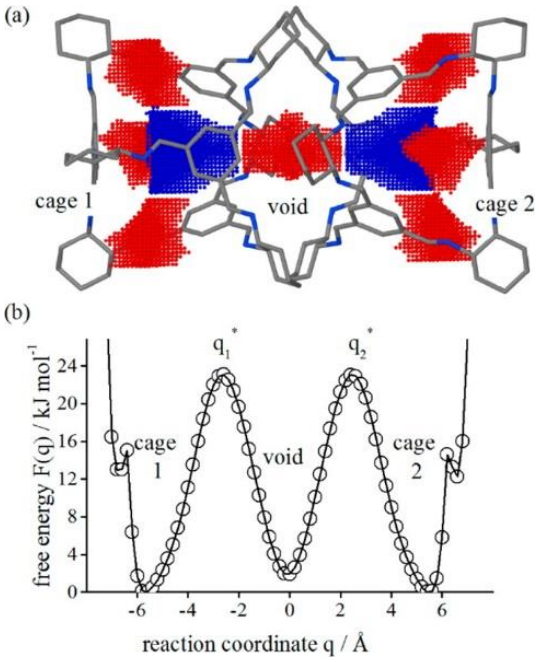

Figure 4.4 (a) Microstates in the rigid experimental CC3 structure revealed by the watershed segmentation algorithm. The 4-coordinated cage sites are shown in blue while the 2-coordinated void sites are shown in red. (b) One-dimensional free energy profile for Xe hopping at 300 K in the rigid experimental structure.

Figure 4.4a shows the result of watershed segmentation of the Xe potential energy grid between adjacent CC3 cages calculated with eq 3 for the experimental CC3 structure. The dividing surfaces that separate the cage microstates (shown in blue) and the void microstates (shown in red) are associated with the transition state free energies $F(q^*)$ on the corresponding 1-dimensional Xe free energy profile shown in Figure 4.4b.

In the rigid experimental structure, the free energy profile, $F(q)$, and associated hopping rates for a given adsorbate can be calculated at arbitrary temperatures by changing T in eqs 3 and 4. Between any two temperatures T_1 and T_2, the cage to void hopping rates $k_{V\rightarrow C}$ can be fit to the Arrhenius equation to yield to activation energy for cage to void hops $E_{A,C\rightarrow V}$:

$$E_{A,C\rightarrow V} = -R\frac{\ln\left(k_{V\rightarrow C}(T_1)\right) - \ln\left(k_{V\rightarrow C}(T_2)\right)}{{T_1}^{-1} - T_2^{-1}} \qquad (5)$$

The Arrhenius prefactor is computed in a similar way. The same procedure is used to determine the activation energy for void to cage hops $E_{A,V\rightarrow C}$.

4.4.2 Flexible Implicit Ligand Sampling TST

To calculate adsorbate hopping rates in the fully flexible CC3 structure, we first applied a variation of the implicit ligand sampling method[45] developed to study diffusion in flexible nanoporous materials.[6] In this approach, hopping rates are determined by TST in an ensemble of rigid frameworks ("snapshots") captured from a fully flexible MD trajectory of the empty porous material. The overall flexible hopping rates are then calculated by averaging the hopping rates calculated in each individual framework conformation. Awati et al. have shown that this approach is quantitatively accurate for diffusion of light gases in small pore zeolites.[46] This method assumes that adsorbates do not significantly influence the motion of the crystalline framework.

Snapshots of the fully flexible CC3 structure were generated by running 20 ps of NPT equilibration at 1 bar followed by a 100 ps NVT production period. Snapshots were saved at 100 fs intervals, producing ensembles of 1000 snapshots at temperatures of 200, 250, and 300 K. The NPT equilibration period results in contraction of the CC3 framework by 5.3% at 200 K, 4.9% at 250 K, and 4.2% at 400 K relative to the experimental structure. The brief molecular dynamics production period used here means that the ensemble of snapshots will capture only local flexible framework vibrations, such as changes in CC3 window sizes, not possible phase transitions that occur at longer time scales. CC3 is not known to undergo any large changes in the structure of its unit cell, but deformations of this type could pose a challenge to applying implicit ligand sampling to materials such as cage crystal 1, which can switch from a porous to a nonporous phase.[47]

In each flexible snapshot, we calculated the free energy profile, *F(q)*, by the same grid method as described above for the rigid experimental structure. The hopping rates in each snapshot were determined by 1-dimensional TST using eq 4. Within each snapshot, we calculated two independent cage to void hopping rates ($k_{\text{cage 1}\rightarrow\text{void}}$ and $k_{\text{cage 2}\rightarrow\text{void}}$) and two independent void to cage hopping rates ($k_{\text{void}\rightarrow\text{cage 1}}$ and $k_{\text{void}\rightarrow\text{cage 2}}$) in each of 16 cage–void–cage pairs that do not cross periodic boundaries. This yielded samples of 32,000 $k_{C\rightarrow V}$

and 32,000 $k_{V \rightarrow C}$ hopping rates when performed over 1000 snapshots. Independent rate samples must be calculated for each adsorbate at each temperature.

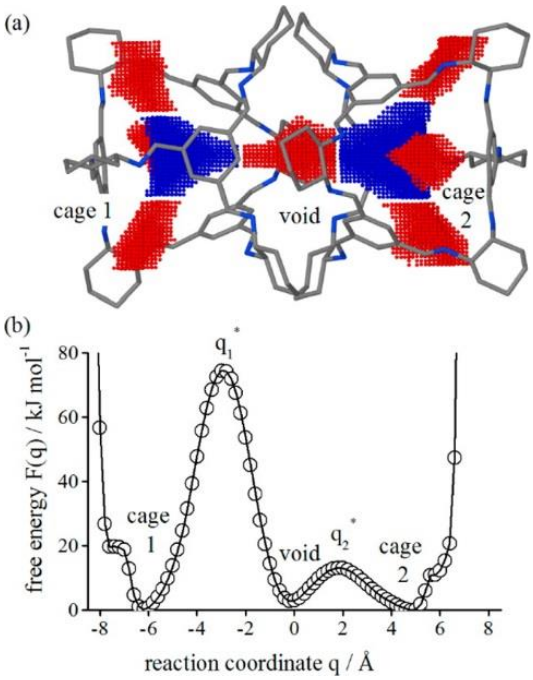

Figure 4.5 (a) Microstates revealed by the Fernand–Meyer algorithm in a snapshot from flexible NVT dynamics at 300 K. (b) Corresponding 1-dimensional free energy profile for Xe hopping.

Figure 4.5a depicts the cage and void microstates in a snapshot collected from NVT dynamics at 300 K using CSFF. The void microstate spanning cage 1 on the left and cage 2 on the right is deformed with respect to the symmetrical void microstate in the rigid experimental structure shown in Figure 4.3a. The cage 1 window coincident with the dividing surface between cage 1 and the void is 3.08 Å in diameter in this snapshot. This leads to a wide dividing surface between cage 1 and the void relative to the CSFF experimental structure, which has windows of 3.62 Å in diameter. In the corresponding 1-dimensional free energy profile in Figure 4.5b, the free energy barrier for Xe hops from cage 1 is 74 kJ mol^{-1} compared to 23 kJ mol^{-1} in the experimental structure. In contrast, the window between cage 2 and the void is 3.92 Å in diameter, leading to a lower free energy barrier (13 kJ mol^{-1}) for hops between cage 2 and the void. The difference in free energy barriers gives a $k_{\text{cage 1}\rightarrow\text{void}}$ hopping rate 10 orders of magnitude slower than k_{cage}

$_{2\rightarrow void}$ in the CC3 framework conformation represented by this snapshot. This enormous difference in rates hints at the important role of framework flexibility in determining the overall diffusivity of species like Xe in CC3.

The overall flexible hopping rates $k_{C\rightarrow V}$ and $k_{V\rightarrow C}$ can be estimated from the sample average of the 32,000 independent rates collected for each adsorbate and temperature condition. This approach can lead to numerical uncertainty, however, when a few exceptionally high rate observations have a dominant influence on the overall hopping rate. For example, in our sample of SF_6 $k_{C\rightarrow V}$ rates collected at 200 K, the two highest observations (out of 32,000) contributed over 80% to the numerator of the sample average. This high sample variance leads to unacceptably large numerical uncertainty in the sample average. We addressed this problem by inferring the structure of the distribution of hopping rates from each hopping rate sample as described below.

Histograms of the various $k_{C\rightarrow V}$ and $k_{V\rightarrow C}$ hopping rate samples collected were not related to an obvious statistical distribution. We investigated further by calculating the activation energies, E_A, underlying each hopping rate in our $k_{C\rightarrow V}$ and $k_{V\rightarrow C}$ samples with eq 5 in the same way as in the rigid experimental structure. In this case, T_1 is the physically meaningful temperature associated with the NVT snapshot collection while T_2 is an arbitrary second reference temperature. The Arrhenius prefactor, A, underlying each rate was computed in a similar way. For example, in the snapshot shown in Figure 4.5, the Xe $k_{cage\ 2\rightarrow void}$ hopping rate is 2.51×10^9 s^{-1} at 300 K, which can be decomposed into a hopping activation energy of 8.28 kJ mol^{-1} and an Arrhenius prefactor of 6.94×10^{10} s^{-1}. This activation energy is significantly lower than in the experimental structure (17.3 kJ mol^{-1}) because of the favorable window geometry between cage 2 and the void in this snapshot.

The resulting histograms of 32,000 hopping activation energies underlying each sample of 32,000 $k_{C\rightarrow V}$ and $k_{V\rightarrow C}$ rates are well described by log-normal distributions. Figure 4.6 shows the distribution of Xe $k_{C\rightarrow V}$ hopping activation energies at 300 K and the associated fit to a log-normal distribution computed with the MATLAB maximum likelihood estimation function.[48] Similar distributions were fit to each of the $k_{C\rightarrow V}$ and $k_{V\rightarrow C}$ rate samples collected for each adsorbate at different temperature conditions. The log-normal location parameter μ and scale parameter σ for each rate sample with associated 95% confidence intervals are given in Appendix B.

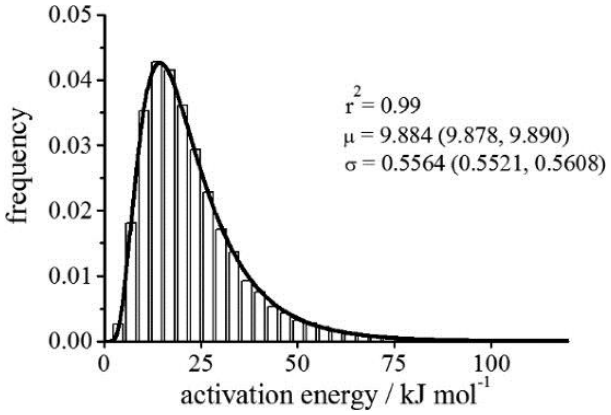

Figure 4.6 Distribution in activation energies for krypton cage to void hops at 300 K with associated fit to a log-normal distribution. Quantities in parentheses indicate 95% confidence intervals.

With the parameters μ and σ for a given rate sample, the expectation value of the overall fully flexible hopping rate $k_{C\rightarrow V}$ or $k_{V\rightarrow C}$ was found by integrating the Arrhenius rate equation over the respective log-normal probability density function:

$$k = \bar{A}\int_0^\infty \frac{1}{x\sqrt{2\pi\sigma}}\exp\left(\frac{-(\ln x-\mu)^2}{2\sigma^2}\right)\exp\left(\frac{-x}{RT}\right)dx \qquad (6)$$

Here, $\bar{A}$ is the Arrhenius prefactor averaged over the 32,000 individual Arrhenius prefactors calculated for each rate sample. This method neglects the effect of covariance between A and the hopping activation energy, which was found to be small (Appendix B). Uncertainties in the overall flexible hopping rates were estimated by recomputing $k_{C\rightarrow V}$ or $k_{V\rightarrow C}$ at the low and high values of μ and σ from the 95% confidence intervals.

When this fitting procedure is applied to fast moving adsorbates such as Kr, the overall flexible hopping rate from eq 6 ($k_{C\rightarrow V} = (6.28 \times 10^9) \pm (0.04 \times 10^9)$ s^{-1} at 300 K) is close to the sample average of 32,000 $k_{C\rightarrow V}$ rates ($(7.33 \times 10^9) \pm (0.25 \times 10^9)$ s^{-1}) . For larger adsorbates, flexible hopping rates from eq 6 have much lower numerical uncertainty than given by simple sample averages. For example, the activation energies for SF_6 cage to void hopping at 300 K were fit to a log-normal distribution with μ = 11.89, σ = 0.4742, and $A = 2.12 \times 10^{11}$ s^{-1}. Here, μ and σ are normalized to a base unit of 1 J mol^{-1} such that e^{μ} yields the most probable diffusion activation energy in units of J mol^{-1}. The expectation value of the $k_{C\rightarrow V}$ hopping rate for this activation energy distribution given by eq 6 is 6500 s^{-1}. The population variance σ^2_k of the rate distribution calculated from the cumulative

distribution function of $k_{C\rightarrow V}$ is 1.662×10^{12} s^{-2} (Appendix B).[49] By the central limit theorem, the sample averages of samples of 32,000 rates randomly drawn from this rate distribution would have a standard deviation of 7200 s^{-1}. Therefore, the means of many such rate samples would be deviate from the expectation value by over an order of magnitude. Using our maximum likelihood estimates of the 95% confidence intervals on μ and σ, we calculate an uncertainty of 6500 ± 500 on the value of the SF_6 cage to void hopping rate at 300 K. This is a significant improvement over sample averages.

4.4.3 Flexible Umbrella Sampling TST

In the implicit ligand sampling method described above, the dynamical problem of rate constant calculations in the fully flexible CC3 structure is reduced to a series of static rate constant calculations in framework conformations sampled from fully flexible dynamics. This method makes the assumption that interactions between adsorbates and the CC3 framework are negligible. This precludes a cooperative diffusion mechanism in which the CC3 window dynamics are influenced by the presence of adsorbates. This assumption may be invalid for adsorbate molecules significantly larger than the size of the CC3 window in the experimental structure.

To overcome this limitation of the implicit ligand sampling method, we must calculate a free energy profile, *F(q)*, that takes into account adsorbate–framework interactions in the fully flexible CC3 structure. This free energy profile could in principle be calculated by histogramming the position of an adsorbate over a very long fully flexible MD simulation to yield the adsorbate probability density function along the reaction coordinate *P(q)*, which is directly proportional to *F(q)*.[50] However, since the adsorbate rarely visits the high energy regions near the transition state, this method would be no more computationally efficient than computing diffusivities with straightforward MD.

Umbrella sampling is a technique to improve computational sampling of systems with high free energy barriers by means of (typically harmonic) restraint forces.[51] At intervals along the reaction coordinate known as “umbrellas”, these spring forces constrain the motion of the adsorbate (or other component of interest) to a region near the spring center. In each umbrella, an independent molecular dynamics simulation is performed to histogram adsorbate positions under the influence of the spring. The unbiased probability

distribution is then reproduced from the overlapping histograms from each window, yielding the free energy along *q*. Umbrella sampling has been used, for example, to study benzene diffusion in rigid[4] and flexible[18] zeolites.

We implemented umbrella sampling using the LAMMPS collective variables library.[52] To reconstruct the free energy profile *F(q)*, we histogrammed the adsorbate position over 24 umbrellas distributed over evenly spaced intervals from q = −6.4 to q = 6.4 Å. The adsorbate atom was confined to each umbrella by a 5 kcal mol^{-1} $Å^{-2}$ harmonic restraint acting along *q*. In each umbrella, the adsorbate atom was placed at the harmonic restraint center, velocities were initialized from the Maxwell–Boltzmann distribution, and a 10 ps fully flexible NVT equilibration was run. Following equilibration, the positions of the adsorbate along *q* within each umbrella was recorded at 0.5 fs intervals over a 100 ps NVT production period.

Figure 4.7a depicts the 24 histograms of Xe positions used to determine *F(q)* for Xe hopping at 300 K. The free energy profile reconstructed was reconstructed from these histograms with the Grossfield et al. Weighted Histogram Analysis Method (WHAM) code.[53-54] The free energy profile for Xe hopping at 300 K is shown in Figure 4.7b. At 300 K, the free energy barrier for Xe cage to void hops (16 kJ mol^{-1}) is lower than in the experimental structure (23 kJ mol^{-1}, shown in Figure 4.3b). This is not surprising; it reflects the observation that hopping is dominated by framework conformations where the hopping barrier is considerably lower than the barrier in the rigid structure. The overall fully flexible hopping rates $k_{C\rightarrow V}$ and $k_{V\rightarrow C}$ from umbrella sampling were calculated with the 1-dimensional TST (eq 4).

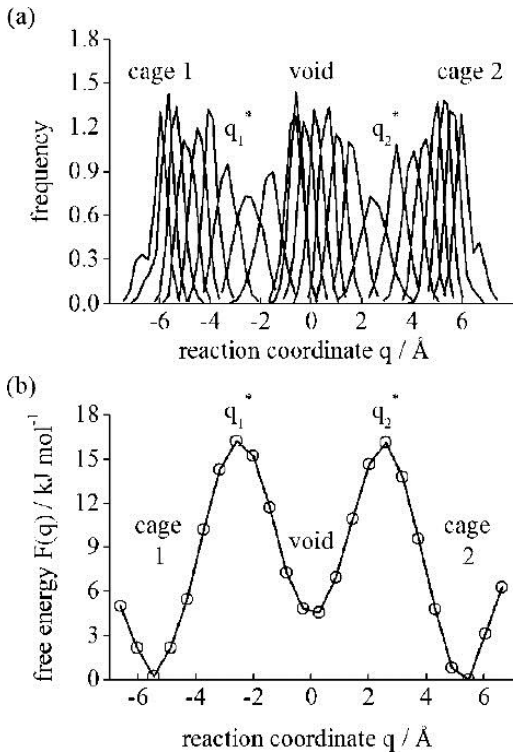

Figure 4.7 (a) Frequency histograms (umbrellas) for a Kr atom confined to 24 windows along the reaction coordinate by harmonic potentials at 300 K.
(b) Reconstructed Kr free energy curve *F(q)*.

TST calculations on polyatomic adsorbates are expensive using the implicit ligand sampling method because many rotational orientations must be integrated over at each grid voxel. In umbrella sampling, these rotational degrees of freedom are efficiently sampled by molecular dynamics. We used umbrella sampling to determine the free energy profile *F(q)* for CO_2 diffusion in CC3 (Figure 4.8). In this case, we found that CO_2 has a qualitatively different diffusion mechanism than spherical adsorbates. The free energy maxima are not located at the windows but significantly closer to the midpoint between the adjacent cages. A similar energy profile is seen for CO_2 diffusion in 8-member ring zeolites. The shallow energy minima at the void results in a high void to cage hopping rate $k_{V \rightarrow C}$.

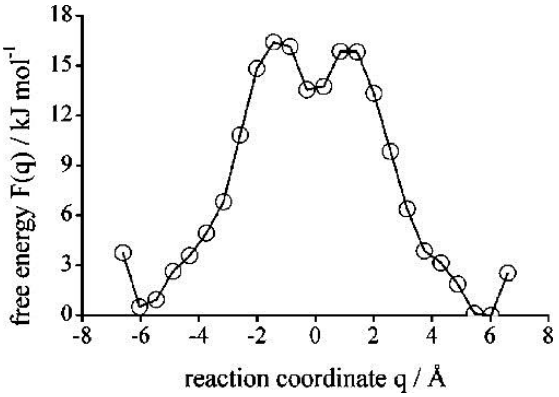

Figure 4.8 Free energy profile *F(q)* for CO_2 diffusion at 300 K by umbrella sampling.

4.5 Results and Discussion

4.5.1 Infinite Dilution Diffusion Coefficients

The three 1-dimensional TST methods we used make different simplifying physical assumptions that reduce computational costs relative to straightforward MD. The simplest method we applied is 1-dimensional TST in the rigid experimental structure. The most physically realistic method we used is umbrella sampling of the fully flexible structure, since this approach considers all degrees of freedom relevant for diffusing molecules. Table 4.1 shows D_s from the two fully flexible TST methods at 300 K and the ratio of the flexible umbrella sampling TST D_s to D_s in the rigid experimental structure. The smallest spherical adsorbates we considered, Kr and CH_4, have flexible umbrella sampling (US) and ILS TST self-diffusivities only a factor of about 2 higher in the flexible CC3 structure versus the rigid CC3 structure at 300 K. In the cases of Xe and Rn, consideration of framework flexibility results in over an order of magnitude faster diffusion than in the rigid experimental structure. CS_2 and SF_6 are essentially immobile in the CSFF rigid experimental structure but diffuse readily in the fully flexible CC3 structure. These results show that considering framework flexibility can be crucial for capturing even the qualitative nature of adsorbate diffusion in CC3.

Table 4.1 D_s for adsorbates in CC3 at 300 K computed by umbrella sampling (US), implicit ligand sampling (ILS), and by TST in the rigid experimental structure.

adsorbate (CAFF)	flexible US D_s at 300 K (cm^2/s)	flexible ILS D_s at 300 K (cm^2/s)	ratio of flexible US D_s to rigid exp. D_s
Kr	1.81×10^{-5}	1.49×10^{-5}	1.7
CH_4	5.08×10^{-5}	3.20×10^{-5}	2.3
Xe	2.87×10^{-6}	1.38×10^{-6}	26
Rn	1.12×10^{-6}	4.03×10^{-7}	170
CS_2	6.02×10^{-8}	1.47×10^{-8}	1.6×10^{6}
SF_6	3.49×10^{-10}	2.53×10^{-11}	10^{16}

Straightforward MD is the most rigorous description of diffusion for a given classical force field because no mechanistic assumptions are made about adsorbate diffusion dynamics. To test the quantitative accuracy of ILS and US TST, we compared each method to MD at 300 K (Figure 4.9). For each adsorbate at 300 K shown in Figure 4.9, the numerical uncertainties in the MD and TST self-diffusivities are on the order of the symbol size.

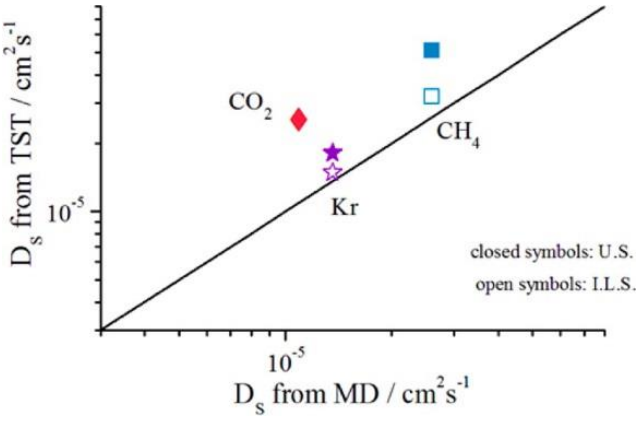

Figure 4.9 Comparison between TST diffusion coefficients from umbrella sampling (filled symbols) and implicit ligand sampling (empty symbols) and straightforward MD at 300 K.

Umbrella sampling and implicit ligand sampling TST overpredict the diffusion coefficients of Kr, CH_4, and CO_2 by factors ranging from 1.2 to 2.4. This systematic

overprediction of Ds makes sense because TST hopping rates uncorrected by a Bennett–Chandler transmission coefficient, κ, should typically be greater than or equal to the true value eq 4. The ILS TST measurements of D_s are consistent with transmission coefficients (κ) of approximately 0.85 for the overall $k_{C\rightarrow V}$ and $k_{V\rightarrow C}$ flexible hopping rates. The higher US TST D_s values are consistent with transmission coefficients (κ) of approximately 0.6, which are reasonable for flexible simulations where the precise location of the dividing surface is expected to vary during the simulation. Overall, both TST methods give reasonable approximations of these fast moving adsorbates.

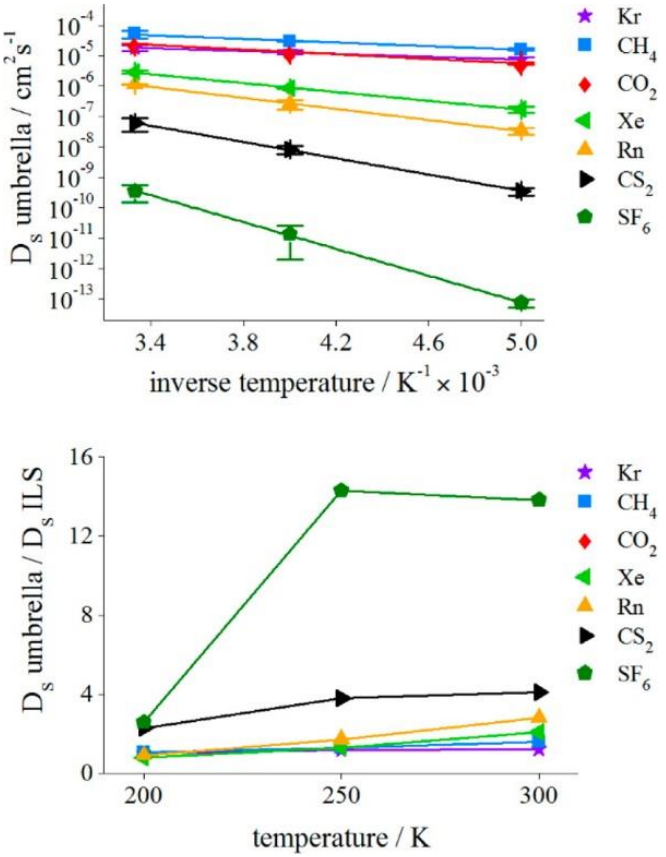

Figure 4.10 (a) Self-diffusion coefficients by umbrella sampling. (b) Ratio of US TST D_s to ILS TST D_s.

Figure 10a shows D_s from US TST for each adsorbate as a function of inverse temperature. Each set of data in CC3 is well described by an Arrhenius equation. Figure 4.10b gives the ratio of D_s from US TST to ILS TST. For adsorbates other than CS_2 and SF_6, there is only an insignificant difference between the US TST and ILS TST measurements. In each case the US TST measurement is higher, but this could be

attributable to a lower Bennett–Chandler correction in the US TST method. For CS_2 and SF_6, umbrella sampling TST gives self-diffusivities significantly higher than ILS TST across the temperatures considered. This discrepancy may be due to the ILS TST assumption that adsorbates do not influence the motion of the CC3 framework. This precludes a concerted mechanism in which CC3 windows expand in the presence of nearby adsorbate molecules, lowering the E_A for adsorbate hopping. TST based on umbrella sampling does not make any assumptions about decoupling between adsorbate and framework motions. By comparing the US derived D_s (which accounts for adsorbate–framework interactions) to the ILS derived D_s (which assumes framework motions are independent of adsorbates), we can quantify the importance of these interactions on diffusion. For SF_6, this factor contributes to up to an order of magnitude difference between US TST and ILS TST derived self-diffusivities. Adsorbate–framework interactions could be more significant for larger polyatomic adsorbates such as aromatics and in materials more flexible than CC3.

4.5.2 Influence of CC3 Window Size on Diffusion

Analysis of individual framework conformations from our ILS TST calculations (Figure 4.5) suggests that adsorbate hopping rates are a strong function of CC3 window size but do not reveal the functional form of the relationship. To investigate this issue, we plotted a large number of Xe $E_{A,C\rightarrow V}$ values calculated with the implicit ligand sampling TST method described above as a function of the corresponding CC3 window sizes (Figure 4.11).

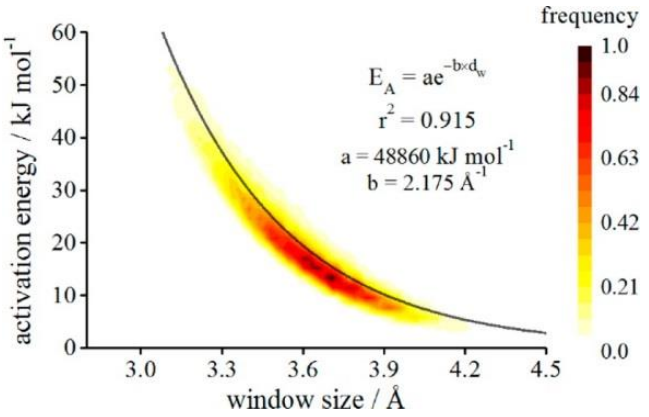

Figure 4.11 E_A for cage to void Xe hops at 300 K as a function of the CC3 windows size. The black curve shows the associated fit to an exponential function.

Figure 11 shows a scatter plot of Xe cage to void hopping activation energies ($E_{A,C\rightarrow V}$) as a function of the CC3 window size. The relationship between E_A and window size is well described by the exponential dependence suggested by Haldoupis et al.[6] The associated fit to $E_A = ae^{-bdw}$, where d_w is the window diameter and a and b are fitted parameters, is show in Figure 4.11. This exponential dependence on the CC3 window sizes (which are normally distributed) explains the log-normal structure of the activation energy distributions collected during our ILS TST calculations (Figure 4.6). As a consistency check, we used the normal distribution fit parameters for CC3 window sizes at 300 K (Figure 4.2 inset) and the exponential fit parameters for the dependence of Xe $E_{A,C\rightarrow V}$ values on window size (Figure 4.11) to reproduce the log-normal μ and σ parameters shown in Figure 4.4. This yields $\mu = 9.89$ and $\sigma = 0.533$, which are within the uncertainty of the direct log-normal fit to the distribution in Xe $E_{A,C\rightarrow V}$ observations from implicit ligand sampling at 300 K.

Comparisons of adsorbate size to pore geometry can often predict the qualitative nature of diffusion through nanoporous materials. In CC3, the arene carbon window size is a reasonable approximation to the pore limiting diameter (PLD) found by Voronoi decomposition with Pore Blazer[55] or Zeo++[27, 56]. Adsorbates larger than the PLD generally diffuse slowly through the tightest constrictions in rigid nanoporous materials. For example, in our TST calculations in the rigid experimental CC3 structure, adsorbates with Lennard–Jones σ parameters (such as Rn, where $\sigma = 4.17$ Å) significantly larger than the experimental window size (3.62 Å) have very slow $k_{C\rightarrow V}$ hopping rates.

In more physically realistic fully flexible CC3 models, a single PLD value calculated from the experimental atomic coordinates does not describe the fluctuating nature of the window that restrict adsorbate diffusions. A more descriptive metric is the "pore limiting envelope", which is the distribution in pore limiting diameters observed in flexible pores similar to the window size distribution. Holden et al. calculated the pore limiting envelope in the flexible CC3 structure with Zeo++.[32] From this distribution, they calculated the percentage of flexible CC3 configurations with a transient PLD that exceeds the size of several light gases. Of the CC3 configurations sampled at 300 K, 58.7% have PLDs larger than the van der Waal diameter of Kr, while only 7.3% of configurations have PLDs which exceed the size of Xe.

Using the pore limiting envelope analysis, the mechanism of adsorbate diffusion is transient pore connectivity which allows for opportunistic gas percolation. Holden et al. found that these transient pore connections are common for adsorbates comparable to the size experimental pore limiting diameter (such as Kr) but infrequent for larger adsorbates (such as Xe and Rn).[31, 41] The pore limiting envelope provides a simple mechanism for the phenomena of "porosity without pores" observed for Xe and Rn diffusion in CC3. However, the quantitative predictions of this approach are strongly dependent on what techniques are used to calculate PLDs and adsorbate radii. This model is not consistent with the treatment of Xe and Rn as soft Lennard-Jones particles rather than hard spheres in MD simulations.

Our results suggest a technique for interpreting our ILS TST calculations which explains opportunistic activated hopping of adsorbate molecules in CC3 without presuming the binary, "open–shut" mechanism characteristic of the pore limiting envelope construct. This analysis quantifies the importance of rare, highly favorable framework configurations to adsorbate diffusion. In our implicit ligand sampling TST calculations described above, the distributions in hopping activation energies $E_{A,C\rightarrow V}$ and $E_{A,V\rightarrow C}$ were fit to log-normal distributions as shown in Appendix B. From these distributions, the expectation values of the overall fully flexible hopping rates were calculated with eq 6. In eq 6, the interval [0, ∞] is integrated over to incorporate contributions from the entire distribution of activation energies to the overall hopping rate $k_{C\rightarrow V}$. If we instead integrate over the finite interval [0, E_A], we find the partial expected value of the overall hopping rate from activation energies

below E_A. The quotient of the overall hopping rate $k_{C\rightarrow V}$ (from eq 6), and this partial expectation gives *f*, the fractional contribution of activation energies below some threshold E_A to the overall rate:

$$f(E_A) = \frac{\bar{A}}{k_{C\rightarrow V}} \int_0^{E_A} \frac{1}{x\sqrt{2\pi\sigma}} \exp\left(\frac{-(\ln(x)-\mu)^2}{2\sigma^2}\right) \exp\left(\frac{-x}{RT}\right) dx \qquad (7)$$

The value of f asymptotically approaches 1 as the E_A approaches ∞. Figure 4.12 shows *f* plotted for the distributions of $E_{A,C\rightarrow V}$ collected at 300 K by implicit ligand sampling. The x-axis of Figure 4.12 is transformed to reflect the standardized scores of $\ln(E_{A,C\rightarrow V})$ in the respective log-normal ILS TST activation energy distribution.

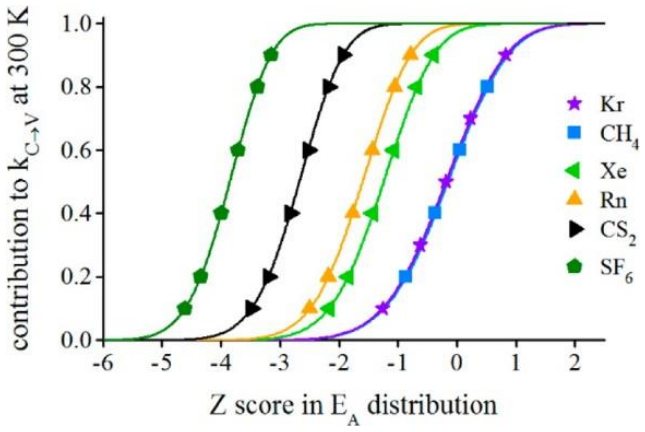

Figure 4.12 Fractional contribution f to the overall ILS TST derived hopping rate $k_{C\rightarrow V}$ for adsorbate hopping at 300 K.

Figure 4.12 demonstrates that as adsorbate size increases from Kr to SF_6, the left tail of the ILS TST activation energy distribution becomes increasingly important to the expected value of the overall hopping rate. For example, about 80% of the overall implicit ligand sampling TST derived $k_{C\rightarrow V}$ for Xe and Rn comes from values of $E_{A,C\rightarrow V}$ more than 1 standard deviation below the mean (Z score < −1) in the respective ILS TST activation energy distribution. A limitation of this analysis is that it neglects the influence of adsorbate–framework interactions on diffusion. We have shown (Figure 4.10) that these interactions may have a considerable influence on the TST derived diffusion coefficients in CC3 for larger adsorbates such as CS_2 and SF_6. Nonetheless, this analysis is a useful to way qualitatively understand the connection between adsorbate size and hopping rate.

4.5.3 Comparison to Published Results

Table 4.2 shows a comparison between our US TST derived values of D_s and MD derived D_s values from Holden[41] and Evans et al.[28] This work and Holden et al. measured these diffusion coefficients at infinite dilution using CSFF and CAFF parameters. Evans et al. measured the diffusion coefficients at 10 bar adsorbate loading and 298 K using universal force field (UFF) parameters to describe all CC3 degrees of freedom.[36]

Table 4.2 Comparison to room temperature diffusion coefficients from Holden et al.[41] and Evans et al.[28]

Gas	Temp. (K)	D_s (cm^2/s)	Reference
CH_4 UA	300	5.08×10^{-5}	this work
CH_4 UA	298	2.1×10^{-6}	Evans et al.
CH_4	300	9.6×10^{-5}	Holden et al.
CO_2	300	2.49×10^{-5}	this work
CO_2	298	3.45×10^{-6}	Evans et al.
CO_2	298	2.6×10^{-5}	Holden et al.
Kr	300	1.81×10^{-5}	this work
Kr	298	2.4×10^{-5}	Holden et al.
Xe	300	2.87×10^{-6}	this work
Xe	298	1.8×10^{-6}	Holden et al.

Our data is in reasonable agreement with Holden et al, despite differences in the implementation of the adsorbate – framework interaction potentials for CO_2 and CH_4. Our results do not agree with the Evans et al. MD results. This discrepancy arises from Evan's use of the Universal Force Field to model framework flexibility in CC3, which causes a significant difference in the CC3 window size distribution.[28] Given the exponential dependence of CC3 diffusion E_A on window size (Figure 4.11), relatively small differences in window size distributions can cause large differences in measured diffusivities.

4.5.4 Computational Cost of Methods

CSFF was implemented in the LAMMPS GPU accelerated molecular dynamics package, which accelerates molecular dynamics by a factor of 3–4 relative to CPU only implementations for systems the size of CC3.[57-59] Each of our umbrella sampling simulations for a particular adsorbate and temperature condition consisted of 110 ps of dynamics in each of 24 independent umbrellas. Therefore, less than 3 ns of molecular

dynamics was needed to measure the diffusivity of SF_6 at 300 K in this work with US TST ($Ds = 3.5 \times 10^{-10}$ cm^2 s^{-1}). In contrast, about 1 μs would need to be run to measure diffusivities on this order of magnitude with straightforward MD. The computational cost of implicit ligand sampling is even lower than umbrella sampling, but we have shown that this method may be inaccurate for larger, slow moving species such as CS_2 and SF_6 where adsorbate–framework interactions could have a significant effect on diffusion. The computational advantages of US TST will become even more marked for slower diffusing species.

4.6 Conclusions

In this work, we have detailed two transition state theory (TST) methods for measuring infinite dilution diffusion in flexible nanoporous materials at time scales inaccessible to straightforward molecular dynamics (MD). These methods were applied to measure the diffusion of light gases in cage crystal 3 (CC3),[23] a promising porous organic cage materials. Simulations[28, 41] and experiments[26] have confirmed that framework flexibility strongly influences adsorbate transport in CC3. We have described the flexibility of CC3 using a variant of the cage-specific force field (CSFF)[37] that does not include intramolecular nonbonded interactions.

When using these TST methods, diffusion in CC3 is modeled as series of uncorrelated hops between 4-coordinated cavity sites (denoted C) inside cage molecules and 2-coordinate void sites (denoted V) between adjacent cages. To calculate the hopping rates $k_{C\rightarrow V}$ and $k_{V\rightarrow C}$ between these microstates, we applied a variant of the implicit ligand sampling (ILS) method first used to study CH_4 diffusion in zeolites[46] and ZIFs[60]. In ILS TST, hopping rates are calculated in an ensemble of rigid framework snapshots captured from a fully flexible MD trajectory of the empty CC3 structure. We have introduced a maximum likelihood estimation approach to calculate the overall flexible rates $k_{C\rightarrow V}$ and $k_{V\rightarrow C}$ from ILS TST that significantly reduces numerical uncertainty relative to simple ensemble averaging. The ILS method assumes the influence of adsorbate–framework interactions on diffusion in CC3 is negligible. An advantage of this method is that in principle it can be used for porous materials for which no reliable force field exists by obtaining snapshots from *ab initio* MD.[6]

The second TST method we applied is umbrella sampling (US). In US TST, the free energy profile is reconstructed from a series of biased MD simulations at intervals along the reaction coordinate known as "umbrellas". The umbrella sampling TST method fully accounts for the influence of cooperative interactions between adsorbates and the CC3 framework on diffusion. Umbrella sampling is readily extendable to polyatomic adsorbates such as CO_2, which was found to have a qualitatively different free energy profile than spherical adsorbates in CC3.

The TST methods described here are orders of magnitude more computationally efficient than straightforward MD for simulating the diffusion of large, slow-moving adsorbates in CC3 and other nanoporous materials. For simulations of larger adsorbates such as aromatic compounds that span the window of CC3, it will likely be important to implement the Bennett–Chandler dynamic correction to account for adsorbate recrossing at the transition state.[61] With the dynamic correction factor, we anticipate that this method could be readily applied to more complex adsorbates known to adsorb in CC3[26-27] and TST simulations of diffusion at finite loadings.

4.7 References

[1] J. Kärger, D. M. Ruthven, D. N. Theodorou, *Diffusion in Nanoporous Materials*, Wiley, **2012**.

[2] S. Keskin, J. Liu, R. B. Rankin, J. K. Johnson, D. S. Sholl, *Industrial & Engineering Chemistry Research* **2008**, *48*, 2355-2371.

[3] R. L. June, A. T. Bell, D. N. Theodorou, *The Journal of Physical Chemistry* **1990**, *94*, 8232-8240.

[4] F. Jousse, S. M. Auerbach, D. P. Vercauteren, *The Journal of Physical Chemistry B* **2000**, *104*, 2360-2370.

[5] D. C. Ford, D. Dubbeldam, R. Q. Snurr, *Diffusion Fundamentals III* **2009**, 459.

[6] E. Haldoupis, T. Watanabe, S. Nair, D. S. Sholl, *ChemPhysChem* **2012**, *13*, 3449-3452.

[7] C. Dellago, P. G. Bolhuis, P. L. Geissler, in *Advances in Chemical Physics*, John Wiley & Sons, Inc., **2003**, pp. 1-78.

[8] H. Eyring, *The Journal of Chemical Physics* **1935**, *3*, 107.

[9] D. Moroni, PhD thesis, Universiteit van Amsterdam **2005**.

[10] D. Dubbeldam, E. Beerdsen, T. J. Vlugt, B. Smit, *Journal of Chemical Physics* **2005**, *122*, 224712.

[11] B. Widom, *The Journal of Chemical Physics* **1963**, *39*, 2808.

[12] G. M. Torrie, J. P. Valleau, *Journal of Computational Physics* **1977**, *23*, 187-199.

[13] J. G. Kirkwood, *The Journal of Chemical Physics* **1935**, *3*, 300.

[14] E. Haldoupis, S. Nair, D. S. Sholl, *Journal of the American Chemical Society* **2010**, *132*, 7528-7539.

[15] G. Henkelman, B. P. Uberuaga, H. Jonsson, *The Journal of Chemical Physics* **2000**, *113*, 9901-9904.

[16] A. Schüring, S. M. Auerbach, S. Fritzsche, R. Haberlandt, *The Journal of Chemical Physics* **2002**, *116*, 10890-10894.

[17] M. K. Abouelnasr, B. Smit, *Physical Chemistry Chemical Physics* **2012**, *14*, 11600-11609.

[18] P. D. Kolokathis, E. Pantatosaki, C.-A. Gatsiou, H. Jobic, G. K. Papadopoulos, D. N. Theodorou, *Molecular Simulation* **2014**, *40*, 80-100.

[19] S. E. Boulfelfel, P. I. Ravikovitch, D. S. Sholl, *Journal of Physical Chemistry C* **2015**, *119*, 15643-15653.

[20] J. R. Holst, A. Trewin, A. I. Cooper, *Nature Chemistry* **2010**, *2*, 915-920.

[21] Y. Jin, Y. Zhu, W. Zhang, *CrystEngComm* **2013**, *15*, 1484.

[22] C. Trolliet, G. Poulet, A. Tuel, J. D. Wuest, P. Sautet, *Journal of the American Chemical Society* **2007**, *129*, 3621-3626.

[23] T. Tozawa, J. T. A. Jones, S. I. Swamy, S. Jiang, D. J. Adams, S. Shakespeare, R. Clowes, D. Bradshaw, T. Hasell, S. Y. Chong, C. Tang, S. Thompson, J. Parker, A. Trewin, J. Bacsa, A. M. Z. Slawin, A. Steiner, A. I. Cooper, *Nature Materials* **2009**, *8*, 973-978.

[24] T. Hasell, S. Y. Chong, K. E. Jelfs, D. J. Adams, A. I. Cooper, *Journal of the American Chemical Society* **2012**, *134*, 588-598.

[25] T. Hasell, M. Schmidtmann, C. A. Stone, M. W. Smith, A. I. Cooper, *Chemical Communications* **2012**, *48*, 4689-4691.

[26] T. Mitra, K. E. Jelfs, M. Schmidtmann, A. Ahmed, S. Y. Chong, D. J. Adams, A. I. Cooper, *Nature Chemistry* **2013**, *5*, 276-281.

[27] L. Chen, P. S. Reiss, S. Y. Chong, D. Holden, K. E. Jelfs, T. Hasell, M. A. Little, A. Kewley, M. E. Briggs, A. Stephenson, K. M. Thomas, J. A. Armstrong, J. Bell, J. Busto, R. Noel, J. Liu, D. M. Strachan, P. K. Thallapally, A. I. Cooper, *Nature Materials* **2014**, *13*, 954-960.

[28] J. D. Evans, D. M. Huang, M. R. Hill, C. J. Sumby, A. W. Thornton, C. J. Doonan, *The Journal of Physical Chemistry C* **2014**, *118*, 1523-1529.

[29] G. Chen, R. P. Lively, C. W. Jones, W. J. Koros, *Industrial & Engineering Chemistry Research* **2014**, *53*, 7113-7120.

[30] F. H. Allen, *Acta Crystallographica Section B: Structural Science* **2002**, *58*, 380-388.

[31] D. Holden, *Understanding the Diffusion of Small Gases in Porous Organic Cages using Molecular Dynamics*, University of Liverpool, **2013**.

[32] T. F. Willems, C. H. Rycroft, M. Kazi, J. C. Meza, M. Haranczyk, *Microporous and Mesoporous Materials* **2012**, *149*, 134-141.

[33] S. L. Mayo, B. D. Olafson, W. A. Goddard, *Journal of Physical Chemistry* **1990**, *94*, 8897-8909.

[34] R. A. Svehla, *NASA: Estimated Viscosities and Thermal Conductivities of Gases at High Temperatures* **1962**.

[35] J. G. Harris, K. H. Yung, *The Journal of Physical Chemistry* **1995**, *99*, 12021-12024.

[36] A. K. Rappé, C. J. Casewit, K. Colwell, W. Goddard Iii, W. Skiff, *Journal of the American Chemical Society* **1992**, *114*, 10024-10035.

[37] D. Holden, K. E. Jelfs, A. I. Cooper, A. Trewin, D. J. Willock, *Journal of Physical Chemistry C* **2012**, *116*, 16639-16651.

[38] H. Sun, *Macromolecules* **1995**, *28*, 701-712.

[39] E. Bitzek, P. Koskinen, F. Gähler, M. Moseler, P. Gumbsch, *Physical Review Letters* **2006**, *97*, 170201.

[40] C. Sholl, *Journal of Physics: Condensed Matter* **2005**, *17*, 1329.

[41] D. Holden, K. E. Jelfs, A. Trewin, D. J. Willock, M. Haranczyk, A. I. Cooper, *The Journal of Physical Chemistry C* **2014**, *118*, 12734-12743.

[42] K. E. Jelfs, A. I. Cooper, *Current Opinion in Solid State and Materials Science* **2013**, *17*, 19-30.

[43] F. Meyer, *Signal Processing* **1994**, *38*, 113-125.

[44] G. Henkelman, A. Arnaldsson, H. Jónsson, *Computational Materials Science* **2006**, *36*, 354-360.

[45] J. Cohen, K. W. Olsen, K. Schulten, *Methods in Enzymology* **2008**, *437*, 439-457.

[46] R. V. Awati, P. I. Ravikovitch, D. S. Sholl, *The Journal of Physical Chemistry C* **2013**, *117*, 13462-13473.

[47] J. T. Jones, D. Holden, T. Mitra, T. Hasell, D. J. Adams, K. E. Jelfs, A. Trewin, D. J. Willock, G. M. Day, J. Bacsa, A. Steiner, A. I. Cooper, *Angewandte Chemie International Edition* **2011**, *50*, 749-753.

[48] N. Buntao, S.-a. Niwitpong, V. Kreinovich, *Estimating Statistical Characteristics of Lognormal and Delta-Lognormal Distributions under Interval Uncertainty: Algorithms and Computational Complexity*, **2012**.

[49] M. Taboga, *Lectures on Probability Theory and Mathematical Statistics - 2nd Edition*, CreateSpace Independent Publishing Platform, **2012**.

[50] S. E. Jee, D. S. Sholl, *Journal of the American Chemical Society* **2009**, *131*, 7896-7904.

[51] B. Roux, *Computer Physics Communications* **1995**, *91*, 275-282.

[52] G. Fiorin, M. L. Klein, J. Hénin, *Molecular Physics* **2013**, *111*, 3345-3362.

[53] A. Grossfield, *WHAM: The Weighted Histogram Analysis Method, version 2.0.9* **2013**.

[54] S. Kumar, J. Rosenberg, D. Bouzida, R. Swendsen, P. Kollman, *Journal of Computational Chemistry* **1992**, *13*, 1011-1021.

[55] L. Sarkisov, *The Journal of Physical Chemistry C* **2012**, *116*, 3025-3033.

[56] M. Pinheiro, R. L. Martin, C. H. Rycroft, M. Haranczyk, *CrystEngComm* **2013**, *15*, 7531-7538.

[57] S. Plimpton, *Journal of Computational Physics* **1995**, *117*, 1-19.

[58] W. M. Brown, P. Wang, S. J. Plimpton, A. N. Tharrington, *Computer Physics Communications* **2011**, *182*, 898-911.

[59] W. M. Brown, A. Kohlmeyer, S. J. Plimpton, A. N. Tharrington, *Computer Physics Communications* **2012**, *183*, 449-459.

[60] E. Haldoupis, S. Nair, D. S. Sholl, *Journal of the American Chemical Society* **2012**, *134*, 4313-4323.

[61] D. Chandler, *The Journal of Chemical Physics* **1978**, *68*, 2959-2970.

5

DIFFUSION OF ARMOATICS IN CAGE CRYSTAL 3

5.1 Introduction

Separation of hydrocarbons of equal or similar molecular weight is an intrinsically difficult problem because of the comparable physiochemical properties, such as boiling points, of these compounds.[1] Separation of paraffins from olefins and xylene isomers from each other was first accomplished at industrial scales by energy-intensive cryogenic distillation.[2] Alternative processes based on adsorption in zeolites such as the UOP OLEX® and PAREX® simulated moving bed processes have since supplanted cryogenic distillation to a large degree.[2-4] These processes could potentially be improved by identification of novel adsorbents with enhanced selectivity for hydrocarbon components over currently used zeolites. For example, Gee et al. performed simulations and experiments to identify metal-organic frameworks that could be more selective for para-xylene than the barium exchanged zeolites currently used in the PAREX® process.[4]

Porous organic cages have shown similar promise in the separation of aromatic hydrocarbons in liquid phase breakthrough experiments.[5] Mitra et al. attributed the adsorptive selectivity of CC3 for 4-ethyltoluene over mesitylene to size exclusion of mesitylene from CC3 cages.[5] Porous organic cages have also shown experimental promise in the separation of rare gases[6], chiral molecules[6], and SF_6 from N_2[7]. Porous organic cages, as discrete molecules, have a practical advantage over extended nanoporous frameworks such as MOF because they can be dissolved in solution. This strategy was used to solution cast porous organic cages into composite membranes [8] and chromatographic stationary phases.[9]

Adsorbate diffusivities are the key predictor of the usefulness of a given adsorbent in separation processes. In the case of membrane based kinetic separations, diffusion coefficients predict which component will preferentially elute. In equilibrium adsorption, adsorbate diffusion coefficients are correlated with the amount of time necessary to complete a single adsorption cycle. In this chapter, we demonstrate the feasibility of

using the TST methods we described in Chapter 4[10] to measure characterize the diffusion of aromatic molecules in flexible nanoporous materials. These methods were applied to C_8 and C_9 aromatic diffusion in porous organic cage crystals CC3. We anticipate that these results can be extended to other aromatic molecules in flexible porous organic cages and metal-organic frameworks.

5.2 Methods and Computational Details

5.2.1 Cage Crystal Structures

The rhomohedral primitive unit cell of the homochiral CC3-R structure (CSD: PUDXES) was used throughout this work.[11] Each rhomohedral primitive cell (a = b = c = 17.536 Å, α = ß = γ = 60°) contains two CC3 molecules compared to eight CC3 molecules in the $F4_132$ symmetry cubic unit cell. All window sizes reported here were calculated by the arene carbon method described by Chen et al.[6] for consistency with window sizes reported in Chapter 4.

5.2.2 Ab-initio Molecular Dynamics

Ab-initio molecular dynamics simulations (AIMD) were performed in the cp2k Born-Oppenheimer MD package using the BLYP exchange correlation functional with the Grimme D3 dispersion correction.[12] The energy cutoff for calculation of the electronic density was set to 65 Ry. During AIMD, the lattice constants were kept at the experimentally derived values. A Nose-Hoover thermostat with a time constant of t = 100 fs was used to maintain temperature at 300 K. Window size distributions were derived from 5 independent 3 ps MD production periods following 1 ps NVT equilibrations.

5.2.3 Force Fields

Cage molecule flexibility was described using the all-atom Optimized Potentials for Liquid Simulations (OPLS) force field of Jorgenson et al.[13] OPLS uses harmonic bond potentials, harmonic angle potentials, and class2 trigonometric torsional potentials to describe bonded interactions. OPLS intramolecular and intermolecular Lennard-Jones interactions were truncated at 10 Å. Atomic point charges were derived from Holden et al. Cage Specific Force Field (CSFF).[14] Coulombic interactions were computed pairwise to

10 Å, and a long-range particle–particle mesh Ewald correction was used thereafter. Intramolecular nonbonded interactions between bonded atoms and second nearest neighbors were ignored, while 1-4 nonbonded interactions were scaled by a factor of 0.5. A full description of OPLS-AA functional forms and parameters is published at http://dasher.wustl.edu/ffe/distribution/params/oplsaa.prm.

Aromatic hydrocarbons (benzene, toluene, xylene isomers, and mesitylene) were described using rigid TraPPE force field united atom models with bond lengths of 1.4 Å and bond angles of 120° between each carbon within aromatic rings.[15] Each CH united atom in these molecules is described by a 12-6 Lennard Jones potential with ε/k_B = 50.5 K and σ = 3.695 Å. All methyl group united atoms (ε/k_B = 98.0 K and σ = 3.750 Å) are separated from aromatic C atoms (ε/k_B = 21.0 K and σ = 3.880 Å) by a bond length of 1.54 Å. All atomic point charges are equal to 0, so there were no electrostatic interactions between adsorbates or between adsorbates and framework atoms. Intermolecular interactions between adsorbates and between adsorbates and cage framework atoms were described by combining OPLS-AA and TraPPE Lennard-Jones parameters with the Lorentz—Berthelot mixing rules.

5.2.4 Classical MD Simulations

Classical molecular dynamics simulations were performed in DL_POLY4[16] to verify that the OPLS-AA force field as correctly implemented for production runs using LAMMPS[17]. Close agreement between the window size distributions at the experimental CC3 lattice constants demonstrated that these codes give equivalent cage framework dynamics.

In NVT umbrella sampling simulations, reaction coordinates were established between the centers of mass of adjacent cage molecules using the LAMMPS collective variable library as described in Chapter 4.[18] To reconstruct the free energy profile *F(q)*, we histogrammed the adsorbate position over 24 umbrellas distributed over evenly spaced intervals from q = −6.4 to q = 6.4 Å. The adsorbate atom was confined to each umbrella by a 5 kcal mol^{-1} $Å^{-2}$ harmonic restraint acting along *q*. Prior to initialization of dynamics, a damped dynamics algorithm was used to energy minimize the adsorbate and cage degrees of freedom. During minimization, adsorbate geometries were kept constant by the use of

very stiff (50,000 kcal/mol Å^2) harmonic bond, angle, and torsional potentials. After minimization, these harmonic restraints were removed and adsorbates were kept rigid using the LAMMPS fix/rigid/nvt/small command. All NVT umbrella sampling simulations were equilibrated for 10 ps followed by a 100 ps production period at 300 K.

During umbrella sampling, the orientation of adsorbate molecules with respect to the reaction coordinate was tracked using the "spinAngle" and "tilt" collective variables defined by the LAMMPS colvar library.[18-19] These collective variables are based on the optimal rotation matrix[19] that best superimposes current adsorbate coordinates onto the starting adsorbate coordinates. The "spinAngle" collective variable calculates the projection of the optimal rotation matrix orthogonal to reaction coordinate. The "tilt" collective variable calculates the projection of the optimal rotation matrix coincident to the reaction coordinate. The "spinAngle" variable is defind from -180:180° while the "tilt" variable is reported in terms of the cosine of the angle (ranging from -1:1) rather than the angle itself.[18]

The Bennett-Chandler transmission coefficients[20] were calculated using the procedure described by Anderson.[21] Briefly, initial configurations are collected by molecular dynamics by confining the adsorbate to the transition state by a very stiff harmonic potential (50,000 kcal/mol Å^2). NVT molecular dynamics is run for 0.5 ns to yield 1,000 adsorbate configurations at the transition state at 500 fs intervals. For each transition state configuration, forward and reverse trajectories (with respect the reaction coordinate) are initialized b removing the spring restraint. Forward trajectories are terminated upon either recrossing the dividing surface or by successfully thermalizing in the product state. Reverse trajectories are terminated upon either successfully returning to the reactant state or thermalizing in the product state. The transmission coefficient is the fraction of trajectories out of the 1,000 initially that meet the successful forward and reverse termination conditions.

5.3 Results

5.3.1 Force Field Validation

The simulations we described in Chapter 4 used the Cage Specific Force Field (CSFF), a bespoke potential parameterized specifically for CC3 by Holden et al.[14] Holden is a collaborator and co-author on this work in progress. Holden and coworkers have found that OPLS is comparably accurate to CSFF, but more broadly applicable to cage materials outside the CSFF training set. To validate our forcefield selection, we ran ab-initio molecular dynamics on the 336 atom CC3 rhombohedral primitive cell and compared these dynamics to classical MD using both the CSFF (Chapter 4) and OPLS force fields. Figure 5.1 shows the window size distributions at the experimental lattice constants at T = 300 K using these different potentials:

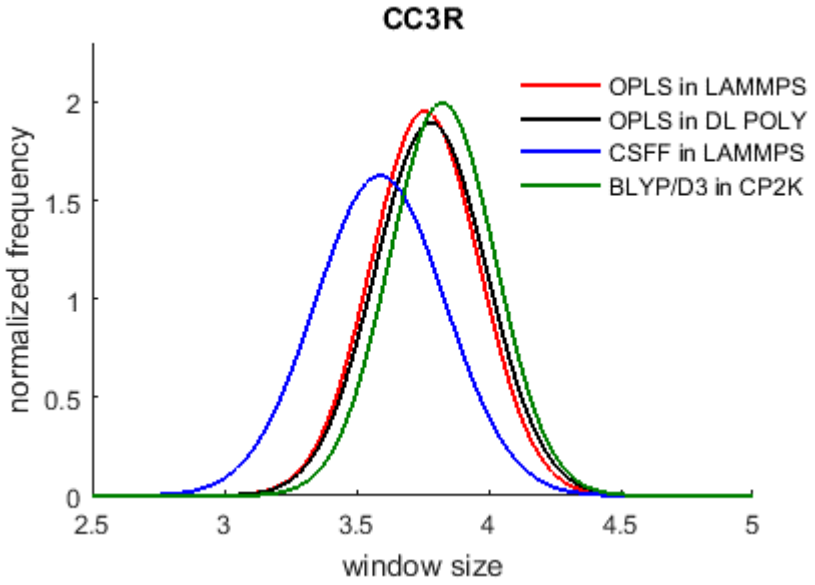

Figure 5.1 Window size distributions for CC3 at 300 K at the experimentally derived lattice constants

Figure 5.1 shows the window size distribution for CSFF is centered closer to the experimental window size of 3.62 Å. However, OPLS (either in LAMMPS or DL_POLY) is in better agreement with the results from AIMD. For this reason, we applied OPLS throughout this work.

5.3.2 Free Energy Profile for Benzene in CC3

Figure 5.2 shows F(q) derived from umbrella sampling and the associated umbrella histograms for benzene in CC3 at 300 K.

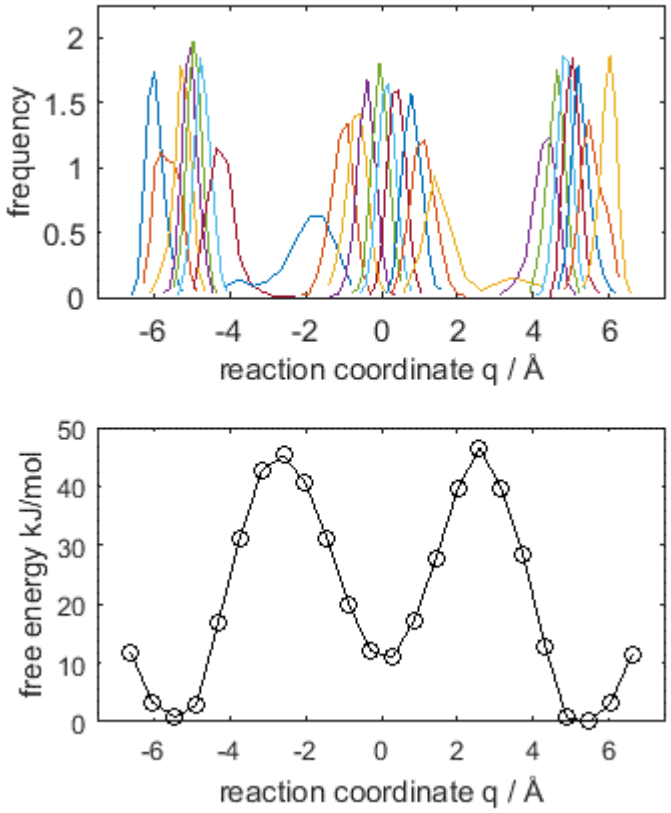

Figure 5.2 Umbrella sampling histograms and the reconstructed free energy profile F(q) for united atom benzene in CC3 at 300 K

Benzene has a free energy profile similar to large, slow moving spherical adsorbates such as Xe in CC3 (Chapter 4). Mechanistically, this means that benzene molecules diffuse by hopping from cage cavities to the voids between adjacent cage molecules. To evaluate the Bennett-Chandler dynamic correction for benzene, we calculated the transmission probability over 1000 independent trajectories initiated from the dividing surface. A total of 571 trajectories successfully thermalized in the product state without recrossing the dividing surface, yielding a correction factor of 0.571. Using this correction, we used 1-dimensional TST (eq 4.4) to find $k_{cage \rightarrow void} = 2.37 \times 10^4\ s^{-1}$ and $k_{void \rightarrow cage} = 1.49 \times 10^6\ s^{-1}$.

In Chapter 4, we derived a simple analytical model (eq 4.2) to relate these hopping rates to infinite dilution adsorbate diffusivities. A key assumption made when using this equation is that successive hops from cage cavities are uncorrelated. For this to be true, adsorbates must be able to reorient within cage cavities on timescale that are fast relative to the characteristic frequency of activated hops. To test whether this is true for benzene, we tracked the "tilt" and "spin" orientations of benzene relative to the reaction coordinate over each 100 ps umbrella sampling simulation. Figure 5.3 shows a box and whisker plot representation of the tilt collective variable over each of the 24 umbrella sampling simulations:

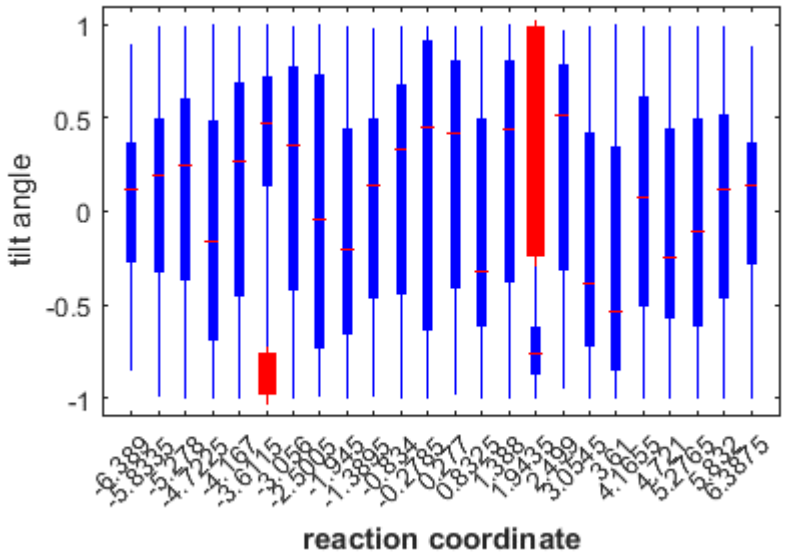

Figure 5.3 Tilt colvar for benzene in CC3. Observations between the 25th and 75th are represented by blue bounding boxes. Outlier observations are shown in red.

Figure 5.3 shows that benzene readily adopts all tilt angles over 100 ps of dynamics. The "spin" angle demonstrates similar free rotation (data not shown). Analysis of recorded molecular dynamics trajectories further confirms that benzene easily reorients within cage cavities. These rotations occur on timescales < 100 ps, which is very fast relative to the average time between hops from cages (~ 10^7 ps). Consequently, we can use eq 4.2 to derive the self-diffusivity of benzene from the cage to void and void to cage hopping rates. This yields $D_s = 5.2 \times 10^{-11}$ cm^2/s for benzene at 300 K. This is about 1 order of magnitude lower than the diffusion coefficient for SF_6 in CC3 at 300 K derived in Chapter 4.

5.3.3 Free Energy Profile for para-Xylene in CC3

Figure 5.4 shows F(q) derived from umbrella sampling and the associated umbrella histograms for para-xylene in CC3 at 300 K.

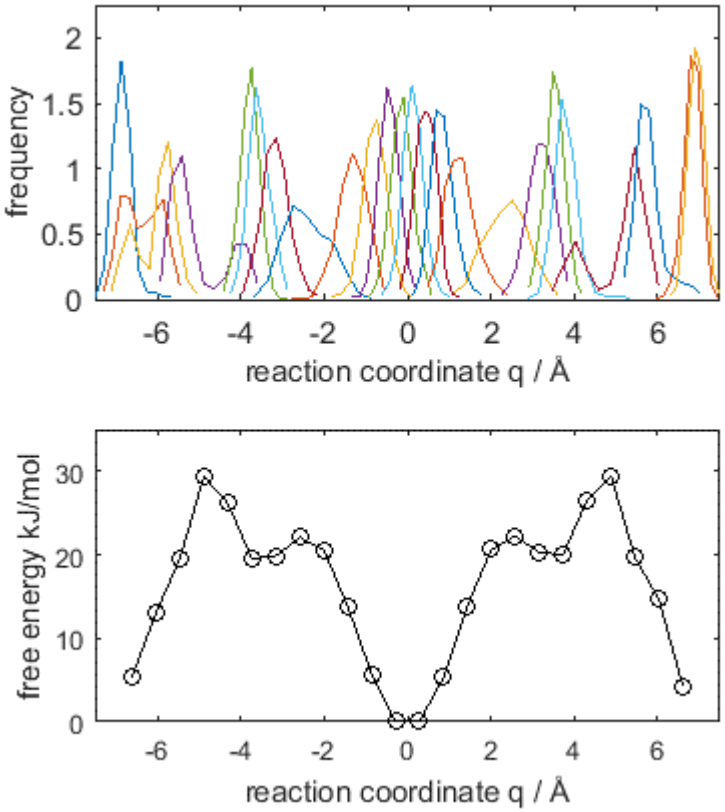

Figure 5.4 Umbrella sampling histograms and the reconstructed free energy profile F(q) for united atom para-xylene in CC3 at 300K

Para-xylene has a qualitatively different free energy profile to benzene and spherical adsorbates in CC3. Para-xylene is at an energetic minimum within the void spanning adjacent cage molecules. This agrees with in-situ crystallographic data showing that para-xylene prefers void cavities to cage cavities.[5] The rate equation derived in Chapter 4 is inapplicable to para-xylene. Instead of diffusing between cage cavity and void cavity microstates, para-xylene hops between the voids separating adjacent cage molecules. This implies that a linear reaction coordinate may not fully capture the diffusion of para-xylene because the adsorbate must negotiate a “turn” within each cage molecule.

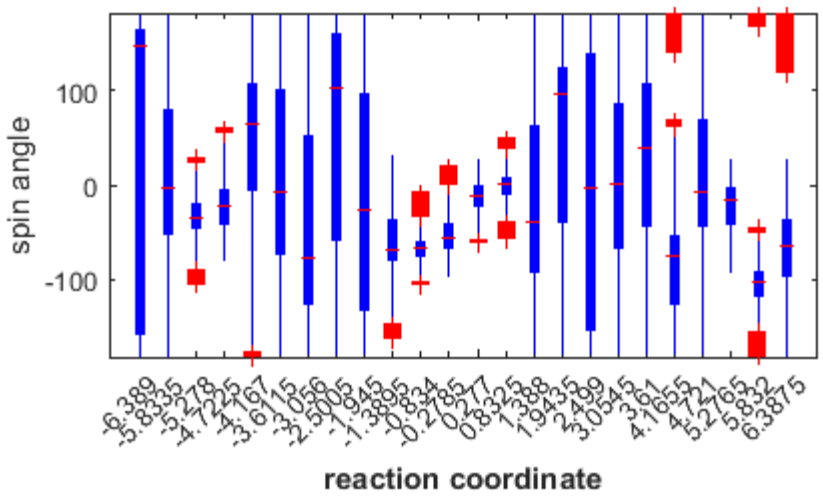

Figure 5.5 Spin colvar for para-xylene in CC3

Figure 5.5 shows that the spin angle of para-xylene is greatly restricted when in the void cavity. Para-xylene is able to freely rotate near the cage windows at q = ± 3.5

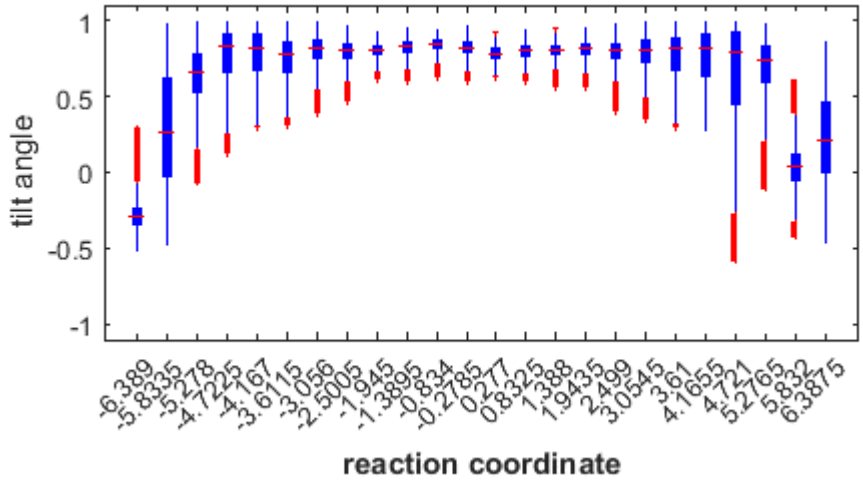

Figure 5.6 Tilt colvar for para-xylene in CC3

Figure 5.6 shows that the tilt angle of para-xylene is greatly restricted within CC3. This means that the major axis of para-xylene prefers to orient parallel to the reaction coordinate.

5.3.4 Free Energy Profile for meta-Xylene in CC3

Figure 5.7 shows F(q) derived from umbrella sampling and the associated umbrella histograms for meta-xylene in CC3 at 300 K.

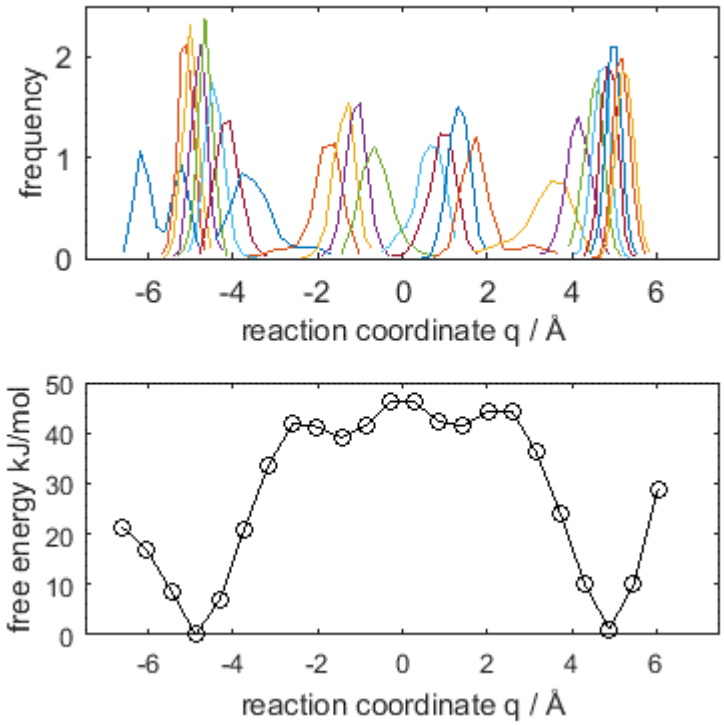

Figure 5.7 Umbrella sampling histograms and the reconstructed free energy profile F(q) for united atom meta-xylene in CC3R at 300K

Meta-xylene has a qualitatively different free energy profile to both benzene and para-xylene in CC3. Meta-xylene exhibits a strong preference for localization within cage cavities rather than voids between cages. This agrees with in-situ crystallographic data showing that meta-xylene molecules are found in CC3 cages .[5] The rate equation derived in Chapter 4 is inapplicable to meta-xylene. Instead of diffusing between cage cavity and void cavity microstates, meta-xylene hops directly between adjacent cage molecules over a diffusive energy barrier spanning the void. Meta-xylene has a transmission coefficient of 0.193, much lower than found for benzene in CC3. Trajectories of up to 100 picoseconds from the dividing surface are necessary to reach thermalization.

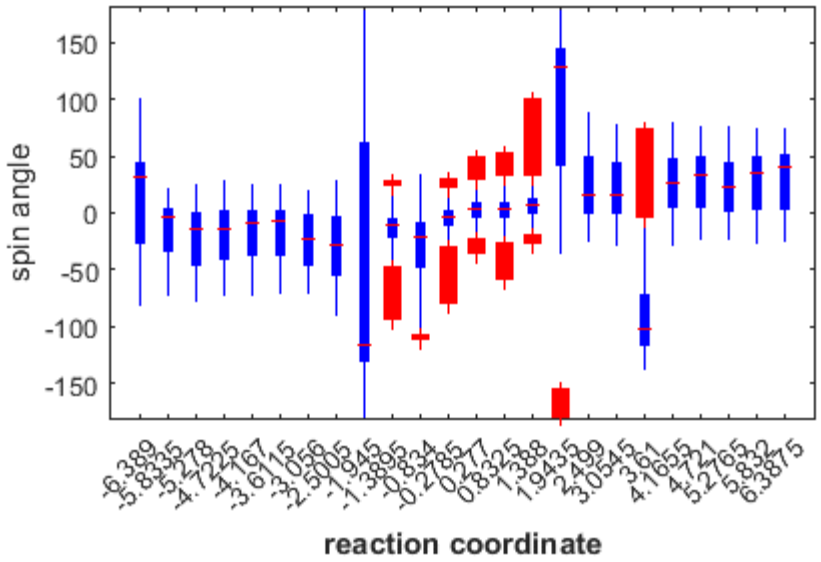

Figure 5.8 Spin colvar for meta-xylene in CC3

Figure 5.8 shows that meta-xylene is more restricted in its spin angle than para-xylene. The tilt collective variable is similar to that shown for para-xylene in Figure 5.6.

5.4 Conclusions and Future Work

We have shown that the methods developed in Chapter 4 can be applied to the diffusion of rigid aromatic molecules in CC3. Initial results show that while benzene diffuses by a similar mechanism to spherical adsorbates in CC3, para-xylene and meta-xylene do not. This necessitates the development of new models to connect hopping rates to diffusivities for these molecules that may incorporate reorientational barriers to diffusion. We plan to derive these models are verify them by performing standard molecular dynamics at high temperatures. Furthermore, we plan to apply this class of methods to study the diffusion of ortho-xylene, mesitylene, and ethyl-toluene diffusion in CC3. These methods may be further applicable to other diamondoid porous organic cages such as cage crystal 3 and co-crystals of cage crystal 1 and cage crystal 3.

5.5 References

[1] C. Y. Lee, Y.-S. Bae, N. C. Jeong, O. K. Farha, A. A. Sarjeant, C. L. Stern, P. Nickias, R. Q. Snurr, J. T. Hupp, S. T. Nguyen, *Journal of the American Chemical Society* **2011**, *133*, 5228-5231.

[2] R. B. Eldridge, *Industrial & Engineering Chemistry Research* **1993**, *32*, 2208-2212.

[3] P. S. Gomes, M. Minceva, A. E. Rodrigues, *Adsorption* **2006**, *12*, 375-392.

[4] J. A. Gee, K. Zhang, S. Bhattacharyya, J. Bentley, M. Rungta, J. S. Abichandani, D. S. Sholl, S. Nair, *The Journal of Physical Chemistry C* **2016**.

[5] T. Mitra, K. E. Jelfs, M. Schmidtmann, A. Ahmed, S. Y. Chong, D. J. Adams, A. I. Cooper, *Nature Chemistry* **2013**, *5*, 276-281.

[6] L. Chen, P. S. Reiss, S. Y. Chong, D. Holden, K. E. Jelfs, T. Hasell, M. A. Little, A. Kewley, M. E. Briggs, A. Stephenson, K. M. Thomas, J. A. Armstrong, J. Bell, J. Busto, R. Noel, J. Liu, D. M. Strachan, P. K. Thallapally, A. I. Cooper, *Nature Materials* **2014**, *13*, 954-960.

[7] T. Hasell, M. Miklitz, A. Stephenson, M. A. Little, S. Y. Chong, R. Clowes, L. Chen, D. Holden, G. A. Tribello, K. E. Jelfs, A. I. Cooper, *J Am Chem Soc* **2016**, *138*, 1653-1659.

[8] S. Jiang, L. Chen, M. Briggs, T. Hasell, A. Cooper, *Chemical Communications* **2016**, *52*, 6895-6898.

[9] A. Kewley, A. Stephenson, L. Chen, M. E. Briggs, T. Hasell, A. I. Cooper, *Chemistry of Materials* **2015**, *27*, 3207-3210.

[10] J. Camp, D. S. Sholl, *Journal of Physical Chemistry C* **2016**, *120*, 1110-1120.

[11] T. Tozawa, J. T. A. Jones, S. I. Swamy, S. Jiang, D. J. Adams, S. Shakespeare, R. Clowes, D. Bradshaw, T. Hasell, S. Y. Chong, C. Tang, S. Thompson, J. Parker, A. Trewin, J. Bacsa, A. M. Z. Slawin, A. Steiner, A. I. Cooper, *Nature Materials* **2009**, *8*, 973-978.

[12] J. Hutter, M. Iannuzzi, F. Schiffmann, J. VandeVondele, *Wiley Interdisciplinary Reviews: Computational Molecular Science* **2014**, *4*, 15-25.

[13] W. L. Jorgensen, D. S. Maxwell, J. Tirado-Rives, *Journal of the American Chemical Society* **1996**, *118*, 11225-11236.

[14] D. Holden, K. E. Jelfs, A. I. Cooper, A. Trewin, D. J. Willock, *Journal of Physical Chemistry C* **2012**, *116*, 16639-16651.

[15] M. G. Martin, J. I. Siepmann, *The Journal of Physical Chemistry B* **1998**, *102*, 2569-2577.

[16] W. Smith, T. Forester, I. Todorov, U. Cheshire.

[17] S. Plimpton, *Journal of Computational Physics* **1995**, *117*, 1-19.

[18] G. Fiorin, M. L. Klein, J. Hénin, *Molecular Physics* **2013**, *111*, 3345-3362.

[19] E. A. Coutsias, C. Seok, K. A. Dill, *Journal of computational chemistry* **2004**, *25*, 1849-1857.

[20] D. Chandler, *The Journal of Chemical Physics* **1978**, *68*, 2959-2970.

[21] J. B. Anderson, *Advances in Chemical Physics*, *91*, 381-432.

6

OUTLOOK

6.1 Impact and Related Work

We have created a database of over 5,000 MOF structures that are immediately useful for atomistic simulations.[1] We demonstrated the utility of the CoRE MOF database by evaluating each structure for its utility in the storage methane for adsorbed natural gas (ANG). During the preparation of the CoRE MOF manuscript, these results became of interest to other researchers within the Department of Energy Nanoporous Materials Genome Center, which funded our work. Simon et al. used our experimentally derived MOF structures as a basis of comparison with over 100,000 hypothetical zeolites, MOFs, zeolitic imidazolate frameworks (ZIFs), and porous polymer networks (PPNs) for performance in methane delivery from 65 to 5.8 bar.[2] These results are shown in Figure 6.1:

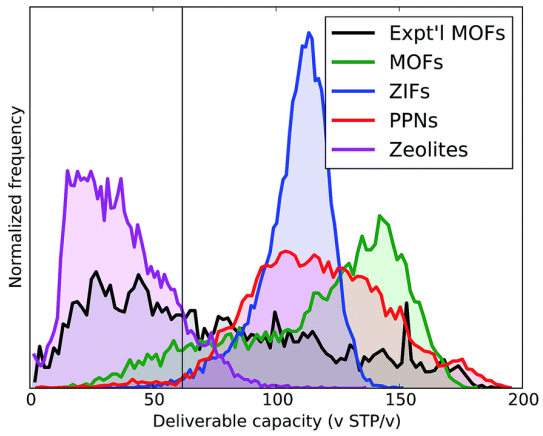

Figure 6.1 Distribution of the deliverable capacity for different materials; the vertical line marks the deliverable capacity of an empty tank (calculated from the density of bulk CH_4 at 65 bar minus that at 5.8 bar using the Peng–Robinson equation of state).

These high-throughput GCMC results demonstrate that experimentally derived MOF structures (from the CoRE MOF database) are predicted to be among the best performing nanoporous materials for adsorbed natural gas storage. The best performing materials in both atomistic simulations and experiments converge on a maximum methane deliverable

capacity of approximately 200 v STP/v.[2] This suggests that top experimentally tested adsorbents may already be approaching thermodynamic and materials performance limits.[2]

The CoRE MOF database has been cited over 25 times since its publication in 2014. For example, the CoRE MOF database has been used to study the influence of higher hydrocarbon co-adsorption on natural gas storage.[3] Ethane and propane were shown to accumulate over successive cycles of adsorption and desorption, which could decrease tank capacity over time. Altintas and Keskin used the CoRE MOF structures to screen for methane-ethane and ethane-ethene separations and identified several MOF as highly selective adsorbents for these separations.[4] The CoRE MOFs have also been used to identify materials selective for xylene separations[5] and xenon over krypton[6]. Outside of adsorption applications, the CoRE MOF structures have been geometrically analyzed for potential electrical conduction through molecular infiltration.[7] Recently, Evans et al. identified 481 potential organic porous molecular crystals using similar methods to those used to create the CoRE MOF database.[8]

During our development of the transition state theory methods to measure diffusion in flexible nanoporous materials described in Chapter 4, related methods were applied by members of the Sholl group to other systems. Verploegh et al. used dynamically corrected TST to measure the loading dependent diffusion of C1-C4 alkanes in the flexible ZIF-8.[9] Simulation results agreed well with experimentally derived diffusion coefficients from the literature. Boulfelfel et al. used path sampling methods to simulate the diffusion of linear alkanes in flexible LTA zeolite.[10] Ongoing work has shown that the results from these path sampling methods agree well with umbrella sampling TST.

Recently, experimental evidence has demonstrated that sulfur hexafluoride readily diffuses through CC3 in room temperature gas phase breakthrough experiments.[11] Our umbrella sampling TST simulations described in Chapter 4 predicted that SF_6 has a self-diffusivity of $D_S = 3.5 \times 10^{-10}$ cm^2 s^{-1} at 300 K in CC3.[12] In contrast, Holden predicted that SF_6 is completely immobile in CC3 by analyzing direct molecular dynamics simulations. This illustrates the potential usefulness of our methods for simulating slow adsorbate diffusion.

6.2 Suggestions for Future Work

6.2.1 Framework Flexibility and Adsorption

MOF framework flexibility can be broadly divided into two classes of phenomena. Most often in the literature, "flexibility" refers to large reversible changes in porosity in response to adsorption, desorption, or temperature changes which is often associated with steps and hysteresis in adsorption isotherms.[13] Materials exhibiting such phenomenon are known as "soft porous crystals".[14] A common example is the transition of MIL-53 from a closed-pore structure to a highly porous open structure with increasing temperature.[15] These phase transitions may be undesirable in applications when porosity is lost, but there are examples where large phase transitions are mechanistically essential to effect a separation. A phase transition in DynaMOF-100 has been shown to result in selective adsorption of styrene from ethyl benzene.[16] Although there are example in the literature of predictions of phase transitions in MOFs with computational tools such as first principles molecular dynamics[17], such approaches are infeasible to perform for thousands of materials in high-throughput screening.

In screening studies, phase transitions are neglected unless the starting database of materials contains an experimentally derived crystal structure for each unique phase. In some instances, this may result in overlooking promising materials with phase transitions. In Chapter 3, we screened several thousand MOFs for use in adsorptive storage of natural gas.[1] The most promising MOF we identified was UTSA-76a, which has a predicted deliverable CH_4 capacity from 65 to 5.8 bar of 189 cm^3 STP cm^{-3}. This prediction closely matched experiments for UTSA-76a, which is not known to undergo phase transitions.[18] Recently, a flexible MOF exhibiting a phase transition at 15 bar was reported to have a higher methane storage capacity of 197 cm^3 STP cm^{-3}.[19] This high pressure phase was not in our starting database of materials, which illustrates one limitation of current computational screening approaches.

The second class of flexibility phenomena in MOFs is associated with local framework vibrations such as rotation of organic ligands, oscillation of coordination bonds, and variation in pore aperture dimensions. This class of framework flexibility was considered in Chapters 4 and 5 in the context of adsorbate diffusion in flexible porous

organic cages.[12] In simulations of adsorption, the influence of local framework flexibility is not well understood. In high throughput simulations of adsorption, the rigid framework assumption may be justified on the basis of studies which have shown good agreement between simulated isotherms and experiments.[20-21] However, there are examples in the zeolite literature where local framework flexibility has a considerable influence on adsorption. Simulations of adsorption in two different experimentally resolved orthosilicalite structures with slightly different lattice constants showed significantly different Henry's constants for benzene.[22] These results suggest that small variations in crystal structures due to framework flexibility may have a substantial influence on thermodynamic observables. Vlugt et al. used a hybrid Monte Carlo approach to directly observe this phenomenon in simulations of isobutane and heptane adsorption in MFI silicalite.[23]

There is a critical need for broader understanding of the influence of MOF framework flexibility on properties of adsorbed phases. In particular, it is important to understand how the standard practice of assuming framework rigidity may influence results in high-throughput simulations of adsorption in MOFs. Here, I outline three research directions for improving our understanding of MOF flexibility with special emphasis on development and application of methods for use on large numbers of materials.

6.2.2 Integration of the CoRE MOF Database with Experimental Adsorption Data

The National Institute of Standards and Technology (NIST) maintains a database of experimental adsorption isotherms for several hundred MOFs.[24] The database serves as a reference for assessing the accuracy of experiments and simulations and could help avoid redundant experiments on known materials. Each isotherm stored in the NIST database is associated with the name of the adsorbent material, but not the crystallographic structural information for these MOFs. Linking the NIST experimental adsorption database to the CoRE MOF structural database would be broadly useful to the MOF community.

For example, the CoRE MOF database contains over 40 CuBTC structures and many duplicate structures for other common MOFs. These structures were originally deposited in the Cambridge Structural Database (CSD)[25] by different research groups using different synthesis, activation, and x-ray diffraction procedures. Researchers working

in collaboration with NIST have observed that minor geometric differences between these CuBTC structures can have a significant influence on simulations of noble gas adsorption.[26] By comparing experimental data from NIST to simulations of adsorption in duplicate CoRE MOF structures, it may be possible to determine which structure is best for use in high-throughput screening. The reverse is also be true – where there is conflicting experimental data in the NIST database, GCMC simulations in CoRE MOF structures could provide insight on which experiments are most reliable.

Linking information between the NIST database and the CoRE MOFs could be accomplished by cross referencing the digital object identifiers used in both databases. This would have the additional benefit of associating CoRE MOF structures with the common names MOFs are called in the literature. Most CoRE MOF structures are currently categorized by a meaningless CSD reference code – an example is "VELVOY" for ZIF-8. This is problematic because there is no obvious way to collate structures that represent different phases of the same material. Associating each CoRE MOF structure with its common name such as "MIL-53" would address this problem.

6.2.3 Benchmark DFT Methods for Modeling MOF Framework Dynamics

The lack of force fields that have been tailored to represent the flexibility of specific MOFs is a significant impediment to understanding of how framework dynamics influence adsorbed phases. Parameterization of custom force fields with DFT calculations for new materials is a labor intensive and computationally expensive process. Software has been written to assist in the process, but users are still required to manually cleave discrete clusters from periodic MOF structures.[27] Transferrable force fields have been proposed for MOFs, but there is not clear evidence that these force fields are applicable outside of their set of training materials.[28]

Ab-initio molecular dynamics (AIMD) is an attractive method for studying MOF framework flexibility without the need to parameterize a classical force field. In AIMD simulations, forces between atoms are computed between each molecular dynamics step by electronic structure calculations (typically DFT). Unfortunately, a significant disadvantage of AIMD is that computing just a few picoseconds of dynamics for a MOF can take over 24 hours on a typical computer cluster. This severely restricts the range of

physical processes that can be feasibly investigated with AIMD. However, local framework flexibility associated with thermal vibration in MOF occurs on these timescales. Experimental and theoretical evidence shows that MOF linker rotations can have a period of less than 10 ps.[29] Simulations of a few picoseconds are long enough to determine the temporal distribution of pore aperture diameters in ZIF-8, which is a key parameter for evaluating the influence of flexibility on adsorbate diffusion.[30]

Researchers using density functional theory for metal-organic frameworks must choose between a wide array of available exchange-correlation functionals, pseudopotential basis sets, and possible dispersion corrections. Nazarian et al. recently minimized the atomic positions and lattice constants of 12 chemically diverse MOF structures with 6 different DFT functionals.[31] All functionals used did a reasonably good job of predicting key geometric features of high-quality experimentally derived crystal structures. However, there is evidence that AIMD calculations using DFT may not be similarly insensitive to choice of functional. The M06-2X and B3LYP functionals give dramatically different energy barriers for rotation of dicarboxylate ligands.[32] Constant pressure AIMD simulations of breathing behavior in MIL-53 (Sc) show that equilibrium cell parameters are sensitive to the choice of dispersion correction.[33]

A benchmark of DFT-based AIMD methods on several well-known MOF structures could provide valuable information on best practices for application of AIMD to large numbers of materials. It is important to understand whether MOF dynamics are highly sensitive to the choice of exchange-correlation functionals and dispersion corrections. To start, a comparison could be made between probability distributions for key structural parameters such as cell volumes, cavity dimensions, and pore aperture diameters over short AIMD trajectories using different pseudopotential basis sets and functionals.

6.2.4 Evaluate the Influence of MOF Framework Flexibility on High Loading Adsorption

Equilibrium based industrial separations typically operate at high pressures, leading to conditions of adsorbate saturation within nanopores. In these liquid-like adsorbed phases, the entropic thermodynamics of adsorbate packing become more important to overall adsorption selectivities than binding energies at the most favorable adsorption

sites.[34] Recently, Gee et al. conducted GCMC simulations of C_8 aromatic adsorption in both a rigid model of MIL-47(V) and an ensemble of snapshots from a molecular dynamics trajectory of the flexible MIL-47(V) framework.[35] Adsorption selectivities for o-xylene over m-xylene and ethylbenzene in the flexible model of MIL-47(V) matched experimental observations much better than simulations in the rigid crystal structure. This may be attributable to lower adsorbate packing efficiency in the flexible material, which has been shown to influence the high-loading adsorption of styrene.[36]

Efforts are underway to apply the method of Gee et al. using AIMD to generate the ensemble of framework snapshots. The aim of this work is to assess whether or not the changes in selectivity observed in MIL-47(V) is a unique feature of this material / adsorbate combination or a more general property of high loading adsorption in flexible MOFs. However, the validity of the Gee et al. methodology should be investigated before general conclusions are drawn. A key assumption of this method is decoupling between framework dynamics the adsorbed phase. Each flexible framework configuration is generated from a molecular dynamics trajectory of the empty material, while adsorption properties are calculated by GCMC at conditions near pore saturation. To investigate this, it would be interesting to compare distributions in structural parameters such as linker torsion angles in AIMD trajectories in a structure pre-loaded with adsorbates versus an empty framework.

6.3 References

[1] Y. G. Chung, J. Camp, M. Haranczyk, B. J. Sikora, W. Bury, V. Krungleviciute, T. Yildirim, O. K. Farha, D. S. Sholl, R. Q. Snurr, *Chemistry of Materials* **2014**, *26*, 6185-6192.

[2] C. M. Simon, J. Kim, D. A. Gomez-Gualdron, J. S. Camp, Y. G. Chung, R. L. Martin, R. Mercado, M. W. Deem, D. Gunter, M. Haranczyk, D. S. Sholl, R. Q. Snurr, B. Smit, *Energy & Environmental Science* **2015**, *8*, 1190-1199.

[3] H. D. Zhang, P. Deria, O. K. Farha, J. T. Hupp, R. Q. Snurr, *Energy & Environmental Science* **2015**, *8*, 1501-1510.

[4] C. Altintas, S. Keskin, *Chemical Engineering Science* **2016**, *139*, 49-60.

[5] J. A. Gee, K. Zhang, S. Bhattacharyya, J. Bentley, M. Rungta, J. S. Abichandani, D. S. Sholl, S. Nair, *The Journal of Physical Chemistry C* **2016**.

[6] C. M. Simon, R. Mercado, S. K. Schnell, B. Smit, M. Haranczyk, *Chemistry of Materials* **2015**, *27*, 4459-4475.

[7] X. Nie, A. Kulkarni, D. S. Sholl, *Journal of Physical Chemistry Letters* **2015**, *6*, 1586-1591.

[8] J. D. Evans, D. M. Huang, M. Haranczyk, A. W. Thornton, C. J. Sumby, C. J. Doonan, *CrystEngComm* **2016**.

[9] R. J. Verploegh, S. Nair, D. S. Sholl, *Journal of the American Chemical Society* **2015**, *137*, 15760-15771.

[10] S. E. Boulfelfel, P. I. Ravikovitch, D. S. Sholl, *Journal of Physical Chemistry C* **2015**, *119*, 15643-15653.

[11] T. Hasell, M. Miklitz, A. Stephenson, M. A. Little, S. Y. Chong, R. Clowes, L. Chen, D. Holden, G. A. Tribello, K. E. Jelfs, A. I. Cooper, *J Am Chem Soc* **2016**, *138*, 1653-1659.

[12] J. Camp, D. S. Sholl, *Journal of Physical Chemistry C* **2016**, *120*, 1110-1120.

[13] A. Allen, L. Espinal, W. Wong-Ng, W. Queen, C. Brown, S. Kline, K. Kauffman, J. Culp, C. Matranga, *Journal of Alloys and Compounds* **2015**, *647*, 24-34.

[14] F. X. Coudert, A. Boutin, M. Jeffroy, C. Mellot-Draznieks, A. H. Fuchs, *Chemphyschem* **2011**, *12*, 247-258.

[15] Y. Liu, J.-H. Her, A. Dailly, A. J. Ramirez-Cuesta, D. A. Neumann, C. M. Brown, *Journal of the American Chemical Society* **2008**, *130*, 11813-11818.

[16] S. Mukherjee, B. Joarder, A. V. Desai, B. Manna, R. Krishna, S. K. Ghosh, *Inorganic Chemistry* **2015**, *54*, 4403-4408.

[17] L. Chen, J. P. Mowat, D. Fairen-Jimenez, C. A. Morrison, S. P. Thompson, P. A. Wright, T. Düren, *Journal of the American Chemical Society* **2013**, *135*, 15763-15773.

[18] B. Li, H.-M. Wen, H. Wang, H. Wu, M. Tyagi, T. Yildirim, W. Zhou, B. Chen, *Journal of the American Chemical Society* **2014**, *136*, 6207-6210.

[19] J. A. Mason, J. Oktawiec, M. K. Taylor, M. R. Hudson, J. Rodriguez, J. E. Bachman, M. I. Gonzalez, A. Cervellino, A. Guagliardi, C. M. Brown, *Nature* **2015**, *527*, 357-361.

[20] D. Dubbeldam, S. Calero, T. J. H. Vlugt, R. Krishna, T. L. M. Maesen, B. Smit, *The Journal of Physical Chemistry B* **2004**, *108*, 12301-12313.

[21] S. Keskin, J. Liu, R. B. Rankin, J. K. Johnson, D. S. Sholl, *Industrial & Engineering Chemistry Research* **2008**, *48*, 2355-2371.

[22] L. A. Clark, R. Q. Snurr, *Chemical Physics Letters* **1999**, *308*, 155-159.

[23] T. J. H. Vlugt, M. Schenk, *The Journal of Physical Chemistry B* **2002**, *106*, 12757-12763.

[24] in *NIST Standard Reference Database Number 205* (Eds.: D. W. Siderius, V. K. Shen, R. D. Johnson, R. D. van Zee), National Institute of Standards and Technology, Gaithersburg, MD.

[25] F. H. Allen, *Acta Crystallographica Section B: Structural Science* **2002**, *58*, 380-388.

[26] K. V. Lawler, Z. Hulvey, P. M. Forster, *Physical Chemistry Chemical Physics* **2015**, *17*, 18904-18907.

[27] L. Vanduyfhuys, S. Vandenbrande, T. Verstraelen, R. Schmid, M. Waroquier, V. Van Speybroeck, *Journal of Computational Chemistry* **2015**, *36*, 1015-1027.

[28] J. K. Bristow, D. Tiana, A. Walsh, *Journal of Chemical Theory and Computation* **2014**, *10*, 4644-4652.

[29] N. C. Burtch, A. Torres-Knoop, G. S. Foo, J. Leisen, C. Sievers, B. Ensing, D. Dubbeldam, K. S. Walton, *Journal of Physical Chemistry Letters* **2015**, *6*, 812-816.

[30] E. Haldoupis, T. Watanabe, S. Nair, D. S. Sholl, *ChemPhysChem* **2012**, *13*, 3449-3452.

[31] D. Nazarian, P. Ganesh, D. S. Sholl, *Journal of Materials Chemistry A* **2015**, *3*, 22432-22440.

[32] S. Pakhira, M. Takayanagi, M. Nagaoka, *The Journal of Physical Chemistry C* **2015**, *119*, 28789-28799.

[33] V. Haigis, Y. Belkhodja, F.-X. Coudert, R. Vuilleumier, A. Boutin, *The Journal of Chemical Physics* **2014**, *141*, 064703.

[34] A. Torres-Knoop, D. Dubbeldam, *ChemPhysChem* **2015**.

[35] J. A. Gee, D. S. Sholl, *Journal of Physical Chemistry C* **2016**, *120*, 370-376.

[36] A. Torres-Knoop, J. Heinen, R. Krishna, D. Dubbeldam, *Langmuir* **2015**, *31*, 3771-3778.

APPENDIX A

1. Topology Analysis Using TOPOS

In order to assign the underlying nets using TOPOS[1], we identified the bonds for all structures using a general cut-off distance specified by the TOPOS software, where atoms below that cut-off were considered bonded. Following the procedure, TOPOS was used to determine ring groups present inside the structure and to separate those ring groups from one another. For example, in IRMOF-1, the organic linker's aromatic ring would be one ring group and the metal corner (including the oxygen atoms from the linker) would be the second ring group. These ring groups are then transformed into a single pseudo-atom placed at the center of the ring. The metal corners from IRMOF-1 would be simplified to pseudo-atom MC (as in Metal Corner) and the organic "rings" simplified down to OL (as in Organic Linker). Connectivity between the simplified groups remains following the procedure, so that the pseudo-atom MC is connected to three other OL's and each OL is connected to two MC pseudo-atoms. The edges do not play an important role, so we have removed them for net simplification purposes and for determining topology. Edges were removed by simply removing all pseudo-atoms with connectivity of 2 or less. For example, the MC corner in IRMOF-1 has a connectivity of 3 (or 6 with periodic boundary conditions), and each OL has a connectivity of 2; thus this simplification method removes the OL pseudo-atoms. Then by taking the locations of the remaining pseudo-atoms, TOPOS can assign a matching topology by comparing that structure to structures within various databases.

In determining the topology using the method described above, the connectivity of atoms plays an important role. Ideally, a standard method would be able to correctly assign bonds to correct pairs of atoms; however, this was not the case for all structures in the CoRE MOF database. Given the large number of structures and the need for an automated procedure for identifying the topologies, the frameworks could not be checked on a structure-by-structure basis to ensure correct bonding of the atoms. Therefore, the structures that failed during the analysis in TOPOS or structures considered to have a "new topology" by TOPOS were discarded from the analysis, leaving a resulting set of ~2,000 structures. A random sampling of ~10 failed and "new topology" structures were checked

for consistency, and we found incorrectly assigned bonding to be the source of failure of the incorrect (i.e., new topology) topological classifications.

2. Synthesis, Activation, and Nitrogen Isotherm Measurements of MIL-53(Al)

We synthesized MIL-53(Al) to test our prediction of a high methane uptake capacity of 267 vol_{STP} vol^{-1} at 298 K and 65 bar. To obtain the highest methane uptake, special care was taken to avoid trapping of ligands and solvents in the pores. The synthesis and activation procedures are described below. This procedure yielded a sample with a surface area and pore volume higher than any reported in the literature for MIL-53(Al)[2], with S_{BET} = 1,530 m^2/g and V_{pore} = 0.56 cc/g.

General experimental procedures

Aluminum nitrate nonahydrate (Aldrich, 98%), terephthalic acid (Aldrich, 98%), N,N-dimethylformamide (DMF) (Macron, 99.8%), ethanol (EtOH) (Macron, 99.8%), and pyridine (Aldrich, 99%) were used as received without further purification. Powder X-ray diffraction (PXRD) patterns of MIL-53(Al) were recorded on a Rigaku ATXG diffractometer equipped with an 18 kW Cu rotating anode, MLO monochromator, and a high-count-rate scintillation detector (measurements made over a range of $2° < 2\theta < 30°$ in 0.05° step width with a 3 deg/min scanning speed). Thermogravimetric analyses (TGA) were performed on a TGA/DSC 1 system (Mettler-Toledo AG, Schwerzenbach, Switzerland), which runs on a PC with STAR software (version 9.10). Samples placed in alumina pans were heated from 25°C to 700°C at 10°C/minute under nitrogen flow. Before sorption measurements, samples were desolvated on a Micromeritics SmartVacPrep. Nitrogen isotherms were measured on a Micromeritics TriStar II 3020 at 77 K. All high-pressure isotherm measurements were performed at the NIST Center for Neutron Research using a computer controlled Sieverts apparatus, details of which have been published elsewhere.[3]

Synthesis of MIL-53(Al)

A sample of MIL-53(Al) was prepared according to a previously reported procedure.[2] This material was designated as $MIL\text{-}53(Al)_{crude}$.

Activation procedure for MIL-53(Al)

A portion of as-synthesized MIL-53(Al)$_{crude}$ (0.200 g) was placed in a vial with 5 mL of pyridine, sealed with a screw-cap and kept in a 100^{0}C oven for 12 hours. Then the liquid phase was separated, and the solid was washed several times with DMF and ethanol to remove pyridine. The sample was then soaked in ethanol for 8 hours. The solid material, MIL-53(Al)$_{EtOH}$, was vacuum filtered, briefly dried in air, and then activated at 200^{0}C for 24 hours under vacuum on a SmartVacPrep. The activated material was designated as MIL-53(Al)$_{desol}$.

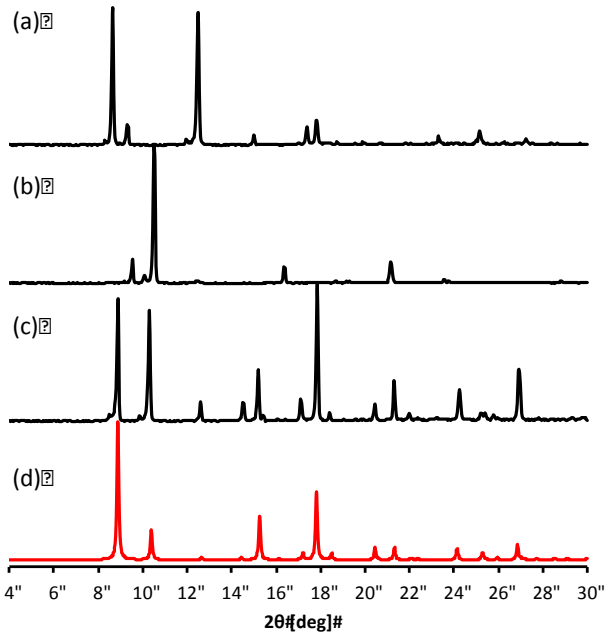

Figure A.1 PXRD patterns of MIL-53(Al): a) desolvated sample MIL-53(Al)$_{desol}$, b) sample before thermal activation MIL-53(Al)$_{EtOH}$, c) as synthesized material MIL-53(Al)$_{crude}$, d) simulated MIL-53(Al).[4]

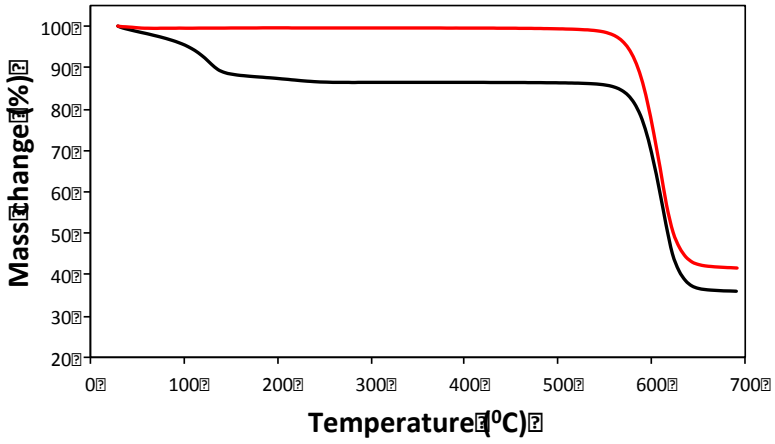

Figure A.2 TGA curves of MIL-53(Al) for sample before thermal activation MIL-53(Al)$_{EtOH}$ (black) and desolvated sample MIL-53(Al)$_{desol}$ (red).

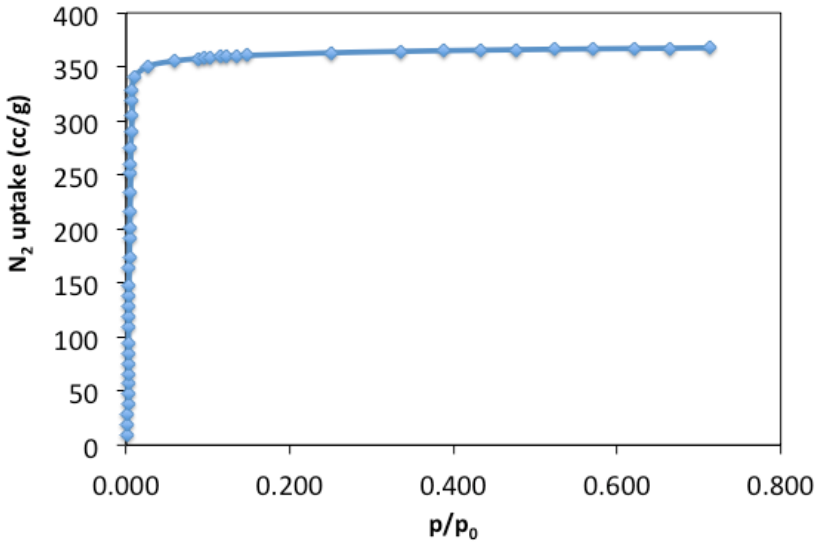

Figure A.3 Nitrogen adsorption-desorption isotherms (T=77 K) for MIL-53(Al)$_{desol}$. Total pore volume of the material is 0.56 cm^3/g.

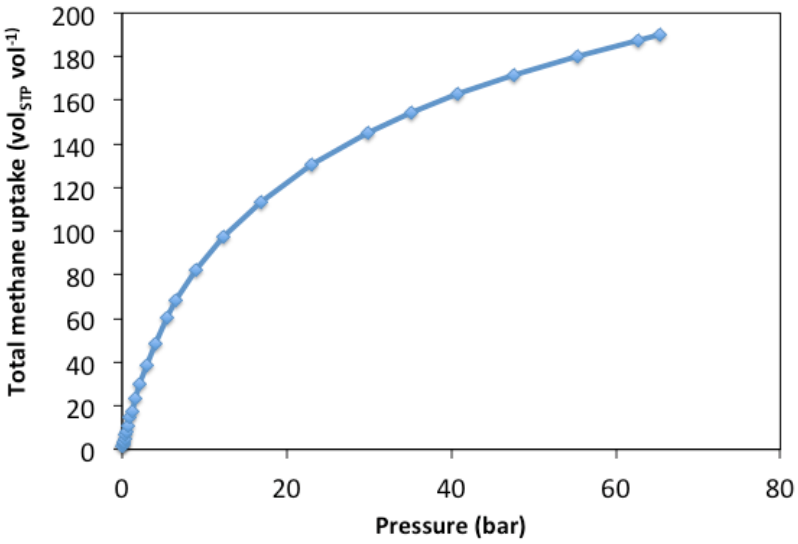

Figure A.4 Measured methane isotherm at 298 K for MIL-53(Al)$_{desol}$. The results obtained from methane adsorption measurements were converted to vol$_{STP}$ vol^{-1} using a framework density of 0.955 cm^3/g and density of gas at STP (T = 273.15 K and P = 101.325 kPa).

3. Discrepancy between MIL-53(Al) GCMC Simulations and Experiment

MIL-53(Al) is a flexible MOF, composed of corner-sharing $AlO_4(OH)_2$ metal clusters and 1,4-benzenedicarboxylate ligands.[5] The structure has two distinct conformational phases – **lp** (large pore) and **np** (narrow pore). MIL-53(Al) has been studied extensively in the past decade because the structure assumes different conformations depending on the current temperature[6], the thermal history of the sample[7], and the type and concentration of adsorbates.[5, 8] Therefore, it is possible that the crystal structures for MIL-53(Al) in the CSD could have different atomic coordinates and lattice constants, which might influence the simulation outcome.

To examine this issue, we searched the CSD to find additional MIL-53(Al) structures based on chemical bond characteristics of MIL-53(Al) and computed the methane uptake at 65 bar in each structure. There are 13 different MIL-53(Al) structures in the CSD, with each having different experimentally resolved lattice parameters and atomic coordinates. We note that not all of these MIL-53(Al) structures were included

during the construction of the CoRE MOF database because many of these structures are highly disordered. Disordered structures were manually repaired and geometrically minimized prior to carrying out these GCMC simulations, and added to the final list of the CoRE MOF database.

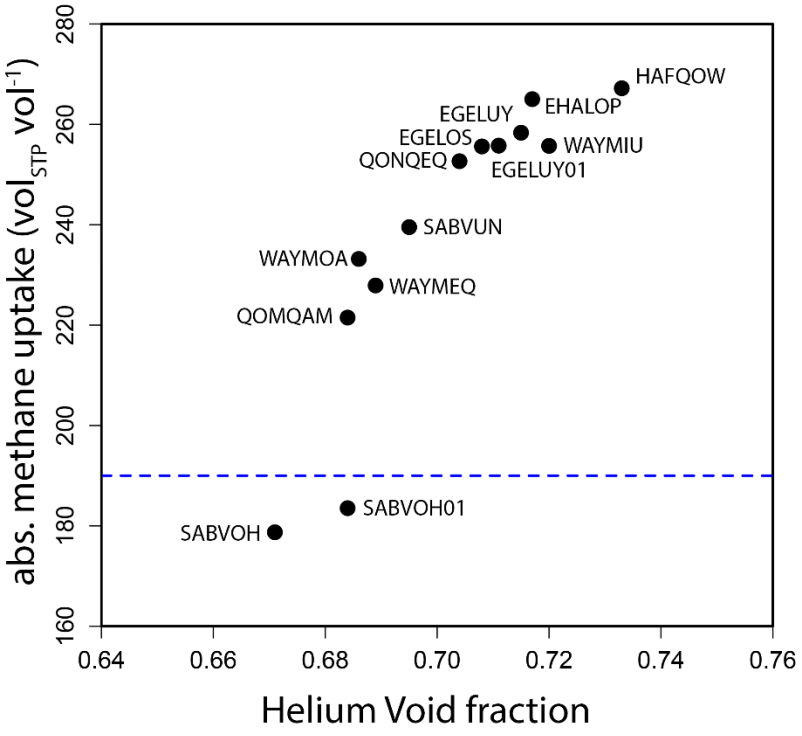

Figure A.5 Simulated absolute methane uptake at 65 bar and 298 K as a function of helium void fraction for a set of 13 MIL-53 (Al) structures obtained from the CSD. The blue dashed line is the experimental methane uptake at 65 bar from Figure A4.

We found that there is a large variation in predicted methane uptake among the 13 different structures, and the variation is correlated with the variation in the helium void fraction among these structures (Figure A5). The helium void fraction varies from 0.67 to 0.74 in these structures (a change of 10%), but the total methane uptake varies from 180 to 267 vol_{STP} vol^{-1} (a change of 48%). Therefore, the variation in helium void fraction alone cannot account for the large variation in methane uptake.

Another possible source for the large variation in methane uptake in these structures is an enthalpy-entropy correlation, which has been studied previously with a simplified 2-D model.[9] In this description, the proximal framework atoms provide potential overlap, which makes more favorable adsorption sites. However, if the proximal framework atoms become too close to one another, repulsion takes effect and the site becomes less favorable for adsorption. As a result, it is possible that changes in the pore size strongly affects the energy landscape inside the pore *in addition* to the free volume for methane occupancy.

Figure A6 shows the density distributions of methane molecules for two different MIL-53(Al) structures at 65 bar. These two structures have the highest and lowest absolute methane loading at 65 bar: SABVOH has the lowest methane uptake (~180 vol_{STP} vol^{-1}), and HAFQOW has the highest methane uptake (~267 vol_{STP} vol^{-1}).

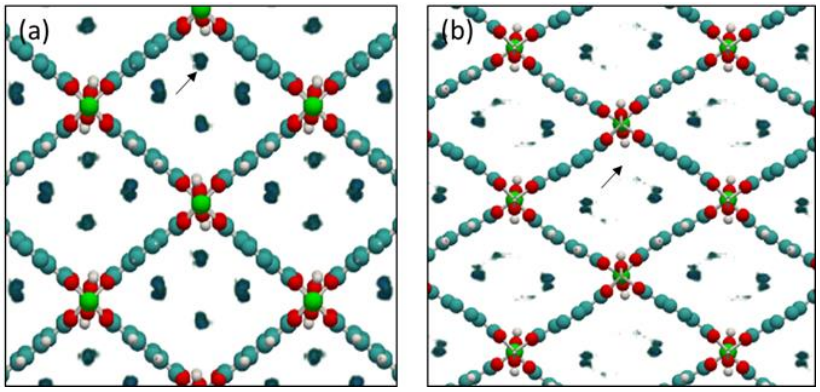

Figure A.6 Probability densities of methane for two MIL-53(Al) structures at 298 K and 65 bar. (a) HAFQOW and (b) SABVOH. White spheres are hydrogen atoms, bright green spheres are aluminum atoms, and light blue spheres are carbon atoms. Arrows in the picture indicate the location of the favorable adsorption site that have disappeared in (b).

Figure A6 (a) shows that there are 4 preferential adsorption sites for methane at 65 bar near the corner-sharing aluminum atoms (bright green spheres) in HAFQOW. Together with other proximal framework atoms, the aluminum atoms create favorable adsorption sites. Figure A.6 (b), on the other hand, shows that there are only 2 preferential adsorption sites for methane in SABVOH.

As discussed above, the proximal framework atoms provide potential overlap, which creates favorable adsorption sites. The energy landscape inside the pore can be affected by subtle changes in the positions of these proximal framework atoms, and if the framework atoms become far apart (instead of too close), the potential overlap decreases. Figure A6 suggests that this leads to the disappearance of the adsorption sites. The framework atoms are located slightly further apart for SABVOH, and the sites that were favorable for methane adsorption in HAFQOW become less favorable for methane adsorption in SABVOH. Based on this, we conclude that the disappearance of favorable adsorption sites in combination with changes in the void fraction are responsible for the large variation of overall methane uptake for nominally similar MIL-53(Al) structures, as shown in Figure A5.

References

[1] V. Blatov, *Structural Chemistry* **2012**, *23*, 955-963.

[2] P. Rallapalli, K. Prasanth, D. Patil, R. S. Somani, R. Jasra, H. Bajaj, *Journal of Porous Materials* **2011**, *18*, 205-210.

[3] W. Zhou, H. Wu, M. R. Hartman, T. Yildirim, *The Journal of Physical Chemistry C* **2007**, *111*, 16131-16137.

[4] M. Vougo-Zanda, J. Huang, E. Anokhina, X. Wang, A. J. Jacobson, *Inorganic chemistry* **2008**, *47*, 11535-11542.

[5] S. Bourrelly, P. L. Llewellyn, C. Serre, F. Millange, T. Loiseau, G. Férey, *Journal of the American Chemical Society* **2005**, *127*, 13519-13521.

[6] A. M. Walker, B. Civalleri, B. Slater, C. Mellot-Draznieks, F. Corà, C. M. Zicovich-Wilson, G. Román-Pérez, J. M. Soler, J. D. Gale, *Angewandte Chemie International Edition* **2010**, *49*, 7501-7503.

[7] Y. Liu, J.-H. Her, A. Dailly, A. J. Ramirez-Cuesta, D. A. Neumann, C. M. Brown, *Journal of the American Chemical Society* **2008**, *130*, 11813-11818.

[8] A. Boutin, F.-X. Coudert, M.-A. Springuel-Huet, A. V. Neimark, G. Férey, A. H. Fuchs, *The Journal of Physical Chemistry C* **2010**, *114*, 22237-22244.

[9] C. M. Simon, J. Kim, L.-C. Lin, R. L. Martin, M. Haranczyk, B. Smit, *Physical Chemistry Chemical Physics* **2014**, *16*, 5499-5513.

APPENDIX B

1. Flexible CC3 Force Field Parameters

The cage specific force field (CSFF) was used to model all nonbonded interactions between atoms in neighboring cage molecules. CSFF uses a 9-6 Lennard Jones potential. A 10 Å cutoff was applied to all CSFF Lennard–Jones interactions:

$$E_{LJ} = \epsilon\left[2\left(\frac{\sigma}{r}\right)^{9} - 3s\left(\frac{\sigma}{r}\right)^{6}\right] \qquad r < 10\text{Å} \quad \textbf{B1}\,.$$

Here, s is a scale factor applied to the dispersive portion of the potential to reproduce the experimental cell volumes at 300 K. We used a scale factor of s = 1.20 following the original CSFF report.

The CSFF atom types and Lennard-Jones σ and ε parameters were adopted directly from the polymer-consistent force field (PCFF)[1-2]:

Table B.1 CSFF nonbonded potential parameters

atom type	description[2]	σ (Å)	ε (kcal mol^{-1})
c	generic sp^3 carbon	4.010	0.0540
c=1	non-aromatic, next to end doubly bounded carbon	4.010	0.0640
cp	sp^2 aromatic carbon	4.010	0.0640
h	generic hydrogen bonded to C	2.995	0.0200
n=	non aromatic end doubly bonded nitrogen	3.800	0.0800

The following 6th power mixing rules were used to describe interactions between different CSFF atom types:

$$\sigma_{ij} = \left[\frac{\sigma_i^6+\sigma_j^6}{2}\right]^{1/6} \quad \textbf{B2} \quad \varepsilon_{ij} = \frac{2\sqrt{\varepsilon_i\varepsilon_j}\sigma_i^3\sigma_j^3}{\sigma_i^6+\sigma_j^6} \quad \textbf{B3}$$

In CSFF, the partial charges on **CC3** framework atoms are the sum of the charge bond increments:

Table B.2 Partial charges

$$q_i = \sum_j \delta_{ij} \quad \textbf{B4}.$$

i	j	δ_{ij}	δ_{ji}
c	c	0.0000	0.0000
c	c=1	0.0000	0.0000
c	cp	0.0000	0.0000
c	h	-0.0530	0.0530
c	n=	0.3000	-0.3000
c=1	cp	0.0000	0.0000
c=1	h	-0.1268	0.1268
c=1	n=	0.3000	-0.3000
cp	cp	0.0000	0.0000
cp	h	-0.1268	0.1268
cp	n=	0.1993	-0.1993

For example, the c=1 carbons which are bonded to 1 aromatic cp carbon, 1 n= nitrogen, and 1 hydrogen have a net partial charge of 0.1732. All Coulombic interactions were computed pairwise to 10 Å and a long range particle-particle mesh Ewald correction was used thereafter. All intramolecular (within an individual cage molecule) nonbonded forces were excluded using the LAMMPS command "neigh_modify exclude molecule all".

CSFF bonded forces are described by a class 2 harmonic potential in units of kcal mol^{-1}:

$$E_{bond} = k_2(r - r_0)^2 - k_3(r - r_0)^3 - k_4(r - r_0)^4 \quad \textbf{B5}.$$

Table B.3 Bonds

i	j	r_0 (Å)	k_2	k_3	k_4
cp	cp	1.3838	465.2720	-1230.7532	1460.3640
cp	h	1.0787	421.9058	-1038.8437	1199.9151
c=1	cp	1.4556	353.8001	-741.6000	936.9211
c=1	h	1.1005	365.7679	-725.5404	781.6621
c=1	n=	1.2700	758.1000	-1675.0000	2153.0000
c	n=	1.4329	324.7000	-218.0000	3030.0000
c	h	1.1010	345.0000	-691.8900	844.6000
c	c	1.5360	250.2000	-503.2000	400.9000

CSFF angle parameters are described by a class2 harmonic potential in units of kcal/mol:

$$E_{angle} = k_2(r - r_0)^2 - k_3(r - r_0)^3 - k_4(r - r_0)^4 \quad \textbf{B6}.$$

Table B.4 Angles

Parameters in bold were missing from the original CSFF report and were adopted directly from PCFF.

i	j	k	θ	k_2	k_3	k_4
cp	cp	h	117.9400	35.1558	-12.4682	0.0000
cp	cp	cp	119.9000	62.0226	-0.9931	0.0000
c=	cp	cp	115.4201	37.1311	0.6510	1.3200
cp	c=	h	117.4110	31.5039	-11.1174	-10.4170
cp	c=	n=	123.6700	57.0000	-20.7468	38.2873
h	c=	n=	122.0000	34.0000	-9.0000	-15.0000
c=	n=	c	110.5000	57.9000	-67.0001	51.7001
c	c	n=	110.0047	53.4431	-59.0000	0.0000
h	c	n=	109.7000	36.0247	1.0032	0.0000

c	c	h	109.7700	37.4530	-10.6040	5.1290
h	**c**	**h**	**107.6600**	**39.6410**	**-12.9210**	**-2.4318**
c	**c**	**c**	**112.6700**	**39.5160**	**-7.4430**	**-9.5583**

CSFF dihedrals are described by a class 2 trigonometric potential in units of kcal mol^{-1}:

$$E_{dihedral}(\varphi) = \sum_{n=1}^{3} \frac{V_n}{2}(1 - \cos(n\varphi - \delta_n)) \qquad \textbf{B7}.$$

The original CSFF report[1] uses a factor of ½ in the coefficient leading the class 2 trigonmetric potential (shown in red above). This is inconsistent with the functional form reported in PCFF, which does not scale the V_n parameters by a factor of ½. This factor was omitted throughout this work because the orginal PCFF functional form better reproduces the reported window size distribution in CC3.[3]

Table B.5 Dihedrals

Parameters in bold were missing from the original CSFF report and were adopted directly from PCFF.

i	j	k	l	V_1	δ_1	V_2	δ_2	V_3	δ_3
h	cp	cp	h	4.5000	0.0000	1.8769	0.0000	0.0000	0.0000
cp	cp	cp	h	-7.0000	0.0000	2.0661	0.0000	0.0000	0.0000
cp	cp	cp	cp	8.3667	0.0000	4.1932	0.0000	0.0000	0.0000
c=1	cp	cp	cp	-3.0000	0.0000	3.3072	0.0000	0.9700	0.0000
c=1	cp	cp	h	7.0000	0.0000	2.5072	0.0000	0.0000	0.0000
h	c=1	cp	cp	1.0000	0.0000	1.1097	0.0000	0.0000	0.0000
cp	cp	c=1	n=	0.0000	0.0000	1.0000	0.0000	0.0000	0.0000
c	n=	c=1	cp	-1.0000	0.0000	7.9000	0.0000	-0.8000	0.0000
c=1	n=	c	h	1.4300	0.0000	-1.1200	0.0000	0.8629	0.0000
c=1	n=	c	c	0.8000	0.0000	0.0400	0.0000	0.2000	0.0000
n=	C	c	n=	0.5071	0.0000	0.2349	0.0000	-0.2349	0.0000

h	C	c	n=	-0.0228	0.0000	0.0280	0.0000	-0.1863	0.0000
h	C	c	h	-0.4432	0.0000	-0.2617	0.0000	-0.1283	0.0000
c	n=	c=1	h	3.0000	0.0000	5.5000	0.0000	0.4000	0.0000
c	**C**	**c**	**c**	**0.0000**	**0.0000**	**0.0514**	**0.0000**	**-0.1430**	**0.0000**
c	**C**	**c**	**h**	**0.0000**	**0.0000**	**0.0316**	**0.0000**	**-0.1681**	**0.0000**
c	**C**	**c**	**n=**	**0.0972**	**0.0000**	**0.0722**	**0.0000**	**-0.2581**	**0.0000**

2. Adsorbate – cage interaction potentials

Cage-adsorption force field (CAFF)[4] parameters were used to model interactions between CC3 framework atoms and Kr, CH_4, Xe, and Rn adsorbate atoms. CAFF uses a 12-6 Lennard Jones potential:

$$E_{LJ} = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right] \qquad r < 10\text{Å} \qquad \textbf{B8}.$$

Interactions between adsorbates and framework atoms are described by the Lorentz–Berthelot mixing rules:

$$\sigma_{ij} = \frac{\sigma_i + \sigma_j}{2} \qquad \textbf{B9}. \qquad \varepsilon_{ij} = \sqrt{\varepsilon_i \varepsilon_j} \qquad \textbf{B10}.$$

Table B.6 Cage-adsorption force field parameters

Atom	**atom type**	**σ (Å)**	**ε/k_B (K)**
Carbon	framework C	3.4730	32.97
Hydrogen	framework H	2.8464	5.27
Nitrogen	framework N	3.2626	26.83
Krypton	Kr	3.69	170.0
methane	CH_4 united atom	3.73	148.0
xenon	Xe	4.10	211.0
radon	Rn	4.17	300.0

Interactions between CC3 framework atoms and united atom models of CS_2 and SF_6 were modeled by combining CAFF parameters for carbon, hydrogen, and nitrogen with parameters from Svehla (1962).[5]

atom	atom type	σ (Å)	ε/k_B (K)
carbon disulfide	CS_2 united atom	4.483	467
sulfur hexaflouride	SF_6 united atom	5.128	222.1

3. Implicit ligand sampling data

In our implicit ligand sampling TST calculations, we fit the activation energy ensemble for each adsorbate and temperature condition to independent log-normal distributions using maximum likelihood estimation. The fit parameters for these distributions are shown below with associated 95% confidence intervals:

Table B.7 Lognormal distribution fits to E_A for Kr cage to void hops

T	μ (95% CI)	σ (95% CI)	mean prefactor A (s^{-1})
200	8.684 (8.678, 8.689)	0.498 (0.495, 0.502)	5.31e+10
250	8.750 (8.744, 8.756)	0.531 (0.527, 0.535)	5.62e+10
300	8.821 (8.815, 8.827)	0.565 (0.560, 0.569)	5.96e+10

Table B.8 Lognormal distribution fits to E_A for Kr void to cage hops

T	μ (95% CI)	σ (95% CI)	mean prefactor A (s^{-1})
200	8.621 (8.616, 8.626)	0.478 (0.474, 0.482)	1.56e+11
250	8.677 (8.671, 8.683)	0.514 (0.510, 0.518)	1.62e+11
300	8.739 (8.733, 8.745)	0.553 (0.548, 0.557)	1.65e+11

Table B.9 Lognormal distribution fits to E_A for CH_4 cage to void hops

T	μ (95% CI)	σ (95% CI)	mean prefactor A (s^{-1})
200	8.737 (8.731, 8.742)	0.489 (0.485, 0.492)	1.24e+11
250	8.799 (8.793, 8.805)	0.522 (0.518, 0.526)	1.31e+11
300	8.866 (8.860, 8.872)	0.557 (0.553, 0.561)	1.39e+11

Table B.10 Lognormal distribution fits to E_A for CH_4 void to cage hops

T	μ (95% CI)	σ (95% CI)	mean prefactor A (s^{-1})
200	8.659 (8.654, 8.664)	0.477 (0.474, 0.481)	3.61e+11
250	8.713 (8.708, 8.719)	0.514 (0.510, 0.518)	3.75e+11
300	8.775 (8.769, 8.781)	0.552 (0.548, 0.557)	3.80e+11

Table B.11 Lognormal distribution fits to E_A for Xe cage to void hops

T	μ (95% CI)	σ (95% CI)	mean prefactor A (s^{-1})
200	9.815 (9.810, 9.820)	0.463 (0.459, 0.466)	5.57e+10
250	9.848 (9.842, 9.853)	0.509 (0.505, 0.512)	5.80e+10
300	9.884 (9.878, 9.890)	0.556 (0.552, 0.561)	6.08e+10

Table B.12 Lognormal distribution fits to E_A for Xe void to cage hops

T	μ (95% CI)	σ (95% CI)	mean prefactor A (s^{-1})
200	9.633 (9.627, 9.638)	0.512 (0.508, 0.516)	1.60e+11
250	9.667 (9.661, 9.673)	0.563 (0.558, 0.567)	1.64e+11
300	9.713 (9.706, 9.719)	0.608 (0.603, 0.612)	1.64e+11

Table B.13 Lognormal distribution fits to E_A for Rn cage to void hops

T	μ (95% CI)	σ (95% CI)	mean prefactor A (s^{-1})
200	10.147 (10.142, 10.152)	0.459 (0.455, 0.462)	4.73e+10
250	10.176 (10.170, 10.181)	0.505 (0.501, 0.509)	4.90e+10
300	10.209 (10.203, 10.215)	0.555 (0.550, 0.559)	5.13e+10

Table B.14 Lognormal distribution fits to E_A for Rn void to cage hops

T	μ (95% CI)	σ (95% CI)	mean prefactor A (s^{-1})
200	9.950 (9.944, 9.955)	0.516 (0.512, 0.520)	1.38e+11
250	9.982 (9.975, 9.988)	0.569 (0.565, 0.574)	1.41e+11
300	10.025 (10.018, 10.032)	0.616 (0.611, 0.620)	1.40e+11

Table B.15 Lognormal distribution fits to E_A for CS_2 cage to void hops

T	μ (95% CI)	σ (95% CI)	mean prefactor A (s^{-1})
200	11.044 (11.039, 11.049)	0.428 (0.425, 0.431)	1.34e+11
250	11.067 (11.062, 11.072)	0.473 (0.469, 0.477)	1.36e+11
300	11.093 (11.087, 11.099)	0.523 (0.519, 0.527)	1.40e+11

Table B.16 Lognormal distribution fits to E_A for CS_2 void to cage hops

T	μ (95% CI)	σ (95% CI)	mean prefactor A (s^{-1})
200	10.793 (10.787, 10.799)	0.515 (0.511, 0.519)	3.95e+11
250	10.822 (10.816, 10.828)	0.569 (0.565, 0.574)	3.96e+11
300	10.863 (10.856, 10.870)	0.617 (0.612, 0.622)	3.87e+11

Table B.17 Lognormal distribution fits to E_A for SF_6 cage to void hops

T	μ (95% CI)	σ (95% CI)	mean prefactor A (s^{-1})
200	11.851 (11.847, 11.855)	0.389 (0.386, 0.392)	2.14e+11
250	11.871 (11.867, 11.876)	0.430 (0.427, 0.434)	2.11e+11
300	11.890 (11.885, 11.895)	0.474 (0.471, 0.478)	2.12e+11

Table B.18 Lognormal distribution fits to E_A for SF_6 void to cage hops

T	μ (95% CI)	σ (95% CI)	mean prefactor A (s^{-1})
200	11.529 (11.524, 11.535)	0.512 (0.508, 0.516)	5.96e+11
250	11.559 (11.552, 11.565)	0.567 (0.563, 0.572)	5.87e+11
300	11.599 (11.592, 11.606)	0.611 (0.606, 0.616)	5.61e+11

Figure B.1 shows the Arrhenius prefactor A as a function of E_A for Xe cage to void hops at 300 K.

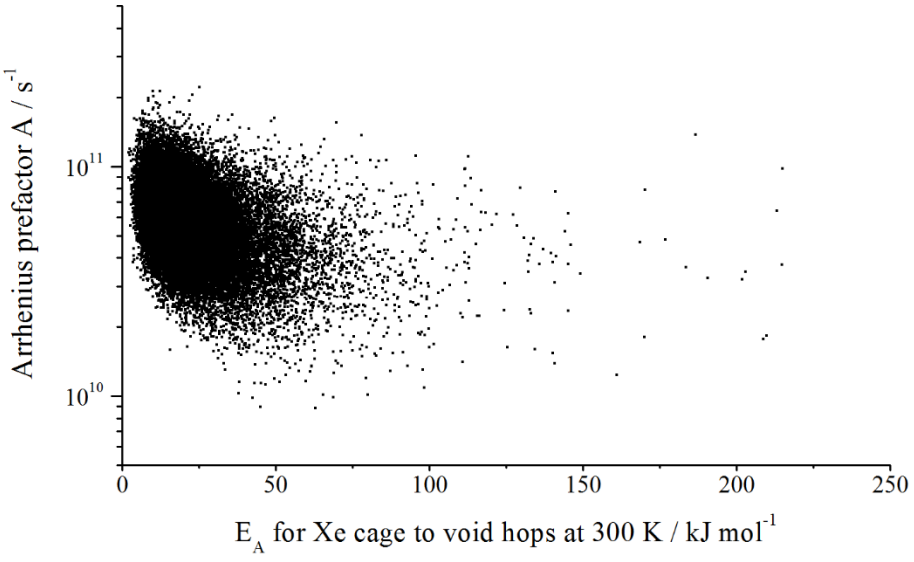

Figure B.1 Arrhenius prefactor vs. E_A for Xe cage to void hops at 300 K

There is no significant covariance here so the mean prefactor of 6.08×10^{10} s^{-1} was used in the hopping rate calculation (Eq. 6 in the text). The covariances between A and E_A were assumed to negligible for the other activation energy distributions shown in Tables S7-S18.

Population variance of ILS hopping rate distributions

The cumulative distribution function for hopping rates (k) as a function of μ, σ, and A is given by:

$$F_k(k) = \frac{1}{2} + \frac{1}{2}\mathrm{erf}\left[\frac{\ln x - \mu}{\sqrt{2}\sigma}\right] \text{ where } x = -RT\ln\left(\frac{k}{\bar{A}}\right) \quad \textbf{B11}.$$

From this cumulative distribution function, the population variance of rates is given by:

$$\sigma_k^2 = 2\int_0^\infty k\big(1 - F_k(k)\big)dk - \left(\int_0^\infty 1 - F_k(k)\,dk\right)^2 \quad \textbf{B12}.$$

4. Direct MD data

The diffusion coefficients of Kr, CH_4, and CO_2 were evaluated by NVT MD at 300 K as described in section Chapter 4 of the text. In each case, the MSD vs. time curve was renormalized to zero by subtracting 4 $Å^2$ from the MSD, which is associated with the ballistic movement of adsorbate atoms within individual cage molecules. The diffusivity was found using the Einstein equation:

$$D_s = \frac{MSD}{6t} \quad \textbf{B13}.$$

The MSD plots for Kr, CH_4, and CO_2 are shown in Figures B.2-B.4 below. Each MSD within the figure insets is given in units of $Å^2\ s^{-1}$. These MD simulations for each adsorbate were averaged over 28 independent simulations. The standard uncertainties in the mean MSDs of each adsorbate were calculated by dividing these 28 simulations into 4 blocks of 7 independent simulations. This yields $8.16\pm0.45\times10^{11}$ $Å^2\ s^{-1}$ ($D_s = 1.36\pm0.076\times10^{-5}\ cm^2\ s^{-1}$) for Kr, $1.56\pm0.016\times10^{12}$ $Å^2\ s^{-1}$ ($D_s = 2.60\pm0.027\times10^{-5}\ cm^2\ s^{-1}$) for CH_4, and $6.55\pm0.30\times10^{11}$ $Å^2\ s^{-1}$ ($D_s = 1.09\pm0.05\times10^{-5}\ cm^2\ s^{-1}$) for CO_2. These uncertainties from block averaging are somewhat larger than suggested by the 95% confidence intervals on fits to the Einstein equation.

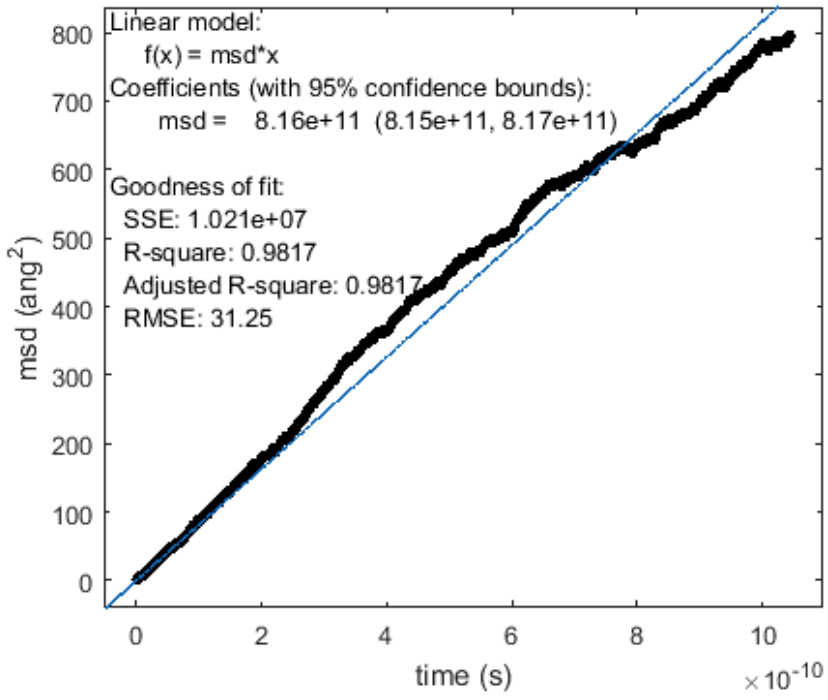

Figure B.2 Kr diffusion by NVT MD at 300 K

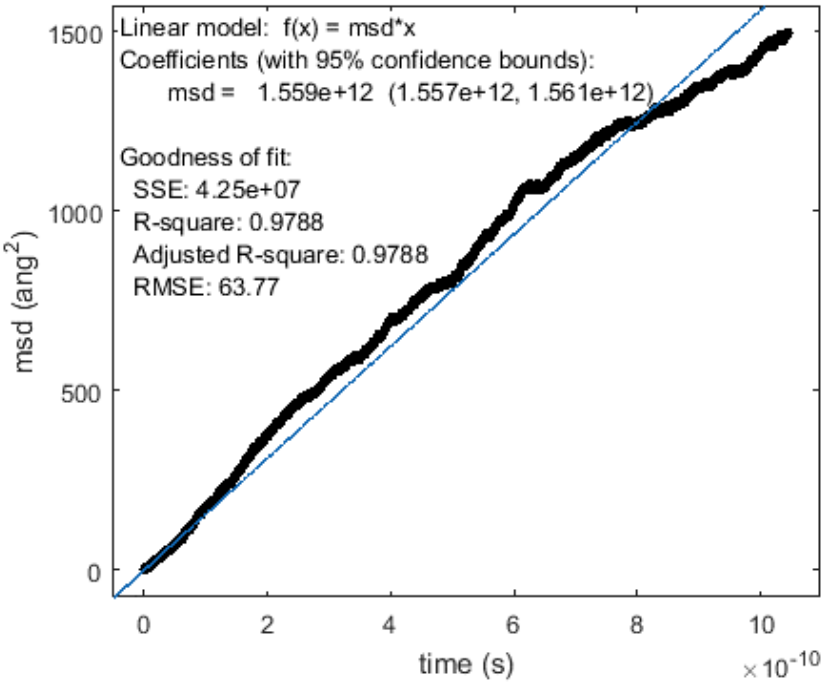

Figure B.3 CH_4 diffusion by NVT MD at 300 K

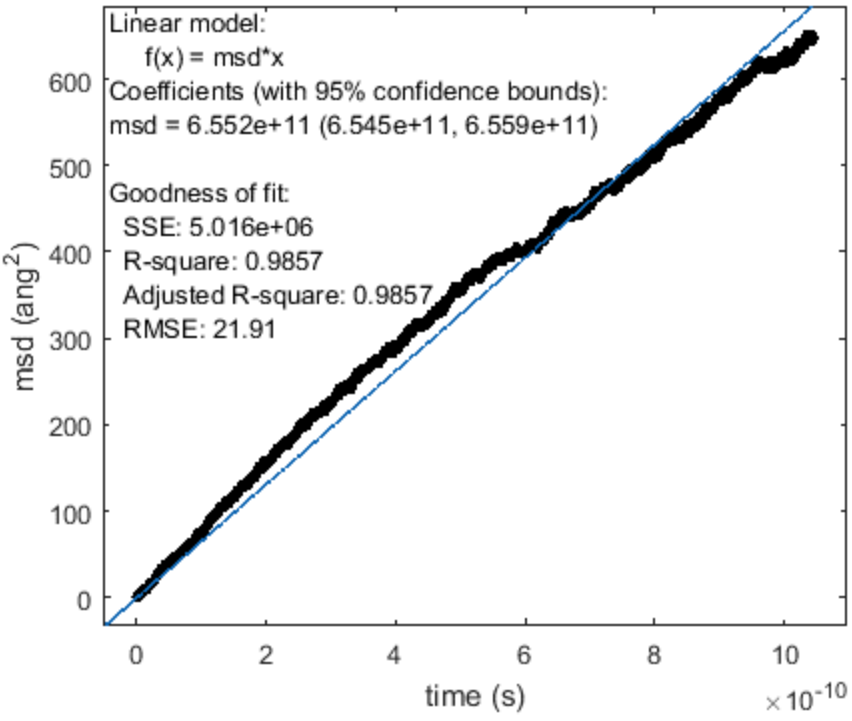

Figure B.4 CO_2 diffusion by NVT MD at 300 K

References

[1] D. Holden, K. E. Jelfs, A. I. Cooper, A. Trewin, D. J. Willock, *Journal of Physical Chemistry C* **2012**, *116*, 16639-16651.

[2] H. Sun, *Macromolecules* **1995**, *28*, 701-712.

[3] D. Holden, K. E. Jelfs, A. Trewin, D. J. Willock, M. Haranczyk, A. I. Cooper, *The Journal of Physical Chemistry C* **2014**, *118*, 12734-12743.

[4] L. Chen, P. S. Reiss, S. Y. Chong, D. Holden, K. E. Jelfs, T. Hasell, M. A. Little, A. Kewley, M. E. Briggs, A. Stephenson, K. M. Thomas, J. A. Armstrong, J. Bell, J. Busto, R. Noel, J. Liu, D. M. Strachan, P. K. Thallapally, A. I. Cooper, *Nature Materials* **2014**, *13*, 954-960.

[5] R. A. Svehla, *NASA: Estimated Viscosities and Thermal Conductivities of Gases at High Temperatures* **1962**.