

# Property Screening of Porous Organic Molecules

Marcin Miklitz

Thesis submitted for the degree of Doctor of Philosophy

**Imperial College  
London**

Department of Chemistry

April 2018

The work presented in the thesis is my own and it is the one upon which I expect to be examined.  
Where information has been derived by other sources it has been referenced accordingly.

Marcin Miklitz

London, April 2018.

The copyright of this thesis rests with the author and is made available under a Creative Commons Attribution Non-Commercial No Derivatives licence. Researchers are free to copy, distribute or transmit the thesis on the condition that they attribute it, that they do not use it for commercial purposes and that they do not alter, transform or build upon it. For any reuse or redistribution, researchers must make clear to others the licence terms of this work.

# Abstract

Porous organic molecules have internal pores readily occupied by gases, solvent or other guests. These molecules can form porous molecular materials with possible application in storage and separation. In this thesis, the properties of isolated porous organic molecules are used as a proxy to the bulk or in-solution applications. As a result, software was first developed for the automated and precise structural characterisation of porous organic molecules. This allows one to easily calculate window diameters to study the thermal window size fluctuations and predict guest diffusion in the bulk. A screening of previously reported porous organic molecules for the application of Xe/Kr separation allowed the most promising material, Noria, to be identified. This was possible with a combination of molecular modelling, electronic structure calculations and structural analysis using the developed software. The experimental Xe/Kr selectivity of Noria, not previously considered for this application, was shown to be comparable with the best performing porous materials. Next, eight synthetically realised porous organic cages were studied as possible C<sub>60</sub> fullerene encapsulants. The relative orientation of the C<sub>60</sub> fullerene in the pore was shown to have little to no effect on the binding energy and the encapsulation of the C<sub>60</sub> during cage formation was determined as the likely mechanism of encapsulation. Lastly, a function-led material design approach was developed. An evolutionary algorithm was used to generate possible C<sub>60</sub> encapsulants from a database of precursors. The resulting porous organic cages are structurally similar to some recently synthetically realised cages and some found in the literature. In summary, presented in this thesis is software, a methodology and results that can further advance the computational function-led materials discovery for specific applications.

# Contents

<b>Abstract</b>	<b>3</b>
<b>Contents</b>	<b>4</b>
<b>Abbreviations</b>	<b>8</b>
<b>Atom Colouring Legend</b>	<b>10</b>
<b>Acknowledgments</b>	<b>11</b>
<b>List of Publications</b>	<b>13</b>
<b>List of Figures</b>	<b>14</b>
<b>List of Tables</b>	<b>17</b>
<b>1 Porous Molecular Materials and Their Applications</b>	<b>18</b>
1.1 Introduction to porous materials . . . . .	19
1.2 Porous molecular materials . . . . .	21
1.2.1 Extrinsic and intrinsic porosity . . . . .	21
1.2.2 Molecular pores and windows . . . . .	22
1.3 Belt-like molecules: cyclodextrins, cucurbiturils and noria . . . . .	22
1.4 Porous organic cages . . . . .	24
1.4.1 Properties of porous organic cages . . . . .	24
1.4.2 Porous liquids . . . . .	28
1.4.3 Possible applications of porous organic cages . . . . .	29
1.5 Computational studies of porous materials . . . . .	32
1.5.1 Role of flexibility in porous materials . . . . .	36
1.6 Adsorption and separation of gases in porous materials . . . . .	38
1.7 The aims and objectives of this PhD thesis . . . . .	40
<b>2 Theoretical Methods</b>	<b>41</b>
2.1 Introduction to computational chemistry . . . . .	42

2.2	Structural characterisation of porous materials . . . . .	43
2.3	Force field potentials . . . . .	46
2.4	Molecular dynamics . . . . .	50
2.5	Metadynamics . . . . .	50
2.6	Density functional theory . . . . .	54
2.7	Grand canonical Monte Carlo . . . . .	56
2.8	Ideal adsorbed solution theory . . . . .	58
2.9	Evolutionary algorithms . . . . .	60
<b>3</b>	<b>Development of pywindow for the Analysis of Structural Properties of Molecular Pores</b>	<b>63</b>
3.1	Introduction . . . . .	64
3.2	Methodology of the <code>pywindow</code> software . . . . .	65
3.2.1	Nomenclature . . . . .	65
3.2.2	Workflow . . . . .	66
3.2.3	Reconstruction of the periodic unit cell . . . . .	67
3.2.4	Structural properties of molecular pores . . . . .	68
3.3	Validation . . . . .	73
3.3.1	Computational details . . . . .	73
3.3.2	Window diameters and number of windows . . . . .	74
3.3.3	Analysis of molecular dynamics trajectories . . . . .	77
3.4	Limitations of <code>pywindow</code> . . . . .	78
3.5	Conclusions . . . . .	79
<b>4</b>	<b>Screening of Molecular Pores for Xenon/Krypton Separation</b>	<b>81</b>
4.1	Introduction . . . . .	82
4.2	Methods . . . . .	83
4.2.1	The Cage Database ( <i>CDB</i> ) . . . . .	83
4.2.2	Structural analysis . . . . .	89
4.2.3	Binding energy calculations . . . . .	90
4.2.4	Pore limiting envelopes . . . . .	91
4.2.5	Free energy barriers to diffusion . . . . .	92
4.2.6	Xe/Kr selectivity calculations . . . . .	93
4.3	Results and discussion . . . . .	93
4.3.1	Molecular <i>vs.</i> solid state behaviour . . . . .	93
4.3.2	Structural analysis . . . . .	95
4.3.3	Guest binding energies . . . . .	95
4.3.4	Pore limiting envelopes . . . . .	100
4.3.5	Free energy barriers to diffusion . . . . .	103
4.3.6	Computational prediction . . . . .	104

4.3.7	Experimental testing . . . . .	105
4.4	Conclusions . . . . .	108
<b>5</b>	<b>C<sub>60</sub> Fullerene Encapsulation in Porous Organic Cages</b>	<b>110</b>
5.1	Introduction . . . . .	111
5.2	Methods . . . . .	113
5.2.1	Generation of the cage set . . . . .	113
5.2.2	DFT and force field C <sub>60</sub> @POC binding energies . . . . .	114
5.2.3	C <sub>60</sub> @ <b>A3</b> complexes with different relative host-guest orientation . . . . .	115
5.2.4	Molecular dynamics study of <b>B4</b> cage and C <sub>60</sub> @ <b>B4</b> complex. . . . .	116
5.3	Results . . . . .	117
5.4	Conclusions . . . . .	123
<b>6</b>	<b>Function-Led Computational Design of New C<sub>60</sub> Encapsulants with an Evolutionary Algorithm</b>	<b>125</b>
6.1	Introduction . . . . .	126
6.2	Methods . . . . .	129
6.2.1	The database of precursors and cage assembly process . . . . .	130
6.2.2	Fitness function . . . . .	132
6.2.3	Fitness function parameterisation . . . . .	133
6.2.4	Evolutionary algorithm calculations . . . . .	135
6.2.5	Post-processing of the results . . . . .	137
6.3	Results . . . . .	137
6.3.1	The database of precursors . . . . .	137
6.3.2	The database of assembled cages . . . . .	137
6.3.3	The parameterisation of fitness function . . . . .	140
6.3.4	The EA calculations . . . . .	142
6.3.5	Derived Cages . . . . .	142
6.4	Discussion . . . . .	147
6.5	Conclusions . . . . .	149
<b>7</b>	<b>Summary and Perspectives</b>	<b>151</b>
	<b>Bibliography</b>	<b>155</b>
	<b>Appendix</b>	<b>175</b>
A.1	The <code>pywindow</code> software window analysis schematics . . . . .	176
A.2	Solid state structures of <b>CC3</b> polymorphs . . . . .	177
A.3	DFT binding energies for <i>CDB26</i> . . . . .	178
A.4	Structural overlays of <b>NC1</b> and <b>NC2</b> . . . . .	180
A.5	Experimental and DFT void diameters for <i>CDB12</i> . . . . .	181

A.6	Free energy barriers for <i>CDB12</i> . . . . .	182
A.7	Experimental details for Chapter 4 . . . . .	184
A.8	Binding energies and cavity diameters for $C_{60}$ @POC complexes . . . . .	192
A.9	The NMR spectra for <b>B</b> -series $C_{60}$ @POC complexes . . . . .	193
A.10	The trialdehyde building blocks and diamine linkers . . . . .	199

# Abbreviations











<b>ASA</b>	Accessible Surface Area
<b>BET</b>	Brunauer–Emmett–Teller
<b>BSSE</b>	Basis Set Superposition Error
<b>CB</b>	Cucurbituril
<b>CD</b>	Cyclodextrin
<b>CIF</b>	Crystallographic Information File
<b>COF</b>	Covalent Organic Framework
<b>COM</b>	Centre Of Mass
<b>CSD</b>	Cambridge Structural Database
<b>CV</b>	Collective Variable
<b>DCC</b>	Dynamic Covalent Chemistry
<b>DF</b>	Diameter Fraction
<b>DFT</b>	Density Functional Theory
<b>DZVP</b>	Double Zeta Valence Potential
<b>EA</b>	Evolutionary Algorithm
<b>FES</b>	Free Energy Surface
<b>GA</b>	Genetic Algorithm
<b>GCMC</b>	Grand Canonical Monte Carlo
<b>GGA</b>	Generalised Gradient Approximation
<b>IAST</b>	Ideal Adsorbed Solution Theory
<b>IUPAC</b>	International Union of Pure and Applied Chemistry
<b>KDE</b>	Kernel Density Estimate
<b>LCD</b>	Largest Cavity Diameter
<b>LDA</b>	Local Density Approximation
<b>MC</b>	Monte Carlo
<b>MD</b>	Molecular Dynamics
<b>MM</b>	Molecular Mechanics
<b>MOF</b>	Metal-Organic Framework
<b>OOP</b>	Object-Oriented Programming
<b>OPLS</b>	Optimised Potentials for Liquid Simulations
<b>PC</b>	Packing Coefficient



<b>PF</b>	Pore Fraction
<b>PLD</b>	Pore Limiting Diameter
<b>PLE</b>	Pore Limiting Envelope
<b>POC</b>	Porous Organic Cage
<b>PMM</b>	Porous Molecular Material
<b>POM</b>	Porous Organic Molecule
<b>PXRD</b>	Powder X-ray Diffraction
<b>QM</b>	Quantum Mechanics
<b>RMS</b>	Root Mean Square
<b>RMSD</b>	Root Mean Square Displacement
<b>SCF</b>	Kohn-Sham Self-Consistent Field
<b>SC-XRD</b>	Single Crystal X-ray Diffraction
<b>TraPPE</b>	Transferable Potentials for Phase Equilibria
<b>TZVP</b>	Triple Zeta Valence Potential
<b>UFF</b>	Universal Force Field
<b>UNF</b>	Used Nuclear Fuel
<b>ZIF</b>	Zeolitic Imidazolate Framework
<b>1D</b>	1-Dimensional
<b>2D</b>	2-Dimensional
<b>3D</b>	3-Dimensional

# Atom Colouring Legend

The adopted in this thesis atom colouring scheme, with the coloured spheres and the corresponding element listed, is as follows:

	Hydrogen
	Boron
	Carbon
	Nitrogen
	Oxygen
	Fluorine
	Sulfur
	Bromine
	Krypton
	Xenon

# Acknowledgments

I would like to express my deepest gratitude to my supervisor Dr Kim Jelfs for being an amazing mentor, for entrusting this PhD project with me and for all the guidance and patience during the difficult parts. This thesis would not be possible without the support.

I would like to thank all past and current members of the Jelfs group. In particular, Valentina Santolini for being there during the whole PhD journey sharing all the ups and down together; James Blood for being a good friend, support and thesis proof reading; Dr Enrico Berardo for the mentoring, discussions and mental guidance that allowed me to succeed. Additionally, Julia Schmidt for making me laugh constantly; Sophie Finnigan and Edward Jackson for the help proof reading this thesis and finally, Lukas Turcani for his coding wisdom and contributions.

I would like to thank Dr João Malhado and Dr Giuseppe Mallia for making my demonstrating experience unforgettable and for all the support.

My thanks to Prof. Andrew Cooper and his group at the University of Liverpool, for successful collaborations. Especially, Dr Rebecca Greenaway and Dr Michael Briggs for the contributions to the work presented in this thesis and for the insightful discussions throughout my PhD; Dr Shan Jiang and Mr. Rob Clowes for the experimental work that also contributed this thesis.

I would like to thank Dr Gareth Tribello (Queen's University Belfast), for the help with understanding the mysteries of metadynamics and Dr Chin Yong (Daresbury Laboratory) for the guidance using DL\_FIELD software.

I am very grateful to Prof. Jerzy Lisowski from University of Wrocław, for introducing me to the world of porous organic cages that led me towards this PhD.

To all those close and dear: Mateusz, Monika, Kamila, Gosia, Damian, Wioletta, Aleksandra, Oskar and Patrycja, thank you for believing in me. These three years would have been much harder without our moments together. Thank you all for always being there for me. There are many others not mentioned here, but who I thank as sincerely.

At last, I would like to express my gratitude to my parents for all the support I have received on their part that made my PhD project successful.

Marcin

# List of Publications

The following published and unpublished work I authored has contributed to the content of this thesis:

M. Miklitz and K. E. Jelfs, pywindow: a Tool for Structural Characterisation of Molecular Pores, *unpublished work*. Github: <https://github.com/marcinmiklitz/pywindow>

M. Miklitz, S. Jiang, R. Clowes, M. E. Briggs, A. I. Cooper and K. E. Jelfs, Computational Screening of Porous Organic Molecules for Xenon/Krypton Separation, *J. Phys. Chem. C*, 2017, **121**, 15211–15222.

V. Santolini, M. Miklitz, E. Berardo and K. E. Jelfs, Topological Landscapes of Porous Organic Cages, *Nanoscale*, 2017, **9**, 5280–5298.

T. Hasell, M. Miklitz, A. Stephenson, M. A. Little, S. Y. Chong, R. Clowes, L. Chen, D. Holden, G. A. Tribello, K. E. Jelfs and A. I. Cooper, Porous Organic Cages for Sulfur Hexafluoride Separation, *J. Am. Chem. Soc.*, 2016, **138**, 1653–1659.

R. Manurung, D. Holden, M. Miklitz, L. Chen, T. Hasell, S. Y. Chong, M. Haranczyk, A. I. Cooper and K. E. Jelfs, Tunable Porosity through Cooperative Diffusion in a Multicomponent Porous Molecular Crystal, *J. Phys. Chem. C*, 2015, **119**, 22577–22586.

# List of Figures

1.1	Examples of porous materials. . . . .	20
1.2	Extrinsic and intrinsic porosity. . . . .	22
1.3	Molecular structures of cyclodextrin, cucurbituril and noria. . . . .	23
1.4	The synthetic route and examples of porous organic cages. . . . .	25
1.5	Porous networks in <b>CC3</b> polymorphs. . . . .	27
2.1	The structural descriptors of a porous network. . . . .	44
2.2	The visualisation of the circumcircle on <b>CC3</b> . . . . .	46
2.3	The Free Energy Surface sampling in metadynamics. . . . .	52
3.1	The workflow schematics of the <code>pywindow</code> software. . . . .	67
3.2	Examples of rebuild periodic unit cells. . . . .	69
3.3	Schematics of finding window diameters. . . . .	71
3.4	Atom sets used to define windows in <b>CC3</b> . . . . .	75
3.5	The <code>pywindow</code> validation. . . . .	76
3.6	Pore limiting envelopes generated using <code>pywindow</code> and circumcircle method. . .	77
3.7	Molecular pore in which the cavity is the pore limiting diameter . . . . .	78
3.8	Spherical window approximation in windows of various shape. . . . .	79
4.1	The key stages of screening molecular pores of promise for Xe/Kr separation. . .	84
4.2	The DFT binding energies calculated for the noble gases in the host molecules of <i>CDB26</i> . . . . .	98
4.3	The molecular pores that form <i>CDB12</i> . . . . .	99
4.4	The pore limiting envelopes for <b>CC3</b> polymorphs and <b>CC3</b> single cage . . . . .	101

4.5	The pore limiting envelopes for <i>CDB12</i> . . . . .	102
4.6	Free energy barriers to diffusion of Xe and Kr for <i>CDB12</i> . . . . .	104
4.7	The experimental Xe and Kr adsorption measurements for <b>CC3<math>\alpha</math>-R</b> , <b>aCC3-R</b> , <b>aNC1</b> and <b>aNC2</b> . . . . .	106
4.8	The space filling models of single Xe and Kr placed in the cavity of <b>CC3</b> , <b>NC1</b> <b>NC2</b> . . . . .	107
4.9	The calculated selectivities for a Xe/Kr mixture uptake. . . . .	108
5.1	Examples of C <sub>60</sub> encapsulants. . . . .	112
5.2	The eight potential POCs for C <sub>60</sub> encapsulation application. . . . .	114
5.3	The DFT and OPLS3 binding energies for the C <sub>60</sub> @POC complexes and the calculated cavity diameters . . . . .	118
5.4	The DFT binding energies for the relative host-guest configurations of C <sub>60</sub> @ <b>A3</b> complex. . . . .	120
5.5	The relative C <sub>60</sub> @ <b>A3</b> host-guest orientation before and after DFT geometry optimisation. . . . .	121
5.6	The relative C <sub>60</sub> @ <b>A3</b> host-guest orientation before and after OPLS3 geometry optimisation and the OPLS3 binding energies . . . . .	122
5.7	The window and cavity diameters of an empty <b>B4</b> cage and C <sub>60</sub> @ <b>B4</b> complex from MD simulations . . . . .	123
6.1	The analysis of the tritopic aldehyde <b>bbs</b> and ditopic amine <b>lks</b> . . . . .	138
6.2	The analysis of the assembled database of cages and the corresponding C <sub>60</sub> complexes. . . . .	139
6.3	The fitness function parameterisation. . . . .	141
6.4	The evolution of the fitness values in the EA calculations. . . . .	143
6.5	The twenty cages with highest fitness value from the combined results of the five EA calculations. . . . .	144
6.6	The features of the final cages in comparison to the features of the database of pre-screened cages. . . . .	147

A.1	Process of identifying windows in molecular pores . . . . .	176
A.2	Process of finding the coordinates of a window true centre. . . . .	176
A.3	Polymorphs of <b>CC3</b> . . . . .	177
A.4	DFT binding energies compared to void diameters . . . . .	179
A.5	Structural overlays of <b>NC1</b> and <b>NC2</b> . . . . .	180
A.6	Free energy barriers <i>vs.</i> void diameters . . . . .	183
A.7	<sup>1</sup> H NMR of <b>NC1</b> . . . . .	185
A.8	<sup>1</sup> H NMR of <b>NC2</b> . . . . .	185
A.9	<sup>1</sup> H NMR of <b>aCC3</b> . . . . .	186
A.10	PXRD of <b>NC1</b> . . . . .	187
A.11	PXRD of <b>NC2</b> . . . . .	187
A.12	PXRD of <b>CC3<math>\alpha</math></b> and <b>aCC3</b> . . . . .	188
A.13	Xe and Kr sorption isotherms for <b>aNC1</b> . . . . .	189
A.14	Xe and Kr sorption isotherms for <b>aNC2</b> . . . . .	190
A.15	Xe and Kr sorption isotherms for <b>aCC3</b> . . . . .	190
A.16	pyIAST adsorption isotherms fitting . . . . .	191
A.17	<sup>1</sup> H NMR spectra for the attempted formation of <b>B1</b> in the presence of C <sub>60</sub> . . .	193
A.18	<sup>1</sup> H NMR spectra for the attempted formation of <b>B2</b> in the presence of C <sub>60</sub> . . .	194
A.19	<sup>1</sup> H NMR spectra for the attempted formation of <b>B3</b> in the presence of C <sub>60</sub> . . .	195
A.20	<sup>1</sup> H NMR spectra for the attempted formation of <b>B4</b> in the presence of C <sub>60</sub> . . .	196
A.21	<sup>1</sup> H NMR spectra of the aromatic regions for the attempted formation of <b>B4</b> in the presence of C <sub>60</sub> . . . . .	197
A.22	<sup>1</sup> H NMR spectra of the aromatic regions for the attempted formation of <b>B1</b> , <b>B2</b> , <b>B3</b> in the presence of C <sub>60</sub> . . . . .	198
A.23	The trialdehyde building blocks used in the assembly of POCs - part 1. . . . .	199
A.24	The trialdehyde building blocks used in the assembly of POCs - part 2. . . . .	200
A.25	The diamine linkers used in the assembly of POCs - part 1. . . . .	201
A.26	The diamine linkers used in the assembly of POCs - part 2. . . . .	202



# List of Tables

3.1	Comparison of different methods for window diameters calculation. . . . .	75
3.2	The window diameters and the total number of windows identified for various molecular pores. . . . .	76
4.1	A list of molecular pores in <i>CDB41</i> . . . . .	85
4.2	Details of the <b>CC3</b> polymorph analysis. . . . .	92
4.3	Structural analysis for <i>CDB41</i> . . . . .	96
4.4	DFT binding energies for single noble gas molecules in <b>CC3</b> polymorphs. . . . .	97
4.5	Candidates identified for Xe/Kr separation ( <i>CDB12</i> ). . . . .	99
4.6	The pore limiting envelope analysis for <b>CC3</b> polymorphs and <b>CC3</b> cage. . . . .	101
4.7	Analysis of pore limiting envelopes for Kr and Xe diffusion for <i>CDB12</i> . . . . .	103
4.8	The measured and literature experimental Xe/Kr selectivities for selected materials. . . . .	108
6.1	The aldehyde <b>bbs</b> and amine <b>lks</b> used to assemble the final population of cages. . . . .	145
6.2	The fitness values and the unscaled properties of the final cages. . . . .	146
A.1	DFT Xe and Kr binding energies for <i>CDB26</i> . . . . .	178
A.2	Experimental and DFT void diameters for <i>CDB12</i> . . . . .	181
A.3	Free energy barriers for <i>CDB12</i> . . . . .	182
A.4	DFT and OPLS3 derived binding energies for C <sub>60</sub> @POC complexes. . . . .	192
A.5	DFT and OPLS3 calculated $D_{void\_opt}$ for empty cages and C <sub>60</sub> @POC complexes. . . . .	192

# Chapter 1

## Porous Molecular Materials and Their Applications

In this chapter, I will give an introduction to the field of porous materials with the main focus on the porous molecular materials that are the subject of this thesis. Porous organic molecules, *i.e.* the belt-like molecules and the porous organic cages, will be introduced and the historical advances, the relevant examples, and the current prospects will be reviewed. Aspects related to the application of these materials in the separation of gases such as porosity, adsorption, and the effects of the materials flexibility on the related properties, will also be discussed.

## 1.1 Introduction to porous materials

The International Union of Pure and Applied Chemistry (IUPAC) defines *porous materials* as solids with pores in the form of cavities and channels.<sup>1</sup> Therefore, *porosity* is a function of a porous material's texture that is dependent on the presence of the pore space. As tiny cavities occur in all solids, it is customary that only materials with pores large enough for applications, meaning that atoms, ions or molecules can fit and access these pores, should be referred to as porous. IUPAC provides a classification of the porosity types that depend on the pore size. Microporosity is a result of the presence of pores not exceeding 2 nm in width, namely micropores. Mesopores fall in the size range of 2 to 50 nm and result in mesoporosity, whereas pores exceeding 50 nm are referred to as macropores and result in macroporosity. The term nanopores is also used and embraces the three mentioned classes of pores collectively, with an upper limit for the pore size of 100 nm.<sup>1</sup>

Traditional porous materials, such as crystalline aluminosilicate zeolites and porous carbons, also known as activated carbon, obtained in the processing of charcoal to introduce pore space, have found use in industry.<sup>2</sup> These are low cost materials that either occur naturally, like zeolites, or are derived from cheap charcoal. The application of these materials can utilise their sponge-like behaviour, where liquid or gas is absorbed by the material, and sieve-like, where separation of molecular mixtures is possible.<sup>3</sup> Zeolites also have catalytic properties.<sup>4</sup> However, the issue of high energy consumption associated with the separation of some liquid and gaseous mixtures, based on distillation, remains. These processes are believed to account for 10-15% of the world's total energy consumption.<sup>5</sup> In search of possible solutions, and to mimic the behaviour of the traditional porous materials, a new generation of synthetic porous materials, such as metal-organic frameworks (MOFs),<sup>6</sup> covalent organic frameworks (COFs),<sup>7</sup> and zeolitic imidazolate frameworks (ZIFs),<sup>8</sup> emerged. Examples of a zeolite, a model of porous carbon, MOF, COF and ZIF are shown in Figure 1.1. These synthetic materials can highly exceed the performance of zeolites and porous carbons in terms of their surface area, specificity expressed by selective binding of mixture components and the topology, defined as the local arrangement of the building blocks that repeats periodically, allowing for well controlled and tunable pore sizes suitable for applications.<sup>9-11</sup> The possible applications in gas storage and

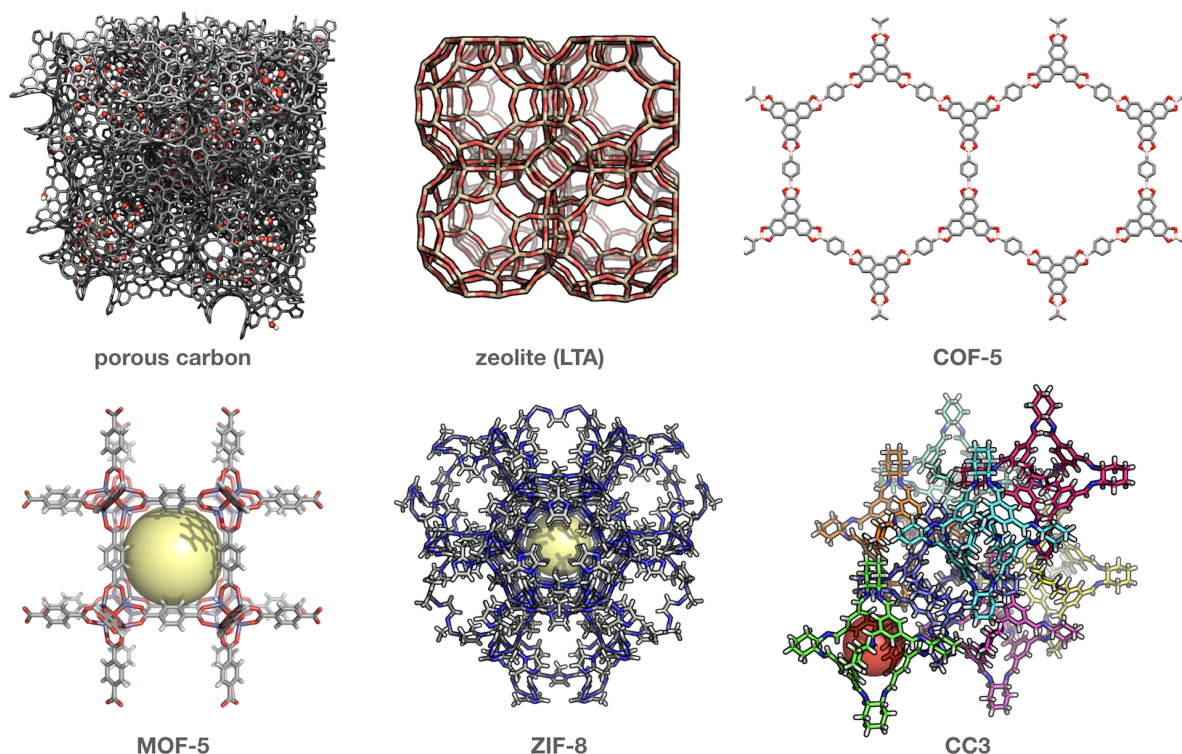


Figure 1.1: Selected examples of porous materials. From top left: porous carbon (with the water molecules adsorbed to the surface),<sup>14</sup> zeolite with LTA topology (silicon - pale yellow),<sup>15</sup> and covalent organic framework (COF-5).<sup>16</sup> From bottom left: metal organic framework with a transparent yellow sphere placed at the pore to represent the void (MOF-5, zinc - violet),<sup>17</sup> zeolitic imidazolate framework (ZIF-8) with a transparent yellow sphere placed at the central pore,<sup>18</sup> and a porous organic cage **CC3**, each discrete unit has different colouring and a red sphere is placed in one of the pores).<sup>19</sup>

separation, catalysis and energy technologies has resulted in a significant effort being put into the development of these materials.<sup>12</sup> The real-world applications include: gas cylinders filled with a MOF for a controlled toxic gas storage and release, or a product, also utilising a MOF, for a controlled 1-methylcyclopropene release that delays fruit ripening.<sup>13</sup>

These historical porous materials are frameworks that extend indefinitely and the porosity results from the pore space present in their topology. Porous *molecular* materials (PMMs), on the other hand, are composed of discrete units, molecules that are not chemically bonded pack together via weak non-bonding interactions.<sup>20</sup> Cavities can form between the molecules as the result of inefficient packing, or the pores can result from the molecules' 3-dimensional (3D) hollow structure. The right arrangement of the molecules in the bulk allows for the formation of channels connecting the pores that results in a porous material. An example of a PMM built from porous organic cage molecules is shown in the bottom right of Figure 1.1.

## 1.2 Porous molecular materials

The adsorption of gases observed in the crystals of  $\beta$ -Co(4-methylpyridine)<sub>4</sub>(SCN)<sub>2</sub> in 1969 and later in Dianin’s compound in 1976, probably mark the beginning of studies of porosity and adsorption in what we would name today PMMs.<sup>21,22</sup> Unlike other molecular materials of that time, known to form inclusion complexes with solvents, these materials did not collapse upon desolvation and retained the porosity. This was the first time a structural property associated with zeolites was observed in crystals built of molecules, thus Barrer and Shanson called the solid made of Dianin’s molecules an *organic zeolite*.<sup>22</sup> The arrangement of Dianin’s molecules in the crystalline form results in the formation of elongated cavities, that resemble hour-glass shape, ca. 11 Å in length and varying diameter at different lengths of roughly 2.8, 4.0 and 6.4 Å.<sup>23</sup> These cavities, first occupied by solvent, are readily available for gas adsorption after the solvent molecules’ removal. The dependency of the size and shape of the guest molecules on adsorption was shown as well. Today, we know that a variety of inorganic and organic molecules can form PMMs.<sup>24,25</sup> In a recent study, Evans *et al.* screened the Cambridge Structural Database (CSD)<sup>26</sup> for potential permanently porous crystals of organic PMMs.<sup>27</sup> These are rare, with only 481 identified, in comparison to >70,000 MOFs found in the CSD,<sup>28</sup> although solvates and co-crystals were excluded.

### 1.2.1 Extrinsic and intrinsic porosity

Discrete molecules that are *extrinsically* porous do not possess any predefined voids, or internal free molecular volume.<sup>29,30</sup> For example, a rigid  $D_{3h}$ -symmetric triptycenetrisbenzimidazolone scaffold can yield a permanently porous crystalline material with a ribbon-like motif and 1-dimensional (1D) channels (see Figure 1.2a).<sup>31</sup> *Intrinsically* porous molecules, on the other hand, contain clefts or cavities. Examples include calixarenes,<sup>32–34</sup> cucurbiturils,<sup>35,36</sup> cyclodextrins<sup>37</sup> and organic cages.<sup>19,25</sup> A requirement of *shape-persistency* often needs to be fulfilled in these materials, meaning a rigid structure that does not collapse upon desolvation, and the observed material’s porosity is inherited from the discrete molecules’ intrinsic pores. PMMs with both intrinsic and extrinsic porosity are also not uncommon with an example shown in Figure 1.2b.<sup>38–40</sup>

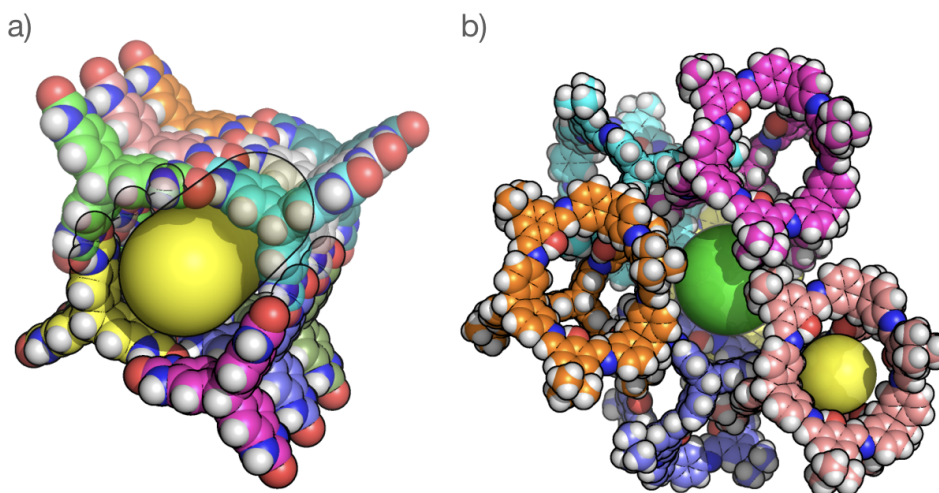


Figure 1.2: a) The example of extrinsic porosity. The assembly of triptycenetrisbenzimidazolone molecules result in 1D channels (yellow tube).<sup>31</sup> b) The assembly of an organic cage molecule with resulting from one-pot [4+6] imine condensation reaction of 2,7,14-trinitrotriptycene and salicylic dialdehyde. Intrinsic porosity is showed as a yellow sphere and resulting from an inefficient packing extrinsic porosity as a green sphere.<sup>39</sup> For clarity, the carbon atoms of each molecule have different colouring, and the two subfigures (a) and (b) are not to scale.

### 1.2.2 Molecular pores and windows

A *molecular pore* is defined as a molecule, and the space confined by it, that meets the requirements of being intrinsically porous.<sup>25</sup> Therefore, from this point onwards *a cavity*, *a void* or simply *a pore*, used interchangeably, will imply that it is the intrinsic space confined by a molecular pore. If the space confined is not fully enclosed by the molecule and is accessible through multiple entry and exit routes, the necking of such a pore channel will be referenced as a *molecular window* or simply a *window*.

## 1.3 Belt-like molecules: cyclodextrins, cucurbiturils and noria

The story of cyclodextrins (CDs) begins in the late 19th century, when it is believed Antoine Villers, the French pharmacist and chemist, extracted  $\alpha$ - and  $\beta$ -cyclodextrin from a fermented potato starch.<sup>37</sup> With  $\gamma$ -cyclodextrin, CDs were shown to readily form inclusion complexes, mostly in solutions, and greatly contributed to the developments of supramolecular host-guest

chemistry. A variety of practical applications of CDs include food, pharmaceutical, cosmetics and chemistry industries. From the structural point of view, CDs are macrocycles built from 5, 6 or 7  $\alpha$ -D-glucopyranoside units with a *belt-like* structure, wider at one side, with two windows and a cavity. Recently, the ability of a  $\beta$ -cyclodextrin derivative to adsorb carbon dioxide ( $\text{CO}_2$ ) in the solid state, was showcased.<sup>41,42</sup> The  $\text{CO}_2$  adsorption capacity was observed to increase with increasing temperature, an unlikely behaviour for conventional adsorbents. The  $\beta$ -cyclodextrin structure is presented in Figure 1.3a.

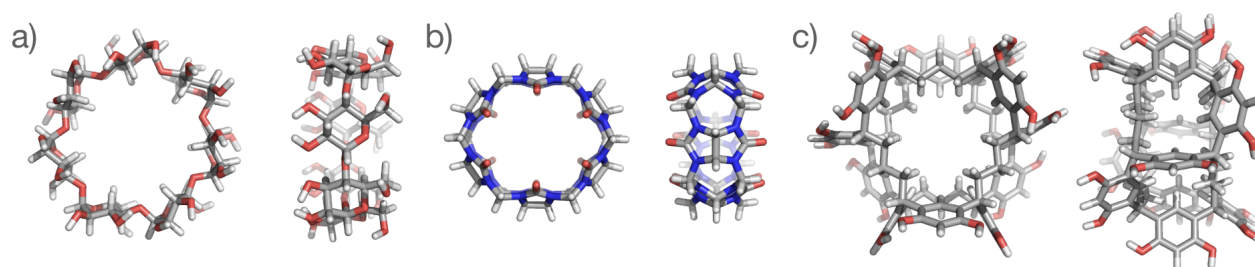


Figure 1.3: The front (left) and side (right) views of a)  $\beta$ -cyclodextrin,<sup>43</sup> b) cucurbit[6]uril,<sup>44</sup> c) noria,<sup>45</sup> molecular structures in stick representation, as extracted from the CSD deposited crystal structures.

Cucurbiturils (CBs) are the condensation product of glycoluril and formaldehyde.<sup>36</sup> CBs also resemble a belt-like structure with a cavity, but unlike CDs, have two symmetrical windows. It is customary to indicate the ring size in square brackets as cucurbit[n]uril or CB[n] depending on the glycoluril's stoichiometry in the condensation reaction. CBs also present a rich host-guest chemistry and application in guest recognition, molecular switches and can impact the distribution of products and reactants in organic synthesis (thermodynamic effects) and the rate of these reactions (kinetic effects).<sup>36</sup> It was shown that cucurbit[6]uril (see Figure 1.3b) can form a porous solid with high thermal stability and permanent porosity with a remarkable acetylene adsorption.<sup>46</sup> An interesting odd-even effect was observed in cucurbit[n]urils ( $n=5-8$ ), where an odd number of glycolurils units in the ring results in an amorphous solid and an even number in a crystalline solid.<sup>44</sup> Nevertheless, most structures have formed either 1D porous channels or 3D porous networks.

In 2006, Nishikubo *et al.* presented a *molecular waterwheel* (noria) molecule from a condensation reaction of resorcinol and an alkanedial.<sup>47</sup> Noria, shown in Figure 1.3c, is a ladder-type oligomer with the structural resemblance of a waterwheel, thus the peculiar name. Studies

on the amorphous solids of noria, carried out by Atwood *et al.*, showed it is porous to gases, with selective sorption of CO<sub>2</sub> over hydrogen (H<sub>2</sub>) and nitrogen (N<sub>2</sub>).<sup>45</sup> Recently, the possible application of noria derivatives as coating materials with extreme ultraviolet resistance was also examined.<sup>48</sup>

## 1.4 Porous organic cages

Porous organic cages (POCs), or *cage compounds*, are polycyclic molecules with the shape of a cage that have 3D structures with three or more molecular windows.<sup>25</sup> This distinguishes them from belt-like molecules, discussed in the previous section, that typically have two windows. The early studies conducted by Cram and Warmuth, on the hemicarceplexes and nanocontainer molecules formed in the condensation reactions of cavitands, laid the foundations for the development of cage compounds.<sup>49–51</sup> Later, in 2008, Skowronek and Gawroński, synthesised a chiral iminospherand in a [4+6] imine condensation reaction (see Figure 1.4a).<sup>52</sup> The same year Mastalerz reported another imine cage (Figure 1.4c) and in 2009 Cooper *et al.* introduced a series of isostructural POCs (Figure 1.4b), including that of Skowronek and Gawronski, known today as Covalent Cage 3 (**CC3**).<sup>19,53</sup> I will now briefly discuss the properties of porous organic cages. For more information regarding POCs, I refer the reader to some excellent reviews in references 26, 55 and 56.

### 1.4.1 Properties of porous organic cages

In the imine condensation reaction, the process of bond formation and breakage stays in equilibrium, a known feature of dynamic covalent chemistry (DCC).<sup>56,57</sup> This feature facilitates the synthesis of cage compounds, circumventing formation of linear oligomers, through the bond “error correction”, that gives the opportunity for the formation of polycyclic molecules with the desired cage-like shape.<sup>19</sup> As a result, experimental realisation of POCs with various stoichiometry, shapes and topologies, was possible.<sup>25</sup> Other DCC reactions that also yield cage compounds include boronic acid condensation,<sup>30</sup> disulfide exchange<sup>58</sup> and alkyne metathesis.<sup>59</sup> Step-wise synthetic routes with irreversible bond formation can also be exploited to form porous



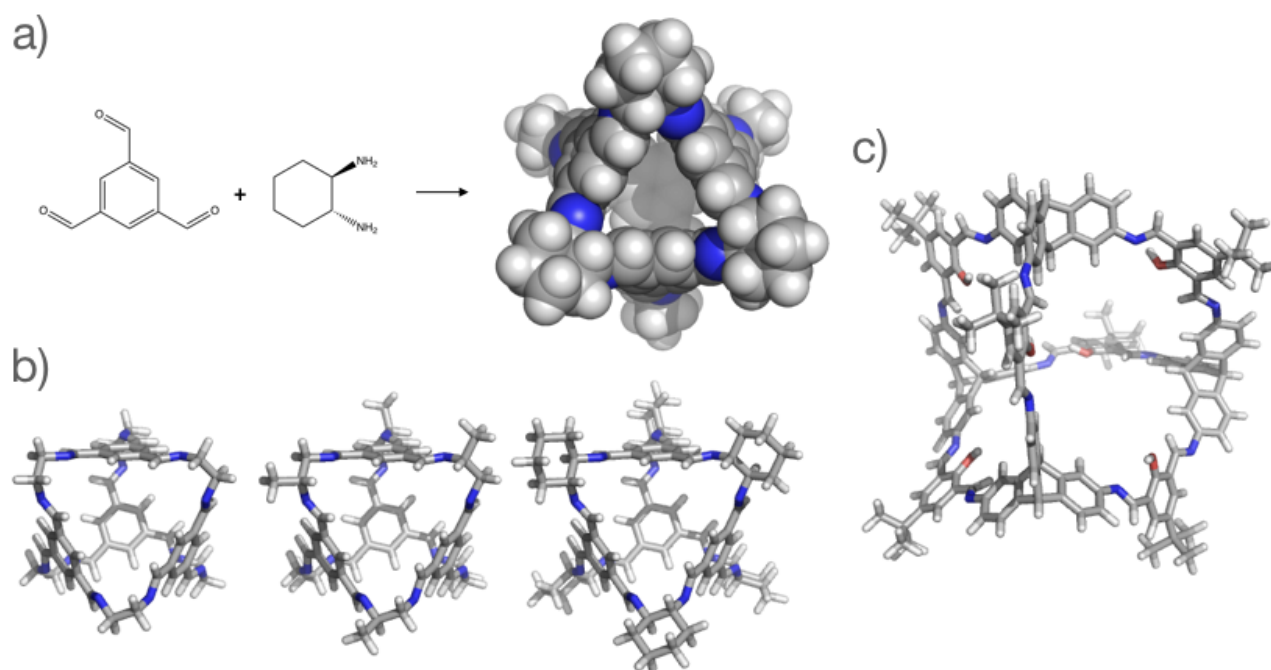


Figure 1.4: a) The simplified reaction scheme and a spacefill structure of a [4+6] iminospherand synthesis reported by Skowronek and Gawroński.<sup>52</sup> b) The stick representation of the molecular structure series of three [4+6] porous organic cages as a result of the reaction of triformylbenzene and 1,2-ethylenediamine (left), 1,2-propylenediamine (centre) and *R,R*-1,2-diaminocyclohexane (right),<sup>19</sup> with common names **CC1**, **CC2** and **CC3**, respectively. c) The stick representation of the the imine cage reported by Mastalerz.<sup>53</sup>

organic cages, as in the synthesis of all-benzene cages with a 3-fold homocoupling reaction.<sup>60</sup> However, the irreversible bonds and multi-step synthesis often result in lower yields, therefore POCs obtained this way are rare.<sup>25</sup>

The discrete character of cage molecules, that results in their enhanced solubility, allows for post-synthetic processing.<sup>61,62</sup> The reduction of reversible imine bonds to irreversible amine bonds can be used to improve the chemical stability of imine POCs by utilising amine chemistry.<sup>53,63</sup> This can, however, impair shape-persistency of these molecules, with greater degrees of freedom for rotation of the amine bond leading to loss of intrinsic porosity. Liu *et al.*, to prevent the possible collapse of cages, have “tied” pairs of the neighbouring amine bonds by binding acetone or formaldehyde to them.<sup>64</sup> This also resulted in an improved stability, high resistance towards acidic and basic conditions in a broad pH range (1.7-12.3), and the same approach improved hydrolytic stability of the reduced and “tied” **CC3**, which could withstand water vapour and boiling for up to 4 hours.<sup>65</sup> The fact that the synthesis of molecular cages can

be separated from the formation of the final material, allows further post-synthetic modifications. Functional groups can be introduced to control the pore size and its chemical character, essentially tailoring the physical properties, such as guest selectivity, of already synthesised cage molecules.<sup>25</sup> Mastalerz *et al.* in their study, showcased that the cage product yield can be greatly improved with the post-synthetic attachment of an ether functional group (yield of 87%), in comparison with 17% product yield of the same reaction carried out with one of the precursors *o*-methylated before the synthesis.<sup>66</sup> Post-synthetic exo-functionalisation that introduces new function on the cage's periphery is also possible.<sup>67</sup> A set of three new organic cages was obtained by attaching phenyl-, xylyl- and naphthyl-acetylene moieties to the secondary amine groups of the pristine cage. The derived cages and their fluorescence have then been tested for the in-solution application of sensing various nitroaromatics.

POCs with their intrinsic porosity, that emerges from the cage-like shape and the internal cavities, are excellent components for the formation of both amorphous and crystalline PMMs.<sup>19,68</sup> For example, the **CC3** cage molecule self-assembles in a window-to-window fashion to yield a permanently porous material, with a 3D diamondoid pore network embedded in the crystal structure, the  $\alpha$  polymorph presented in Figure 1.5b. This polymorphism, the ability of molecules to crystallise in more than one form,<sup>69</sup> results from the molecular character of organic cages. For example, the **CC3 $\beta$**  polymorph (see Figure 1.5c) has a different window-to-arene molecules orientation in comparison to **CC3 $\alpha$** .<sup>40</sup> This molecular character can be utilised and the physical properties of the solid altered via straightforward recrystallisation or single-crystal-to-single-crystal transformation. Jones *et al.* presented a study where the material's porosity was switched "on" and "off" through a single-crystal-to-single-crystal transformation between porous and non-porous **CC1** polymorphs.<sup>70</sup> The porosity control was achieved by treating the material with ortho-xylene and dichloromethane vapours and the process was reversible. The modular character of organic cages can also be exploited in the formation of multicomponent co-crystals, that could give rise to multifunctional materials simply by mixing building component of various functions.<sup>71</sup>

In 2014, **CC3 $\alpha$**  was shown to have almost four times higher Xe/Kr selectivity, in a kinetic separation of a gas mixture, than a leading MOF at the time, Ni/DOBDC, in the same experi-

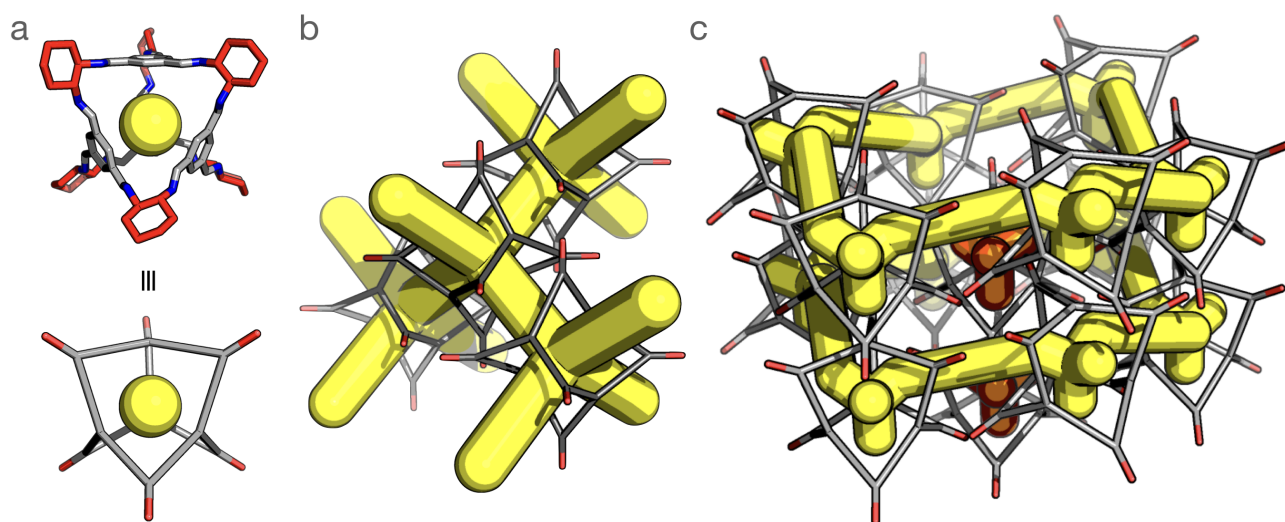


Figure 1.5: a) On the left, the stick representation of the **CC3** molecular structure with the following colour scheme: the vertices - red, the imine bond - blue and the carbon backbone - grey, with hydrogens omitted for clarity. On the left is the simplified wireframe model of the **CC3** molecular structure. In both cases the intrinsic cavity is presented as a solid yellow sphere. b) The **CC3 $\alpha$**  polymorph crystal structure simplified using the wireframe model from a); with the 3D diamondoid porous network coloured yellow. b) The **CC3 $\beta$**  polymorph crystal structure; with the non-accessible parts of the porous network coloured orange.

mental conditions (20.4 versus 7.3<sup>72</sup>). The kinetic selectivity was determined with breakthrough experiments, that allows one to measure the retention time of the gaseous mixture components passing through a sample, and yields the material's selectivity for the pair of gases.<sup>73</sup> The Xe uptake capacity was also shown to be about 2 times higher ( $\sim 11$  mmol kg<sup>-1</sup>) than that of Ni/DOBDC (4.8 mmol kg<sup>-1</sup>). The observed high Xe/Kr selectivity was rationalised, using molecular modelling, as the result of the fine size match between Xe and the **CC3** cavity. In the high-throughput computational study of 137000 MOFs for Xe/Kr separation application, a similar conclusion was made, that MOFs with pores just large enough to fit a single xenon atom, and tubular pores with uniform width, are predicted to perform best in Xe/Kr separation application.<sup>74</sup>

In their work, Chen *et al.* also studied the homochiral **CC3 $\alpha$ -R** in the separation of 1-phenylethanol enantiomers from their racemic mixture.<sup>73</sup> This ability arises from the **CC3** molecule's chirality, where the higher affinity of **CC3-R** towards the (*S*)-1-phenylethanol and **CC3-S** towards (*R*)-1-phenylethanol was exploited to perform the separation. A subsequent study has shown the ability of **CC3 $\alpha$**  to also separate C8 and C9 aromatics, an important mixture in the petrochemical industry, by means of shape and size sorting.<sup>75</sup> The C8 and C9

aromatics, ortho-, meta-, and para-xylene and mesitylene have similar chemical properties, therefore a separation based on the varying position and number of the methyl groups, which results in different shapes and sizes of the guests, could provide a viable alternative to the distillation-based industrial separation.

## 1.4.2 Porous liquids

The idea of *porous liquids* emerged in 2007, where it was hypothesised that certain properties of microporous materials can be combined with the transport properties of a fluid.<sup>76</sup> Three types of these hypothetical porous liquids have been proposed. Type I - a neat liquid of shape-persistence molecular pores, type II - a solution of molecular pores and type III - a dispersion of microporous nanoparticles. In 2012, Giri *et al.* attempted to synthesise a type I porous liquid.<sup>77</sup> Through attaching long alkyl chains to the **CC3** molecule's vertices, the melting point would lower enough for the material to stay liquid at room temperature, however the alkyl chains hindered the presence of the void space and a complete loss of porosity was observed.<sup>77</sup> Later, in 2015, Giri *et al.* decided to focus on type II porous liquids.<sup>78</sup> The solution of a **CC3** derivative with bulky crown ether functionalisation on the cage's vertices allowed this cage to be dissolved in crown ether solution with a high 1:12 cage to solvent ratio. This has resulted in a first porous liquid (type II) and permanent porosity was preserved, as the solvent was too bulky to penetrate the cage's interior cavity. This porous liquid would sorb different gasses, however the proposed porous liquid would not match the gas uptake of conventional porous materials. Therefore, the proposed application is the separation of gas mixtures, rather than gas storage, where the advantage of the liquid phase could be utilised industrially in flow processes.<sup>79</sup>

A different approach of "scrambled" cages was also proposed by Giri *et al.*, where two different diamine linkers and trialdehyde building block were used as precursors, resulting in a thermodynamic distribution of multicomponent [4+6] cages with varying ratios of the diamine linkers on the vertices.<sup>78</sup> Effectively, that mixture of 10 different final cage products and the significant asymmetry of the cages significantly improved solubility in hexachloropropane (1:31 cage to solvent molar ratio), and type II porous liquid emerged. Since then, the "scrambled" cages approach was studied for various sets of precursors, a proof of concept of this approach

to producing a type II porous liquid.<sup>80</sup>

### 1.4.3 Possible applications of porous organic cages

The possible future applications of porous organic cages have already been suggested throughout the introduction, with many examples of studies towards the gas storage and mixture separation applications mentioned. A lot of emphasis has been put towards the separation of binary gas mixtures, where molecular pores exhibit moderate to excellent selectivities, such as CO<sub>2</sub>/CH<sub>4</sub>,<sup>39,81,82</sup> CO<sub>2</sub>/N<sub>2</sub>,<sup>45,63,82–84,84–87</sup> H<sub>2</sub>/N<sub>2</sub>,<sup>68,70</sup> SF<sub>6</sub>/N<sub>2</sub><sup>88</sup> and Xe/Kr gas mixtures.<sup>73,89,90</sup> The ability of the relatively bulky SF<sub>6</sub> to diffuse in **CC3** could enable SF<sub>6</sub> gas recovery and the environmental release control of this highly potent greenhouse gas.<sup>91</sup> The potential of some of molecular pores for Xe and Kr capture and Xe/Kr separation application could lead to applications in the sequestration of the unspent nuclear fuels, reducing the negative impact on the environment and possibly providing a valuable source of Xe gas for commercial applications.<sup>92,93</sup> Air purification using molecular pore embedded membranes or filters, which take advantage of the irreversible binding of formaldehyde by the reduced amine form of **CC3**, also seems to be a viable application.<sup>64</sup>

These applications require POCs in their solid form to be permeable to gases (or other guests, depending on the application), a property shared with other porous framework materials. Therefore, for POCs to successfully compete with the porous framework materials for applications, their modular character is often exploited, a trait that framework materials lack.<sup>88</sup> Not without importance is the organic chemistry behind molecular pores, that can lead to significant flexibility of these pores, with the role of the flexibility discussed in more detail in Section 1.5.1. I will now present a selection of studies involving other possible applications of POCs, and molecular pores in general, however many of these are at the very initial stage of scientific discovery. Nevertheless, these studies put into perspective the types of applications these materials are believed to be suitable for and where they could find commercial or industrial uses. Recent advances in batch and flow synthesis of organic cages, that can significantly reduce the costs and allow scalability, also put these possible applications a step closer to realisation.<sup>94,95</sup>

For future applications, the material's stability and tailoring of properties towards a desired outcome is of great interest. Some PMMs remain highly crystalline upon multiple sorption-desorption cycles, with no observable decrease of the material's gas uptake capacity.<sup>45</sup> Several approaches to improve chemical stability of POCs, such as the imine to amine chemistry conversion or 'tying' the amine bonds together that result in enhanced stability, were mentioned in Section 1.4.1. At the nanoscale, the size and morphology of nanocrystals can be controlled through the crystallisation conditions by varying the rate of mixing and temperature.<sup>96</sup> Finally, the modularity of molecular pores allows a variety of forms that these molecules can be cast into, apart from the crystalline and amorphous solids, as membranes additives,<sup>97</sup> thin-films and coatings,<sup>48,98</sup> and as solutes and porous liquids.<sup>67,78,99</sup> Uemura *et al.* also showed that **CC3** can be used in a controllable and flexible vinyl polymerisation within the pore space, where the flexible structure and molecular assembly respond to the quantity and type of accommodated vinyl monomers, making the polymerisation highly specific.<sup>100</sup>

Utilisation of PMMs in the separation of closely related chemical species could provide a cost-effective alternative for the energy intensive industrial fractional distillation of chemical feedstock. Further, the ability to use molecular pores as a stationary phase in gas chromatography provides a viable route to exploit the properties of these materials in the separation of small organic molecules, their mixtures, and in case of stationary phases made of chiral porous organic cages, the separation of racemates.<sup>101,102</sup> Alternatively, many studies show an excellent potential for use of molecular pores as additives to enhance membranes performance, while retaining their high permeability.<sup>61,62,97,98</sup> Recently, Kong and Jiang, studied computationally the performance of membranes composed of POCs, **CC1**, **CC2**, **CC3**, **CC16** and **CC17**, for the application of water desalination using reverse osmosis.<sup>103</sup> They show that all POC-based membranes exhibit 100% salt rejection and the membrane water permeability is highly affected by the level of pore connectivity. The crystalline POC membranes usually offer higher water permeability, however the crystalline **CC1** membrane is impermeable to water, due to lack of interconnected pores. The amorphous POC membranes are more easily obtained and up to one-fold increase of the water permeability can be achieved in mixed POCs membranes, which allows one to precisely control the membrane performance. The permeability of amorphous

**CC3**, **CC16** and **CC17** membranes outperform currently used reverse osmosis membranes, therefore their potential in this particular application is demonstrated.

Several studies have shown the possible application for in-solution sensing by molecular pores, where the fluorescence quenching was used to sense and identify nitroaromatics such as picric acid, trinitrotoluene, 2,4-dinitrophenol and others.<sup>67,99,104</sup> In 2012, a porous organic cage was used as a highly potent affinity material in a sensing application with quartz crystal microbalances.<sup>105</sup> The high affinity was expressed towards a set of aromatic solvent vapours and the size and shape of the internal cavities strongly influenced the binding selectivity. The ability to detect  $\gamma$ -hydroxybutyric acid (GHB) and  $\gamma$ -butyrolactone (GBL), two psychotropic substances used in drug facilitated sexual abuse, was shown.<sup>106</sup> The specific supramolecular interaction of the GHB and GBL substances and the functionalised cage internal cavity, importantly in the presence of ethanol and water, was responsible for high sensitivity, and it is the first material to exhibit direct affinity for GBL sensing.

Recently, **CC3**-derived materials exhibited efficient protonic conductivity.<sup>107</sup> The usual proton conductive materials express a directed 1- and 2D conductivity, as often observed in more widely studied MOFs. Whereas, thanks to the particular packing motif (equivalent to that of **CC3 $\alpha$**  shown in Figure 1.5) and the 3D porous network that does not restrict protons to diffuse directionally a 3D conductivity was observed. The measured conductivity would range between  $\sim 1.0 \times 10^4$  S cm<sup>-1</sup> and  $\sim 1.1 \times 10^3$  S cm<sup>-1</sup> for 30 and 95% relative humidity, respectively. The lower value is comparable with the commercially available Nafion 117, a widely used proton conductive material, and the upper limit is comparable with the best performing MOFs. The study has focused on hydrated materials, but the ability of **CC3 $\alpha$**  to bind other small molecules in its pore space could lead to anhydrous proton conduction and find a possible application as a proton exchange membrane in the fuel cells.

Porous organic cages and other molecular pores might not be as competitive in gas storage applications as framework porous materials like MOFs, that often exhibit much higher gas uptakes and strong guest binding. However, the high flexibility of PMMs, narrow channels and well controlled pore space open to fine tuning, result in high selectivities for many important binary gas mixtures. The application of porous liquids is still uncertain and more studies are

necessary to allow further advances in property tuning, toxicity control and reducing porous liquids' viscosity to allow industrial applications. Recently, liquid MOFs have also emerged, a reason for a likely boost in porous liquids development.<sup>108</sup> Possible applications of porous liquids could take advantage of the mobility of the liquids phase and the uptake of liquids is significantly enhanced with the presence of the pore space.<sup>76</sup> Computational studies provide an invaluable insight towards the kinetics and thermodynamics of the possible gas storage application in porous liquids.<sup>79</sup> This requires novel approaches such as the single molecule's property extrapolation to the experimentally observed phenomena.

## 1.5 Computational studies of porous materials

The development of computational methods have significantly increased the number of tools available to scientists to advance the science of porous materials. A selection of methods used in this thesis is discussed in detail in theoretical methods section, however, I will now present a variety of computational studies and their importance in understanding and predicting the properties of porous materials. The first mention of computer simulations used in early zeolite catalysis research and methane adsorption studies is from the late 1980s.<sup>109,110</sup> Computer simulations, much advanced since the discovery of MOFs two decades ago,<sup>6</sup> are now used routinely to study gas adsorption and separation in porous materials,<sup>111</sup> their catalytic properties,<sup>112</sup> and in computer-aided materials design.<sup>113</sup> In the early computational characterisation of MOFs, much effort was put towards simulating adsorption isotherms.<sup>111</sup> Statistical methods, such as grand canonical Monte Carlo (GCMC), allow the evaluation of adsorption energetics by calculating heats of adsorption and provide a molecular level insight into the probability density of the adsorbent at various adsorption sites. Conclusions drawn from these insights would often be used to sketch sets of generic rules for the materials to meet to perform well in CH<sub>4</sub> and H<sub>2</sub> storage or CO<sub>2</sub> adsorption. Robust studies would often be carried out on different MOFs, to establish their general applicability, as in the H<sub>2</sub> purification process by pressure swing adsorption study.<sup>114</sup> It had been concluded that MOFs can indeed compete with the commercially used zeolites and activated carbons, and exhibit a range of affinities towards gases considered as impurities in the H<sub>2</sub> purification. However, a computational characterisation often cannot



provide a quantitative comparison of a specific material's performance, but rather suggest a set of well performing candidates.

As a result of ever growing computational resources and the development of new algorithms, high-throughput computational studies of materials are possible. A large scale screening of the methane storage capacity in 137000 hypothetical MOFs was performed by Wilmer *et al.*, where MOF structures were first computationally generated and their CH<sub>4</sub> adsorption isotherms then simulated using GCMC simulations. However, the frameworks were treated as rigid, neglecting potential effects of the framework's flexibility on the storage capacity.<sup>115</sup> Nevertheless, this screening identified large number of promising candidates ( $\sim 300$ ) and one of these candidates, namely NOTT-107, was synthesised and compared to the best performing MOF at the time, PCN-14. The storage quantities have been predicted to be 213 vol<sub>STP</sub>vol<sup>-1</sup> and 197 vol<sub>STP</sub>vol<sup>-1</sup>, respectively, with NOTT-107 predicted to slightly outperform PCN-14. Compared to the experimental value from a different study, NOTT-107 would however have lower capacity than PCN-14 by  $\sim 8\%$ . This study is an example of a high-throughput computational screening of a large number of structures by simulating the adsorption isotherm curves. In 2011, Haldoupis *et al.* analysed the pore sizes of  $>250000$  hypothetical zeolites focusing on the structural analysis rather than on the adsorption data.<sup>116</sup> The structural analysis was used to pre-screen the set of structures to identify the most promising candidates ( $\sim 8000$ ) for which the adsorption energy profiles were analysed by calculating Henry's constants for CH<sub>4</sub> adsorption. This was a valuable study where the structural analysis of pore sizes and pore channels limiting diameters were used to significantly reduce the computational expense and to significantly narrow down the number of promising candidates.

The structural analysis of porous materials, in terms of pore topology and roughness, channel dimensionality and pore size is often seen as a viable insight into the source of the arising properties and the effect the structure has on the adsorption and energetics. The nanoconfinement affects the behaviour of the fluid phase that can experience more extreme pressures than in the bulk, a phenomenon driven by pore topology and roughness.<sup>117</sup> The understanding of the fluid behaviour in cylindrical pores can advance the development of better separation processes.<sup>118</sup> The effect of morphology on the adsorption CH<sub>4</sub>/H<sub>2</sub> mixtures was shown.<sup>119</sup> The

amount of molecules that can pack in the pore, the way they pack into layers and therefore the pore morphology that maximises these features, cylindrical pores, was concluded to maximise the material's selectivity. The same effect of morphology will apply to gas mixtures, but also in the separation of molecular species such as alkylaromatics. The difference of packing, based on the shape of the guest molecules and their interaction with pore walls strongly impact the selectivity. Mecozzi and Rebek have proposed a general rule that a guest molecule confined in a nanopore will experience the maximum interaction when it roughly occupies  $55 \pm 9\%$  of the pore volume, a "55% solution" rule.<sup>120</sup>

Unsurprisingly, computational methods are widely used in the design and characterisation of porous molecular materials,<sup>20,121</sup> and screening studies can be used to identify new applications or even new examples of PMMs hidden, and previously ignored, in large databases of crystal structures.<sup>27,122</sup> Computational studies have been a vital part of the design and characterisation of porous organic cages. Jelfs *et al.* in their study used the relative energies of a series of imine porous organic cages to propose an "odd-even" effect observed experimentally.<sup>123</sup> In a different study, from the crystal energy landscapes generated using a rigid molecule force field lattice energy minimisation for a series of POCs, it was possible to successfully identify and reproduce the crystal structures of polymorphs observed experimentally.<sup>124</sup> Computational studies can be vital in the evaluation of the amorphous phases where the material's structure cannot be elucidated precisely by X-ray diffraction.<sup>68,125</sup> It is possible to model these amorphous phases, according to the experimental data at hand, to suggest further alterations in the structure to enhance the sorption capabilities facilitating the design of new materials.

The complex gas sorption behaviour of PMMs, related to the high flexibility of the molecular pores can often be rationalised with molecular simulations. Gas diffusion of gases with different molecular sizes, H<sub>2</sub>, N<sub>2</sub>, CO<sub>2</sub>, CH<sub>4</sub>, Kr, Xe and SF<sub>6</sub>, have been studied using molecular dynamics in **CC3** $\alpha$ . These gases differ not only in the overall size but also shape, from completely spherical atomic gases Kr and Xe, rod-like H<sub>2</sub>, N<sub>2</sub>, CO<sub>2</sub>, and tetra- and octahedral shapes of CH<sub>4</sub> and SF<sub>6</sub>, respectively. The molecular dynamics analysis of the pore network dynamic behaviour combined with a novel void space histogram, an overlay of multiple framework topologies generated with molecular dynamics, gave a graphical representation. The flexible

host was used to enable transient channel formation, that was necessary to rationalise the diffusion of larger species like CH<sub>4</sub>, Kr, Xe, which are larger than the average pore limiting diameter of the **CC3** $\alpha$  static crystal structure. The SF<sub>6</sub> was shown not to be able to diffuse. In work during my PhD (but not included in this thesis), I carried out a study, where molecular dynamics was used in combination with metadynamics to enforce a higher energy configuration on the system (see Section 2.5). The analysis of a two dimensional energy surface of the cage-SF<sub>6</sub> and the relative window-SF<sub>6</sub> orientation allowed us to rationalise the experimentally observed SF<sub>6</sub> uptake in **CC3** $\alpha$ . The mechanism of diffusion required window widening and a specific SF<sub>6</sub> rotation, where the triangular face of the octahedral SF<sub>6</sub> would be able to pass through the triangular window of **CC3** molecule, followed by 60 degree rotation and passing of the rest of the molecule through the window.<sup>91</sup>

In the thermodynamic analysis of Xe/Kr selectivity in over 137000 hypothetical MOF structures, the optimal performance was related to the pores size matching the dimension of Xe and Kr atoms and the cylindrical morphology of these pores.<sup>126</sup> In a similar study of over 670000 porous material structures for Xe/Kr separation application, a combined machine learning and molecular simulation approach was chosen.<sup>93</sup> The structural descriptors were generated for the whole set and molecular dynamics calculations performed for a training set of 20000 structures. For the remaining structures, Xe/Kr selectivity was predicted using machine learning. Many structures were predicted to outperform the previously mentioned organic cage **CC3** and its high Xe/Kr selectivity. The two best candidates, an aluminophosphate zeolite analogue<sup>127</sup> and a calcium coordination network<sup>128</sup> were not previously tested for Xe/Kr separation. This is an excellent example where screening large databases of materials using computational analysis can identify promising materials among those not previously considered and therefore never tested. My own screening of the molecular pores for the Xe/Kr separation application is presented in Chapter 4 of this thesis. A single molecule structural analysis and flexibility study is combined with the Xe/Kr binding affinity and free energy barriers for diffusion. The results are used as a proxy to the prediction of the bulk material performance.

### 1.5.1 Role of flexibility in porous materials

In computational studies of zeolites and MOFs, especially in high-throughput screenings, the “rigid host” approximation is often used, meaning the material’s framework is kept rigid with fixed atomic positions.<sup>129</sup> The framework’s flexibility is often considered as a negligible effect on the studied properties or as a trade-off between the computational cost and time required and the accuracy, which is especially true if the high-throughput studies of thousands of structures are to be performed in a reasonable timescale. However, the flexibility can play a significant role even in zeolites and their sorption properties,<sup>130</sup> and the role of flexibility is recognised in MOFs, with considerably greater motion of the organic linkers.<sup>129</sup>

In ZIF-8, adsorption of the gas molecules  $N_2$  and  $CO_2$ , which are relatively bulky with respect to the ZIF-8 pores, was observed.<sup>131</sup> This was rationalised using computational simulations with the considerable flexibility of the ZIF-8 structure and the “gate-opening” effect, where seemingly narrow channels open wide enough for the passage of larger guests. A joint experimental and computational study of ZIF-8, using GCMC simulations with additional DFT parametrisation, allowed for relatively good reproduction of the experimental isotherms. The complex gas adsorption behaviour and a step wise isotherm was rationalised with the structural change that ZIF-8 undergoes during gas adsorption.<sup>132</sup> A similar “breathing” effect associated with the pore flexibility can be observed in MIL-53. A seemingly collapsed structure undergoes a transformation step upon gas adsorption where the pore space is gradually filled with gas that stabilises the open pore structure upon the system reaching appropriate gas pressure. In this computational study, *ab initio* molecular dynamics simulations were used to rationalise MIL-53 structure transformation.<sup>133</sup> Some flexibility in the structure is inherent to all materials due to the thermal motion of atoms, but it can also originate from external stimuli: host-guest interactions, change of external pressure, and from a photo-response.<sup>134</sup> In their study, Witman *et al.* linked the observed discrepancy of the computationally predicted Xe/Kr selectivities in MOFs to the framework flexibility.<sup>135</sup> In MOFs with pore sizes commensurate to those of Kr and Xe, the selectivity would diminish with increasing flexibility.

Recently, an intriguing discrepancy in the nitrogen uptake in porous organic cage co-crystals was observed.<sup>71</sup> The binary co-crystal **CC1/CC4-R** had almost twice the uptake of nitrogen

in comparison to **CC1/CC3-R**, 7.9 and 4.5 mmol g<sup>-1</sup> respectively. These two co-crystals share the packing motif and are build of organic cages with comparable size and the same octahedral shape. Interestingly, a smooth relation between the **CC3-R** to **CC4-R** ratio, in the ternary co-crystal **CC1/CC3-R/CC4-R**, to the experimentally measured nitrogen uptake was also showcased. In a subsequent computational study, performed by myself and co-workers (not disclosed in this thesis), we have rationalised this discrepancy in nitrogen uptake with the effect of flexibility on the adsorption profile in these binary co-crystals.<sup>136</sup> The channels connecting the pore network were found to form spontaneously from seemingly inaccessible voids (when considering static crystal structures). This allowed us to explain the higher nitrogen uptake observed for **CC1/CC4-R** in comparison to **CC3-R**. The latter has more bulky vertex functionalisation (cyclohexanes versus cyclopentanes in **CC4-R**), that possibly prevent the formation of wide enough channels for the nitrogen to access the additional void space, thus the lower uptake.

Host flexibility plays an important role in porous molecular materials, due to the fully organic composition of these molecules with no covalent bonding between them when packed in a solid.<sup>121</sup> For example, a seemingly non-porous PMM made of a **Me,Et,SiMe<sub>2</sub>** cavitand (the R,R',Y simplified nomenclature describes the R, R' exo-functionalisation and Y bridged rim of the cavitand's bowl), where the pores are classified as 0-dimensional with no apparent connectivity between them, shows experimental permeability to water at elevated temperatures.<sup>34</sup> The crystal structure stays unchanged upon desolvation, which leads to the conclusion that the ability of the molecules to temporarily adjust their packing, in response to the guest presence, allows for guest diffusion and results from the material's significant flexibility. Barbour refers to this phenomenon as "porosity without pores".<sup>137</sup> The judgement of a material's porosity will depend upon the guest size and the temperature. Holden *et al.* recently divided the porosity in PMMs into static, dynamic and cooperative.<sup>138</sup> Static porosity is observed regardless of the presence of the guest or consideration of the material's flexibility. Dynamic porosity requires host motion to be considered, for example the breathing of the host's molecular windows, and cooperative porosity is the result of cooperative interactions between both the guest and host to induce window opening. These mechanisms were rationalised using molecular dynamics

simulations and the analysis of the **CC2** window size fluctuations and the accessibility of the pores.

## 1.6 Adsorption and separation of gases in porous materials

The *sorption* of gases is here discussed in terms of *physisorption*, which is a general phenomenon explained as the interaction of gases brought in contact with the surface of a solid governed by the intermolecular forces responsible for the condensation of vapours and the imperfection of real gases.<sup>1</sup> *Chemisorption*, which I will not deal with in this thesis, involves chemical bond formation. The simplest model to describe sorption is that of Langmuir, that considers the formation of a continuous monolayer of sorbate on the solid surface.<sup>139</sup> A required assumption is that the sorbate behaves as an ideal gas in the isothermal conditions and that the surface is completely flat, with binding sites each allowed to be occupied by one particle of the sorbate. The interactions between the sorbate molecules in adjacent sites, however, are here neglected. The Brunauer–Emmett–Teller (BET) theory provides a more realistic model of sorption with a formation of a multilayer of sorbate that does not interact chemically with the material’s surfaces.<sup>140</sup> The multilayer is believed to form in a process similar to condensation, however at lower than saturation pressure. The material’s surface interactions with the sorbate result in the formation of the first sorption layer, with the sorbate-sorbate interactions necessary to form further layers.

The BET theory is routinely applied to experimentally quantify a porous material’s *specific surface area*, usually defined in terms of the total surface area of the material per unit of mass ( $m^2 g^{-1}$ ), derived from the sorption isotherms of gases near the boiling point. The measured specific surface area is a scale-dependent property, that arises from the sorbate cross-section function of the size and shape, with the  $N_2$  sorption at 77 K most commonly used.<sup>141</sup> Porous materials, with a specific surface area well exceeding that of  $100 m^2 g^{-1}$ , are often viewed for possible application in gas storage. MOFs are considered as particularly promising, with surface areas ranging typically from 1000 - 10000  $m^2 g^{-1}$ , high pore volumes and the possibility to tailor the structures to maximise the favourable interactions between the framework and the adsorbate.<sup>12</sup> As they are built of fully organic discrete units, PMMs usually achieve lower

surface areas, yet the highest measured to date, for a crystalline solid made of POCs, has an astounding BET surface area of  $3758 \text{ m}^2 \text{ g}^{-1}$ .<sup>142</sup>

The separation of gases using porous materials can be achieved by either a thermodynamic or kinetic route.<sup>143</sup> The thermodynamic separation approach utilises the variation in strength of the interactions between the material's surface and the components of the gaseous mixture at equilibrium. The strength of these interactions can be quantified as the isosteric heat of adsorption, measured as the change of enthalpy for the single component adsorption isotherms at different temperatures.<sup>144</sup> The difference in the isosteric heats of adsorption of mixture components can be then used in the pressure or temperature swing adsorption separation systems, that by alternating sorption and desorption cycles enable one to effectively separate and purify components of gaseous mixtures.<sup>145</sup> The kinetic route, on the other hand, relates to the differential rate for *diffusion* of gases in porous material.<sup>143</sup> It relates to the sieve-like behaviour of porous materials with pores close in size with that of the sorbate. The mixture components are allowed to enter the pore space of the material, that effectively plays a role of a membrane. Components of the mixture smaller than the limiting width of the pore network channels can diffuse freely through the material, whereas components close in size to that limiting dimension will experience obstruction in their diffusion.

The degree of the translational and rotational freedom of material's structural substituents and the thermal motion results in various flexibility of materials.<sup>146</sup> In the more flexible materials a dynamic change of the pore network can affect the guest diffusion and lead to a kinetic guest trapping. The diffusion rates are also affected by the affinity of the adsorbate to the material's surface or functional groups, but the kinetics of diffusion is the driving force of the separation process. For practical adsorption applications both the kinetics and thermodynamics play a role. The performance of the material is usually assessed in breakthrough experiments, where the retention time for the mixture components is evaluated and the *selectivity* is determined as the ratio of the retention time of two components. The kinetics and thermodynamics of separation processes can be controlled by the match of the sorbate to the size and shape of the pores,<sup>75</sup> the chirality of the material,<sup>73</sup> its flexibility and topology,<sup>147</sup> or by introducing functional groups to maximise the sorbate affinity to the materials surface.<sup>148</sup>

## 1.7 The aims and objectives of this PhD thesis

The following are the main aims and objectives that are in the scope of this PhD thesis:

- To present the process of screening of porous organic molecules, and porous organic cages in particular, for properties and possible applications with molecular modelling and structural analysis techniques. This includes screening of both the synthetically realised porous organic molecules extracted from the literature and the Cambridge Structural Database, as well as the hypothetical organic cages obtained with the methodology discussed in this thesis. The ability for a single molecule analysis to act as a proxy for the prediction of the properties in the bulk is also showcased.
- To present a novel software for the structural analysis of porous organic molecules for their structural features and dynamical behaviour that can be used to rationalise the observed properties in the bulk and in the design of novel materials.
- Lastly, to present a novel methodology for an evolutionary algorithm assisted function-led materials discovery. The presented scheme enables one to design new porous organic cages with desired properties from a set of initial precursors, chosen stoichiometry and topology and chemistry.



# Chapter 2

## Theoretical Methods

In this chapter, I will describe the theoretical background to the main methods of computational chemistry used in this thesis. The discussion of structural characterisation tools for chemical applications will be followed by a description of relevant classical mechanics methods, from the perspective of this thesis, including force field potentials, molecular dynamics and metadynamics. I will also give a brief introduction to the widely used quantum chemical method - density functional theory. The statistical mechanics methods used to simulate adsorption isotherms with grand canonical Monte Carlo and the ability to calculate binary mixture selectivities using the ideal adsorbed solution theory will also be discussed. I will also present the principles of evolutionary algorithms and how these can be applied to find solutions to complicated chemical problems in the vast chemical space.

## 2.1 Introduction to computational chemistry

All computational methods have roughly the same purpose; to give a mathematical description of the system to yield energetics and to allow for the derivation of modelled system properties. Whereas most experiments measure properties of materials in the bulk, molecular modelling can provide an insight to the system at the molecular and atomistic level. This can often facilitate the understanding of processes hard to study experimentally, but possible to realise in computational simulations, or that requiring an atomistic insight. Apart from aiding the experiments, the growing importance of computation can be observed in materials design approaches, where models that can reproduce the experimental results to a satisfying degree can lead to a relatively fast and cheap property prediction of whole classes of materials, often allowing to the reduction of the time and costs of experiments by focusing on the materials predicted as the most promising.

The principles of different molecular modelling vary substantially, and one way to group them is by the level of theory at the method's foundation:

- Quantum mechanics (QM): where a system is described in terms of 'wave functions' and each observable property has an equivalent operator that when applied on the wave function returns a value for that property. The description of the system is derived directly from theory, thus these methods are often referenced to as *first principles* or *ab initio*.
- Molecular mechanics (MM): where a system is described using force field potentials. The force field parameters can originate both from *ab initio* methods and experiments.
- Statistical mechanics: which operates with a random probability distribution of system states that allows one to obtain the thermodynamic system equilibrium in a stochastic way.
- Structural characterisation: where a combination of mathematical methods and visualisation allows one to quantify structural features with the aim of rationalising the observed material's kinetic or thermodynamic properties.

The use of any of these methods depends on the system, models available and the properties

or processes of interest. In this work, a selection of methods is applied to study the structural properties, the thermodynamics and kinetics of processes in PMMs. The general properties of interest in modelling of porous materials may include, but are not limited to, gas capacity, kinetic and thermodynamic selectivity, guest binding and diffusion. I will now provide the theoretical background of the methods used throughout this thesis.

## 2.2 Structural characterisation of porous materials

The development of the structural characterisation tools for computational chemistry is as important as the advances in the molecular modelling techniques. The bulk properties of materials with porosity, in addition to the chemical composition, will be greatly influenced by the structural variations at the atomistic level. The size and shape of pores, and the connectivity between them, will have an impact on the properties that rely on the geometrical features.<sup>115,116,126,149</sup> The structural characterisation of porous materials, in addition to established molecular modelling techniques, can often yield the connection between the structure and the observed properties. Such an approach is common in computer-aided materials design.<sup>121,125,150,151</sup> The structural descriptors often used to describe porous networks in porous materials, although the terminology may vary in different studies, are the largest cavity diameter (LCD), the largest sphere that can be inserted in the porous material without overlapping with any of the atoms, and the pore limiting diameter (PLD), the largest sphere that can freely diffuse through the porous material without overlapping with any of the atoms.<sup>116</sup> Figure 2.1 presents a simplified 2D model of a porous network with LCD and PLD marked as 1 and 3, respectively.

Various openly available software exist that allow the characterisation of porous materials. An early example, the HOLE package,<sup>154</sup> focuses on the ion channel dimensionality evaluation and visualisation. This is possible with the Monte Carlo simulated annealing procedure for a spherical probe, that finds the spheres of variable size that can squeeze along the channel. Similarly, Monte Carlo integration algorithm is used in the “surface area program”,<sup>155</sup> where a sampling sphere of a given size is used to measure the surface area of the porous material, by virtually “rolling” over the materials surface. This methodology makes the comparison of the

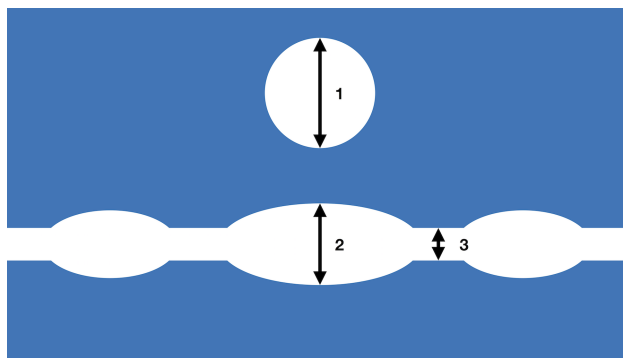


Figure 2.1: The structural descriptors for porous networks are presented. 1) The largest cavity diameter (LCD) or the largest included sphere, 2) The largest included sphere along the free sphere path. 3) the pore limiting diameter (PLD) or the largest free sphere.<sup>116,152,153</sup>

returned accessible surface area (ASA) of the material, with the BET surface area determined experimentally, valid. Further examples, that use Monte Carlo based algorithms to determine ASA, include *ZEOMICS*,<sup>156</sup> *MOFomics*<sup>157</sup> and *Poreblazer*.<sup>158</sup> The *Zeo++* package combines the Monte Carlo approach with the Voronoi tessellation method, that is used to determine channels interconnectivity.<sup>152</sup> This software also performs the structural analysis in terms of the largest included sphere (equivalent to LCD), the largest free sphere (equivalent to PLD) and the largest included sphere along free sphere path, marked as 1, 3 and 2 in Figure 2.1, respectively. The largest included sphere along free sphere path is the largest sphere that can be inserted within the porous network.<sup>153</sup> The utility to determine channel interconnectivity is important, as it makes possible to determine the accessible and non-accessible surface area (NASA) from the perspective of the probe being able to diffuse through the channels to access the surface.

*Zeo++* was used to study the dynamic pore network connectivity, for a probe of fixed diameter, where the short channel openings have rationalised an increased ASA determined computationally for a multicomponent porous molecular crystal.<sup>136</sup> This material was both, intrinsically and extrinsically porous, where apart from the porous network of interconnected molecular pores, it had excluded voids between molecules. The feature to study the channel interconnectivity, in *Zeo++*, allowed for a dynamic void analysis of the molecular dynamics trajectory. This allowed the authors to elucidate the especially high experimentally determined BET surface area of the material, with short windows of interconnectivity between the porous network and the excluded voids. This is an example of the dynamic porosity described by Holden *et al.*,<sup>138</sup> also described in Section 1.5.1. Software like *pyMolDyn*,<sup>159</sup> can offer graphical

user interfaces and the analysis of the cavity shape in terms of asphericity, acylindricity, and relative shape anisotropy.

The mentioned software are usually not limited to zeolite and MOF structural characterisation and can often be used for other porous systems; however, the aim of these packages are framework materials, therefore, special requirements on the input often must be met, such as a periodic system. Nevertheless, these tools can aid the visual inspection of porosity, which is difficult for large systems with 3D porous networks. Other approaches aim on providing the tools for the pore network assessment in a visual but also a quantitative and qualitative manner.<sup>134</sup> In a study of Theisen *et al.*, the description of the void space topology was approached in an abstract fashion, with convoluted ‘chemical hieroglyphs’, used to describe the 2D pores and connectivity of the local environment. The growing number of approaches and available software only stresses the need for newer and better tools to aid the description of the pore network topology and related structural features in an approachable fashion.

A methodology to determine a circular window diameter of an organic cage, with a circumcircle, was proposed by Holden *et al.*<sup>160</sup> The circumcircle method requires defining a triad of atoms, that the circumcircle passes through, as shown in Figure 2.2a on an organic cage examined in that study. **CC3** has 4 windows, therefore four such triads has to be assigned manually to calculate the circular diameter for all windows. The window diameter of a static **CC3** extracted from the crystal structure was too small (3.62 Å) to allow nitrogen diffusion (3.64 Å) in contradiction to the experimental observations. The circumcircle method was used to analyse the molecular dynamics trajectory of an isolated **CC3** molecule and projected as a distribution of windows diameters as a function of time to yield the pore limiting envelope (PLE), shown in Figure 2.2b. The consideration of the dynamic window size fluctuations shows that for 67% period of time the cage’s window is bigger than nitrogen, explaining the experimental findings. A possible drawback of the circumcircle method is the fact that the set of atoms for the circumcircle is usually chosen from a random cage configuration and fixed throughout the generation of PLE, that can lead to an inaccurate representation of the window, and the need for the visual inspection of the cage structure. Other approaches to identify intrinsic porosity in molecules are developed. Garcia *et al.*, reported the analysis of a single molecule’s surroundings

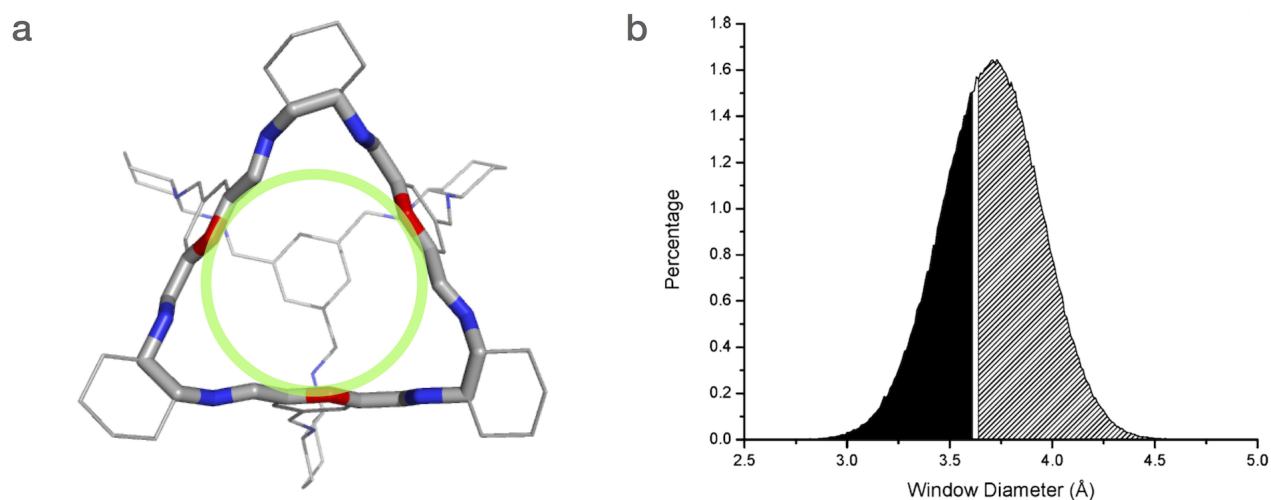


Figure 2.2: a) The circumcircle (pale green) defined by the atom positions of three carbon atoms (coloured red) that envelopes one of the **CC3** windows, as reported by Holden *et al.*<sup>160</sup> b) The pore limiting envelope (PLE) for **CC3**. The black section shows the proportion of the simulation time when the window diameter is smaller, and the cross-hatched section shows the proportion of the simulation time when the window diameter is bigger, than the average window diameter (3.62 Å) calculated for **CC3** as extracted from the crystal structure. Subfigure b extracted from reference 161.

and the projection of void regions within the perimeter of a molecule.<sup>122</sup> The method allowed to identify six new molecules in the PubChem<sup>161</sup> repository as intrinsically porous.

## 2.3 Force field potentials

The molecular modelling of large systems, with a significant number of atoms, is possible with *force field* potentials. The movement of the electron clouds, according to the Born-Oppenheimer approximation, is instantaneous with respect to the movement of the atomic nuclei. Thus, in accordance to molecular mechanics principles, the *classical* representation of the system treats atoms as point charges and bonds as springs. The role of the force field is to provide a detailed system description, in the form of parameters, and the functional for the determination of the total potential energy ( $E_{potential}$ ), as a function of the atomic positions. The form of the functional can differ with respect to the force field used, however, the most basic consideration requires five terms; three for the intramolecular potentials, the bond stretching ( $E_{bonds}$ ), the angle bending ( $E_{angles}$ ), the bond rotation ( $E_{torsions}$ ), and the non-bonding interactions terms for the description of the Coulomb electrostatic interactions ( $E_{Coulomb}$ ) and the van der Waals

forces ( $E_{vdw}$ ), that are the interatomic potentials. Thus, the functional form is given by:

$$E_{potential} = E_{bonds} + E_{angles} + E_{torsions} + E_{Coulomb} + E_{vdw} \quad (2.1)$$

and allows one to determine the force field-dependent system energy.

The intramolecular term for the bond stretching can be written as the sum over all bonded atom pairs and is often expressed by the harmonic potential:

$$E_{bonds} = \sum_{bonds} \frac{k_i}{2} (l_i - l_{i,0})^2 \quad (2.2)$$

where  $l_i$  is the bond length,  $l_{i,0}$  the bond reference value at equilibrium defined by the force field, and  $k_i$  is the force field constant for a harmonic spring for the bond type.

The angle bending can also be described in the form of a harmonic potential, and the sum over all angles in the system:

$$E_{angles} = \sum_{angles} \frac{k_i}{2} (\theta_i - \theta_{i,0})^2 \quad (2.3)$$

with  $\theta_i$  and  $\theta_{i,0}$  representing the calculated and equilibrium angles, respectively, and  $k_i$  is the harmonic constant for the angle type.

The last intramolecular term is for the description of torsions and often takes a cosine series expansion form:

$$E_{torsions} = \sum_{n=0}^N \frac{V_n}{2} (1 + \cos(n\omega - \gamma)) \quad (2.4)$$

here  $\omega$  is the torsion angle,  $n$  is the multiplicity that defines the number of minima in the cosine function and  $\gamma$  is the phase factor for where the torsion passes through its minimum value. The  $V_n$  constant is correlated to the relative energy penalty for the rotation around the bond, however, the ‘barrier’ height for rotation is a product of multiple terms of the cosine series expansion, in higher functionals, and the non-bonding interactions.

The intramolecular potentials presented so far account for the specific bond relationship between atoms. However, atoms within the same molecule (usually separated by 4 or more bonds) and between molecules, within some maximum distance, also interact through the non-

bonding forces that act through space. These can arise from the electronegativity of atoms, the fact that some elements attract electrons more strongly than others, resulting in the uneven charge distribution in the molecule (the electrostatic interactions), or from the electron cloud's fluctuations (the van der Waals interactions). A common approach, in the classical representation, is to express the charge distribution in the form of *partial atomic charges* localised at the atomic nuclei. Coulomb's law then allows one to define the sum of electrostatic interactions between pairs of point charges according to the equation:

$$E_{electrostatics} = \sum_{i=1}^{N_A} \sum_{j=i+1}^{N_B} \frac{q_i q_j}{4\pi\epsilon_0 r_{ij}} \quad (2.5)$$

where  $N_A$  and  $N_B$  are the numbers of point charges,  $q_i$  and  $q_j$  are the partial atomic charges for pair of atoms separated by distance  $r_{ij}$ , and  $\epsilon_0$  is the permittivity of vacuum.

The van der Waals interactions, as a combination of the dispersion and repulsion energies, are best known in the form of the Lennard-Jones potential:

$$E_{vdw} = \sum_{i=1}^{N_A} \sum_{j=i+1}^{N_B} 4\epsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r_{ij}} \right)^{12} - \left( \frac{\sigma_{ij}}{r_{ij}} \right)^6 \right] \quad (2.6)$$

which is the sum of interactions between all atom pairs in the system, where  $\epsilon_{ij}$  is the depth of the interatomic potential well,  $\sigma_{ij}$  is the atomic distance at which attraction and repulsion exactly balance,  $r_{ij}$  the distance between atoms and  $N_A$  and  $N_B$  are the numbers of atoms.

The force field parameters can be based on an atom's element, hybridisation and connectivity, as in more generic force fields that intend to reproduce the solid state structure of molecules, and are traditionally obtained from the crystal structures.<sup>162,163</sup> The validation process of these force fields often results in the comparison of the generated structures to the benchmark molecules' crystal structures. The force fields designed to reproduce the properties of liquid phases, rather than the properties of solids, can have their parameters derived from the Monte Carlo simulations that generate data on the liquid's density, the heats of vapourisation, or the vapour-liquid coexistence curves.<sup>164,165</sup> Validation of these force fields requires the comparison of the generated data to the thermodynamic properties of liquids for which the force field is parametrised. The choice of appropriate force field is important, as it should



describe the studied system as accurately as possible. The number of available force fields is considerable, including more general ones such as DREIDING<sup>162</sup> or Universal Force Field (UFF),<sup>163</sup> that derive their parameters from crystallographic data on small molecules, or more specialised ones like Optimised Potentials for Liquid Simulations (OPLS),<sup>164</sup> and Transferable Potentials for Phase Equilibria (TraPPE),<sup>165</sup> that were parametrised to represent bulk solvent thermodynamic properties. Force fields like CHARMM,<sup>166</sup> tailored for an accurate description of biomolecules, are also available.

Throughout this thesis the OPLS\_2005<sup>167</sup> and OPLS3<sup>168</sup> force fields are used. OPLS\_2005 was shown to give an accurate description of the structural and thermodynamic properties for a range of imine POCs.<sup>123</sup> The generated structures of single POC molecules showed small deviations from these observed in crystal structures. The OPLS\_2005 derived formation energies showed the same “odd-even” effect of cage formation with an alternating [2+3] and [4+6] topologies, the effect that relates to the number of backbone carbon atoms in the organic linker, that was observed experimentally. The OPLS3 force field, a newer implementation of OPLS\_2005, was developed with the enhanced protocol and a dataset of 11845 molecules used in the parametrisation, in comparison to 5700 molecules used in the parametrisation of OPLS\_2005.<sup>167,168</sup> High quality quantum chemical calculations were used to more accurately describe valence and torsional parameters, with 2 orders of magnitude more explicitly fit torsional parameters (48142) than OPLS\_2005 (1576). The number of parameters for stretches (1187) and bends (15236) is also much higher in OPLS3 than in OPLS\_2005, 1054 and 3997, respectively. The validation showed that the root mean square (RMS) error in comparison with over 11000 quantum mechanics torsion profiles is smaller than 1 kcal mol<sup>-1</sup> for 88 % of torsions benchmarked with OPLS3, while OPLS\_2005 scores considerably lower at 33 % of torsional profiles with the RMS error within 1 kcal mol<sup>-1</sup>. An improved charge model for the computation of partial charges was applied and extensive testing of the models against larger data sets of small molecules was part of the enhanced protocol. A high level of accuracy across performance benchmarks that assess small molecule conformational properties and solvation was shown. The OPLS3 force field is only available commercially and is used with the MacroModel Software (Schrödinger, LLC), whereas, OPLS\_2005 is used with DL\_POLY2.20 package.<sup>169</sup>

## 2.4 Molecular dynamics

Molecular dynamics (MD) is a type of simulation where the numerical solution of the Newton's equation of motions is used to generate the dynamics, in the form of a trajectory of the system's coordinates, of a time-dependent system using force-field potentials. The velocities acting on the system, between two states separated by a timestep  $dt$ , need to be calculated. The two common approaches to calculate these forces are the velocity Verlet algorithm, that yields the atom positions, velocities and accelerations at time  $t$ , and the velocity Leapfrog algorithm, where the velocities are first calculated at time  $t + \frac{1}{2}dt$  and these are used to calculate the atomic positions at time  $t + dt$ .<sup>170</sup> In that fashion, the positions *leap* over the velocities and the velocities *leap* over the position. All the forces in the Newton's equation of motions need to relate to the potential energy of the system, this way the total energy of a system  $E_{total}$  is conserved in form of:

$$E_{total} = E_{kinetic} + E_{potential} \quad (2.7)$$

An important aspect of MD simulations is the ensemble in which the system is kept. If the energy  $E$  of the system is conserved, and the volume  $V$  and the number of particles  $N$  also kept constant, it is referred to as the *microcanonical* ensemble or  $NVE$ . In the case when a precise temperature  $T$  has to be constant, the isothermal canonical ensemble, or  $NVT$ , is used where the number of particles, volume and the temperature are fixed. Other ensembles with constant pressure ( $NPT$ ) or chemical potential ( $\sigma VT$ ) are also used. Where the temperature or pressure has to be constant, algorithms are used to keep these parameters fixed with a selection of thermostats and barostats available to choose from, such as Evans<sup>171</sup>, Berendsen<sup>172</sup> or Nosé-Hoover.<sup>173,174</sup>

## 2.5 Metadynamics

Molecular mechanics simulations with force fields allow for longer timescales than *ab initio* methods as a trade-off between the speed and accuracy. However, the tens of nanoseconds obtained routinely in modern MD computation, might not suffice to sample processes that require orders of magnitude longer timescales, the so-called *rare events*. Additionally, the

favourable configurations necessary to overcome the energetic barriers to sample the potential energy surface efficiently, might not be accessible through thermal fluctuations. Enhanced sampling techniques were developed over time to address the timescale limitations of regular MD simulations.<sup>175</sup> In *metadynamics*, an external bias potential is applied on a selection of degrees of freedom, defined as the collective variables (CVs).<sup>176</sup> Hence, the system is able to overcome the energetic barriers and escape local minima. The bias potential works in the CV space along the system trajectory. As a result, the system is discouraged to revisit the configurations that have already been sampled, thus accelerating the rare events. The sum of Gaussians deposited in form of the bias potential allows one to reconstruct the free energy surface (FES).

The schematic representation of reconstructing the FES by sampling a one-dimensional CV along its path  $S$  using metadynamics is shown in Figure 2.3. The initial configuration of the system, with a specific value for the CV represented with a blue dot, is located in the local minimum **1**. Over the simulation time, the external bias potential is gradually applied to the system and slowly fills the local minimum, with the deposited Gaussians that allows the system to explore the less energetically favourable configurations further from the local minimum **1** up to time  $t_1$ . The sum of the potential at time  $t_1$  is represented by a thin wavy line. At some point after  $t_1$ , the system accumulates enough potential to overcome the free energy barrier **a** and moves to a new local minimum **2**, a process indicated with a dotted arrow. Now the new system state (a green dot) is located in a not yet sampled space along the CV path. Again, the process of accumulating the potential through times  $t_2$  and  $t_3$ , where at time  $t_3$  the system is able to freely oscillate between the two local minima **1** and **2**, results in passing over the barrier **b**. This way the system is able to sample a new space of  $S$  along minimum **3** (the red dot). Through a constant deposition of the bias potential, the well **3** is slowly filled with the potential and sampled through time  $t_4$ , up to the moment where the accumulated potential allows the system to diffuse along the whole CV path at time  $t_5$ . The FES, a thick black line in the Figure 2.3, can be reconstructed as the inverse of the sum of the Gaussians deposited. The benefit of metadynamics, shown by the described process of FES sampling, is that it inherently samples the low regions first and no prior knowledge of the FES landscape is required.

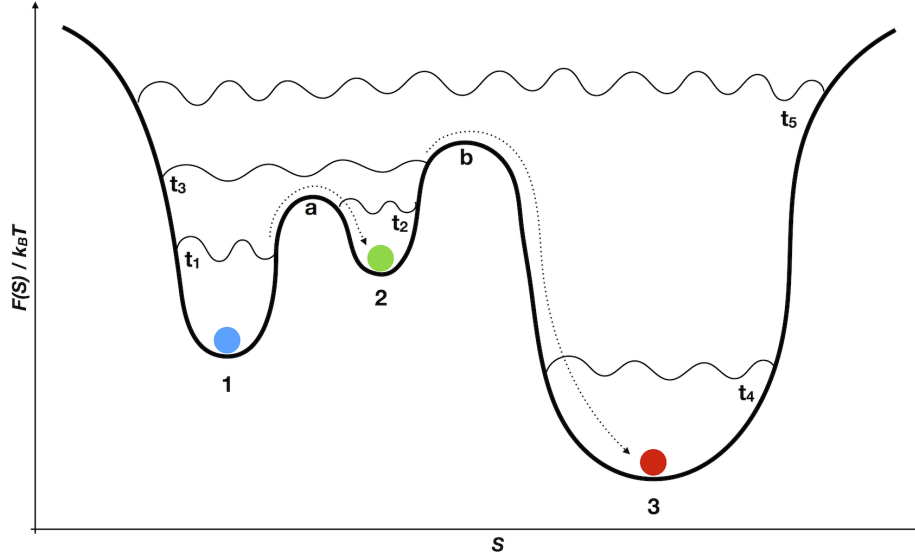


Figure 2.3: The schematic representation of the continuous filling of the underlying potentials by Gaussians deposited in the collective variable path. The process of overcoming the energetic barriers and sampling successively different phases of the CV space is marked by the configuration evolution with blue, green and red dot, respectively. The reconstructed FES and the current potential levels at times  $t_1 - t_5$  are shown as thick black line, and thin wavy line, respectively.

In metadynamics, the bias is applied continuously during the MD simulation and directly on the coordinates of the system. If we define the CVs as a set  $S$  of  $d$  functions of the microscopic system coordinates  $R$ :

$$S(R) = (S_1(R), \dots, S_d(R)) \quad (2.8)$$

The bias potential  $V_G$  applied to the system at time  $t$  is given by the formula:

$$V_G(S, t) = \int_0^t dt' \omega \exp \left( - \sum_{i=1}^d \frac{S_i(R) - S_i(R(t'))^2}{2\sigma_i^2} \right) \quad (2.9)$$

where  $\sigma_i$  is the Gaussian width for the  $i^{\text{th}}$  CV and  $\omega$  is the constant energy rate defined as the relation of the Gaussian height ( $W$ ) to the deposition stride ( $\tau_G$ ):

$$\omega = \frac{W}{\tau_G} \quad (2.10)$$

The applied potential can then be used to reconstruct the FES with the relation:

$$V_G(S, t \rightarrow \infty) = -F(S) \quad (2.11)$$

Therefore, the FES as a function of the CVs can be obtained through this equation:

$$F(S) = -\frac{1}{\beta} \ln \left( \int dR \delta(S - S(R)) \exp[-\beta U(R)] \right) \quad (2.12)$$

where  $\beta = (k_B T)^{-1}$ ,  $k_B$  is the Boltzmann constant and  $T$  is the temperature of the system and  $U(R)$  its potential energy.

The number of CVs should be low due to the efficiency of the method scaling exponentially with respect to the number of dimensions involved, as well as, the difficulty of visualising multidimensional FES. Additionally, the CVs have to fully describe the studied phenomenon. The region of the phase space accessible for the CVs can be limited, if necessary, with a restraining potential. The lower limit for the boundary condition is often referenced as the ‘lower wall’ and the upper limit as the ‘upper wall’. The restraining potential acts on the system if the CV exceeds its value over the set limit and is given by:

$$\sum_i \kappa_i \left( \frac{x_i - a_i + o_i}{s_i} \right)^{e_i} \quad (2.13)$$

$(a_i)$  is the set upper or lower limit,  $(o_i)$  is the offset from this value,  $\kappa_i$  is the energy constant,  $s_i$  the rescaling factor and  $e_i$  the exponent determining the power of the restraining potential.

Metadynamics simulations have found application in materials science, predicting equilibrium crystal structures and chemical reaction pathways, and in understanding the dynamics of small biomolecules and peptides by exploring their conformational FES.<sup>176</sup> Most recently, in work I have co-authored (not included in this thesis), we used metadynamics to characterise the mechanism of SF<sub>6</sub> diffusion through a narrow triangular window of **CC3**.<sup>88</sup> It required reconstructing a 2D FES obtained for **CC3**-SF<sub>6</sub> system in isolation, using metadynamics to accelerate the otherwise rare event of SF<sub>6</sub> diffusion through **CC3** window. The derived mechanism, for the **CC3**-SF<sub>6</sub> system in isolation, is a likely proxy for the SF<sub>6</sub> diffusion mechanism in the crystalline **CC3** $\alpha$  due to its particular crystalline assembly, where the narrow **CC3** windows are the limiting spatial factor for the diffusion.

## 2.6 Density functional theory

In quantum mechanical methods, the molecular system is fully described by the time-independent Schrödinger equation:

$$\hat{\mathcal{H}}\Psi = E\Psi \quad (2.14)$$

where  $\Psi$  is the wave function that contains all the information on the studied molecular system. The solution to the *Hamiltonian* operator ( $\hat{\mathcal{H}}$ ) acting on the wave function  $\Psi$  is the system total energy ( $E$ ).<sup>177</sup> The difficulty arises with the  $3N$  spatial coordinates and  $N$  spin coordinates of the electrons that the wave function depends on, making the Schrödinger equation unsolvable for systems containing more than one electron without further approximations. In quantum chemical calculations, the Born-Oppenheimer approximation allows the movement of the much larger nuclei to be considered so slow, in comparison to the small and fast electrons, that the electrons can be treated as moving in a potential defined by static nuclei.<sup>178</sup> The kinetic energy of the static nuclei is zero and the nuclei-nuclei repulsion energy is constant, therefore, the total Hamiltonian can be expressed as the sum of the electronic and the nuclear-nuclear interactions. With the latter being a constant value, one can focus on the solution of the electronic interactions.

Density functional theory (DFT) allows one to circumvent directly solving the Schrödinger equation with a set of further approximations.<sup>177</sup> In the previous paragraph, we arrived at the conclusion that the total energy depends only on the positions and atomic numbers of the nuclei and the total number of electrons. Therefore, the total energy can be derived from electron density ( $\rho$ ) instead. The electron density relates directly to the total number of electrons in the system, the electron density maxima to the nuclei positions and the nuclei atomic numbers can also be elucidated from these maxima and the spherically averaged density of the electrons around these maxima. In DFT, the the attraction to the nuclei exerted on the electrons is described by an ‘external potential’. Thus, it can be derived directly from the electron density which is dependent on the attraction of the electrons to the nuclei. This, however, still leaves the unsolvable many-body problem of the electron-electron correlations in any system of more than one electron.

The breakthrough came with the Kohn-Sham self-consistent field (SCF) methodology.<sup>178</sup> In 1965, Kohn and Sham realised that the Hamiltonian constructed for non-interacting electrons can be used as a proxy to the exact solution. For SCF to work, one has to construct a *fictitious* system with non-interacting electrons that in their ground-state density have the same density as the *real* system of interest, where electrons do interact.<sup>177</sup> As the electron density determines the nuclear positions and atomic numbers, these are therefore identical in the fictitious and real systems. As a result, the electron density based energy functional ( $E[\rho(\mathbf{r})]$ ), for the real system, can be written as a sum of specific components:

$$E[\rho(\mathbf{r})] = T_{ni}[\rho(\mathbf{r})] + V_{ni}[\rho(\mathbf{r})] + V_{ee}[\rho(\mathbf{r})] + \Delta T[\rho(\mathbf{r})] + \Delta V_{ee}[\rho(\mathbf{r})] \quad (2.15)$$

where  $T_{ni}[\rho(\mathbf{r})]$  is the kinetic energy of the non-interacting electrons,  $V_{ni}[\rho(\mathbf{r})]$  the nuclear-electron interaction,  $V_{ee}[\rho(\mathbf{r})]$  the electron-electron repulsion,  $\Delta T[\rho(\mathbf{r})]$  the correction to the kinetic energy in respect to the electron-electron interactions and  $\Delta V_{ee}[\rho(\mathbf{r})]$  the sum of all corrections to the electron-electron repulsion energy. The last two terms are often referenced to as the exchange-correlation potential ( $E_{XC}$ ). This includes the effect of quantum mechanical exchange and correlation, the correction for the classical self-interaction energy, but also the difference in kinetic energy for the fictitious system of non-interacting electrons and the real system. In DFT, the dependence of the  $E_{XC}$  on the electron density is often expressed as the interaction of the electron density and the ‘energy density’ ( $\epsilon_{XC}$ ):

$$E_{XC}[\rho(\mathbf{r})] = \int \rho(\mathbf{r})\epsilon_{XC}[\rho(\mathbf{r})]d\mathbf{r} \quad (2.16)$$

where  $\epsilon_{XC}$  is a sum of individual exchange and correlation contributions.

If the exact form of the exchange-correlation potential was known, solving the Schrödinger equation would be possible. As this is not the case, it is required to approximate the form of  $\epsilon_{XC}$ . The simplest approximation of  $\epsilon_{XC}$  is given by the local density approximation (LDA).<sup>177</sup> In LDA, the value of  $\epsilon_{XC}$  at some position  $\mathbf{r}$  is derived from the value of the electron density at that position. This requires that the electron density is single valued at every position, in reality this means deriving the electron density for a uniform electron gas, and in result

the electron density is the same at every point. This approach is a good model for simple metals, however for molecules in which the density varies rapidly, this representation is not sufficient. Therefore, the LDA was surpassed with the generalised gradient approximation (GGA), where the additional term for gradient of the electron density is added to the LDA functional to account for the electron non-homogeneity in molecules. In GGA, the  $\epsilon_{XC}$  is dependent on both the local electron density and the extent to which the electron density is locally changing. Generally, the GGA are more accurate than the LDA and different GGAs are nowadays available, for example PBE or BLYP.<sup>179,180</sup> Nowadays, meta-GGAs and hybrid GGAs provide accuracy at the chemical level, such as B3LYP and B97.<sup>178</sup> However, these are computationally expensive and were not used in this thesis.

In DFT, to construct  $\rho(\mathbf{r})$  a linear combination of function mathematical representation of the atomic orbitals, in form of the basis set, is used.<sup>178</sup> A minimal basis set consists of a linear combination of singular functions describing each orbital. These function can be atom centred (Gaussian type), plane waves or a mixture of both. Double- (DZV) or triple-zeta (TZV) basis sets use combinations of two or three functions respectively to describe the orbitals. The DZV and TZV potentials, used in this thesis, are augmented with the polarisation functions for higher unoccupied atomic orbitals, allowing for the occupied orbitals to distort from their original atomic symmetry and adopt to the molecular environment.<sup>181</sup> In the hybrid Gaussian and plane waves method, the atom centred Gaussian type basis sets are used to describe the wave functions, but the electron density is described with an auxiliary plane wave basis. The electron density is then represented as plane waves or on a regular grid. This allows a linear scaling of the computational expense with the system size in this type of calculations.

## 2.7 Grand canonical Monte Carlo

Monte Carlo (MC) techniques are widely used in determination of the equilibrium state properties of molecular systems in a stochastic and time-independent manner.<sup>182</sup> In MC methods, to calculate an average system property, a large set of independent random configurations is generated. Each generated random state is then accepted or rejected according to the MC algorithm with certain probability. The use of random number generators and the use of probability



that allows to derive equilibrium properties are the reason for the naming of MC methods; due to Monte Carlo's fame for its casino games. As MC methods allow one to calculate an average system's property, the experimental results and MC generated properties can often be directly compared, where the experimental measurements also result from averaged properties of a bulk system.

The adsorption of gases and gas mixtures on porous solids can be simulated using MC methods, with the inter- and intramolecular interactions of the system described with force field potentials.<sup>182</sup> The grand canonical assemble, where the chemical potential of the system, the system volume and temperature are fixed, allows the direct simulation of the adsorption isotherms with the replication of the experimental system behaviour. Thus, grand canonical Monte Carlo (GCMC) is nowadays used routinely to study the gas adsorption of porous solids, such as zeolites, MOFs and PMMs.<sup>183–186</sup> In GCMC, the gas particles are manipulated in the system with 'moves' of particle creation/deletion, particle displacement, and in the case of gas mixtures simulations, the particle identity exchange.<sup>182</sup> Each such move has a given probability ( $P$ ) of occurring. For example, the acceptance of translation or rotation move of a gas particle is often determined with the following probability:

$$P = \min\{1, \exp(-\beta\Delta U)\} \quad (2.17)$$

where  $\beta = 1/kT$  ( $k$  is the Boltzman constant), and  $\Delta U$  is the change of the system's potential energy ( $U$ ). The probability of a particle insertion can be expressed as follows:

$$P = \min\left\{1, \frac{\beta f V}{N + 1} \exp(-\beta\Delta U)\right\} \quad (2.18)$$

where  $N$  is the number of particles present in the system and  $f$  the fugacity of the gas-phase adsorptive. The use of fugacity, instead of pressure, allows the consideration of a real gas behaviour rather than ideal gas, that could lead to measurement disparity from the experimental values. Similarly, the probability of a particle deletion can be written as:

$$P = \min\left\{1, \frac{N}{\beta f V} \exp(-\beta\Delta U)\right\} \quad (2.19)$$

In gas mixtures, the probability of the identity swap move, where the the identity of a sorbate is changed from one component of a mixture to another, is often expressed with the equation:

$$P = \min\left\{1, \frac{f_b N_A}{f_a (N_B + 1)} \exp(-\beta \Delta U)\right\} \quad (2.20)$$

where  $N_A$  and  $N_B$  is the number of particles of species  $A$  and  $B$  in the system and  $f_a$  and  $f_b$  their fugacities.

In MC methods, for the simulated properties to be statistically meaningful, hundreds of thousands of moves usually have to be considered.<sup>182</sup> After each move, the probability of the move acceptance or rejection has to be calculated, preceded by the evaluation of the potential energy of the system. Therefore, MC simulations can often be time consuming and requiring considerable resources. Additionally, to replicate the experimental adsorption isotherms with GCMC, simulations at different pressures are required to replicate the adsorption curve, which multiplies the requirement for computational resources.

## 2.8 Ideal adsorbed solution theory

The Ideal adsorbed solution theory (IAST) developed by Myers and Prausnitz in 1965, is a theoretical model for gas co-adsorption.<sup>182,187</sup> IAST requires only a prior knowledge of the single-component gas adsorption isotherms and allows one to derive the adsorbed amount of each simulated mixture component at constant temperature ( $T$ ) and pressure ( $p$ ). The assumption is that the pure-component adsorption isotherms are measured at the same temperature as the gas mixture of interest, and that the gas mixture behaves as *an ideal solution*. This means that the strength of interactions of all the mixture species are the same, the adsorbent is inert and the same surface area is accessible to all adsorbates.

The equilibrium between the adsorbed phase and the gas phase is equivalent to the vapour-liquid equilibrium, as in Raoult's law for the gas phase, where the partial pressure  $p_i$  of the  $i^{th}$  mixture component is equal to the mole fraction  $x_i$  of the  $i^{th}$  component in the mixture

multiplied by the vapour pressure of the pure component  $p_i^\circ(\pi, T)$ :

$$p_i = x_i p_i^\circ \quad (2.21)$$

The hypothetical pressure of the bulk phase of the pure component  $p_i^\circ(\pi, T)$  can be derived from the pure component isotherm equation, that gives the spreading pressure ( $\pi_i$ ). From the assumption that  $\pi_i$  of each component is the same as the spreading pressure of the mixture ( $\pi$ ):

$$\pi = \pi_i = \pi_j \quad i \neq j \quad (2.22)$$

and from the Gibbs adsorption equation, the following is true:

$$\frac{A\pi}{RT} = \frac{A\pi_i}{RT} = \int_0^{p_i^\circ} \frac{n_i}{p_i} dp_i \quad (2.23)$$

where  $n_i$  is the adsorbed amount of the  $i^{\text{th}}$  component of the mixture, as derived from the single-component isotherm,  $A$  is the area of the adsorbent and  $R$  the ideal gas constant. If we consider the total amount of the mixture adsorbed ( $n_T$ ) is:

$$\frac{1}{n_T} = \sum \frac{x_i}{n_i} \quad (2.24)$$

and keep the mole balance:

$$\sum_{i=1}^N x_i = 1 \quad (2.25)$$

the amount adsorbed ( $n_i$ ) of the  $i^{\text{th}}$  component can be derived from the total amount of the mixture adsorbed ( $n_T$ ) and the component mole fraction in the mixture:

$$n_i = x_i n_T \quad (2.26)$$

The main benefit of IAST is that the multi-component adsorption isotherms can be derived from single-component adsorption isotherms at the same temperature. A beneficial approach compared to laborious mixed-gas adsorption laboratory measurements. Often, cal-

culating the IAST multicomponent isotherms from the GCMC simulated single component adsorption isotherms is faster than GCMC co-adsorption simulations of gas mixtures.<sup>188</sup> IAST is a versatile technique and is widely used in the calculation of mixture component adsorption isotherms in the study of gas adsorption in MOFs.<sup>189</sup>

## 2.9 Evolutionary algorithms

Evolutionary algorithms (EAs) are a type of computer code that mimics evolutionary processes to solve global minimisation problems.<sup>190–192</sup> EA methods, which utilise genetic algorithms (GAs), make use of ideas and terminology usually used to describe biological processes, yet at their core use mathematical and logical methods to fit the optimal solution to complicated chemical problems. In these methods, information is carried by the ‘genetic code’ made of ‘genes’, where genes are variables to be optimised and can be expressed in form of a string. The quality of a genetic code, its ‘fitness’, is dependent on the ‘fitness function’ that scores the solution. The starting point of any EA is the assembly of random solutions into an ‘initial population’ of individuals, where each ‘member’ of the population has a specific genetic code and fitness associated with it. After calculating the quality of each of the members within the population, evolutionary manipulations are performed on that population to obtain an extended population of members. This process, in cycles, renders subsequent ‘generations’ of individuals until the optimal solution, with respect to the fitness function, is found.

In order to understand the operation of a GA, let us assume that the starting genetic code is a string made of four letters of the alphabet in uppercase ‘ABCD’. The fitness function compares the similarity of that string to another string ‘QXYZ’, which is the expected solution. The algorithm uses the evolutionary operations until the fitness of the solution stops increasing, meaning an optimal solution with respect to the fitness function is found. The first evolutionary operation is ‘mutation’. The mutation operation causes a random change in the genetic code, for example the exchange of one of the letters with a new letter of the alphabet, such as ‘X’ to yield ‘AXCD’. Now, the similarity of the new string ‘AXCD’ to the solution ‘QXYZ’ is greater, thus its fitness will be higher. This operation can be repeated until the exact solution, the string ‘QXYZ’, is replicated, or the solution with the highest similarity is found. The role of

a gene mutation is to enrich the population with new genetic information, therefore increasing population diversity.

The second type of GA manipulation is ‘crossover’. This operation is directly inspired by the biological process of mating, where the genetic information carried by two parents is recombined to produce an offspring. This way offspring shares the genetic features of parents. The role of crossover is different from mutation and does not enrich the population with new data. Instead, it enables a reshuffle of the genetic information already encoded within a population. This enables it to sample different regions of parameter space simultaneously and to find better individuals through the consideration of all the features shared within the population. For example, if we consider the genetic code of two parent individuals ‘ABCD’ and ‘EFGH’, a crossover can result in ‘ABGH’ and ‘EFCD’ offspring. Now the offspring shares the genetic code of the parents.

Each time a population is generated, each individual is evaluated with the fitness function. The mutation and crossover operations are then applied to the selected individuals. A common selection process of the candidates is the ‘roulette wheel’. The individual is selected for crossover or mutation if its’ fitness value  $f(x)$  follows the  $f(x) > y$  relation, where  $y$  is a randomly generated value between 0 and 1. Therefore, the higher the fitness of a population member, the higher the likelihood of a member contributing the generation of new individuals. One of the common rules in EAs is to keep the size of a population fixed during the evolutionary process. The candidates for the new generation are a sum of current population members and the individuals resulting from the crossover and mutation operations. To keep the population size constant, only the fittest candidates with the highest fitness propagate to the next generation.

The efficiency of EAs depends on many factors, such as the specific set up of the calculations and the problem investigated. For example, additional selection criteria can be applied such as ‘elitism’, that ensures the propagation of the fittest candidate to the new generation. These are carried out until a certain number of generations is reached or the exit rules are met. For example, when a candidate meeting the fulfilment criteria is found and the EA cycle is ended, or when the average fitness of the population converges to the expected value.

EAs have been widely employed in drug design studies. Chemoinformatics used in drug

discovery often represent molecules as vectors, which makes EAs easy to implement.<sup>193</sup> In materials design studies, EAs can be applied to solve crystal structure prediction problems,<sup>194–196</sup> and in the design of molecules and materials with desired properties and function for applications.<sup>197,198</sup> In Chapter 6 of this thesis, I will present the application of an EA in the computer-aided design of new porous organic cage molecules for application.

## Chapter 3

# Development of `pywindow` for the Analysis of Structural Properties of Molecular Pores

In this chapter I will present the `pywindow` Python package; an automated tool for the structural analysis of molecular pores. I will discuss the methodology and workflow of the code followed by the validation process, the examples of application, and the discussion of the possible limitations.

The `pywindow` source code is available on GitHub:

<https://github.com/marcinmiklitz/pywindow>

## 3.1 Introduction

The motivation behind developing the `pywindow` package was the lack of readily available software to carry out the structural analysis of molecular pores; the subject of this thesis. Molecular pores, introduced in Section 1.2.2, have a cage- or belt-like shape with intrinsic voids and windows. These molecules are the building units of the previously described porous molecular materials (PMMs) that can be amorphous or crystalline, where multiple polymorphs are not uncommon.<sup>68,199</sup> Additionally, as the result of their discrete character, the molecular pores can find in-solution applications<sup>99</sup> or be incorporated into materials like mixed-matrix membranes.<sup>61,62</sup> Therefore, the structural descriptors of individual molecular pores, such as the void size and window diameter, and their relationship to properties, might be more informative where the isolated molecules are studied, or the solid state assembly and the void connectivity have insignificant effect on the property. The structural characterisation tools, mentioned in Section 2.2, allow one to describe the extended porous networks in terms of structural descriptors such as pore limiting diameter (PLD) and largest cavity diameter (LCD), the porous network dimensionality and connectivity, or the specific surface area. The described tools can also be used to study the structure and porosity of PMMs, however, these do not characterise the underlying features of the individual molecular pores, but rather of the pore network as a function of the molecular pores' assembly.

The void size and window diameters yield the general description of the individual molecular pores, in the same fashion as the LCD and PLD that of the porous network, and in some particular molecular assemblies these can be the same. The manual measurements of window diameters, within the available visualisation software, is laborious, prone to human error, and unfeasible for the analysis of molecular dynamics (MD) trajectories often counted in thousands frames. The circumcircle method,<sup>160</sup> described in Section 2.2, facilitates the measurement of window diameters, however, a prior visual inspection of the molecule is necessary to assign triads of atoms describing the windows. The future considerations of window size calculation in materials design, described in Chapter 6, required a construction of an automated tool, that I have called `pywindow`. Furthermore, the designed software is a transferable tool that does not depend on the given topology and chemistry. Although in this chapter I will focus on



the application of the structural analysis of molecular pores, porous organic cages (POCs) and belt-like molecules, other materials such as coordination polyhedra or representative fragments of metal-organic frameworks and zeolites can also be analysed. The source code of this software was made open source and is freely available, and I will now present its functionality.

## 3.2 Methodology of the *pywindow* software

The *pywindow* software was developed using the Python3 scripting language,<sup>200</sup> and the paradigm of object-oriented programming (OOP).<sup>201</sup> This results in a flexible workflow that does not limit the code application to a single scheme, and allows code flexibility and merging with other software. The OOP methodology requires the information to be stored in the form of containers, called *objects*, and to create methods that work on the object and can be invoked if necessary. This gives full control over the analysis procedure to use only the functionality of the code that is of interest. OOP codes support the use of interactive programming tools like IPython<sup>202</sup> or Jupyter.<sup>203</sup> Pre- or post-processing methods can then be easily applied on the object, even if they are not part of the initial software workflow.

The atomic weights ( $m$ ), the van der Waals atomic radii ( $r_{vdw}$ ) and the covalent radii ( $r_{cov}$ ), used in the *pywindow* software were taken from the ‘CSD Elemental Radii’ spreadsheet as found on the Cambridge Structural Database (CSD) website.<sup>204</sup> The code has no underlying units and inherits that of the input file, however, all molecular coordinates are assumed to be in Angstroms [Å]. This condition is necessary for calculations that depend on the  $r_{vdw}$  with the unit of Å.

### 3.2.1 Nomenclature

Before discussing the workflow of the *pywindow* software, I will first introduce the necessary nomenclature that is intended to facilitate the understanding of the basic functionality and the code structure. The package is built upon *modules* that store *classes* and independent *functions*. A module is referenced to by its name in *italics*. A class is denoted by the class name, in `code` formatting, with no spaces between separate words that each starts with a capital letter and the

class name is followed by brackets. An object created by the class inherits the class name but without brackets. The functions and class methods are in the `code` formatting with lower case letters only, where words are separated with underscores, and the function or method name is followed by brackets. Therefore, a class `ClassExample()` of module *moduleexample* creates an object `ClassExample` and contains the methods like `method_example()` that can perform an operation on the `ClassExample` object.

The *molecular system* is here defined as the structural and chemical information on either a single molecule or an assembly of molecules, in the form of the Cartesian coordinates, the elements or force field atom types, and the unit cell information (angles and cell dimensions for periodic systems). A *molecule* refers to each discrete molecule, regardless of whether it is a molecular pore or not, that is part of the molecular system. A *trajectory* is a set of *frames* containing the structural information equivalent to the *molecular system*.

### 3.2.2 Workflow

In Figure 3.1 the simplified schematic of the *pywindow* code structure and possible workflow is presented. The input data can be extracted from one of the commonly used file extensions (XYZ, PDB, MOL), containing the Cartesian coordinates and atom types, through the `MolecularSystem()` class of the *molecular* module, ported directly from the `RDKit` molecule object also through the `MolecularSystem()` class, or from the trajectory file loaded with one of the *trajectory* module classes. The latter method requires the pre-processing step of extracting each individual frame from the trajectory.

If additional pre-processing is necessary, such as the assignment of elements to the force field atom types, or the rebuilding of a unit cell to facilitate the presence of the periodic boundary and the fact that the analysis can only be performed on whole molecules (see Section 3.2.4), the `decipher_atom_keys()` or `rebuild()` methods of the `MolecularSystem()` class can be used. The structural analysis is then carried out for each discrete molecule in the molecular system. These are extracted with the `make_modular()` method that returns each molecule as a `Molecule` object. Finally, the `Molecule()` class methods allow one to measure the structural features.

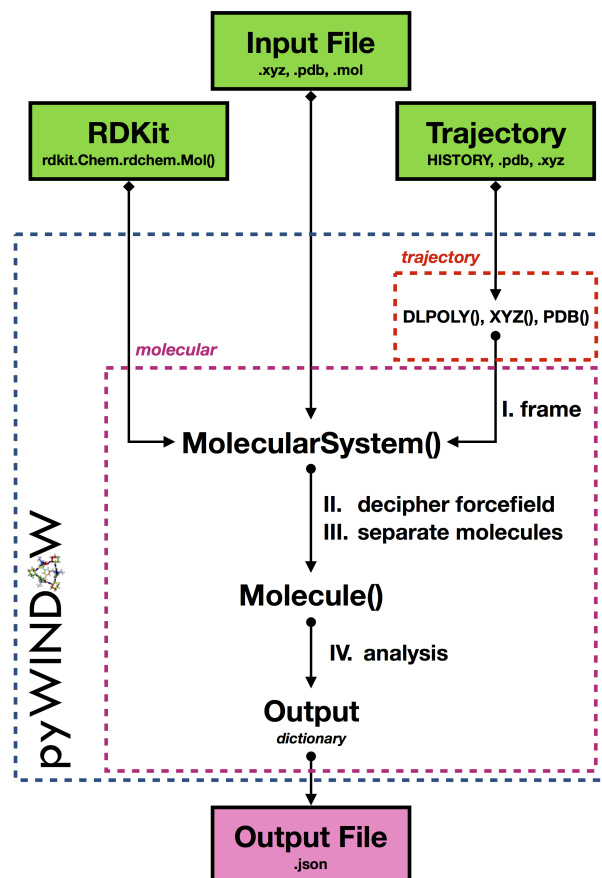


Figure 3.1: The Workflow schematics of the *pywindow* software.

### 3.2.3 Reconstruction of the periodic unit cell

As a result of periodicity, the atom positions enclosed in the unit cell will often cause the molecules to cross through periodic boundaries. The analysis requires the reconstruction of the molecules based on the atomic positions and the connectivity between atoms. Therefore, a simple algorithm was implemented in *pywindow* that allows to rebuild the systems (the `rebuild()` method of the `MolecularSystem()` class). First, a  $3 \times 3 \times 3$  supercell is constructed resulting in 27 symmetry representations of the system, with the initial unit cell in the centre of such supercell. For all atomic positions (excluding some elements<sup>1</sup>) the one closest to the origin is picked as the seed for the individual molecule reconstruction. With a formula to determine if a bond is present between two atoms A and B we iterate through all atoms in the system.

<sup>1</sup>H, Cl, Br, F, He, Ar, Ne, Kr, Xe, Rn

The atom pair A and B is connected if the distance ( $d_{A,B}$ ) meets the follow criterion:

$$d_{A,B} < 2 * r_{cov,max} + t \quad (3.1)$$

where  $r_{cov,max}$  is the largest covalent radius of all the atom types in the molecular system and  $t$  is a user defined cut-off. In each iteration, the atoms that are determined as bonded to any of the atoms in the previous set are sampled for their atomic neighbours. This is performed for the initial unit cell only, the role of the atom representations from the supercell is to allow cross boundary connections. The atoms from the supercell area can populate the search set. However, none of the atoms from the supercell area can act as a starting seed. Only one occurrence of an atom, out of the 27 symmetry equivalences, can be assigned to a molecule.

For each molecule found, the centre of mass (*COM*) is determined and if it is in the initial unit cell, this molecule is then returned as part of the rebuilt molecular system. The performance of the algorithm is presented in Figure 3.2 with an example of a cubic, monoclinic and triclinic symmetry periodic unit cell. The individual molecules for the analysis are then extracted using the `make_modular()` method, of the `MolecularSystem()` class, that returns each molecule as the `Molecule` object. This is done on the connectivity criteria based on the individual atom pair distances and the covalent radii.

### 3.2.4 Structural properties of molecular pores

The structural descriptors that can be calculated using *pywindow* for a molecular pore are:

- *COM*: the centre of mass of a molecule.
- $D_{max}$ : the maximum diameter of a molecule.
- $D_{void}$ : the intrinsic void diameter of a molecule.
- $V_{void}$ : the intrinsic void volume of a molecule.
- $D_{void\_opt}$ : the optimised intrinsic void diameter of a molecule.
- $V_{void\_opt}$ : the optimised intrinsic void volume of a molecule.

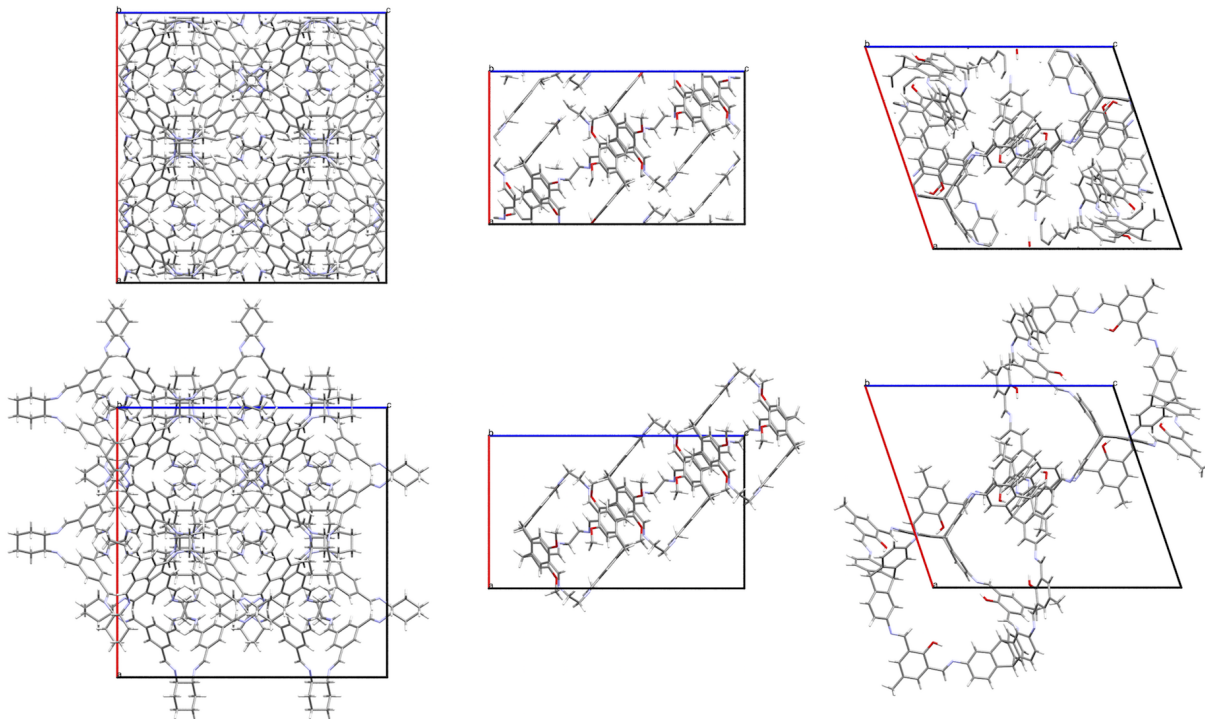


Figure 3.2: Three examples of the periodic unit cells before (top) and after (bottom) the rebuild process. The example on the left is of the cubic symmetry unit cell containing eight discrete molecules, the middle example of a monoclinic unit cell containing three discrete molecules and the right example of a triclinic unit cell containing two discrete molecules. Any solvent molecules have been omitted for clarity.

- $D_{window}$ : the circular diameter of an  $x^{th}$  window of a molecule.

The *COM* of a molecule is calculated according to the equation:

$$COM = \frac{1}{MW} \sum_{i=1}^n m_i v_i \quad (3.2)$$

where  $MW$  is the molecular weight of the molecule,  $v$  is a vector containing the Cartesian coordinates of an atom and  $m$  is the atomic mass for the element. The formula yields *COM*, which is a vector that corresponds to the coordinates of the centre of mass of the molecule.

The molecule's maximum diameter ( $D_{max}$ ) is determined as the distance between two furthest atoms in the molecule. It is obtained using a Euclidean distance matrix between all atoms in the molecule. From the obtained matrix of distances, the largest value found is between the coordinates of the two furthest atoms. The distance value is then corrected for the corresponding  $r_{vdw}$  for the pair of elements, by adding it to the calculated distance yielding the  $D_{max}$ .

The intrinsic void diameter ( $D_{void}$ ) is determined as the distance between the *COM* of a molecule and the closest atom. This is calculated using the euclidean distance matrix between the *COM* vector and the positions of all elements in the molecule. The smallest value in the set of distances is then corrected by subtracting the appropriate  $r_{vdw}$  of the determined closest element and multiplying it by 2 to yield the  $D_{void}$ .

The intrinsic void volume ( $V_{void}$ ) of a molecule is calculated with a formula for a sphere volume, where the void radius  $R_{void}$  equals to  $\frac{1}{2}D_{void}$ . The function for the optimised intrinsic void diameter  $D_{void\_opt}$  facilitates the possibility where the assumed overlap of the *COM* with the pore centre, as in  $D_{void}$  and  $V_{void}$ , does not occur in the non-symmetrical molecule. This was addressed by introducing an optimisation step in the void centre determination before calculating the void diameter. The *COM* of the molecule is used as a starting guess for the coordinates of the pore centre in the minimisation function that calculates the negative value of the the  $D_{void}$  as a function of coordinates that are being minimised. This allows one to find the optimal coordinates of the intrinsic void and the  $D_{void\_opt}$  is yielded in the same fashion as the  $D_{void}$ . The  $V_{void\_opt}$  is equivalent to  $V_{void}$  with the difference of using  $R_{void\_opt}$  as half of the obtained  $D_{void\_opt}$ .

The process of calculating the window diameter ( $D_{window}$ ) for each of the windows in the molecule, determined as the channel necking that connects the intrinsic void with molecule's surrounding, is presented in Figure 3.3 and is performed as following:

1. From the Cartesian coordinates of a single molecule, the *COM* is calculated and the molecule is shifted to the origin ( $O$ ) of the Cartesian system by subtracting the *COM* vector from the molecule's coordinates.
2. On a sphere, with radius equal to  $D_{max}$  and the centre of the sphere located at  $O$ , a set of *sampling points* is generated with the Vogel's method<sup>205</sup> for a spiral distribution of points on a disc, adopted for a sphere using cylindrical coordinates. The approach allows one to obtain a uniform distribution of points on a sphere surface (see Figure 3.3a).
3. The number of sampling points is system dependent and calculated according to the

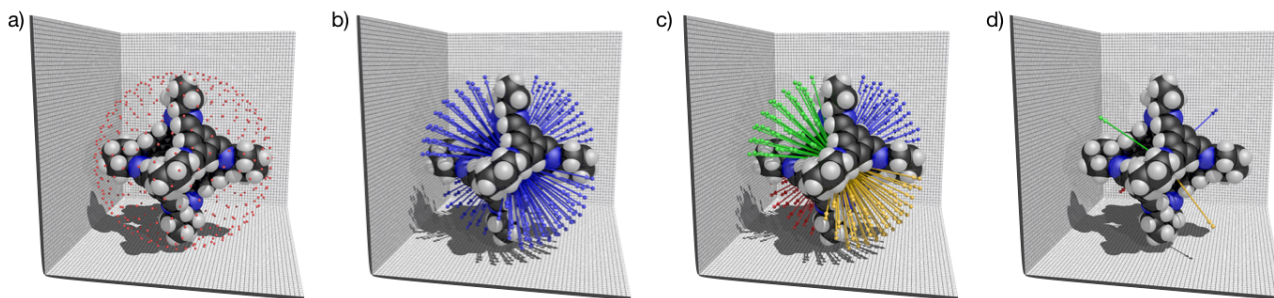


Figure 3.3: The steps of determining the window diameter. a) a sphere of evenly distributed sampling points is generated around a molecule b) vectors connecting the centre of mass of a molecule and the sampling points are analysed for the overlap with the van der Waals atoms spheres of the molecule’s atoms c) the vectors that do not overlap with any of the molecule’s atoms van der Waals spheres (vectors passing through windows) are clustered into distinct windows. d) for each window the vector with the largest included sphere along the vectors path is chosen and the window’s circular diameter calculated.

formula:

$$n = a * \log_{10}(A) * 250 \quad (3.3)$$

where  $A$  is the surface area of the sphere in  $\text{\AA}^2$ . The logarithmic scaling is used that prevents oversampling of larger molecules, which is the bottleneck for the algorithm performance, and the 250 factor proved in the design and validation processes to give a sufficient sampling level. Further, the scaling factor  $a$  allows further control on the sampling level. For each sampling point a *sampling vector* is defined that starts at the  $O$  and the end point are the coordinates of a given sampling point. A set of coordinates along the vector in increments is calculated and defined as the ‘vector path’. At each point along the vector’s path the distance of the point to the position of the closest atom is calculated and corrected for the appropriate  $r_{vdw}$  of an element by subtracting it from the distance.

4. Each vector with a set of only positive distances, meaning the vector does not cross through the van der Waals sphere of any atom in the molecule, are then clustered using the density-based spatial clustering algorithm (DBSCAN) from the `scikit-learn` package (see Figure 3.3b, c).<sup>206</sup> The number of returned clusters defines the number of windows found in the molecule.
5. The found clusters are analysed separately. For each such cluster the vector’s paths

- sets are compared for its minimum, therefore the smallest distance in the set, and the vector with highest value for that minimum is determined as the rough estimate of the window centre and the coordinates of this minimum and the distance calculated is used to determine the window circular plane perpendicular to the vector (see Figure 3.3d).
6. The diameter of a window ( $D_{window}$ ) is determined by the largest circle that can be circumscribed into the window. The radius of that circle is defined by the distance between the centre of the window and the closest atom. From that radius the appropriate  $r_{vdw}$  of that element is subtracted and it finally is multiplied by 2 to yield  $D_{window}$ . First the rotation of the molecule's coordinates is performed to align the sampling vector with the  $Z$  axis with the  $COM$  of the molecule kept at  $O$ . In such an arrangement, the plane of the window is perpendicular to the  $XY$  plane of the coordinate system. The two step minimisation, where the coordinates of the window centre are used as variables and  $D_{window}$  as the output is performed. The  $x$  and  $y$  components are minimised by using them as the variables in the `minimize` function from `SciPy` package<sup>207</sup> and the negative of  $D_{window}$  is returned. This ensure that the correct centre is found, *i.e.* one that has the largest distance to the closest atom in the window plane. Next, the optimisation of the  $z$  coordinate is performed in the  $Z$  axis direction also using the `minimize` function of `SciPy` package, with the  $D_{window}$  value as output. This ensures the correct distance of the window centre from the cage  $COM$  at which the window is the narrowest. For this optimised centre of the window, the  $D_{window}$  value is recalculated. At this point, the reverse rotation and initial molecule's  $COM$  translation to the  $O$  is applied to the window centre coordinates by applying the appropriate rotation matrix and adding the initial  $COM$  vector.
  7. The  $D_{window}$  for each window, and the central coordinates for each window with respect to the original molecule's input coordinates are returned as output.



## 3.3 Validation

The validation process included the structural analysis of a set of molecular pores. For one of these, the obtained values were first compared to the other known methods to calculate window diameters presented in the literature, the circumcircle described in Section 2.2 and the PLD output of Zeo++. Then, for the set with the varying number of windows, the `pywindow` output was compared with the visual inspection of the determined values. Lastly, the MD trajectory of an isolated POC was analysed using `pywindow` and the circumcircle method and the generated pore limiting envelopes were then compared.

### 3.3.1 Computational details

#### Analysis of the individual molecular pores

The crystallographic information files (CIFs) containing the crystal structures of the molecular pores presented and analysed in this section were obtained from the Cambridge Structural Database (CSD).<sup>26</sup> The molecular pores are referenced by the CSD refcodes under which the CIFs were found in the CSD. The Cartesian atomic coordinates of a single molecular pore from the obtained CIF were extracted using the Materials Studio software.<sup>208</sup> The analysis of the window diameters was performed using the `pywindow` package. The circumcircle method, from reference 161, is presented for comparison. The LCD and PLD of **CC3** were calculated using the Zeo++0.2.2 package<sup>152</sup> and the CIF (CSD refcode: PUDXES) as input.

#### Analysis of the MD trajectory

The MD simulations was performed using the DL\_POLY2.20 software.<sup>169</sup> The Cartesian coordinates of a single molecular pore were extracted directly from the CIF (CSD refcode: PUDXES) and the OPLS\_2005 allatom force field<sup>209</sup> atom types were obtained from the Schrödinger LLC's MacroModel `ffld_server` script. The DL\_POLY2.20 software input files were prepared using the DL\_FIELD3.3 software<sup>210</sup> and the OPLS\_2005 force field. A 1 ns production run and 0.2 ns equilibration was performed, with a timestep of 0.5 fs, at 300 K and 0.0001 Pa. The Coulomb summation for the electrostatic interactions was used. The Leapfrog Verlet al-

gorithm<sup>211</sup> was used to assign forces and velocities. No periodic boundary condition was used with the image conversion set to 0 in the CONFIG input file. The obtained HISTORY trajectory file was then analysed using the `pywindow` package and the `HISTORY()` class of the `trajectory` module. Each frame was extracted as the `MolecularSystem` object, and with the `make_modular()` method passed to a `Molecule` object where the structural analysis to measure the window diameters was performed. The pore limiting envelopes were generated using the kernel density estimate (KDE) function of the `SciPy` package.

### 3.3.2 Window diameters and number of windows

The first step in validation of the `pywindow` package is the comparison of the calculated window diameters to other methods. First, a system has to be chosen carefully for the comparison. The system with a CSD refcode PUDXES, which is a porous organic cage also known as **CC3** is chosen for few specific reasons: i) **CC3** was previously studied using the circumcircle method to calculate window diameters, which will allow for a direct comparison with that method; ii) The particular window-to-window assembly of the **CC3** molecules in the PUDXES crystal structure, also known as polymorph  $\alpha$ . The window diameter happens to be the narrowest point in the pore in this case and therefore it is possible to compare the PLD output by `Zeo++` to the window diameter; iii) the intention of the `pywindow` software is to analyse such molecules, where the intrinsic void of this porous organic cage is significantly bigger than the window diameters.

For the analysis with `Zeo++` I used directly the CIF acquired from the CSD. For the circumcircle method and structural analysis we use an XYZ file of a single discrete molecule extracted directly from the CIF. I visually inspected the molecule to choose four atom triads each describing an individual window. These atom triads lie in a common plane and correspond to the narrow point in a window's pore (see Figure 3.4). This operation is performed for all four windows in this molecule. The result of a comparison of the `pywindow` package analysis output ( $D_{window}$ ) to the window diameters determined with the circumcircle method and the PLD output by `Zeo++` software is given in Table 3.1. The circumcircle method is also equivalent to a manual measurement in any software with a graphical user interface. The results for all three

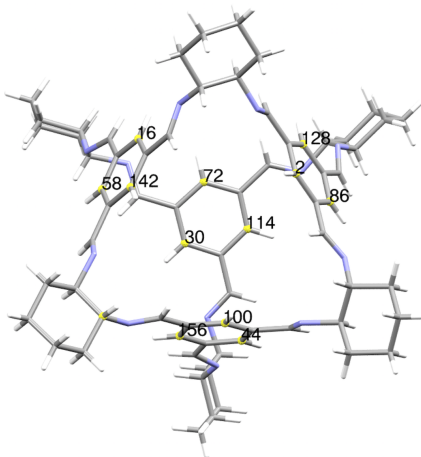


Figure 3.4: The atoms chosen in the PUDXES molecules to represent the window diameters. For each window, three carbon atoms that lie in the same plane and describe the narrowest point of the window channel necking were chosen.

methods are comparable. The Zeo++ results give the largest window diameters of 3.66 Å. The circumcircle method gives the consistent value of 3.63 Å and the pywindow based analysis an average value of 3.64 Å.

Table 3.1: The comparison of window diameters determined for PUDXES using pywindow and circumcircle method, as well as the PLD output of Zeo++. There is only one PLD generated for the Zeo++, however, from the symmetry considerations we report the value for each window separately. The values are for a single static snapshot of the molecule’s coordinates as extracted from the CIF.

<i>Methods</i>	$D_{window,1}$ [Å]	$D_{window,2}$ [Å]	$D_{window,3}$ [Å]	$D_{window,4}$ [Å]
pywindow	3.64	3.64	3.64	3.63
<b>circumcircle</b>	3.63	3.63	3.63	3.63
Zeo++0.2.2	3.66	3.66	3.66	3.66

Next, a set of molecular pores that meet the definition of being intrinsically porous and to possessing a variety of number of windows (with an exception of the control molecule  $C_{60}$  fullerene) were analysed with pywindow. These are shown in Figure 3.5. The BATVUP, NUXHIZ, PUDXES, and REYMAL molecules contain 2, 3, 4 and 6 windows, respectively. The YAQHOQ has no windows and was used as a control. The total number of windows identified and the calculated  $D_{window}$  value for each window is presented in Table 3.2. The total number of windows identified for each molecule equals the number of windows identified from

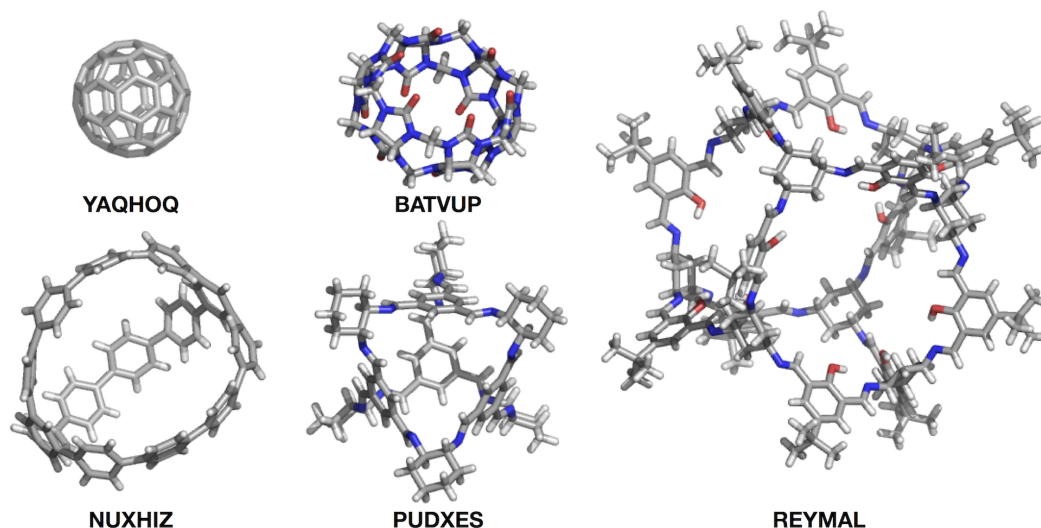


Figure 3.5: The five molecular pores chosen for the validation of `pywindow`. These were taken from the CSD database, with refcodes of the CIF that the Cartesian coordinates were extracted from below each molecule.

the visual inspection of Figure 3.5. The PUDXES has a very consistent window size, whereas the BATVUP, NUXHIZ and REYMAL molecules have windows of various sizes. From visual inspection and manual measurements this is indeed the case. The obtained result for the BATVUP molecule might seem counterintuitive at first, as from the symmetry consideration both windows should be identical. However, the fact that the molecule was extracted from the crystal structure directly might explain the difference. To confirm that, I performed a force field based minimisation (`MacroModel`, `OPLS_2005`, standard setup) that led to a structure with two identical windows with a circular diameter of 4.15 Å each. The control molecule YAQHOQ returned no value as no windows were identified.

Table 3.2: The window diameters and the total number of windows identified using the `pywindow` package for various molecular pores.

<i>System</i>	<i>Number of windows</i>	$D_{window}$ [Å]
<b>YAQHOQ</b>	0	-
<b>BATVUP</b>	2	3.73, 3.34
<b>NUXHIZ</b>	3	6.50, 7.90, 7.27
<b>PUDXES</b>	4	3.64, 3.64, 3.64, 3.63
<b>REYMAL</b>	6	9.06, 9.17, 9.18, 9.16, 9.19, 9.05

### 3.3.3 Analysis of molecular dynamics trajectories

So far, the applicability of the `pywindow` package to measure automatically window diameters of various molecular pores has been presented. Here, the MD trajectory of a molecular pore is analysed to create the pore limiting envelope (PLE), a distribution of window diameters over the trajectory. The resulting PLE generated using the `pywindow` package is compared to the PLE generated using the `circumcircle` method with a similar methodology to that presented in reference 161. Results are presented in Figure 3.6.

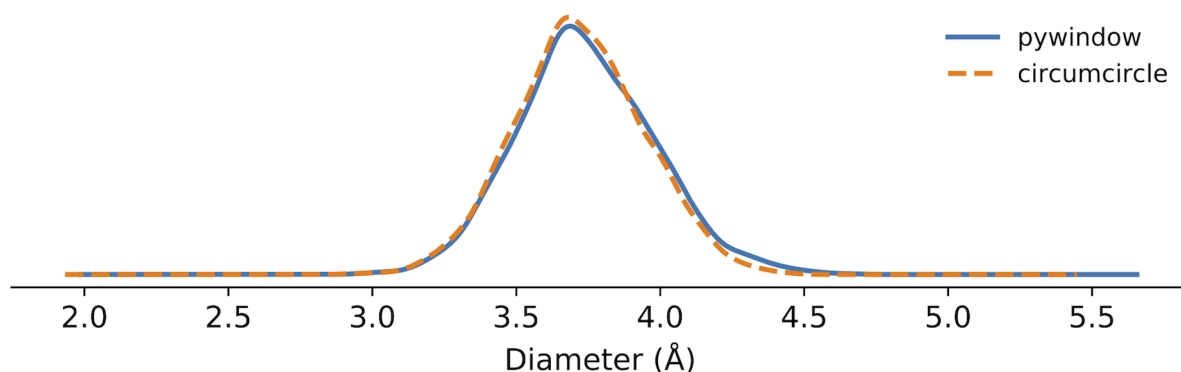


Figure 3.6: The PLEs generated using the `pywindow` package (blue line) and `circumcircle` method from reference 161 (orange dashed line).

The PLE generated with the `pywindow` software is very similar to the one generated using the `circumcircle` method. The advantage of the `pywindow` package is that there is no need for the initial visual inspection, as required in the `circumcircle` method (see Section 3.3.1). Additionally, the three-atom representation of the window is not always ideal, with different shapes of the windows in molecular pores (see discussion in Section 3.4). The computational expense required to calculate the circular window diameter using `pywindow` software is significantly larger than that of the `circumcircle` method. The analysis of a single **CC3** molecule (168 atoms, the maximum diameter of 22.4 Å) on a single CPU processor (2.4 GHz Quad Core Intel Xeon) takes `pywindow` ca. 3.5 s in comparison to 11 ms required by the `circumcircle` method. However, the necessary visual inspection of the molecule for the second method greatly exceeds the time difference in favour of `pywindow`. That difference will be more significant in the analysis of the trajectory. Therefore, the process of trajectory analysis is parallelised so that multiple frames can be analysed simultaneously. Thus, the analysis is still cheap and easily achievable over

a matter of hours or days at most, with the whole process automated for the extraction of the coordinates, reconstruction of the unit cell, and determination of structural properties of individual discrete molecules.

### 3.4 Limitations of *pywindow*

The *pywindow* package can be unfit for the analysis of some molecular pores even from the group of POCs. The definition of the window that has been described earlier states that it is the necking in the molecule's structure that connects the intrinsic void and the cage's exterior. However, the example of a cage can be found where such a necking is not present, or to be more accurate, it is the intrinsic void that is the narrowest point (see Figure 3.7).

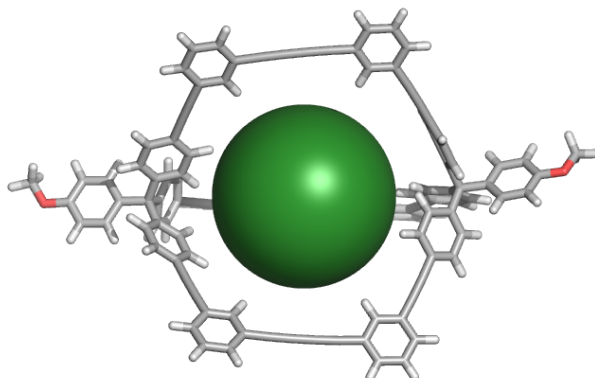


Figure 3.7: The example of a molecular pore where the intrinsic void (green solid sphere) is simultaneously the channel necking of the windows.

Another example of possible faulty analysis is found in the cyclodextrin family. Their specific belt-like structure contains two windows. However, one is wider than the intrinsic void and one is smaller. Here, the larger window does not meet the window definition and during the minimisation step, the  $z$  coordinate of the window centre will be minimised to the void's centre at the origin. Additionally, the assumption that the window is of a circular shape will result in variety of range of underestimations of the actual window's area. For the PUDXES molecule, in the middle of Figure 3.8, the window is actually of a triangular shape. However, the assumption of the circular window shape is intentional for its simplicity. In some extreme cases, such as cryptophane-A (far right in Figure 3.8), very narrow windows close to the shape

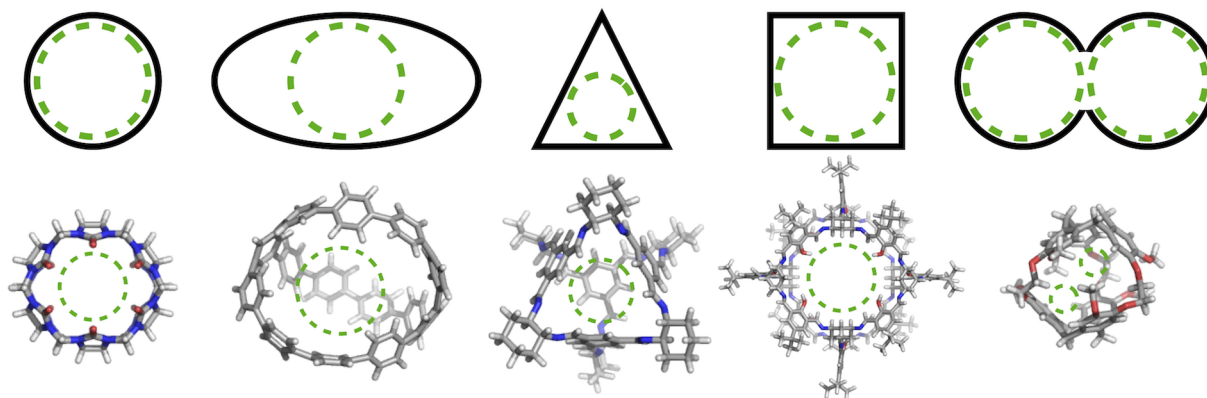


Figure 3.8: Examples of molecular pores with various windows shapes and the possible fit of the spherical window approximation (green dashed circle) as assumed in the `pywindow` package window diameter calculation. Molecules not shown to scale.

of number 8 can result in each of the circles to being identified as a separate window and the wrong number of windows output by `pywindow`.

## 3.5 Conclusions

The `pywindow` Python package, its structure and functionality, has been presented and it is now available and open-source. This tool allows for the determination of the following molecular pores structural descriptors: the maximum diameter of a molecule, the intrinsic void diameter and its volume, and the window diameter. The procedure to prepare the input, decipher the force field atom keys, reconstruct molecules in the periodic unit cell, and to extract individual discrete molecules, was presented. The possibility to use the code to determine number of windows and their diameters, in molecules of various shape and chemistry, has also been demonstrated. The tool can be used to determine whether a material has predisposition to be porous and whether the windows are big enough to allow for guest diffusion through the material. I have presented the functionality to analyse molecular dynamics trajectories, in an automated and straightforward fashion. The pore limiting envelope allows one to relate the dynamic change of window diameter, that relates to the flexibility of the material, to the diffusion of guests. The possibility of guest molecules to bind within the pores can be assessed with the calculation of the cavity size.

In summary, the benefits of the novel approach to determine various structural descriptors of molecular pores, especially the circular window diameter, and of the adopted OOP methodology in the code design, are as follows:

- There is no requirement to visually assess the molecules prior to the analysis.
- No predefined information about the molecule is required.
- The ability to analyse isolated molecules, as well as periodic materials made of discrete units. Possibility to extend the functionality to other types of materials such as coordination polyhedra.
- Fairly quick analysis of single molecules.
- A variety of input files accepted and the ability to read and extract data from trajectories.
- A fully automated and parallelised process of trajectories analysis.
- Stand-alone character of the code that does not require any third party libraries apart from commonly available ones.
- The object-oriented character of the code and modularity makes it easily extendable.

The `pywindow` package was compared with the two other available approaches to determine the window diameter, the circumcircle method and the `Zeo++` software. The validation showed a good agreement in the determined values. The fact that the code is modular and there is no pre-processing required allowed `pywindow` package to be implemented into another software for the computational high-throughput design of novel POCs using the evolutionary algorithms, as presented in Chapter 6. The `pywindow` software is equipped with novel features allowing for faster and automated analysis of molecular pores and can facilitate the faster computational discovery of new materials. The planned extension of `pywindow` functionality, methods to determine other structural properties such as the shape of the molecule, in terms of asphericity and acylindricity, shape of the cavities and windows, can make it a powerful tool for the future. This will make it possible to relate the structural properties of molecules and the observable properties of materials with a higher accuracy.



# Chapter 4

## Screening of Molecular Pores for Xenon/Krypton Separation

In this chapter, a computational screening of previously reported PMMs for potential Xe/Kr separation application, is presented. The analysis of single molecules and their complexes, is evaluated, in comparison to the results obtained for the solid state structures. The bulk properties estimated from the single molecule analysis significantly decrease the computational resources required. The analysis of the structural features, the DFT guest binding energies, and the effect of the hosts' flexibility studied with MD and metadynamics are presented. The **CC3** molecule, previously reported as high performing for Xe/Kr separation, is concluded as the most promising of this class of materials reported to date. The noria and noria-Boc molecules are also found to be promising, and they were therefore synthesised by the synthetic collaborators and tested for their performance for Xe/Kr separation.

*The content of this chapter resulted in the following published work:*

M. Miklitz, S. Jiang, R. Clowes, M. E. Briggs, A. I. Cooper and K. E. Jelfs, Computational Screening of Porous Organic Molecules for Xenon/Krypton Separation, *J. Phys. Chem. C*, 2017, **121**, 15211–15222.

The experimental details were provided by Dr. Shan Jiang, Mr. Rob Clowes and Dr. Michael Briggs, from the University of Liverpool.

## 4.1 Introduction

The separation of noble gases, Xe and Kr in particular, is of industrial relevance due to their application in the lighting industry,<sup>212</sup> anaesthesia,<sup>213</sup> medical imaging<sup>214</sup> and scientific research.<sup>215</sup> Currently, Xe is extracted directly from the air, where it is present as only a small component (0.086 ppm),<sup>216</sup> but the nuclear industry is a possible alternative source.<sup>92</sup> The separation of Xe from used nuclear fuel (UNF) would significantly lower the storage cost of the remaining <sup>85</sup>Kr, as well as providing a new source for Xe. However, the cryogenic distillation of Xe from air or UNF is costly from both a financial and energetic perspective. Therefore, separation-based processes are an attractive alternative.<sup>92</sup> The extraction of Xe from gaseous mixtures can be achieved through a thermodynamic or kinetic separation route employing porous materials. The thermodynamic equilibrium-based separation of gases results from the different binding energies for components of the mixture. As Xe ( $r_{vdw} = 2.05 \text{ \AA}$ ) is larger than Kr ( $r_{vdw} = 1.845 \text{ \AA}$ ) and has greater polarisability and thus potential for favourable long range interactions, a majority of porous materials will selectively sorb Xe over Kr.<sup>217</sup> The kinetic separation of gases relates to differential diffusion rates of the two gases through a material. The overall selectivity of porous materials for Xe/Kr is usually a result of a combination of both these thermodynamic and kinetic effects.

Several recent studies have investigated the performance of porous materials such as zeolites, COFs, ZIFs and MOFs for noble gas separation.<sup>186,218</sup> Banerjee *et al.* discussed MOF candidates and their potential in Xe/Kr separation for reprocessing of UNF, showing that MOFs can exhibit high selectivities and are a viable alternative to cryogenic distillation.<sup>92</sup> Molecular pores as well were shown to be competitive for Xe/Kr separation, where the organic cage **CC3** exhibited exceptional performance in breakthrough experiments.<sup>73</sup> **CC3** selectivity for an industrially relevant mixture of Xe (400 ppm) and Kr (40 ppm) balanced with common air components at 1 bar, room temperature and flow conditions was reported as 20.4. Chen *et al.* demonstrated the host flexibility effect on **CC3** and its experimentally derived Xe/Kr selectivity.<sup>73</sup> The **CC3** window diameter of 3.6  $\text{\AA}$  in the single-crystal X-ray diffraction structure at 293 K is too small for either Xe or Kr to diffuse through. MD simulations were able to explain this apparent discrepancy with the experimental observations; by calculating pore limiting envelope (PLE),

a histogram of window sizes over an MD simulation at 298 K, which demonstrated that the windows were large enough for Xe diffusion for 7.3% of the time and for Kr 58.7% of the time.<sup>73</sup>

In this chapter, I present a computational screening of previously reported molecular pores in search of feasible candidates for Xe/Kr separation application. In the light of the reported Xe/Kr selectivities of **CC3** in the solid state, other PMMs not considered for Xe/Kr separation application could be identified as promising. A few other studies have previously investigated the interaction of porous molecular materials with noble gases, in particular Xe. This includes a DFT study of noble gas binding in the belt-like cucurbit[6]uril,<sup>219</sup> and X-Ray diffraction studies of the  $\alpha$ -cyclodextrin inclusion complexes with Kr,<sup>220</sup> and of the extreme confinement of Xe in 111-cryptophane in the solid state.<sup>221</sup> These studies indicate a positive affinity of several molecular pores towards the Xe and Kr noble gases discussed here. Molecular pores can also be constructed into membranes,<sup>61,98</sup> demonstrating that they have the potential to compete with the best performing MOFs.

## 4.2 Methods

A database of molecular pores was compiled from the crystal structures deposited in the Cambridge Structural Database (CSD).<sup>26</sup> A single molecule analysis for various structural and physical properties, including host-guest binding energies and free energy barriers for diffusion, to better understand the host-guest interactions and structural correlations to the Xe/Kr selectivity, was performed. The structural properties, such as the void diameters and the windows diameters, to create PLEs based on the MD trajectories, were generated with the `pywindow` package introduced in Chapter 3. The schematics of the key stages in this screening is presented in Figure 4.1. The identified promising candidates were tested in the laboratory by collaborators.

### 4.2.1 The Cage Database (*CDB*)

The literature on porous molecules was reviewed to find X-ray diffraction crystal structures of intrinsically porous molecules that had been deposited in the CSD as of February 2016. This

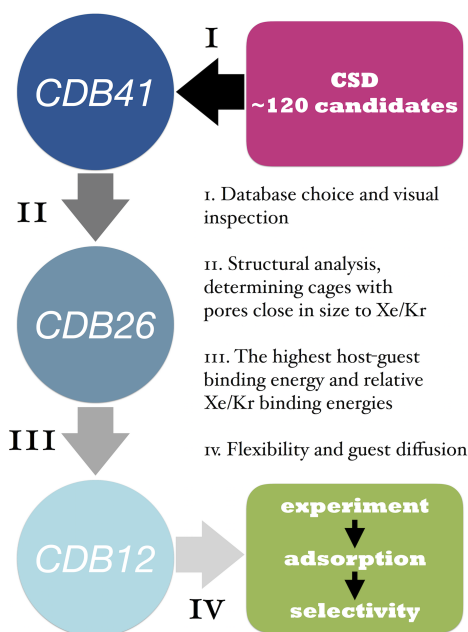


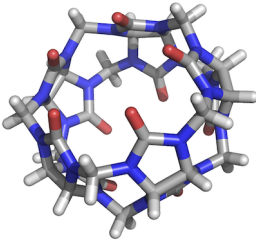
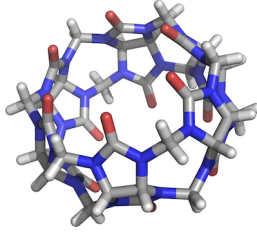
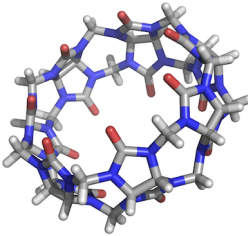
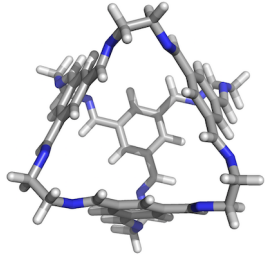
Figure 4.1: Schematic showing the key stages applied in the screening for molecular pores of promise for Xe/Kr separation. *CDBXX* stands for “Cage Database XX”, where ‘XX’ is the number of entries in the database at that stage.

focused on the reviews in references 55, 56 and 89, as these systems can have multiple names in the literature and thus searches in databases, such as Web of Science, are liable to missing structures. The crystallographic information files (CIFs) were then extracted for the molecules of interest to form the Cage Database (*CDB*). Whilst the primary focus is on molecules that are shape-persistent, molecules that may potentially lack shape-persistence where a solvate crystal structure was reported, but *in silico* solvent removal suggests an intrinsic void, were also included. Macrocyclic molecules such as cucurbiturils,<sup>219</sup> cyclodextrins<sup>43</sup> and cryptophanes,<sup>221</sup> which have the potential for guest binding or encapsulation were also added to the *CDB*. Any charged molecules were disregarded. The modularity of these materials allows one to extract a single host molecule from the crystal structure for further analysis. The initial search resulted in ~120 candidates, from which 41 were chosen based on the criteria just described to form *CDB41*.

The naming system for the molecular pores in this chapter is the following: the base name is typically two letters followed by a number denoting the chronological order of the molecule being reported in the literature. The first letter stands for the first letter of the surname of the group leader where the cage was reported, such as “C” - Cooper, “M” - Mastalerz, “D” -

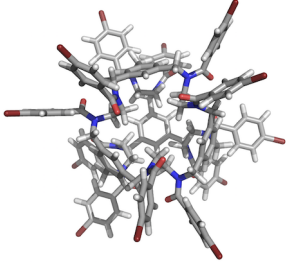
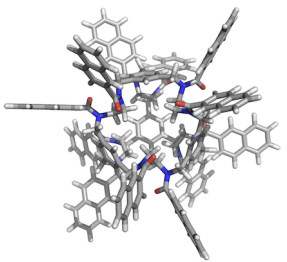
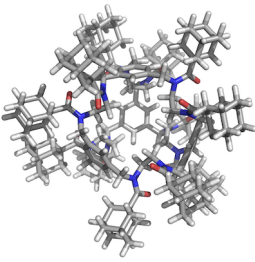
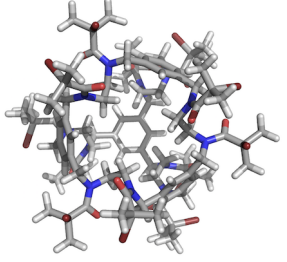
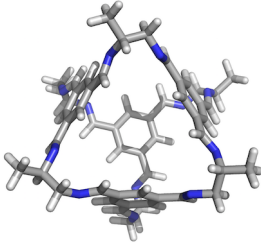
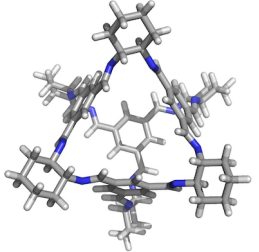
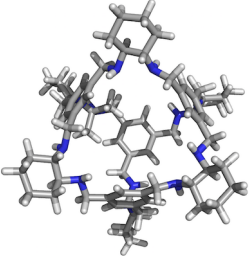
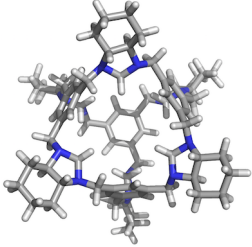
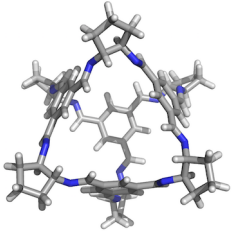
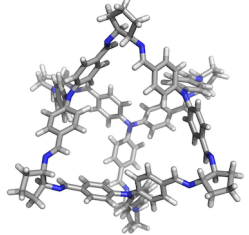
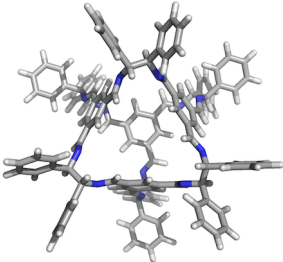
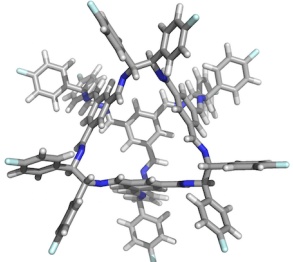
Doonan, “**G**” - Gawroński, “**N**” - Nishikubo and so on. The second letter “**C**” stands for “cage”. For non-cage compounds, such as macrocycles, I use a separate system, where “**CB**” stands for cucurbiturils with the following number describing their size, “**CD**” stands for cyclodextrins, followed by a number representing the Greek letter’s position in the Greek alphabet as used to describe them in the literature (*e.g.*  $\alpha$  is 1) and “**CP**” stands for cryptophanes. For reduced cages (for example where a related imine cage has been reduced to an amine cage), a single capital letter “**R**” is added in front of the code name. In the case of the **RCC1** family, where a reduced cage **CC1** was exo-functionalised, an “a, b, c, d” naming system was used, as in the work reporting these cages. A preceding Greek letter is used to refer to the known polymorphs of the crystalline form, followed by a dash and “*R*” or “(*R,S*)” to describe an enantiopure co-crystal or a racemate possible for the chiral **CC3**, respectively. A full table relating the presented naming system to any literature alternatives and the references of first reports of molecules for *CDB41* is provided in Table 4.1.

Table 4.1: A list of molecular pores in *CDB41*.

<i>CDB ID</i> <sup>a</sup> (alt. names) <sup>b</sup> reference no. <sup>c</sup> CSD refcode <sup>d</sup>	image <sup>e</sup>	<i>CDB ID</i> <sup>a</sup> (alt. names) <sup>b</sup> reference no. <sup>c</sup> CSD refcode <sup>d</sup>	image <sup>e</sup>
<b>CB5</b> (cucurbit[5]uril) 223, 224 FUYHEN		<b>CB6</b> (cucurbit[6]uril) 36, 45 BATVUP	
<b>CB7</b> (cucurbit[7]uril) 223, 224 FUYHIR		<b>CC1</b> - 20 PUDWUH	

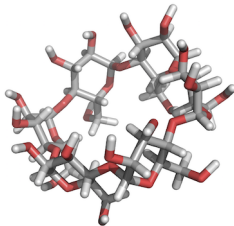
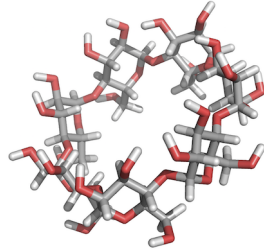
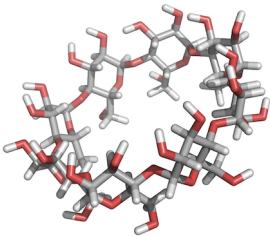
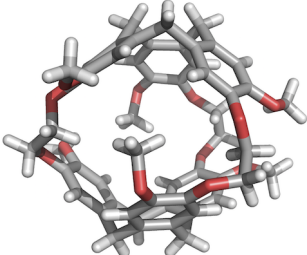
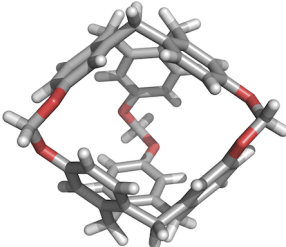
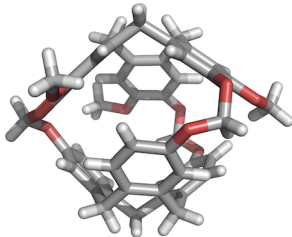
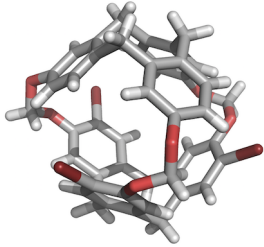
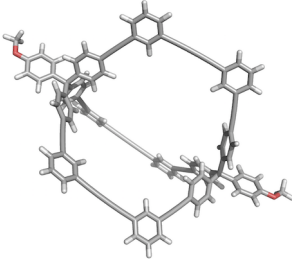
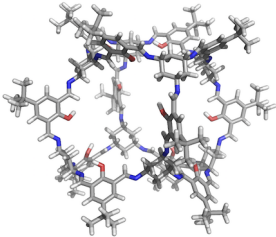
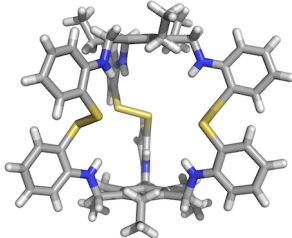
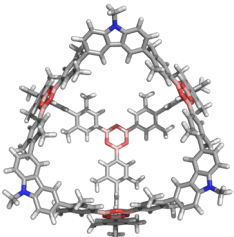
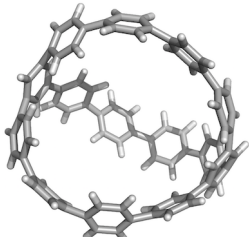
*Continued on next page*

Table 4.1 – Continued from previous page

<b>RCC1a</b> - 225 FIFTAR		<b>RCC1b</b> - 225 FIFTEV	
<b>RCC1c</b> - 225 FIFTIZ		<b>RCC1d</b> - 225 FIFTOF	
<b>CC2</b> - 20 PUDXAO		<b>CC3</b> - 53, 20 PUDXES	
<b>RCC3a</b> (AT-RCC3) 65 VOLZON		<b>RCC3b</b> (FT-RCC3) 65 VOMPAQ	
<b>CC4</b> - 226 OZECAY		<b>CC5</b> - 227 UTEVOF	
<b>CC9</b> - 39 GANDAC		<b>CC10</b> - 39 GANDUW	

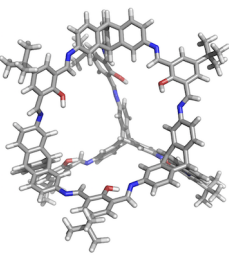
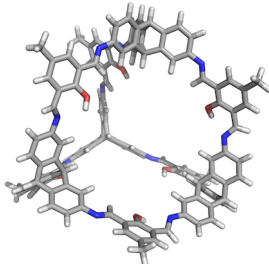
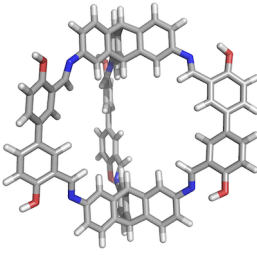
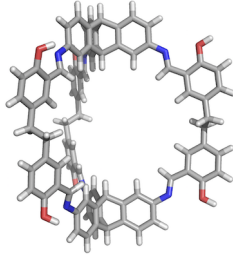
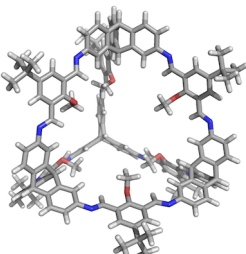
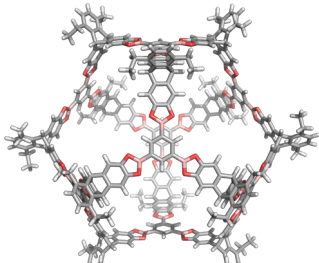
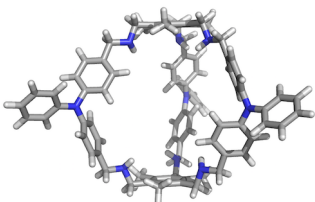
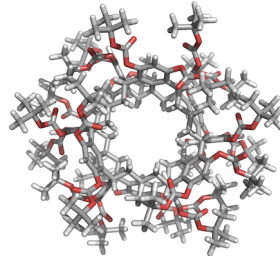
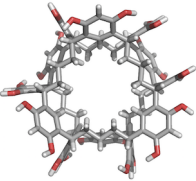
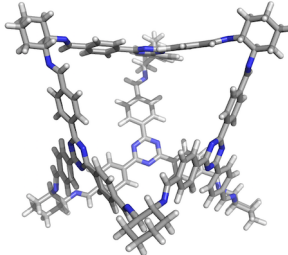
Continued on next page

Table 4.1 – Continued from previous page

<p><b>CD1</b> (<math>\alpha</math>-cyclodextrin) 38, 228 ACDMFM</p>		<p><b>CD2</b> (<math>\beta</math>-cyclodextrin) 38, 44 DUCMUL</p>	
<p><b>CD3</b> (<math>\gamma</math>-cyclodextrin) 38, 229 NUNRIX</p>		<p><b>CP1</b> (cryptophane-A) 230, 231 OJITOR</p>	
<p><b>CP3</b> (111, cryptophane-111) 230, 222 NOVNAP</p>		<p><b>CP4</b> ((MeO)<sub>3</sub>-111, cryptophane-111) 230, 232 FOQTOW</p>	
<p><b>CP5</b> (Br<sub>3</sub>-111, cryptophane-111) 230, 232 FOQTEM</p>		<p><b>DC1</b> (C1) 60 REQXES</p>	
<p><b>GC1</b> - 233 REYMAL</p>		<p><b>HC1</b> - 59 MAVVAI</p>	
<p><b>IC1</b> - 234 YUKHOD</p>		<p><b>IC2</b> (carbon nanocage) 61 NUXHIZ</p>	

Continued on next page

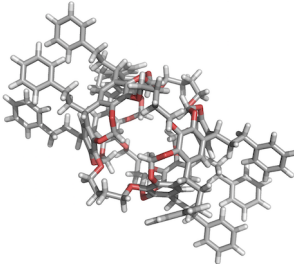
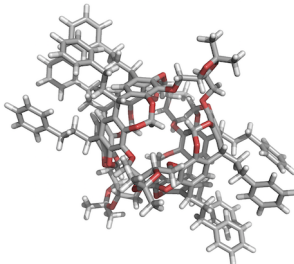
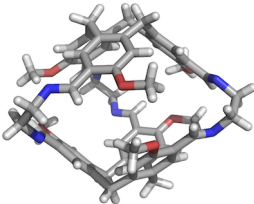
Table 4.1 – Continued from previous page

<p>MC1 - 54, 40 EKUKUR</p>		<p>MC2 - 235 TATVER</p>	
<p>MC3 - 236 SATJAA</p>		<p>MC4 - 236 SATJEE</p>	
<p>MC5 - 67 FEQXAC</p>		<p>MC6 - 143 ZIRCIO</p>	
<p>MC7 - 105 FOMLUQ</p>		<p>NC1 (noria-Boc) 48 MESTUA</p>	
<p>NC2 (noria) 48, 46 GUMCIB</p>		<p>WC1 cage 1 87 TOVWUY</p>	

*Continued on next page*



Table 4.1 – Continued from previous page

<p><b>WC2</b> (hemicarcerand) 237 LUXVAB</p>		<p><b>WC3</b> (hemicarceplex) 51 PAQFES</p>	
<p><b>WC4</b> (hexaimine cryptophane) 238 EPIRUR</p>			

<sup>a</sup> The identification of the molecular pore used in *CDB*.

<sup>b</sup> The alternative identification or name found in literature.

<sup>c</sup> The references are typically for the work where the molecule was first reported, however, when it is not the source of the X-ray diffraction crystal structure used in this work, an additional reference is given.

<sup>d</sup> The 6-letter CSD ‘refcode’.

<sup>e</sup> The molecules are not shown to scale.

## 4.2.2 Structural analysis

The Cartesian coordinates of a single molecular pore for each PMM in *CDB41* were extracted from the CIF and analysed with *pywindow*. The structural parameters that were then calculated include: the maximum diameter of a molecule ( $D_{max}$ ), the internal void diameter ( $D_{void}$ ) and the internal void volume ( $V_{void}$ ). These have been introduced in Section 3.2.4. Additionally, I define new structural features related to  $D_{void}$  and  $V_{void}$ , and the Xe and Kr  $r_{vdw}$  and resulting van der Waals volumes of Kr ( $26.31 \text{ \AA}^3$ ) and Xe ( $36.09 \text{ \AA}^3$ ) derived from their  $r_{vdw}$ :

1. The diameter fraction ( $DF$ ), defined as the diameter of the noble gas divided by the host molecule’s  $D_{void}$ . This was calculated for both Kr and Xe ( $DF_{Kr}$  and  $DF_{Xe}$  respectively)
2. The pore fraction ( $PF$ ), defined as the guest’s volume divided by the host’s  $V_{void}$  for both Kr and Xe ( $PF_{Kr}$  and  $PF_{Xe}$  respectively).

The definition of the  $PF$  proposed here is equivalent to the “packing coefficient” ( $PC$ ) from

Mecozzi and Rebek, which has previously been shown to be optimal for guest encapsulation when equal to  $55 \pm 9\%$ .<sup>120</sup>

### 4.2.3 Binding energy calculations

The binding energies between the host and a single Kr or Xe atom placed in the centre of a single host molecule’s cavity were calculated using the QUICKSTEP module<sup>181</sup> of the CP2K software.<sup>238</sup> The calculations were performed with the Gaussian plane-wave scheme<sup>239</sup> and the Goedecker-Teter-Hutter pseudopotentials.<sup>240</sup> Grimme’s D3 dispersion correction<sup>241</sup> and a TZVP (for the host) and DZVP (for Kr and Xe) MOLOPT basis sets<sup>242</sup> were used. A cut-off of 350 Ry was applied for all calculations, chosen as a trade off value between the cost and the accuracy of the calculations. The PBE functional<sup>179</sup> was used for all the DFT-D3 calculations. For each self-consistent-field cycle the electronic structure was optimised to the accuracy of  $10^{-7}$  Hartree. This set up is similar to that used by Chen *et al.* for isolated **CC3** Xe/Kr binding energies calculations.<sup>73</sup> Each cage was placed in the centre of a cubic cell, with the cell lengths equal to the  $D_{max}$  of the cage plus 8 Å on each side. Later, single point energy calculations with a cubic reference cell were performed, with cell vectors lengths equal to 40 Å, to ensure consistent grid density across the systems. The basis set superposition error (BSSE) correction using the counterpoise method was determined<sup>243</sup> and then the resulting binding energies ( $E_b$ ) calculated according to:

$$E_b = E_{host...guest} - E_{host} - E_{guest} - E_{BSSE} \quad (4.1)$$

where  $E_{host...guest}$  is the energy of the geometry optimised host/guest pair,  $E_{host}$  is the energy of the geometry optimised host and  $E_{guest}$  is the energy of the isolated guest atom.

For analysis of binding energies in the solid state, a single unit cell of **CC3** $\alpha$ -R (CSD refcode: PUDXES,  $a = b = c = 24.8$  Å) was used and the periodic DFT calculation with the same set up as described for isolated molecules was run with a 280 Ry energy cut off, chosen based on the parameters for similar calculations found in the literature.<sup>40</sup> A dispersion correction cut-off of 10 Å was applied. A single Xe or Kr atom was placed in the intrinsic void

of one randomly selected cage or in the extrinsic void in between two adjacent cages. For all DFT-D3 calculations of this type, an estimated error is of the order of a few  $\text{kJ mol}^{-1}$ .<sup>244</sup>

#### 4.2.4 Pore limiting envelopes

MD simulations were performed in order to calculate the PLE for each isolated molecular pore. The MD package `DL_POLY2.20`<sup>169</sup> was used, with input files generated using `DL_FIELD3.5`<sup>210</sup> and the all-atom OPLS force field,<sup>209</sup> as it has previously been shown to reproduce well both the structures and energetics of POCs.<sup>123</sup> `DL_FIELD3.5` will assign partial charges according to the charge increments for the all-atom OPLS force field. The performance of the OPLS force field on new systems in the *CDB* was checked by overlaying the structures of geometry optimised molecules with the ones extracted from crystal structures. No interactions were applied between the periodic images of the cell and the direct Coulomb summation was used for electrostatic interactions. The simulations were run in the NVT ensemble using the Hoover thermostat and a temperature of 300 K, equivalent to likely operating temperatures for a separation. A 0.7 fs timestep was used, with equilibration of 50 ps followed by a production run of 100 ns. The structures were sampled every 1.4 ps during the production run. To generate the PLEs, first the trajectories were analysed with the `pywindow` package to measure  $D_{window}$  at each trajectory step. Then, a curve was fitted to a normalised distribution of window diameters for the whole trajectory, using the KDE function in the `SciPy` python package,<sup>207</sup> to generate the PLE in the same fashion as in the example in Section 3.3.3 and the methodology described in Section 3.3.1.

To analyse the influence that crystal packing effects might have on the PLEs in the solid-state, a set of solid state simulations were performed for **CC3**. A supercell was constructed for each of the two known polymorphs,  $\alpha$  and  $\beta$  of the homochiral **CC3-*R***, and for the **CC3 $\alpha$ -(*R,S*)** racemate (see Table 4.2). The periodic calculation in the NPT ensemble was run at 300 K using the Hoover thermostat and 0.0001 Pa pressure using the Nosé barostat. A 0.7 fs timestep was used with an equilibration period of 0.2 ns, followed by an NVT ensemble run with an equilibration period of 0.1 ns and 1 ns production run at the same temperature. The Ewald summation algorithm was used for the electrostatic interactions and the trajectory was sampled

Table 4.2: Details of the **CC3** polymorph analysis.

Polymorph	Supercell ( $a \times b \times c$ )	Supercell dimensions ( $\text{\AA}$ )	Cell symmetry	Reference	CSD refcode
<b>CC3</b> $\alpha$ - <i>R</i>	2 x 2 x 2	49.6 x 49.6 x 49.6	cubic	20	PUDXES
<b>CC3</b> $\alpha$ -( <i>R,S</i> )	2 x 2 x 2	49.2 x 49.2 x 49.2	cubic	<i>a</i>	-
<b>CC3</b> $\beta$ - <i>R</i>	2 x 2 x 4	50.3 x 50.3 x 44.1	trigonal	200	PUDXES02

<sup>a</sup> Private communication with Dr. Samantha Y. Chong and Dr. Marc Schmidtman from the University of Liverpool.

every 1.4 ps to generate the PLE with `pywindow`. The structures of the **CC3** polymorphs are shown in Figure A.3 in Appendix.

### 4.2.5 Free energy barriers to diffusion

To determine the free energy barriers to diffusion of the noble gases out of the molecular pores, metadynamics simulations were performed using PLUMED2.2<sup>245</sup> combined with the DL\_POLY2.20 package, with MD simulations at 300 K with the same setup as the previous section. The collective variable (CV) for the metadynamics was defined as the distance between the centre of mass of a cage and the guest noble gas atom. An upper wall was set up for the CV at a distance of 30  $\text{\AA}$  from the centre of the host’s cavity (35  $\text{\AA}$  for **NC1** system), with a wall force energy of 150.0 kcal mol<sup>-1</sup> and the exponent determining the power of the wall, the rescaling factor and the offset kept at default values of 2, 1.0 and 0  $\text{\AA}$ , respectively. The grid used to store the bias has a spacing of 0.2  $\text{\AA}$  and the lower and upper bounds for the grid were set to -5  $\text{\AA}$  and 50.0  $\text{\AA}$ . The Gaussian width was set to 2  $\text{\AA}$  and the Gaussian height to 1.2 kcal mol<sup>-1</sup>, with the pace of depositing the Gaussians set to 500 and the biasfactor equal to 5. The output containing the distance of the CV and the energy value for the external potential was collected at 100 step intervals. The production runs were between 300 and 1100 ns, depending on the length of time required to reach satisfactory convergence for each system; this was determined as the point when the maximum fluctuations in the barrier height were less than 2 kJ mol<sup>-1</sup>. For the **NC1** system with Kr, the upper wall was set at a 35  $\text{\AA}$  distance with the force energy of the wall set to 200 kcal mol<sup>-1</sup> and the upper bound for the grid set at

100 Å, whereas for **NC1** with Xe the upper wall was set to 35 Å and the force energy of the wall set to 250 kcal mol<sup>-1</sup> with the upper bound for the grid set at 90 Å. This was due to the fact that **NC1** is much bulkier ( $D_{max} = 29.6$  Å) than the other analysed molecules.

### 4.2.6 Xe/Kr selectivity calculations

The Xe/Kr selectivities were calculated using single component adsorption isotherms. The selectivities ( $S$ ) at infinite dilution of Xe and Kr were calculated using Henry's constants ( $K_H$ ), extracted from the isotherms using pyIAST software,<sup>188</sup> according to the equation:

$$S = \frac{K_{H,Xe}}{K_{H,Kr}} \quad (4.2)$$

where  $K_{H,Xe}$  and  $K_{H,Kr}$  are Henry's constants for the Henry's Law region for Xe and Kr respectively. The selectivities at 1 bar pressure and 20/80 mol % Xe/Kr mixture were also calculated using the IAST as implemented in pyIAST. In all cases, the Langmuir analytical model was used to fit the experimental single component adsorption isotherms for Kr, whereas linear interpolation was used to fit the Xe isotherms. The second approach does not require the analytical model to be fitted. The selectivity is calculated as:

$$S = \frac{n_{T,Xe}}{n_{T,Kr}} * \frac{p_{Kr}}{p_{Xe}} \quad (4.3)$$

where  $n_{T,Xe}$  and  $n_{T,Kr}$  are the total moles of gas adsorbed derived from the pure component adsorption isotherms using IAST and  $p_{Xe}$  and  $p_{Kr}$  are the partial pressures of the Xe/Kr mixture components.

## 4.3 Results and discussion

### 4.3.1 Molecular *vs.* solid state behaviour

Intrinsically porous molecular systems present a unique opportunity for computational screening as a result of their inherently modular nature. This means that assessment of their properties

can potentially be carried out by analysing the individual molecule's properties alone, without the requirement of a representative solid state structural model, or the necessity for more expensive bulk simulations, which will invariably involve simulations with larger numbers of atoms. This can provide the opportunity to explore the intrinsic potential performance of the individual molecule, particularly for applications based on host-guest supramolecular chemistry, regardless of the final solid state structure, which could include multiple crystalline polymorphs or an amorphous phase. However, of course by assessing the single molecules alone, there is a potential loss of the system characteristics that emerge as a result of packing effects in the solid state. The relative orientations of the molecules in the solid state will certainly influence the longer range pore topology, for example **CC3** can pack in a window-to-window orientation that provides a diamondoid pore topology, but in the amorphous state and other crystalline polymorphs, this pore topology is lost. Furthermore, by assessing the dynamic motion of the hosts as single molecules or in calculating binding energies of single host-guest complexes, there is an influence of the surrounding host molecules in the solid state that is not considered.

There is precedent in the literature for guest and solvent controlled crystallisation that gives control over the shape of porous network,<sup>70,199</sup> crystal retro-engineering,<sup>87</sup> and pre- and post-solution processing<sup>66,81</sup> of POCs. Thus there is potential to control the solid state form of POCs and thus tune properties that might depend critically on the packing and not the single molecule's potential alone. Focusing on the fundamental properties arising from single molecules can also be a better approach in studying recently reported porous liquids based upon POCs.<sup>79,80</sup> This molecular approach was successfully used in the past in work carried out by myself, but not included in this thesis, to rationalise the diffusion behaviour of sulfur hexafluoride ( $\text{SF}_6$ ) in **CC3**, which had been observed experimentally but poorly understood at the molecular level.<sup>88</sup> Here, a further examination of the extent to which ignoring the solid state packing of the host porous molecule results in significant errors in the assessment of properties relevant to Xe/Kr separation for the **CC3** system is performed.

### 4.3.2 Structural analysis

In the first stage of the assessment of the porous molecular candidates for Xe/Kr separation, a structural analysis of the individual molecules was conducted. *CDB41* consists of molecular pores with a range of void diameters spanning from 3.3 to 21.9 Å. In Table 4.3 the parameters for *CDB41* are presented. No obvious correlation between the size of the molecules and the internal cavity diameters is observed, which is to be expected as they feature a variety of different peripheral functionalisations (*e.g.* compare **NC1** and **NC2** in Table 4.1). I have then excluded molecular pores in which the internal cavity is either too small to accommodate a single Xe atom or large enough to accommodate more than one Kr atom ( $DF_{Kr} > 0.50$  and  $DF_{Xe} < 1.02$ ). This resulted in *CDB26*, a refined version of the *CDB41* which was then analysed for the binding energies of Kr and Xe molecular pore complexes at the DFT level of theory.

### 4.3.3 Guest binding energies

The binding energies between the host molecules and a single noble gas guest allow to measure the strength of binding in the host-guest complexes. Firstly, I compared the binding energies for the noble gases in a single molecule of **CC3** with the equivalent ‘cage-cavity’ site in the solid state **CC3 $\alpha$ -R** structure (see Table 4.4). There is a small difference, with the binding energy of -25.1 compared to -20.3 kJ mol<sup>-1</sup> for Kr and -27.3 compared to -28.1 kJ mol<sup>-1</sup> for Xe between the single molecule and solid state structures, respectively. For the alternative binding site in the solid state, the extrinsic ‘window-cavity’ site between two adjacent cages in a ‘window-to-window’ orientation, which was found to be occupied from previous crystallographic analysis,<sup>73</sup> the binding energies are more favourable than the ‘cage-cavity’ site by 3 kJ mol<sup>-1</sup> (Kr) and 6 kJ mol<sup>-1</sup> (Xe), most likely as a result of the smaller size of the ‘window-cavity’. This demonstrates a dependence upon the binding site, as expected by the difference in the environment in the cavity and window sites, but also an influence of the surrounding solid state structure on the binding energy of the single molecule site. However, the difference in the relative Xe/Kr binding energy in the cage-cavity site of 5.5 kJ mol<sup>-1</sup> is also arguably within the error of the DFT calculations, thus it is important to not place too great a significance on

Table 4.3: Structural analysis for *CDB41*.

<i>CDB41</i>	$D_{max}$ (Å)	$D_{void}$ (Å)	$V_{void}$ (Å <sup>3</sup> )	$DF_{Kr}$	$DF_{Xe}$	$PF_{Kr}$	$PF_{Xe}$
<b>RCC3a</b>	22.6	3.33	19.3	1.11	1.23	1.36	1.87
<b>CP4</b>	14.8	3.40	20.5	1.09	1.21	1.28	1.76
<b>CP5</b>	14.2	3.69	26.4	1.00	1.11	1.00	1.37
<b>CD1</b>	16.0	3.84	29.5	0.96	1.07	0.89	1.22
<b>CB5</b>	12.8	3.95	32.2	0.93	1.04	0.82	1.12
<b>RCC1c</b>	25.8	4.06	35.0	0.91	1.01	0.75	1.03
<b>RCC1a</b>	28.3	4.09	35.9	0.90	1.00	0.73	1.01
<b>WC2</b>	29.4	4.11	36.4	0.90	1.00	0.72	0.99
<b>RCC1b</b>	29.8	4.23	39.7	0.87	0.97	0.66	0.91
<b>CP3</b>	13.3	4.38	44.1	0.84	0.94	0.60	0.82
<b>HC1</b>	19.1	4.52	48.4	0.82	0.91	0.54	0.74
<b>RCC1d</b>	22.2	4.53	48.6	0.82	0.91	0.54	0.74
<b>WC3</b>	28.8	4.63	52.0	0.80	0.89	0.51	0.69
<b>CP1<sup>a</sup></b>	15.8	4.64	52.3	0.80	0.88	0.50	0.69
<b>NC2<sup>a</sup></b>	20.5	4.66	53.1	0.79	0.88	0.50	0.68
<b>NC1<sup>a</sup></b>	29.6	4.79	57.4	0.77	0.86	0.46	0.63
<b>CB6<sup>a</sup></b>	14.8	4.84	59.3	0.76	0.85	0.44	0.61
<b>MC3</b>	21.5	5.23	74.7	0.71	0.78	0.35	0.48
<b>WC4<sup>a</sup></b>	16.0	5.30	78.0	0.70	0.77	0.34	0.46
<b>CC2<sup>a</sup></b>	20.1	5.40	82.6	0.68	0.76	0.32	0.44
<b>CC10</b>	26.7	5.42	83.6	0.68	0.76	0.31	0.43
<b>CC9<sup>a</sup></b>	25.4	5.43	83.7	0.68	0.76	0.31	0.43
<b>CC3<sup>a</sup></b>	22.6	5.47	85.9	0.67	0.75	0.31	0.42
<b>CC1<sup>a</sup></b>	17.6	5.52	88.2	0.67	0.74	0.30	0.41
<b>RCC3b<sup>a</sup></b>	22.5	5.71	97.5	0.65	0.72	0.27	0.37
<b>CD2<sup>a</sup></b>	18.0	5.77	100	0.64	0.71	0.26	0.36
<b>CC4<sup>a</sup></b>	21.6	6.09	118	0.61	0.67	0.22	0.30
<b>CB7</b>	16.0	6.26	128	0.59	0.66	0.21	0.28
<b>MC4</b>	22.2	7.00	180	0.53	0.59	0.15	0.20
<b>CD3</b>	19.1	7.32	205	0.50	0.56	0.13	0.18
<b>MC7</b>	25.3	7.43	215	0.50	0.55	0.12	0.17
<b>IC2</b>	18.9	7.66	235	0.48	0.54	0.11	0.15
<b>MC5</b>	31.1	7.95	263	0.46	0.52	0.10	0.14
<b>CC5</b>	28.3	8.43	313	0.44	0.49	0.084	0.12
<b>DC1</b>	34.4	9.63	468	0.38	0.43	0.056	0.077
<b>MC1</b>	31.7	10.3	568	0.36	0.40	0.046	0.064
<b>MC2</b>	29.5	10.4	582	0.36	0.40	0.045	0.062
<b>WC1</b>	32.9	10.4	585	0.36	0.40	0.045	0.062
<b>GC1</b>	34.3	12.7	1074	0.29	0.32	0.025	0.034
<b>IC1</b>	34.4	14.5	1584	0.26	0.28	0.017	0.023
<b>MC6</b>	42.8	21.9	5513	0.17	0.19	0.005	0.007

<sup>a</sup> *CDB12* entries that were chosen for detailed analysis



Table 4.4: DFT binding energies and the relative binding energies in  $\text{kJ mol}^{-1}$  for single noble gas molecules in **CC3**.

Site	$E_{b,Kr}$	$E_{b,Xe}$	$E_{b,rel}$
<b>CC3</b> - <i>R</i> single molecule cage-cavity	-20.3	-28.1	7.7
<b>CC3</b> $\alpha$ - <i>R</i> cage-cavity	-25.1	-27.3	2.2
<b>CC3</b> $\alpha$ - <i>R</i> window-cavity	-28.0	-33.1	5.1

differences of a few  $\text{kJ mol}^{-1}$ .

The results presented in Figure 4.2 for *CDB26* present an insight into the potential separation capability for Xe/Kr at equilibrium (numerical values in Table A.1 in Appendix). No simple correlation between the relative binding energies and the pore sizes is observed (see Figure A.4 in Appendix). Weaker binding energies are generally observed for hosts with significantly larger pores than the guest size, as the guest cannot maximise favourable intermolecular interactions with all sides of the cavity, as it typically sits on one side of the cavity only. If the guest is larger than the cavity size, this forces strain into the host to accommodate the guest, which weakens the binding energy. However, this correlation is undermined by the influence of the local chemical structure and potential for optimisation of intermolecular interactions between the host and the guest should the cavity size be neither excessively small or large for Xe and Kr. An example of this is **MC3**; it has a pore size close to that of **WC4**, however the binding energies for Xe ( $-12.6 \text{ kJ mol}^{-1}$ ) and Kr ( $-15.2 \text{ kJ mol}^{-1}$ ) are much weaker than that for **WC4** ( $-27.7$  and  $-37.7 \text{ kJ mol}^{-1}$ ). If those host molecules were to be examined in more detail, the cryptophane **WC4** provides ‘better’ encapsulation than **MC3** (see Table 4.1), which has a more open framework. This means that **WC4** has a much greater opportunity for favourable intermolecular interactions with the guest noble gases. In the case of **WC3**, binding energies would indicate a reversed affinity for Xe/Kr binding, with the slightly higher binding energy for Kr and the relative difference of approximately  $2 \text{ kJ mol}^{-1}$ . A repeated calculation for **WC3** resulted in a similar value. However, as the structurally similar **WC2** system does not result in a similar effect, and because the estimated error for this type of DFT-D3 calculation of the order of few  $\text{kJ mol}^{-1}$ , this result was treated as an anomaly.

The highest observed relative binding energy is  $13.6 \text{ kJ mol}^{-1}$  for **NC1**, almost twice as

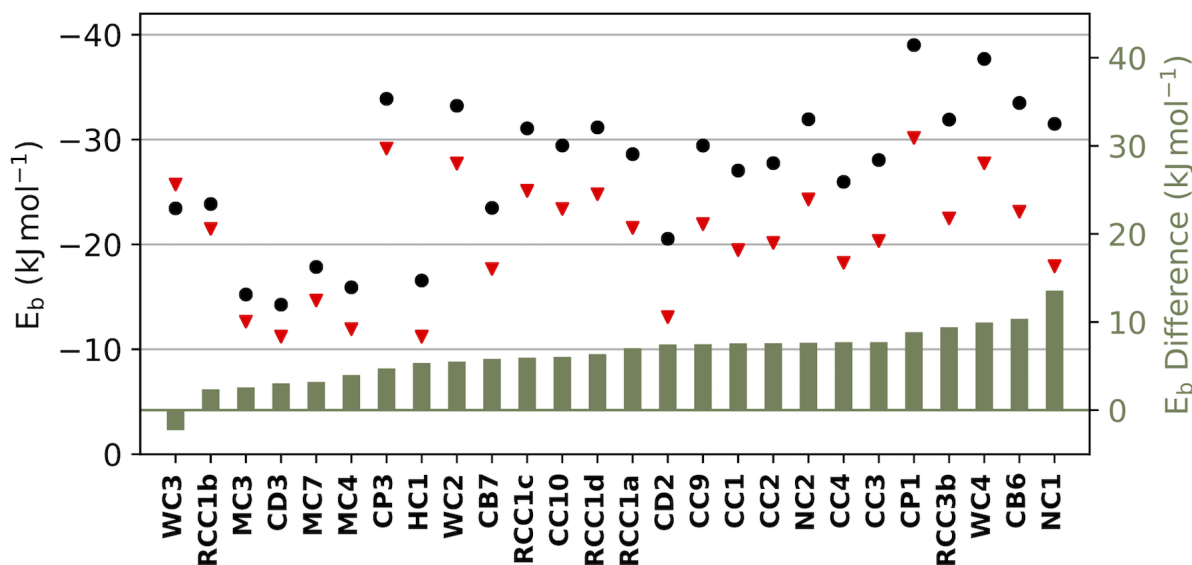


Figure 4.2: The DFT host-guest binding energies calculated for the noble gases in the molecular pores of *CDB26* (Xe - black dots and Kr - red triangles, values with reference to the left hand axis). The green bar plot shows the relative host-guest binding energy between the complexes with Xe and Kr, in increasing order (right hand axis).

high as for **NC2** ( $7.6 \text{ kJ mol}^{-1}$ ). Interestingly, both **NC1** and **NC2** have the same structural core, with  $D_{void}$  values that are only slightly different;  $4.79 \text{ \AA}$  for **NC1** and  $4.66 \text{ \AA}$  for **NC2**. However, **NC1** has a significant external functionalisation from the BOC protecting groups, which influences the structure of the core cage and thus the pore environment compared to the DFT geometry optimised molecular structures (see Figure A.5 in Appendix). There could be an additional effect of neighbouring molecules in the bulk that could further affect the pore's environment, although this was not observed in other porous molecular systems. The  $D_{void}$  values are significantly smaller than the void of  $5.47 \text{ \AA}$  for **CC3**, yet the binding energy difference for **CC3** ( $7.7 \text{ kJ mol}^{-1}$ ) is very similar to that of **NC2**.

Based on the calculated binding energy difference, I have further refined the set of the studied molecular pores to form *CDB12* where the  $E_b$  difference  $> 7 \text{ kJ mol}^{-1}$ . The final set of 12 molecular pores sufficiently represent the diversity of the initial *CDB41* with examples of a cucurbituril, a cyclodextrin, cryptophanes, two noria molecules and a set of isostructural POCs (see Figure 4.3 and Table 4.4). All molecular pores that constitute *CDB12* have significant binding affinity to both Kr and Xe, but significantly greater for the latter which should positively

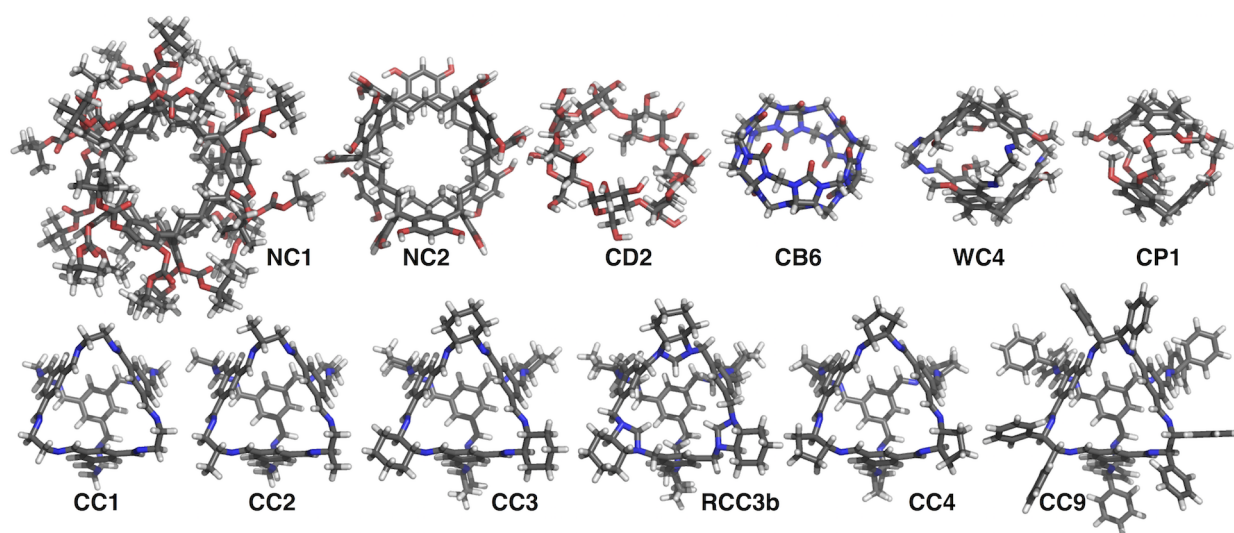


Figure 4.3: The porous organic molecules that form *CDB12*. First row from left: noria-Boc,<sup>47</sup> noria,<sup>45</sup>  $\beta$ -cyclodextrin,<sup>43</sup> cucurbit[6]uril,<sup>44</sup> hexaimine cryptophane,<sup>237</sup> cryptophane-A.<sup>230</sup> Second row from left: The family of Cooper [4+6] cages with tri-topic 1,3,5-triformylbenzene building unit and ditopic amine linkers: 1,2-diethylamine,<sup>19</sup> 1,2-dipropylamine,<sup>19</sup> 1,2-diaminocyclohexane,<sup>19</sup> 1,2-diaminocyclohexane (reduced and tied),<sup>64</sup> 1,2-diaminocyclopentane,<sup>225</sup> 1,2-diphenylethylenediamine.<sup>38</sup> Next to the structures their *CDB* names are given.

affect the overall Xe/Kr selectivity, the final measure of these materials performance. The identification of new promising materials with high Xe/Kr selectivity is a goal of this work.

Table 4.5: The twelve best candidates identified here for Xe/Kr separation (*CDB12*).

<i>CDB12</i>	<i>Alternative literature name</i>	<i>CSD refcode</i>
<b>CB6</b>	cucurbit[6]uril <sup>35</sup>	FUYHEN
<b>CC1</b>	Covalent Cage 1 <sup>19</sup>	PUDWUH
<b>CC2</b>	Covalent Cage 2 <sup>19</sup>	PUDXAO
<b>CC3</b>	Covalent Cage 3 <sup>52</sup>	PUDXES
<b>RCC3b</b>	<b>FT-RCC3</b> <sup>64</sup>	VOMPAQ
<b>CC4</b>	Covalent Cage 4 <sup>225</sup>	OZECAY
<b>CC9</b>	Covalent Cage 9 <sup>38</sup>	GANDAC
<b>CD2</b>	$\beta$ -cyclodextrin <sup>37</sup>	DUCMUL
<b>CP1</b>	cryptophane-A <sup>229</sup>	OJITOR
<b>NC1</b>	noria-BOC <sup>47</sup>	MESTUA
<b>NC2</b>	noria <sup>47</sup>	GUMCIB
<b>WC4</b>	hexaimine cryptophane <sup>237</sup>	EPIRUR

### 4.3.4 Pore limiting envelopes

The PLE allows one to consider the potential for dynamic porosity due to pore breathing behaviour. The `pywindow` code used for the MD trajectory analysis allowed for automated and rapid window diameter measurements. Then, the generated PLE can be compared to the size of the guests that would be diffusing through the pores and the percentage of simulation time during which the window diameter is large enough for a guest to diffuse can be calculated.

To justify the single molecule analysis approach for the determination of PLEs, I first compared the PLEs calculated for a single **CC3-R** molecule to that of:

- enantiopure **CC3 $\alpha$ -R**, with window-to-window packing into a diamondoid pore topology in a cubic cell.
- racemate **CC3 $\alpha$ -(R,S)** that crystallises isostructurally to **CC3 $\alpha$ -R**, with the same window-to-window packing.
- **CC3 $\beta$ -R**, a different polymorph of **CC3-R** with a trigonal cell, less efficient packing and some cage windows having a window-to-face orientation.

with the molecules packing and the resulting pore networks, for the described **CC3** polymorph, shown in Figures 4.4 and A.3 in Appendix. The generated PLEs, pore topologies and cage packing for these systems is presented in Figure 4.4. The observation can be made that the single **CC3** molecule has a PLE very close to that of the symmetrically packed enantiopure **CC3 $\alpha$ -R** system, which has the ‘ideal’ window-to-window packing. The **CC3 $\alpha$ -(R,S)** has slightly smaller cell dimensions (see Table 4.2), which results from the previously reported stronger interaction between heterochiral **CC3** pairs compared to homochiral pairs.<sup>96</sup> This difference, whereby the cages are packed together more tightly, clearly results in the windows being more open than in the enantiopure system. **CC3 $\beta$ -R** packs differently, with lower symmetry and this results in a slightly broader distribution of window diameters, including to smaller values than observed in the other systems. This comparison was only performed for the **CC3** single molecule, as it has been reported in several crystalline polymorphs.

Chen *et al.*, have previously reported times that **CC3** windows are open for Kr (58.7%) and Xe (7.3%), for **CC3 $\alpha$ -R** polymorph, using the circumcircle method.<sup>73</sup> The windows are

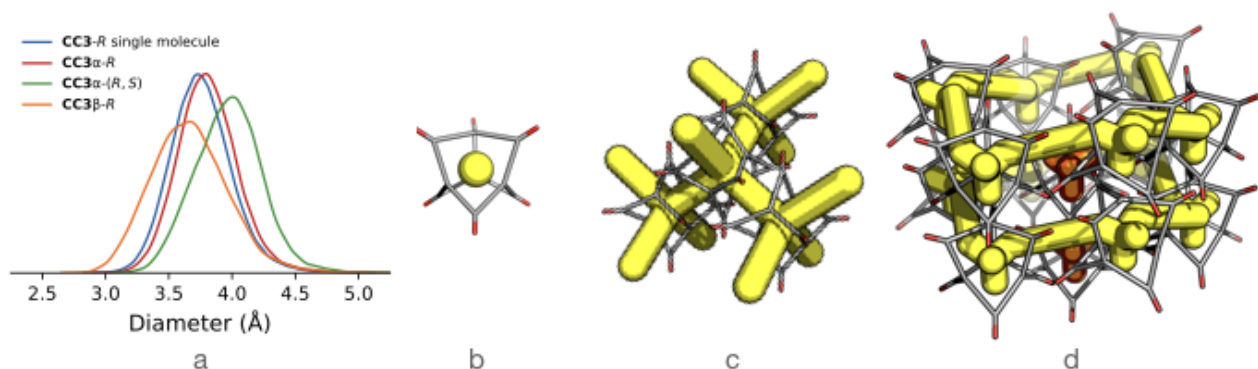


Figure 4.4: a) The pore limiting envelopes (PLEs) calculated for the **CC3-R** single molecule and three solid state **CC3** polymorphs: b) the simplified wireframe model of a **CC3** molecule (analogous to Figure 1.5) c) **CC3 $\alpha$ -R** and **CC3 $\alpha$ -R,S** crystal packing with the diamondoid shaped pore network shown as yellow tubes and **CC3** replaced with the simplified wireframe model d) **CC3 $\beta$ -R** crystal packing with the honey-comb shaped interconnected porous network considered as ‘accessible’ showed as yellow tubes and the intrinsic excluded intrinsic voids in orange that are considered as ‘non-accessible’.

considered opened when the circular window diameter is bigger than the van der Waals diameter of the noble gas of interest. These differ from the values presented here 68.6% and 10.7% for Xe and Kr, respectively. In Section 3.3.3, I have presented a comparison of a PLE generated with `pywindow` and the circumcircle method for an isolated **CC3** single molecule with a very close match of the two methods. The difference could be associated with a use of a different, in-house parametrised force field or the small differences in the MD simulation conditions. On the other hand, for a single molecule, the calculated times that **CC3** windows are open, Kr (59.8%) and Xe (7.9%) to diffuse compare well with those of Chen *et al* (see Table 4.6).

Table 4.6: Analysis of the percentage time a window can be considered open for Kr and Xe diffusion based on PLEs for the **CC3-R** single molecule and three solid state **CC3** polymorphs.

<i>System</i>	<i>Kr (%)</i>	<i>Xe (%)</i>	<i>Ratio Kr:Xe</i>
<b>CC3-R</b> single molecule	59.8	7.9	7.6
<b>CC3<math>\alpha</math>-R</b>	68.6	10.7	6.4
<b>CC3<math>\alpha</math>-(R,S)</b>	84.9	30.5	2.8
<b>CC3<math>\beta</math>-R</b>	44.8	7.8	5.7

The morphology of the PLEs for **CC3-R** single molecule and **CC3 $\alpha$ -R** is very similar, validating the `pywindow` single molecule analysis approach. To summarise, the observed deviations will most likely be caused by solvent occupying the pores, external functionalisation of neigh-

bouring molecules or unfavourable orientation of neighbouring molecules, therefore each type of system has to be treated on case-by-case basis. Overall, despite the small differences across the systems as a result of their crystal packing, these results show that the isolated molecule can represent the PLEs of the solid state structures surprisingly well.

The PLEs for *CDB12* are shown in Figure 4.5 and Table 4.7. As expected from the smaller size of Kr, it would diffuse more readily than Xe, with windows being open for greater percentages of time. With the exception of **CP1** and **WC4**, all the host molecules have considerable times when the window is open for Xe diffusion, presuming a solid state structure with an interconnected pore network linked via the calculated pore windows. Some of the host molecules have windows of more than one size, for example **NC2** has both small windows (with a diameter smaller than that of molecular hydrogen and thus physically inaccessible) and larger windows. **CD2**'s PLE is an elongated curve as a result of its belt-like, highly flexible structure and windows of two different topologies that can adopt multiple sizes.

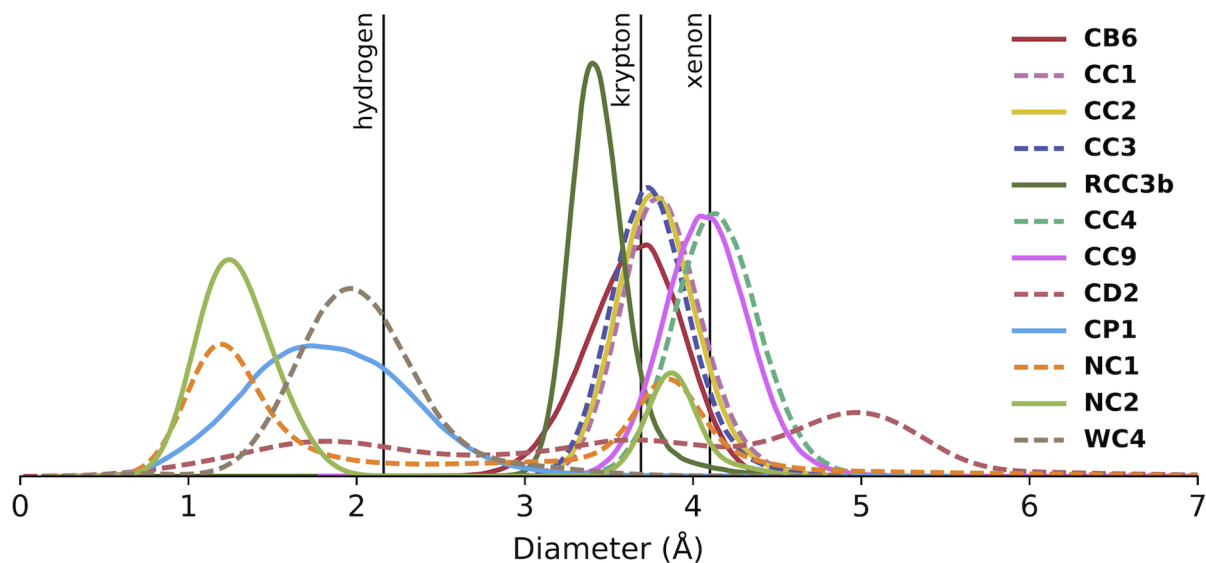


Figure 4.5: PLEs for *CDB12*. The vertical lines represent the minimum molecular diameters of hydrogen (2.18 Å), krypton (3.69 Å) and xenon (4.10 Å).

It is known from the experimental reports of Chen *et al.*,<sup>73</sup> that both Xe and Kr are able to diffuse through the **CC3-R** system and thus a window opening time of  $\sim 8\%$  for Xe is sufficient to allow diffusion, albeit slower than that for Kr (opening time of  $\sim 60\%$ ). This in combination with the stronger binding energy for Xe over Kr is the rationale for the observed separation

Table 4.7: Analysis of the percentage time a window can be considered open for Kr and Xe diffusion based on PLEs for *CDB12*. If a host molecule has more than one size window, only the larger window is considered.

<i>CDB12</i>	Kr (%)	Xe (%)	Ratio Kr:Xe
<b>CB6</b>	45.2	4.5	10.0
<b>CC1</b>	69.4	12.3	5.6
<b>CC2</b>	65.4	10.2	6.4
<b>CC3</b>	59.8	7.9	7.6
<b>RCC3b</b>	9.1	1.1	8.3
<b>CC4</b>	96.7	56.1	1.7
<b>CC9</b>	94.1	46.3	2.0
<b>CD2</b>	68.6	58.0	1.2
<b>CP1</b>	0.6	0.2	3.0
<b>NC1</b>	65.4	21.7	3.0
<b>NC2</b>	92.5	10.3	9.0
<b>WC4</b>	0.8	0.2	4.0

outcome of the breakthrough experiments of Chen *et al.* Therefore, the best porous molecular material for Xe/Kr separation would likely be expected to have a narrow PLE distribution that sits between the diameters of Kr and Xe, allowing for Kr to diffuse freely and for Xe to diffuse much more slowly; this should ensure longer retention times for Xe and thus optimal separation of the pair. Such a distribution is observed for **NC1** and **NC2**, so from the PLE analysis, these appear the most promising new candidates for Xe/Kr separation.

### 4.3.5 Free energy barriers to diffusion

The free energy barriers to diffusion of Xe and Kr through the host molecule windows of *CDB12*, as calculated with metadynamics simulations, are shown in Figure 4.6a. The calculated results for **CC3** compare well with those reported from an umbrella sampling study with Xe atom diffusing between a flexible pair of **CC3** molecules by Camp and Sholl;<sup>246</sup> their barrier for Xe diffusion was 16 kJ mol<sup>-1</sup>, compared to the value reported here of 13 kJ mol<sup>-1</sup> with a single isolated, but flexible, **CC3** molecule. In Figure 4.6b, the mean molecular window size, as measured during the 100 ns MD simulations at 300 K, is compared to the barrier heights to diffusion for *CDB12*. There is naturally a trade off that must be considered; whilst a large difference in the barriers to diffusion of Xe and Kr is desired for a good kinetic separation, if the

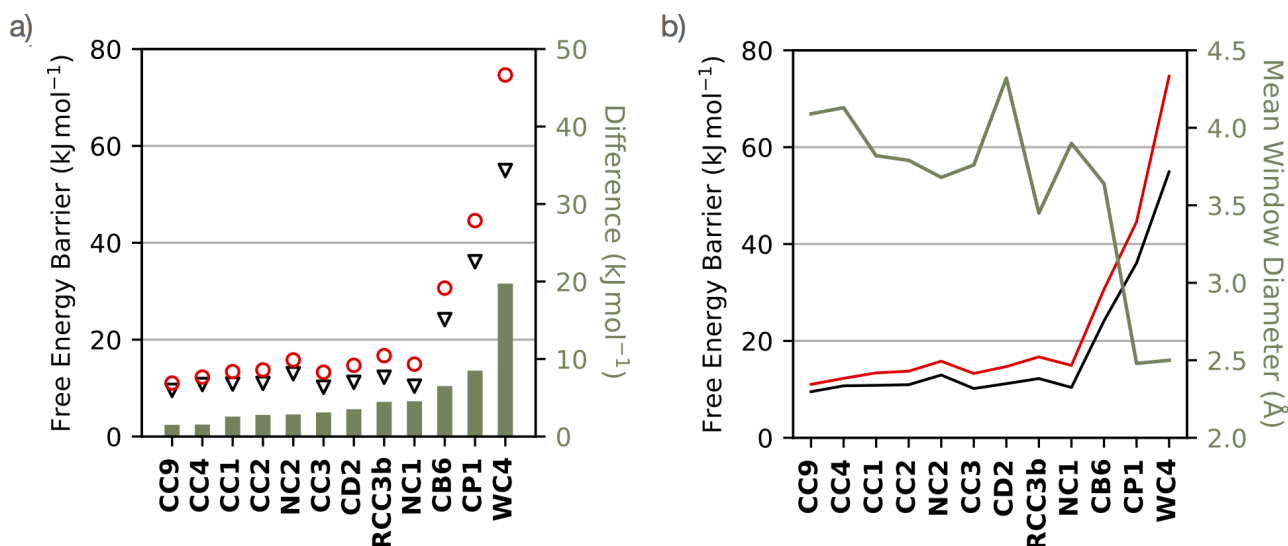


Figure 4.6: (a) Free energy barriers to diffusion of Xe and Kr out of the internal cavities of the *CDB12* molecules. Red opened circles and black opened triangles show barriers for Xe and Kr, respectively, using the left-hand Y-axis. The overlaid bar plot (green) shows the difference between the Xe and Kr barriers, using the right-hand Y-axis. b) The free energy barriers for Xe (red) and Kr (black) plotted against the mean window diameter from 100 ns MD (green).

magnitude of the barriers themselves is too great, as observed for **CB6**, **CP1** and **WC4**, this could limit the diffusion rates to too great an extent, hindering diffusion entirely for both gases, ruling out any separation applications. For the rest of the molecules, quite moderate barriers that are largely correlated to the window size are observed. The barriers span from 10 - 14 kJ mol<sup>-1</sup> for Kr and from 10 - 18 kJ mol<sup>-1</sup> for Xe. The observed difference between the barriers never exceeds 5 kJ mol<sup>-1</sup>. Whilst this might seem small, the difference reported for **CC3** of 3 kJ mol<sup>-1</sup> is sufficient to result in an efficient separation in a breakthrough experiment.

### 4.3.6 Computational prediction

The 12 most promising candidates for Xe/Kr separation (*CDB12*) are given in Table 4.5. From the computational screening, the most promising new candidates (beyond **CC3**) for Xe/Kr separation are identified as **NC1** and the structurally similar **NC2**. The former as it had the highest binding energy difference from the considered *CDB12* set. The latter, had a positioning of the PLE that would result in an optimal diffusion of Kr and high possibility for kinetic trapping of the larger Xe (window considered open for 92.5 and 10.3% of the total simulation time, respectively) maximising the breakthrough selectivity, however still allowing



for an effective diffusion of both species. Based upon this finding, the synthesis and experimental assessment of **NC1** and **NC2** was undertaken by our experimental collaborators. Whilst my research was being carried out, there was a report in the literature of noria (**NC2**), originally synthesised in 2006,<sup>47</sup> being tested for Xe/Kr separation, with it exhibiting selectivity of 9.4 (calculated with IAST theory for a 20/80 molar ratio mixture at 1 bar pressure and 298 K from single component isotherms).<sup>89</sup> In the following section my own findings for the selectivity of **NC1** and **NC2** are reported.

### 4.3.7 Experimental testing

The whole of the experimental work was performed by Dr. Shan Jiang, Mr. Rob Clowes and Dr. Michael Briggs. **NC2** and **NC1** were synthesised following the procedures of Kudo *et al.*<sup>47</sup> **NC2** was isolated as an amorphous pale yellow powder<sup>45</sup> that was shown to be poorly crystalline by powder X-ray diffraction (PXRD). Previous attempts by Tian *et al.* to isolate a highly crystalline desolvated phase of **NC2** were unsuccessful, with reports of amorphisation on desolvation or grinding. The attempts here to isolate a bulk crystalline phase of **NC1** from chloroform/hexane also proved unsuccessful, with only amorphous powders obtained. To assist in determining the importance of long range order on the separation performance of **NC2** and **NC1**, an amorphous sample of **CC3** was prepared to allow a direct comparison between the performance of its amorphous phase and the reported high Xe/Kr selectivity of crystalline **CC3**. For clarity, the bulk amorphous phases have a prefix of small letter 'a' before the *CDB* name, that is **aCC3-R**, **aNC1** and **aNC2**.

From the single adsorption isotherm data of pure Xe and Kr (Figure 4.7), the highest capacity was exhibited by crystalline **CC3 $\alpha$ -R**, with an uptake of 2.43 mmol g<sup>-1</sup> and 0.93 mmol g<sup>-1</sup> of Xe and Kr, respectively. The amorphous phase **aCC3-R** has significantly lower uptake of Xe and slightly smaller Kr uptake of 2.07 mmol g<sup>-1</sup> and 0.71 mmol g<sup>-1</sup> (at 0.92 bar; 0.75 mmol g<sup>-1</sup> when extrapolated to 1 bar using the Langmuir model) respectively. It is known that the loss of crystallinity in the **CC3** system produces additional extrinsic voids between the cage molecules and can therefore increase the BET surface area considerably.<sup>96</sup> However, clearly this does not translate to enhanced noble gas uptake, presumably due to the

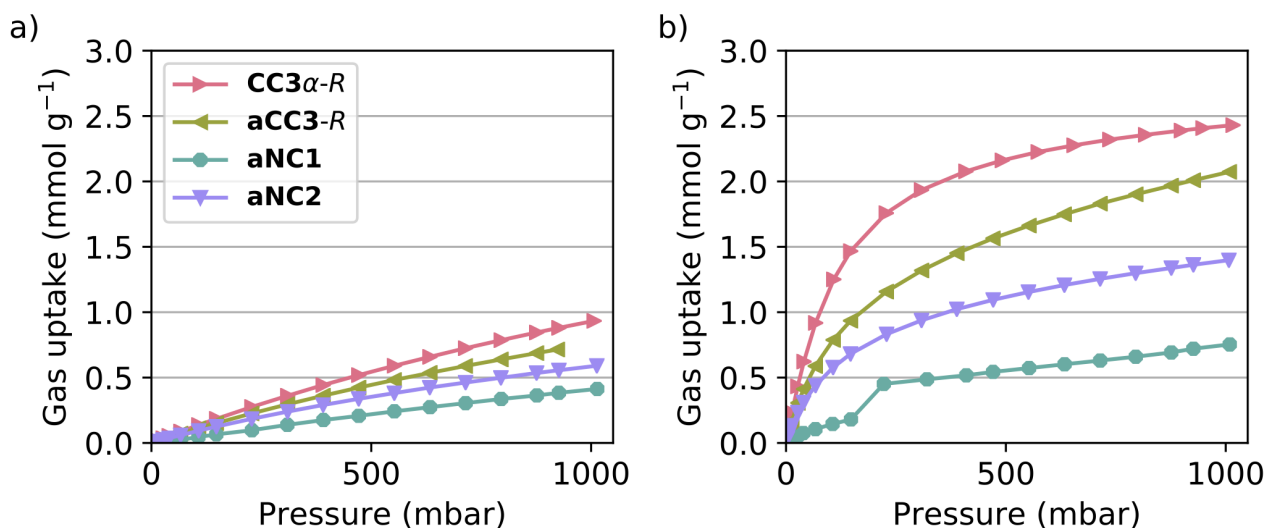


Figure 4.7: The experimental adsorption measurements for (a) Kr and (b) Xe in crystalline **CC3 $\alpha$ -R** and amorphous **aCC3-R**, **aNC1** and **aNC2**.

pore network being less accessible for these gases when the ordered diamondoid pore network is lost.

As only the amorphous phases for **NC1** and **NC2** were obtained, this is likely to reduce the noble gas capacities compared to what could potentially be achievable in a highly ordered crystalline phases with extended connectivity of pores. The uptake of **aNC2** is approximately half that of **CC3 $\alpha$ -R** (1.40 for Xe and 0.59 mmol g<sup>-1</sup> for Kr at 298 K and 1 bar), whereas for **aNC1** it is even lower, at 0.75 for Xe and 0.41 mmol g<sup>-1</sup> for Kr. The Xe isotherm for **aNC1** displays a marked step in the adsorption curve and a pronounced hysteresis on desorption (Figure A.13 in Appendix). This could be due to the BOC groups gating access to the intrinsic cavities. If we compare the gas adsorption per host molecule, **NC1** holds more gas molecules than **NC2**, with 1.7 molecules of Kr and 3.1 molecules of Xe adsorbed per **NC1** molecule in **aNC1** in comparison to 1.0 and 2.4 in **aNC2**, respectively. This difference can be attributed to the bulky BOC-protecting groups that cause an increase in mass, from 1704 g mol<sup>-1</sup> for **NC2** to 4104 g mol<sup>-1</sup> for **NC1** (see Figure 4.8). Patil *et al.* in their recent work reported that at 1 bar and 298 K, **aNC2** absorbs 1.55 mmol g<sup>-1</sup> of Xe and 0.64 mmol g<sup>-1</sup> of Kr, which is slightly more than the values reported here.<sup>89</sup> These minor differences are likely attributable to batch-to-batch variability and experimental error.

The calculated selectivities for the 20/80 molar ratio mixture of Xe and Kr are shown with

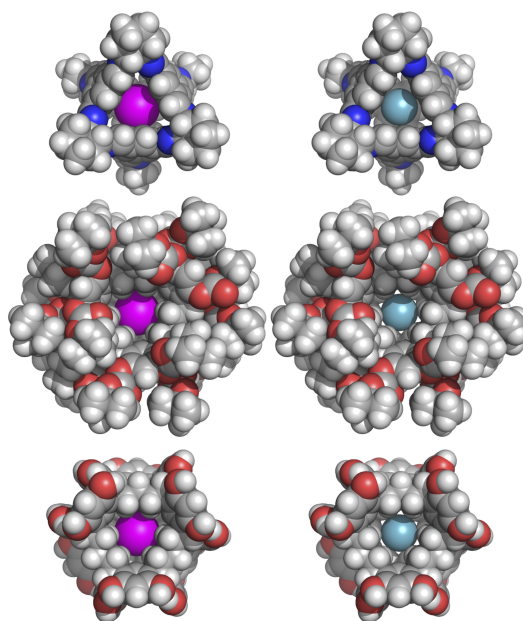


Figure 4.8: The space filling models of single Xe (left) and Kr (right) placed in the cavity of **CC3** (top), **NC1** (middle) **NC2** (bottom).

respect to the Xe uptake in Figure 4.9. Of the systems presented in this work, the highly crystalline **CC3 $\alpha$ -R** exhibits the highest selectivity for a porous molecular material of 14, a little less than previously reported SBMOF-1 ( $\sim 16$ ).<sup>247</sup> However, in terms of the Xe uptake, **CC3 $\alpha$ -R** remains unbeaten with a value of  $1.4 \text{ mmol g}^{-1}$ , 40% higher than that of SBMOF-1 ( $1.0 \text{ mmol g}^{-1}$ ). The materials with a good balance between Xe uptake and Xe/Kr selectivity would be favourable; therefore, the **CC3 $\alpha$ -R** system might look more promising for real life applications with only slightly worse selectivity, but significantly higher uptake. Both **aCC3-R** and **aNC2** showed similar selectivity of  $\sim 10$ , however, **aNC2** exhibits much lower uptake for Xe. **aNC1** exhibits the smallest uptake per unit of mass and a selectivity of only  $\sim 3.8$ . The infinite dilution selectivities calculated from Henry's coefficients extracted from the single component isotherm and these are shown in Table 4.8 and Figure 4.9. The presented results suggest that if it were possible to form crystalline solid state structures of **NC1** and **NC2**, with highly interconnected pore networks, then these should have an improved performance relative to the amorphous phases tested here. Of the two noria molecules, **NC1** is the more promising candidate, as its lower molecular weight and reduced peripheral bulk should allow for the better uptake capacity of the guest molecules.

Table 4.8: The calculated Xe/Kr selectivities ( $S$ ) for molecules reported here, as well as high performing materials from the literature. **aNC2<sup>a</sup>** is the noria sample reported by Patil *et al.*<sup>89</sup>

system	20/80 mol Xe/Kr (%)			Infinite dilution		
	$n_{T,Kr}$ $mmol\ g^{-1}$	$n_{T,Xe}$ $mmol\ g^{-1}$	$S$	$K_{H,Kr}$ $mmol(bar\ g)^{-1}$	$K_{H,Xe}$ $mmol(bar\ g)^{-1}$	$S$
<b>CC3<math>\alpha</math>-R</b> <sup>73</sup>	0.406	1.430	14.1	1.329	17.377	13.1
<b>aCC3-R</b>	0.339	0.928	10.3	1.104	12.812	11.6
<b>aNC1</b>	0.287	0.270	3.8	0.419	1.952	4.6
<b>aNC2</b>	0.270	0.656	9.7	0.904	10.142	11.2
<b>aNC2<sup>a</sup></b>	0.307	0.710	9.2	0.933	8.730	9.4
<b>SBMOF-1</b> <sup>247</sup>	0.258	1.016	15.8	2.374	38.424	16.2

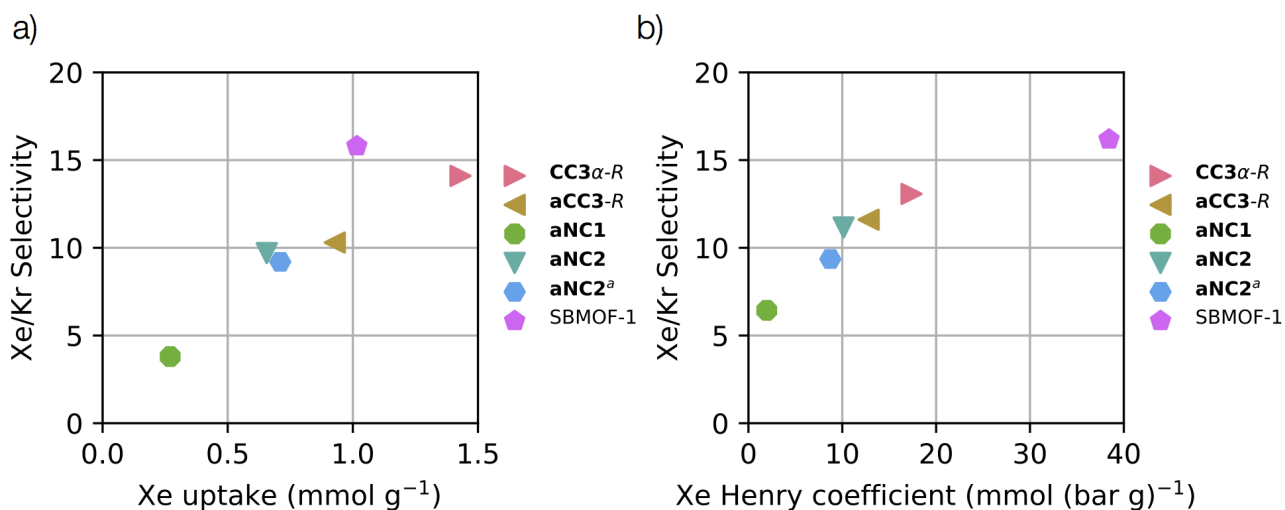


Figure 4.9: The calculated selectivities for a) a 20/80 Xe/Kr mixture at 1 bar and 298 K against the Xe uptake b) Henry's region (infinite dilution) for Xe and Kr at 298 K against the Xe Henry's coefficients at such conditions. The data is as reported in this work except for SBMOF-1,<sup>247</sup> **CC3 $\alpha$ -R**<sup>73</sup> and **aNC2<sup>a</sup>**, where the latter is the noria sample as reported by Patil *et al.* in their work<sup>89</sup>.

## 4.4 Conclusions

I presented a systematic computational screening study of previously reported porous molecular materials for Xe/Kr separation. Including the use of the recently developed `pywindow` software, a single molecule analysis approach was applied that included screening initially based on simple structure descriptors such as pore diameter and pore fraction, followed by DFT guest binding energies, analysis of pore limiting envelopes and metadynamics simulations of the free energy barriers to guest diffusion. It was shown for the well-known **CC3** porous organic cage, that the

single molecule approach, whilst losing some information as a result of neglecting consideration of solid state packing, overall captures the key features relevant to Xe/Kr separation and allows one to uncover a ‘best case scenario’ prediction for experimental separation performance. The reported structural analysis for these molecular pores could also be useful for discovering other separation applications, of different sized guest molecules, that these systems have potential for.

The results highlighted the potential for noria molecules (**NC1** and **NC2**) to perform well in Xe/Kr separation and they were therefore synthesised and tested accordingly by synthetic collaborators. Whilst both were found to have good performance (selectivity at 1 bar and 298 K of 3.8 for **aNC1** and 9.7 for **aNC2**), this was considerably less than that previously reported 14.1 for **CC3 $\alpha$ -R**. In my opinion, some of this weaker performance relative to **CC3** can be attributed to the amorphous nature of the solids forms for the noria molecules, which inevitably influences its pore structure. Indeed, when an amorphous form of **CC3** was tested, it had a lower selectivity of 10.3 compared to the crystalline state. However, when also considering the higher uptake capacity for **CC3**, and the ease with which it can be processed into a crystalline form, with respect to the previously reported porous molecular materials, **CC3** has the best Xe/Kr separation performance. Thus in this case experimental testing alone, with some degree, perhaps, of serendipity, has uncovered the best intrinsically molecular material reported thus far for Xe/Kr separation.

The single molecule screening approach for molecular pores, here applied to previously reported porous molecular materials, can now be automated for high-throughput screening of a much vaster number of hypothetical cage molecules, a type of molecular pores that have attracted much interest in recent years, which have a near infinite chemical space. I believe the methodology provides the right tools for fast screening of the chemical space of organic cages that could significantly accelerate materials discovery.

# Chapter 5

## C<sub>60</sub> Fullerene Encapsulation in Porous Organic Cages

In this chapter, a computational study investigating the potential of eight recently synthesised porous organic cage (POC) molecules for encapsulation of C<sub>60</sub> fullerene is presented. A density functional theory (DFT) and force field evaluation of the host-guest interactions is performed and the results of the two methods compared. The effect of the relative orientation of the C<sub>60</sub> fullerene in the host cavity on the binding energy is also studied. These findings were used as the validation step for the screening of hypothetical POCs for the C<sub>60</sub> fullerene encapsulation application in the following chapter.

The studied POCs were proposed, synthesised and the NMR spectra measured and analysed by Dr. Rebecca Greenaway (University of Liverpool). The coordinates of the lowest energy conformers were provided by Valentina Santolini (Imperial College London).

## 5.1 Introduction

Since their discovery in 1985, fullerenes have been of great interest and earned Kroto, Curl and Smalley the Nobel prize in chemistry in 1996.<sup>248</sup> The application of fullerenes span over multiple areas of biomedicine<sup>249</sup> and materials science,<sup>250</sup> for example in organic photovoltaics devices and superconductive materials.<sup>251,252</sup> The most representative fullerene (buckminsterfullerene),  $C_{60}$ , formed of 60 carbon atoms connected into five and six-membered rings, resembling the pattern of a football, is only one out of a diverse group of fullerenes.<sup>248</sup> The common source of fullerenes, fullerene soot, is a mixture of the most abundant  $C_{60}$ ,  $C_{70}$ , and some higher fullerenes with other forms of carbon present. Therefore, a lot of effort has been applied to research into the selective binding of different species of fullerenes for the purification process.<sup>253</sup> The immobilisation of the fullerenes in their complexes enables controlled property tuning and selective formation of fullerene adducts.<sup>254,255</sup> Fullerenes can also act as templates and drive the macrocycle formation towards desired supramolecular architectures.<sup>256</sup>

The common mechanism of fullerene encapsulation is to maximise the non-specific van der Waals interactions between the host molecule that “wraps” itself around the fullerene, as in the “buckycatcher” (see Figure 5.1).<sup>257</sup> The host-guest interactions of fullerenes,  $C_{60}$  in particular, have been studied in complexes with various hosts, such as macrocycles, cyclic porphyrins or metallocages.<sup>258</sup> Organic cages, however, have been mostly absent in fullerene host-guest supramolecular chemistry. The only two examples that have been proposed as possible fullerene hosts: a sandwich-like cage and a porphyrin cage (**COP-5**) shown in Figure 5.1, are both a product of an alkyne metathesis reaction.<sup>259,260</sup>

Grimme *et al.* performed an in-depth computational evaluation of some density functional theory (DFT) generalised gradient approximation (GGA) functionals, B97, BLYP, and PBE.<sup>261</sup> These functionals were used to study the host-guest interactions in a multi-shell “hyperfullerene”, where  $C_{60}$  is encapsulated within a larger  $C_{240}$  shell ( $C_{60}@C_{240}$ ). The reported binding energies ( $E_{binding}$ ) were -770 and -606 kJ mol<sup>-1</sup> for the dispersion corrected B97-D/TZVP and PBE-D/TZVP calculations, respectively. By contrast, the non-corrected PBE/TZVP and BLYP/TZVP would indicate a repulsive interaction of 4 and 51 kJ mol<sup>-1</sup>. Indeed, the major contribution to the host-guest interactions comes from the weak van der Waals forces that DFT

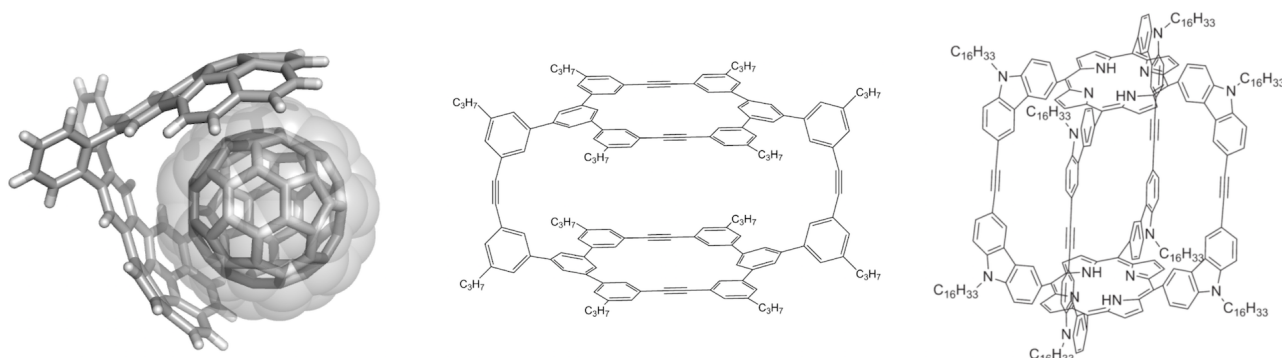


Figure 5.1: Examples of C<sub>60</sub> host molecules. From left: C<sub>60</sub>@buckycatcher complex with the C<sub>60</sub> van der Waals spheres semi-transparent overlay,<sup>257</sup> a tetrameric cage with two macrocyclic panels<sup>259</sup> and a porphyrin cage.<sup>260</sup>

does not describe and this demonstrates, unsurprisingly, the necessity of the dispersion correction. In a subsequent study of the C<sub>60</sub>@buckycatcher (see Figure 5.1), they established the supporting role that the  $\pi - \pi$  interactions play in stabilising the “van der Waals complexes” of conjugated systems.<sup>262</sup> Here, the interaction is calculated as  $-186 \text{ kJ mol}^{-1}$  in the B97-D/TZVP calculation, which is significantly less than that for the C<sub>60</sub>@C<sub>240</sub>. However, the latter should be seen as the physical upper limit for the C<sub>60</sub> interactions with a host that evenly and optimally encapsulates it from each direction. In C<sub>60</sub>@buckycatcher the C<sub>60</sub> encapsulation by the two corannulene bowls results in less interactions.

In this chapter, I examined eight possible C<sub>60</sub>@POC complexes, proposed by the experimental collaborators from recently synthesised POCs. The interaction of C<sub>60</sub> with the host molecules were analysed at the DFT level and compared to the force field obtained energies. I looked at the correlation of the observed  $E_{binding}$  and the C<sub>60</sub> relative orientation in the host’s cavity. Lastly, MD calculations were performed to indicate whether diffusion of C<sub>60</sub> through the POC windows is possible for the most promising C<sub>60</sub>@POC complex, for which formation was confirmed experimentally.



## 5.2 Methods

### 5.2.1 Generation of the cage set

A set of building blocks (triotopic amines) and linkers (di- and tritopic aldehydes) were proposed by Dr. Rebecca Greenaway (University of Liverpool) to form cages with [4+6] (**A1**, **A2**, **B1**, **B2**) and [4+4] (**A3**, **A4**, **B3**, **B4**) topologies, presented in Figure 5.2, and subsequently recently realised experimentally.<sup>1</sup> The [4+6] and [4+4] topologies result from the imine condensation reaction of 4 aldehydes and 6 or 4 amines, respectively. These particular organic cages were chosen as their cavities were assessed to be large enough to encapsulate C<sub>60</sub> fullerene. As there was no single crystal X-ray diffraction (SC-XRD) data available for the proposed cages at the time of this study, the Cartesian coordinates for the molecules were generated.

The cages were assembled using an in-house python script. The lowest energy conformers were then obtained by Valentina Santolini (Imperial College London). To find the lowest energy conformers for the cages, high temperature MD simulations were performed in Schrödinger LLC's MacroModel software (Release 2015-3). The MD simulations were run at 1000 K for 100 ns, with a time step of 0.7 fs. The OPLS3 all-atom force field<sup>168</sup> was used in the MD simulations, as it provides more parameters than the older OPLS\_2005 version (see discussion in Section 2.3). A set of randomly chosen configurations from the MD trajectory were then geometry optimised with OPLS3. To obtain a more reliable energetic scoring, an additional DFT geometry optimisation of each structure was carried out in CP2K.<sup>238</sup> A PBE functional<sup>179</sup> with Grimme's D3 dispersion correction<sup>241</sup> and TZVP basis set<sup>242</sup> was used. The plane wave cut-off was set to 350 Ry in all DFT calculations chosen as a trade off value between the cost and the accuracy of the calculations. The PBE-D3/TZVP set up, and plane wave cut-off, was found before to provide a good balance between the computational cost and accuracy.<sup>263</sup> The MD simulation and DFT geometry optimisation was then repeated for each lowest energy conformer to ensure no lower energy conformers were to be found with the described procedure.

---

<sup>1</sup>Unpublished work, manuscript in preparation.

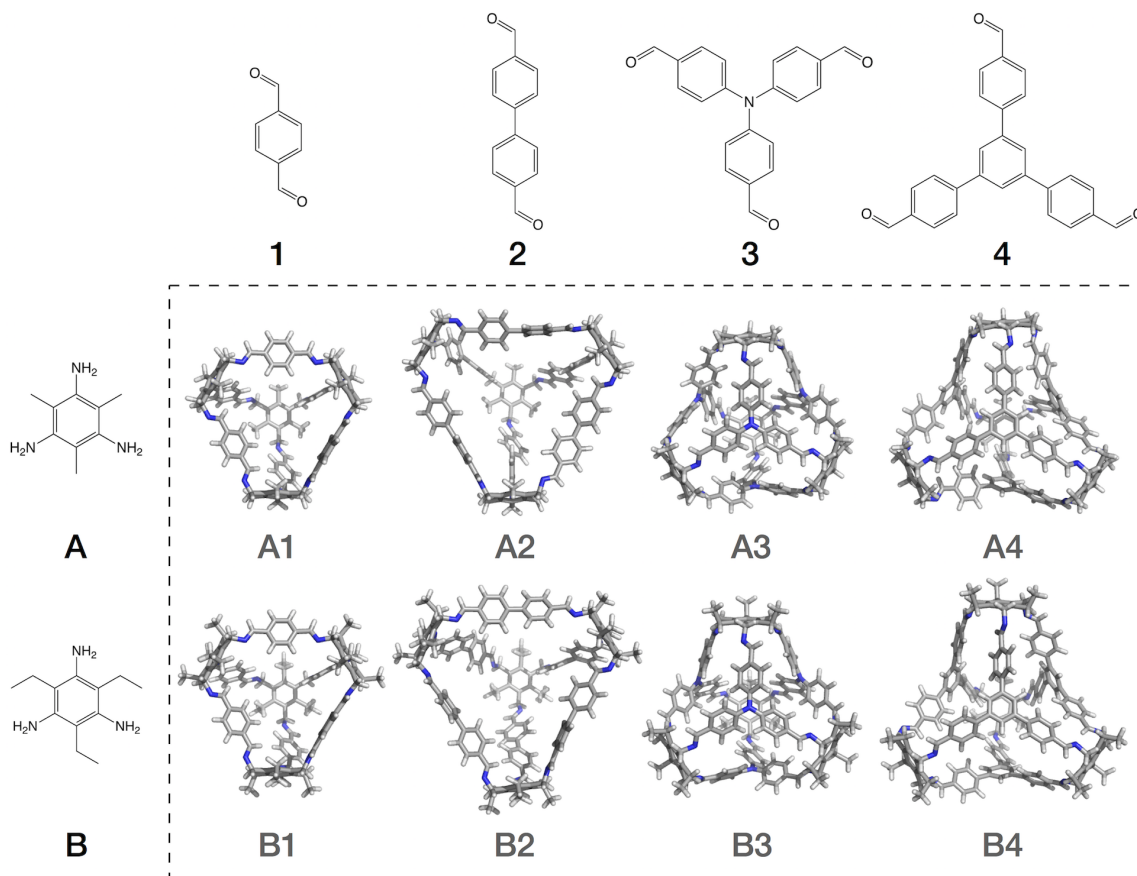


Figure 5.2: The two triamines: **A** - (2,4,6-Trimethylbenzene-1,3,5-triyl)trimethanamine, **B** - (2,4,6-Triethylbenzene-1,3,5-triyl)trimethanamine, and the aldehyde linkers: **1** - terephthalaldehyde, **2** - 4,4'-biphenyldicarboxaldehyde, **3** - tris(4-formylphenyl)amine, **4** - 5'-(4-Formylphenyl)-[1,1':3',1''-terphenyl]-4,4''-dicarbaldehyde, used in the imine condensation reaction to form eight porous organic cages. The resulting cage molecules from the combination of presented amines and aldehydes are shown in the area enclosed by dashed lines.

### 5.2.2 DFT and force field $C_{60}$ @POC binding energies

The Cartesian coordinates of the lowest energy conformer for each of the eight cages were used as the initial configuration in all calculations. The DFT energies of an isolated empty cage ( $E_{host}$ ), an isolated  $C_{60}$  fullerene ( $E_{C_{60}}$ ), and a  $C_{60}$ @POC complex where the  $C_{60}$  was placed at the centre of the cage cavity ( $E_{C_{60}@host}$ ), were calculated. Each such set was run in a unit cell of cubic symmetry with a consistent unit cell dimension, determined as the maximum dimension of a cage plus 8 Å in each direction. The DFT set up as described in the previous section was used. The binding energy was calculated according to the following equation:

$$E_{binding} = E_{C_{60}@host} - E_{host} - E_{C_{60}} \quad (5.1)$$

To evaluate the contribution from the van der Waals interactions to the  $E_{binding}$  each calculation was also performed without the dispersion correction applied (PBE/TZVP). From the  $E_{binding}$  obtained from PBE-D3/TZVP calculations, the  $E_{binding}$  obtained from PBE/TZVP calculations was subtracted, to obtain the van der Waals contribution.

The force field derived  $E_{binding}$  was obtained in a similar fashion from the OPLS3 geometry optimised structures of an empty cage, an isolated  $C_{60}$  fullerene and a  $C_{60}@POC$  complex using Schrödinger LLC's MacroModel (Release 2016-2). The OPLS3 force field was used throughout and the cutoff for van der Waals interactions and electrostatics set to 7 and 12 Å, respectively, the default values for cutoff in MacroModel. The charges were assigned by MacroModel for OPLS3. All calculations were performed *in vacuo*, where specification of the enclosing unit cell is not necessary. The  $E_{binding}$  was derived from equation 5.1. The van der Waals contribution to the  $E_{binding}$  was calculated from the the total energy decomposition into specific contributions. Using equation 5.1 and the energy originating from the van der Waals interactions, instead of the total energy of the system, the van der Waals interactions originating only from the host-guest interactions was calculated.

To obtain the optimised void diameter ( $D_{void\_opt}$ ) of the empty cages and cages from  $C_{60}@POC$  complexes with  $C_{60}$  deleted from the cavity, the `pywindow` software, introduced in Section 3, was used. The  $D_{void\_opt}$  was determined for the Cartesian coordinates taken from the geometry optimised structures from both DFT and OPLS3 calculations.

### 5.2.3 $C_{60}@A3$ complexes with different relative host-guest orientation

An in-house python script was developed to look at the effect of the relative host-guest orientation on the binding energy. The guest molecule was rotated around two of the three XYZ axes with respect to the host molecule kept fixed. The **A3** cage was chosen as the one with the strongest host-guest interactions with  $C_{60}$  out of the eight studied organic cages. Firstly, the  $C_{60}$  was placed at the centre of the **A3** cavity to yield a  $C_{60}@A3$  complex. The  $C_{60}$  was then aligned with respect to its symmetry axes so that both the 5-fold symmetry axis passing through the centre of one of the 5-membered rings would be identical to the Z axis and the 2-fold rotation axis, perpendicular to the 5-fold symmetry axis, was aligned with the X axis.

From this initial configuration, a rotation around the Z axis would be performed in  $n * 9^\circ$  step, where  $n$  is a number from 0 to 7 in ascending order. These values of  $n$  are due to the 5-fold symmetry, which results in the symmetry equivalence of the  $0^\circ$  and  $72^\circ$  relative orientations of  $C_{60}$  with respect to the Z axis ( $n$  equal to 0 and 8, respectively). Therefore, the number of unique configurations can be reduced by a factor of 5, in comparison with a full  $360^\circ$  rotation around the Z axis with  $9^\circ$  step, by considering only  $n = 0 - 7$ . For each  $n$ , 20 subsequent rotations around the X axis with a  $m * 9^\circ$  step were generated ( $m = 0 - 19$ ), taking advantage of the 2-fold symmetry axis. This procedure resulted in a total of 160 symmetrically unique configurations, equivalent to a full rotation around the Z and X axes with a  $9^\circ$  step, that would otherwise render 1600 non-unique configurations. For each generated structure, a geometry optimisation using DFT and OPLS3 calculations was performed by the same procedure as described in Section 5.2.2.

#### 5.2.4 Molecular dynamics study of B4 cage and $C_{60}@B4$ complex.

Based on the experimental results (see Section A.9 of Appendix)  $C_{60}$  was found to have a templating effect on the formation of **B**-series cages.<sup>2</sup> However, no  $C_{60}@POC$  complexes were identified experimentally, except for a potential mixture of **B4** and  $C_{60}@B4$  complex (see Figure A.21 in Appendix). To study the potential of  $C_{60}$  diffusion through the cage's windows and the dynamic effect on the pore space, further evaluation of **B4** and  $C_{60}@B4$  using an MD simulation was carried out. The Cartesian coordinates for the **B4** cage and  $C_{60}@B4$  complex were taken from the DFT geometry optimised structures. The OPLS\_2005 all-atom force field<sup>209</sup> parameters were generated with Schrödinger LLC's MacroModel `ffld_server` script. The DL\_POLY2.20 software<sup>169</sup> input files were prepared using the DL\_FIELD3.3 software.<sup>210</sup> The choice of OPLS\_2005 is dictated by the fact that the OPLS3 parameters are only available for the calculations performed in Schrödinger LLC's MacroModel, whereas here I used the DL\_POLY2.20 software. The calculations consisted of a 100 ns production run preceded by 0.2 ns equilibration. A timestep of 0.5 fs was used and configurations were sampled every 2 ps. All calculations were performed at 298 K and 0.0001 Pa. The Leapfrog Verlet algorithm<sup>211</sup>

---

<sup>2</sup>Unpublished work, manuscript in preparation.

was used to assign forces and velocities and the electrostatic interactions treated with the Coulomb summation. No periodic boundary condition was used and the cut-off for the short-range van der Waals interactions was set to 15 Å. The obtained trajectory was then analysed using the `pywindow` package to obtain the window diameters and  $D_{void\_opt}$ . The pore limiting envelopes (PLEs) and the distribution of cavity diameters, were generated using the kernel density estimate (KDE) function of the `SciPy` package as described in Chapter 3.<sup>207</sup>

## 5.3 Results

There are two main objectives for the computational analysis of the proposed set of porous organic cages as possible  $C_{60}$  encapsulants. Firstly, using the structural analysis of the void diameters I tried to determine whether these POCs could potentially fit  $C_{60}$  in their cavity and if the binding affinity renders the  $C_{60}$ @POC complexes likely. Secondly, I derived the binding energies from the electronic structure calculations at the DFT level of theory and from the force field potential using OPLS3. The two methods differ fundamentally. The DFT geometry optimisation considers the electronic structure of the system and was used here with the GGA PBE functional that derives the parameters directly from theory. The DFT refined geometries are often considered as more accurate in comparison with force field potential optimised geometries. However, in these  $C_{60}$ @POC complexes, the majority of host-guest interactions arise from the van der Waals forces. Therefore, it is necessary to apply an empirical dispersion correction in DFT calculations. In contrast, in force field methods the van der Waals interactions are directly derived from the Lennard-Jones potential and the family of OPLS force fields was shown previously to provide a good agreement of the geometries with the SC-XRD crystal structures of some porous organic cages.<sup>123</sup> Therefore, I provide a comparison of the two methods with the focus on the binding energies and the calculated cavity diameters. This chapter is a validation step for the following chapter where an evolutionary algorithm-assisted material design is described, and where OPLS3 is used to derive the  $C_{60}$ @POC complex binding energies as a faster and cheaper method than DFT.

The  $C_{60}$ @POC complex binding energies for the DFT (PBE-D3/TZVP) and OPLS3 calculations are presented in Figure 5.3. The contribution from the van der Waals forces in the DFT

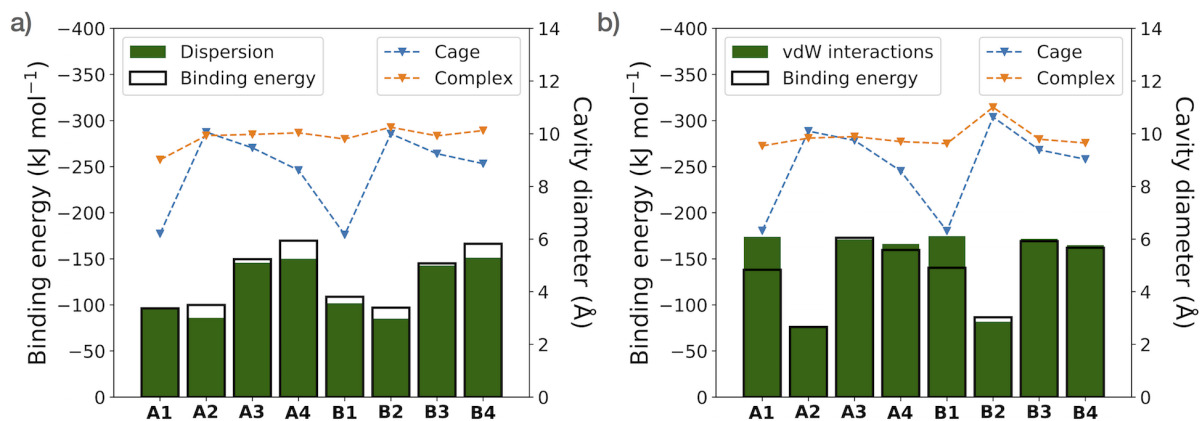


Figure 5.3: The binding energies for C<sub>60</sub>@POC complexes derived from DFT (PBE-D3/TZVP) - plot a, and OPLS3 - plot b. The binding energy is presented as an empty bar plot with black edges, whereas the van der Waals interactions is estimated from the dispersion correction (DFT) or the system total energy decomposition contribution (OPLS3) as overlaid green bars, with respect to the left hand side axes. The calculated cavity diameters, for the optimised structure of an empty cage (orange) and for the cages extracted from the optimised structures of the C<sub>60</sub>@POC complexes (blue), are presented with respect to the right hand side axes.

and OPLS3 derived  $E_{binding}$  is also evaluated. Lastly, the calculated cavity diameters for the optimised structures of empty cages and for the C<sub>60</sub>@POC complexes, with the C<sub>60</sub> removed from the pore, are also given. The range of the binding energies for both DFT and OPLS3 is similar and falls between -96 to -170 kJ mol<sup>-1</sup> and -76 to -169 kJ mol<sup>-1</sup>, respectively. However, the binding energy trend is not consistent between the DFT and OPLS3 calculations. This is acceptable for this study, as the evaluation of these two methods is one of the objectives. The binding energies for the **A**- and **B**-series of discussed C<sub>60</sub>@POC complexes are consistent in both methods. The small structural difference in the external functionalisation with methyl (**A**-series) or ethyl (**B**-series) groups, that point outwards from the cage cavity, seem to have a small effect on the resulting  $E_{binding}$ . Therefore, I will only discuss the results from the perspective of the **A**-series C<sub>60</sub>@POC complexes, where the discussion of the results for the **A**-series complexes is also applicable to the **B**-series complexes for both DFT and OPLS3. The calculated cavity diameters are very similar between DFT and OPLS3 geometry optimised empty cages and cages extracted from the C<sub>60</sub>@POC complexes and follow the same trend. The exact numerical values for  $E_{binding}$  and  $D_{void\_opt}$  can be found in Section A.8 of Appendix.

The highest DFT derived binding energies are for C<sub>60</sub>@**A4** (-169.6 kJ mol<sup>-1</sup>) and C<sub>60</sub>@**A3** (-149.6 kJ mol<sup>-1</sup>) complexes. However, the reverse is observed for the OPLS3 derived values,

with  $-172.8 \text{ kJ mol}^{-1}$  for  $\text{C}_{60}@A3$  and  $-159.6 \text{ kJ mol}^{-1}$  for  $\text{C}_{60}@A4$ . The calculated cavity sizes of empty **A3** and **A4** are 9.7 and 8.6 Å for the DFT optimised geometries, and 9.5 and 8.6 Å for the OPLS3 optimised geometries. Therefore, some expansion (between 7 - 21%) of the cavities of empty cages is necessary to accommodate  $\text{C}_{60}$  ( $d = 10.4 \text{ Å}$ ) in both **A3** and **A4**. Both the DFT and OPLS3 methods rank **A3** and **A4** as better hosts for  $\text{C}_{60}$  encapsulation than the **A1** and **A2** cages. However, the ranking between **A3** and **A4** is different in the two methods. An empty **A1** cage has the smallest cavity (6.3 Å) that has to undergo the most significant distortion of 51% for the  $\text{C}_{60}$  fullerene to fit in the cavity. On the other hand, **A2** has the largest cavity from the four **A**-series POCs (10.1 Å) and that results in a small contraction (2%) over  $\text{C}_{60}$  encapsulation. The OPLS3 derived binding energy for the  $\text{C}_{60}@A1$  complex is significantly larger ( $-138.1 \text{ kJ mol}^{-1}$ ) than that derived from DFT ( $-96.3 \text{ kJ mol}^{-1}$ ). The non-bonding van der Waals interactions for  $\text{C}_{60}@A1$  complex in OPLS3 are a significant contributing factor to the  $E_{binding}$ , that in fact exceed the strength of the resulting binding affinity. It is possible, the strain that the **A1** cage experiences when  $\text{C}_{60}$  is present in its cavity is responsible for that difference.

In summary, the OPLS3 ranking of **A1**, seen as an extreme case of fitting  $\text{C}_{60}$  in too small a cavity, is worse than that from the DFT. Here, the two extremes **A1** and **A2**, are both ranked as poor candidates. The dispersion correction to the binding energies derived from DFT is significant and accounts for over 85% of the  $E_{binding}$  in all cases. In the OPLS3 derived binding energies, the non-bonding interactions constitute all of the resulting binding affinity, or in some cases exceed it. Some correlation can be observed for the DFT binding interaction and the cavity size of empty cages. Small cavities results in lower  $E_{binding}$ . Similarly, larger cavities, as in **A2**, also results in small binding energies. The cavities of **A4** and **A3** are possibly of an optimal size for  $\text{C}_{60}$  encapsulation and thus the higher binding energies. The same trend is observed for OPLS3, with the exception of **A1**.

From the previous considerations, I chose the  $\text{C}_{60}@A3$  complex, which had the highest  $E_{binding}$ , to look at the effect of the orientation of the  $\text{C}_{60}$  fullerene in the POC cavity on the binding energy (see Figure 5.4), and where this orientation effect could be most significant. Although  $\text{C}_{60}$  can often be approximated to a perfect spherical shape, at the atomistic level

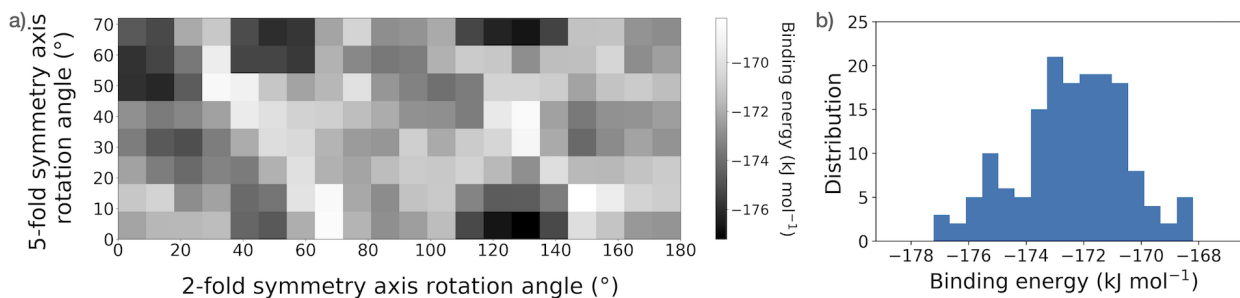


Figure 5.4: a) The heat-map of the DFT derived  $E_{binding}$  for a set of configurations with a relative host-guest orientation of  $C_{60}@A3$  complex. b) the distribution of the DFT derived  $E_{binding}$  for the same set.

it has a “rough” surface with cones and planes present. To quantify the effect this structural roughness might have on the  $E_{binding}$ , a systematic study of the relative host-guest orientation was performed. The rotation of the guest, with host coordinates frozen in space, is possible in 3 dimensions. However, to reduce the number of considered configurations, only the rotation around the high symmetry  $C_5$  axis passing through one of the 5-membered rings and the  $C_2$  axis perpendicular to it were considered, resulting in 160 distinct configurations. If the rotation around the  $C_3$  symmetry axis were also included, this would result in over 2000 configurations. Moreover, if no  $C_{60}$  symmetry was considered and a full rotation around the XYZ axes with a  $9^\circ$  step were performed, it would result in 64000 configurations, although not symmetrically distinct. Therefore, as the result of the used DFT set up (PBE-D3/TZVP), the compromise between the number of configurations considered and the possibility of some information on the 3-fold symmetry axes hidden, had to be made.

The presented heat-map, in Figure 5.4a, shows the  $E_{binding}$  in respect to the rotation of the  $C_{60}$  around  $C_5$  and  $C_2$  symmetry axes. The Figure 5.4b shows the distribution of the DFT binding energies for the 160  $C_{60}@A3$  configurations. Overall, the range of binding energies spans from -168 to -177 kJ mol<sup>-1</sup> giving an approximate 9 kJ mol<sup>-1</sup> window. There are two possible explanations for the observed  $E_{binding}$  distribution. The relative host-guest orientation even for a highly spherical molecule, that  $C_{60}$  is, should result in small variations in the binding energy. Simultaneously, the observed distribution of binding energies can be attributed to the accuracy of the PBE-D3/TZVP setup. Previously, in a Grimme *et al.* study for various van der Waals complexes and different DFT set ups, the PBE-D3/TZVP functional was associated



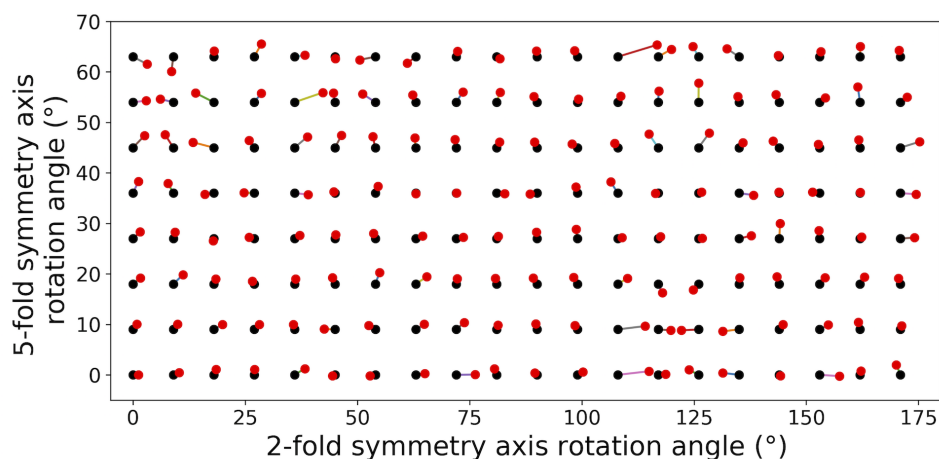


Figure 5.5: The scatter plot of the initial rotation of  $C_{60}$  relative to the  $C_5$  and  $C_2$  symmetry axes (black dots) and the relative rotation after the DFT geometry optimisation (red dots), paired using lines of various colours.

with the mean absolute error of around  $10.5 \text{ kJ mol}^{-1}$ .<sup>244</sup> This value is close to the one observed here. In conclusion, it is likely that both the effect of the relative  $C_{60}$  orientation and the error expected for this type calculations is a cause for the observed range of the binding energies. The specific contributions of the two effects cannot be easily distinguished, therefore it can be concluded that the relative orientation for molecules of such high sphericity is not a significant factor, and in future considerations, the relative  $C_{60}@POC$  orientation can be ignored. This assumption is not true for other non-spherical guest molecules, where the orientation of the guest will have moderate to large effect on the binding affinity and will have to be considered, significantly increasing the computational expense of such studies. In Figure 5.5, the change in the relative orientation of the  $C_{60}$  with respect to the cage, between the input and output coordinates of the geometry optimisation, is presented. The freedom of  $C_{60}$  to diverge from the initial configuration is small, as a result the heat-map in Figure 5.4a can be considered as for the initial relative host-guest orientation.

The same set of 160 initial configurations of relative host-guest orientation was used to generate the distribution of OPLS3 binding energies. In Figure 5.6a, the change of the initial relative host-guest orientation during the force field geometry optimisation is presented. The change of the configuration is much more significant, in comparison with DFT geometry optimisations (see Figure 5.5), as a result of the much larger flexibility of force fields to overcome small energy barriers required for rotations. Therefore, the structures tend to group into local

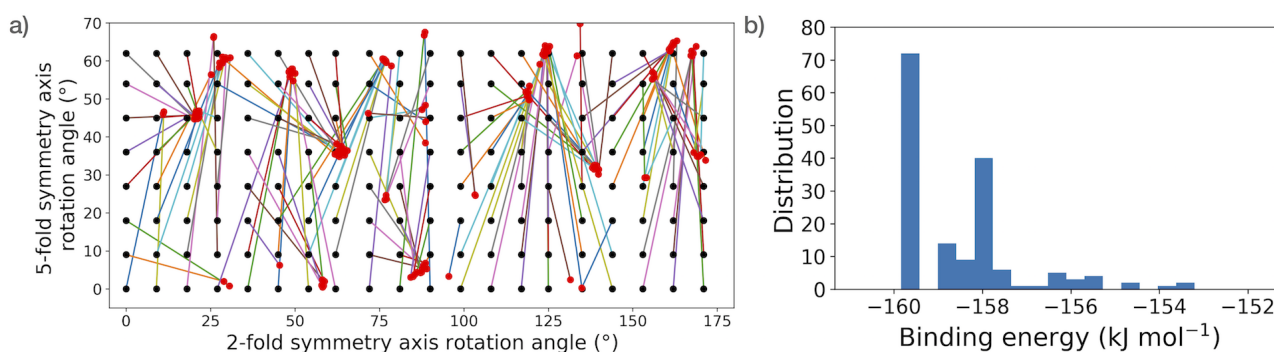


Figure 5.6: a) The scatter plot of the initial rotation of  $C_{60}$  relative to the  $C_5$  and  $C_2$  symmetry axes (black dots) and the relative rotation after the OPLS3 minimisation (red dots), paired using lines of various colours. b) the distribution of the OPLS3 derived  $E_{binding}$  for the  $C_{60}@A3$  complex.

minima, as can be seen in Figure 5.6a. A heat-map could not be generated due to the change of the relative host-guest orientation before and after geometry optimisation. The highest binding energy, shared between the largest number of configurations, is around  $-160 \text{ kJ mol}^{-1}$ , this is  $\sim 12 \text{ kJ mol}^{-1}$  lower than the averaged value derived from DFT. In summary, the ability of the force field to overcome small barriers and find minima in the geometry minimisation process is higher than that of DFT. Therefore, it can be more practical to use force fields in computationally generated complexes. This could increase the likelihood of finding the global minimum configuration as a result of the guest reorientation with respect to the host.

The NMR spectroscopy experiments, performed by Dr. Rebecca Greenaway (University of Liverpool), show several of the **B**-series POCs form at lower concentrations only in the presence of  $C_{60}$  (see Section A.9 of Appendix). This suggests that  $C_{60}$  has a templating effect on the formation of the POCs. Additionally, in the case of the **B4** cage, an additional cage product formed during synthesis in the presence of  $C_{60}$ , suggesting formation of the  $C_{60}@B4$  complex. To determine if the  $C_{60}@B4$  complex is a result of the  $C_{60}$  diffusion into the cavity of already formed cage, or whether  $C_{60}$  is encapsulated during the cage formation, MD calculations were performed. The computed distributions of cavity diameters and the PLEs for the empty **B4** and  $C_{60}@B4$  complex are presented in Figure 5.7. As expected, the distribution of cavity diameters is shifted to larger values in the  $C_{60}@B4$  complex, closer to the assumed  $C_{60}$  diameter of  $10.4 \text{ \AA}$ . In comparison, the empty **B4** has much broader distribution of the cavity diameters shifted to smaller values. The window diameters (Figure 5.7b) exceed the diameter of  $C_{60}$  ( $d = 10.4$

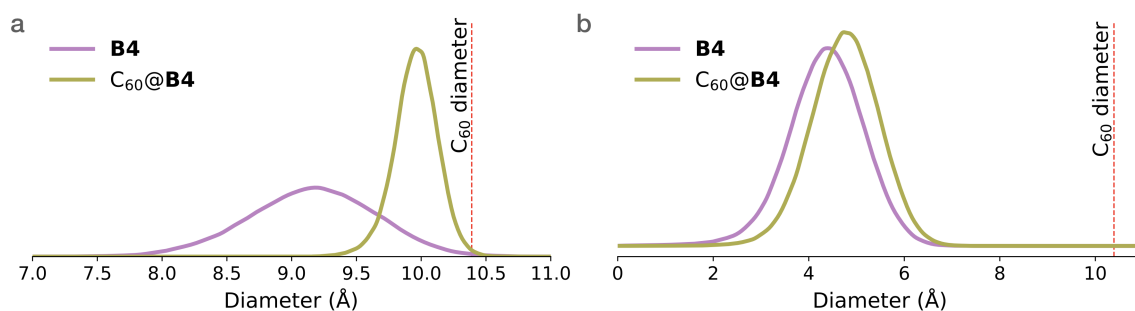


Figure 5.7: a) The distribution of the cavity diameters of an empty **B4** cage and  $C_{60}@B4$  complex plotted from MD simulations. The plot was truncated to show a range between 7 - 11 Å. b) The pore limiting envelopes of an empty **B4** cage and  $C_{60}@B4$  complex plotted from MD simulations. The plot was truncated at 11 Å.

Å) in both an empty **B4** and  $C_{60}@B4$  complex, however, for very small periods of time. In both cases the window is being considered “open”, meaning the window diameter is bigger than the  $C_{60}$  diameter for only 0.0016% of the total simulation time. This suggests diffusion of  $C_{60}$  into the cavity of an already formed cage is unlikely, and that the  $C_{60}$  is encapsulated during the **B4** formation.

## 5.4 Conclusions

The evaluation of eight synthetically realised POCs for application as  $C_{60}$  encapsulants was presented. The analysis of the DFT and OPLS3 derived binding energies for the eight  $C_{60}@POC$  complexes and the structural effect of the  $C_{60}$  encapsulation on the cavity, suggests **A3**, **A4**, **B3**, and **B4** as the most promising candidates. The other proposed cages also show affinity towards  $C_{60}$  binding. The overall agreement of the DFT and OPLS3 derived energies was discussed, with a possible explanation for the observed differences given. The binding energy values derived from both methods fall into the expected range of values from the literature, with an upper limit of the interaction on the order of hundreds of  $\text{kJ mol}^{-1}$ . These findings were later supported by the experimental study of the templating effect of  $C_{60}$  presence on the formation of the discussed POCs. The  $^1\text{H}$  NMR experimental analysis of **B4** formation in the presence of  $C_{60}$  suggest a formation of  $C_{60}@B4$  complex, although further experimental evaluation would be necessary to confirm the finding and to exclude formation of other products. This result was followed by computational MD calculations. Results suggested that diffusion of  $C_{60}$  through the

**B4** into the cavity through the windows as unlikely. The possible binding of the  $C_{60}$  in cavity during the **B4** formation could be confirmed, if the formation of  $C_{60}@B4$  complex was further studied experimentally. This work shows the supporting role that computational studies can have to the experiments.

The comparison of the DFT and OPLS3 derived binding energies give an insight into the accuracy of the latter method. In summary, the OPLS3 force field was found to possibly overestimate the binding energies for extreme cases of  $C_{60}$  confinement as in the  $C_{60}@A1$  and  $C_{60}@B1$  complexes. However, good agreement overall for the best candidates proves the force field derived binding energies are a good compromise between computational expense and accuracy. The systematic study of the effect of the relative host-guest orientation in the case of highly spherical  $C_{60}$  fullerene, suggests a relatively small effect on the resulting DFT binding energies, with a narrow distribution. This effect was measured to be of similar magnitude as the expected DFT accuracy ( $\sim 10 \text{ kJ mol}^{-1}$  for PBE-D3/TZVP). Therefore, it is uncertain to what extent the scale of the binding energies distribution can be attributed to the relative host-guest configuration or the uncertainty of the method. On the other hand, the rotational freedom of the guest molecule in the force field geometry optimisation process was shown to be higher. This is expected for the force fields methods, where the ability to overcome small energy barriers for configurational changes is higher than in DFT. However, this could be advantageous in the automated computational high-throughput studies, such as the one presented in the following chapter. For the purpose of automation, the guest molecule is inserted into the centre of the cavity in a random orientation. Consideration of additional configuration will have a linear scaling effect on the computational expense of such high-throughput screening. As discussed in this chapter, for a non-symmetrical and non-spherical guest molecule, a rotation around all three Cartesian axes with a modest step of  $9^\circ$  could result in 60000 distinct configurations. For the computational high-throughput screenings, other approaches to significantly reduce the number of considered configurations are advised, such as limiting the routine to a small number of random configurations or short high temperature MD simulations. Additionally, the relatively large hops on the potential energy surface, that were shown here to be characteristic of a force field geometry optimisation, could increase the likelihood of finding lower minima.

## Chapter 6

# Function-Led Computational Design of New C<sub>60</sub> Encapsulants with an Evolutionary Algorithm

In this chapter I present an evolutionary algorithm approach to the computational design of porous organic cages with tailored properties. The cages are constructed using software developed in our group specifically for this purpose that uses an evolutionary algorithm to effectively sample the large chemical space of organic cages assembled from selected precursors. The solutions to the fitness function are then effectively identified. The resulting cages are tailored to fit C<sub>60</sub> in their cavity and exhibit strong binding affinity towards C<sub>60</sub>. The presented approach is easily transferable to other databases of precursors and can be adjusted to aim at different properties. The results show a successful function-led computational design of porous organic cages for applications.

The Supramolecular Toolkit (STK) software used in this chapter was developed by Lukas Turceni and Dr Enrico Berardo (Imperial College London). The database of precursors was selected by Dr Enrico Berardo. The assembly and geometry optimisation of porous organic cages for the purpose of fitness function parameterisation was carried out by Lukas Turceni.

## 6.1 Introduction

Computation is increasingly important in the discovery and development of advanced materials.<sup>264</sup> The ability to design materials with desired function *in silico*, meaning using computation instead of laborious experiments, is slowly revolutionising science and industry. In 2011, the U.S. government founded the Materials Genome Initiative in order to accelerate advanced materials discovery.<sup>265</sup> The initiative highlights that the time necessary for newly discovered materials to reach the market from the laboratory is far too long, usually 10 to 20 years. This is partially due to the discovery being a result of trial and error experimentation, with much of the design and testing of new advanced materials performed in a time consuming and repetitive loop of experiment and characterisation. Furthermore, new discoveries are still more often a result of serendipity rather than of intentional planning. The Materials Genome Initiative emphasises that the development of new materials can be accelerated through advances in computational techniques. Thus, part of the experiment and characterisation of newly discovered materials could be performed computationally, resulting in the synthetic realisation of only the most promising materials. Although the Materials Genome Initiative is intended to advance materials discovery in the U.S., without doubt the objectives of the initiative and the proposed solutions can accelerate advanced materials discovery worldwide.

The high-throughput computational screening of existing or hypothetical compounds can facilitate the discovery of new materials (examples of high-throughput screenings were discussed in Section 1.5).<sup>266</sup> These studies can often identify promising candidates, or help formulate principles for further rational materials discovery. However, computational high-throughput screenings are still brute force searches, highly accelerated by computational power. Although the equivalent search would not be feasible experimentally, the cost of finding a few promising candidates often means screening thousands of ineffective materials. An alternative to computational high-throughput screenings is to improve the properties of a working material by creating its' analogues.<sup>267-269</sup> This can be achieved by alternative substitutions and changes to the structure, or by varying building blocks and metal nodes. But, this approach can be largely biased towards the choice of the initial material. As a result, materials that have non-trivial structures, are considered hard to make, or are discarded for any other reason, can easily be

omitted.

The lack of an equivalent approach in the synthesis of porous organic cages (POCs) to the reticular synthesis of metal-organic frameworks (MOFs), which allows synthesis of MOF analogues by varying the precursors without alternating its topology, leads to a far smaller amount of POCs realised experimentally to date.<sup>270</sup> Therefore, the possibility of performing experimental or computational high-throughput studies of POCs is still out of reach. The screening of existing POCs for the Xe/Kr separation application I described in Chapter 4 included roughly 40 POCs identified as promising, with only a few times this amount of organic cages realised synthetically to date. A different approach is necessary to design novel POCs with specific functions. The examples of the computational studies that focus on the design of advanced materials discussed throughout this thesis suggest a way forward.

The computational design of organic cages is often limited to one or a set of similar molecules based on previously realised cases to maximise the synthesisability.<sup>123,125,271</sup> However, without limiting it to the previously realised POCs and their analogues, which could greatly bias the search space, the chemical hyperspace of hypothetical POCs from organic precursors is vast. This results in the opposite problem compared to the very limited number of POCs realised synthetically. As these molecules are usually composed of two organic building blocks the number of possible combinations is enormous. Fink *et al.* estimated the chemical space of chemically stable and synthetically feasible small organic molecules up to 11 atoms and molecular weight below  $160 \text{ g mol}^{-1}$  to be  $\sim 14$  million.<sup>272</sup> In a different study, screening of small molecules up to 17 atoms made of C, N, O, S and halogens resulted in finding 166 billion molecules.<sup>273</sup> In contrast, van Deursen and Reymond estimate the chemical space of all possible organic molecules below  $500 \text{ g mol}^{-1}$  to be in the range of  $10^{20}$ - $10^{200}$ .<sup>274</sup> If only molecules feasible synthetically and with specific functionality were to be considered, for example tritopic aldehydes and ditopic amines that yield imine cages, the number of possible precursors is still too great for brute force screenings. It can be argued that generating large numbers of hypothetical POCs is still possible. However, sampling a chemical space that is diverse enough to avert bias would be challenging with poor candidates possibly outnumbering the optimal solutions by orders of magnitude. The cost of finding good candidates would require a lot of computational power.

Thus, novel approaches are necessary to help navigate this vast chemical hyperspace in an effective manner.

Using the properties of the already sampled candidates to help navigate the chemical hyperspace of possible solutions should allow a more feasible sampling. The right tool would be efficient and allow one to jump between chemically distinct spots on this hyperspace to allow a fast screening of structures with diverse properties to find optimal solutions. Evolutionary algorithms (EAs) are such a tool and have already found some application in computational materials discovery. Davies *et al.* used an EA to optimise the structure of tubular carbon nanoclusters.<sup>275</sup> Bao *et al.* used an EA for the structure optimisation of MOFs for methane storage capacity.<sup>151</sup> This study differs fundamentally from the previously mentioned computational high-throughput studies of MOFs for methane storage applications. Here, the property of the material is part of the feedback loop that allow the EA to optimise the MOF structure by varying precursors and topologies. Recently, Huwig *et al.* developed a genetic algorithm (GA) approach, a subtype of EA, to design dye molecules for the dye-synthesised solar cells with exceptional performance.<sup>276</sup> Their work included optimisation of properties such as HOMO-LUMO gap and the sunlight absorption of organic molecules. Starting from a simple benzene molecule and considering different functional groups, the GA efficiently and quickly optimised the properties of the potential dye molecules. Although the synthesisability was not a considered factor, resulting in many unfeasible structures, the patterns found in the optimal solutions could aid the rational design of synthetically possible dyes for solar cells.

In 2015, a “function-led” computational materials discovery was proposed by Slater and Cooper.<sup>270</sup> They discuss this approach as a four step process, using the well-studied POC **CC3** as an example. The first step requires the assembly of a hypothetical material’s building units, discrete molecules, from precursors. Then, to study the properties of the molecules in bulk, the likely packing motifs need generating with methods such as crystal structure prediction. Finally, using computational chemistry methods, the thermodynamic and kinetic properties of the material can be predicted and the most promising materials realised experimentally. Examples of each of these steps are already present in the literature, however, performed individually.<sup>73,123,124,226,271,277,278</sup> The systematic approach of the function-led material discovery



proposed by Slater and Cooper requires a modular and flexible software capable of performing all these steps, which has not yet been developed. Depending on the desired solution, for example if this is a periodic material with particular mechanical or electronic properties, or a discrete molecule for in-solution guest binding application, the four steps of the proposed function-led computational discovery should be adjusted accordingly.

In this chapter, I present a function-led computational design of  $C_{60}$  encapsulants using evolutionary algorithm software developed in our group. This work is related to the preceding chapter, where the binding interactions between synthetically realised POCs and  $C_{60}$ , with DFT and the OPLS3 force field were evaluated. Here I use a combination of software that allows for the assembly of POCs from precursors and an evolutionary algorithm design process that is capable of propagating the structural information in the population to optimise the solutions. The structural properties of POCs and the corresponding  $C_{60}$ @POC complexes were calculated with a fully automated `pywindow` functionality ported with the software, that I developed and presented in Chapter 3. The promising results and the knowledge obtained can be further utilised in a rational design of POCs with good performance for  $C_{60}$  encapsulation.

## 6.2 Methods

The operation of an EA and the related terminology is described in Section 2.9. The Supramolecular Toolkit (STK) software used in this chapter was developed in our group by Lukas Turcani and Dr Enrico Berardo (Imperial College London). This multi-purpose tool allows the assembly of porous organic cages from their precursors and the construction of a tailored fitness function that can be minimised with a EA. The application of STK for the design of novel  $C_{60}$  encapsulants in this chapter provided feedback and testing to assist in the development of STK.

The goal of this work is to find new  $C_{60}$  encapsulants from computationally generated organic cages. This study is a proof of concept that uses a small custom database of precursors and to reduce the vast chemical space of possible precursors, the following objectives were applied. All cages are assembled in single 4:6 stoichiometry ([4+6]) using imine chemistry. This stoichiometry requires four aldehydes with three functional groups (tritopic) and six amines with

two amine functional groups (ditopic). The [4+6] imine condensation results in a tetrahedral topology of the generated cages.

To simplify the description of the derived POCs and  $C_{60}$ @POC complexes, a generated POC is simply referenced to as the “cage” and the corresponding  $C_{60}$ @POC complex as the “complex”. The final population of cages were assigned code names of type **CX**, where **X** is a number in ascending order and **C1** is a cage of highest fitness value. Lastly, “**CX** complex” corresponds to the  $C_{60}$ @**CX** complex.

### 6.2.1 The database of precursors and cage assembly process

The database of precursors used in this study was selected by Dr Enrico Berardo (Imperial College London). This includes 43 tritopic aldehydes (trialdehydes), presented in Figures A.23 and A.24, and 90 ditopic amines (diamines), presented in Figures A.25 and A.26, in Section A.10 of Appendix. The trialdehydes are also referred to as building blocks (**bbs**) and diamines as linkers (**lks**), based on their positioning on the template geometry in the cage assembly process (trialdehydes on the vertices and diamines on the edges). These precursors were selected from previously reported organic cages that were successfully realised synthetically, and other molecules suggested by our experimental collaborators at the University of Liverpool, who are experts in the synthesis of porous materials. This database is intentionally limited in size to allow quick screening for the purpose of the fitness function parameterisation and to limit the computational expense of this proof-of-concept study.

The cages were assembled with the STK software by placing the **bbs** on the vertices and **lks** on the edges of a template tetrahedral geometry. Initially, the building blocks are placed with a significant spacing between each other that corresponds to a distance of few times greater than the dimensions of the precursors. Then, the appropriate atoms from the functional groups are deleted and the connection between the residues of the aldehyde and amine functionality is created to yield the imine bond. The resulting cage molecule is then minimised to reduce the length of the imine bond to a physically acceptable value, defined by the force field. The assembly process of the related  $C_{60}$  complexes is virtually identical. The only difference is the placement of the  $C_{60}$  at the centre of the template tetrahedral geometry at the very beginning

of the assembly process before geometry optimisation.

For the purpose of the fitness function parameterisation, a database of cages and their  $C_{60}$  complexes resulting from the combination of all possible **bbs** and **lks** was generated. The combination of 43 trialdehydes and 90 diamines resulted in 3870 distinct organic cage molecules and the equivalent number of complexes. All possible cages from a combination of the available precursors were generated as result. This is a step that would not normally be performed in an EA-assisted material design, as it obviously contradicts the purpose of using an EA. This is further discussed in Section 6.4. However, in this proof-of-concept work, the whole database is constructed to allow development of a customised fitness function parameterisation methodology and to compare the final results with all combinations of cages.

The following procedure for finding the lowest energy conformer was performed on the empty cages and their complexes using the OPLS3 all-atom force field<sup>168</sup> in Schrödinger LLC's MacroModel (Release 2016-2):

1. A cage and a complex is assembled with the described procedure.
2. A geometry optimisation is performed. The convergence of the optimisation is defined as a gradient change smaller than 0.05 Å and the maximum number of iterations allowed is 2500. All bonds, apart from those created during the assembly step (imine bonds), are restricted during the geometry optimisation.
3. The optimisation is followed with an MD run at 700 K and with a 1 fs timestep. A 10 ps equilibration is followed by a 200 ps production run. From the resulting trajectory, 20 random configurations were sampled every 10 ps. Each configuration was then geometry optimised with the procedure from step 2. The configuration with the lowest energy at this stage is selected and evaluated with the fitness function.

For each geometry optimised cage and its complex, the following properties were calculated:

1. The  $C_{60}$  binding energy in the complex ( $E_{binding}$ ).
2. The cavity diameter of the cage extracted from the complex ( $D_{complex}$ ).
3. The asymmetry of the cage extracted from the complex ( $A_{complex}$ ).

4. The strain of the cage extracted from the complex ( $ST_{complex}$ ).
5. The cavity diameter of the empty cage ( $D_{cage}$ ).
6. The asymmetry of the empty cage ( $A_{cage}$ ).
7. The strain of the empty cage ( $ST_{cage}$ ).

The binding energy is calculated with the formula:

$$E_{binding} = E_{complex} - E_{cage} - E_{C_{60}} \quad (6.1)$$

The total energy ( $E_{C_{60}}$ ) of an isolated  $C_{60}$  is obtained through the same process of finding the lowest energy conformer as for cages and their complexes. The cavity diameter ( $D$ ) and the window diameters used to calculate the asymmetry of a cage ( $A_{complex}$ ) and its complex ( $A_{cage}$ ) were obtained with `pywindow` (implemented as a part of the `STK` software) as described in Section 3.2.4. The asymmetry ( $A$ ) is defined as the difference between all window diameters in a cage. First, the window diameters are calculated. Then, the asymmetry is calculated as a sum of the differences in all window diameters. The more comparable the window diameters are, the lower the asymmetry of a cage. This is a good indication of the high structural symmetry observed in shape-persistent and non collapsed cages of tetrahedral topology. The strain ( $ST$ ) is calculated as the root mean square displacement (RMSD) of building blocks extracted from the cages and the corresponding building blocks from the database. The deviation from the minimised building blocks in the database of precursors and the building blocks of the cage gives a relative value of the “strain” experienced by the cage structure. The RMSD for a **bb** or **lk** and its counterpart from the precursors database is calculated with the `RDKit` software.<sup>279</sup>

### 6.2.2 Fitness function

The fitness function ( $f(x)$ ) was constructed in `STK` and calculates the fitness of a cage and its  $C_{60}$  complex. The fitness value ranks the cages as possible  $C_{60}$  encapsulants. First, the parameters 1-7, described in previous section, have their values normalised. This ensures that all the parameters are positive, as for example the  $E_{binding}$  can be positive or negative. For each

parameter, the lowest value in the population is found and then this value (+1%) is added to this parameter of the entire population, ensuring all the values are greater than 0. Then, all the values are normalised by dividing them with the mean value of a given parameter within the population. Each parameter can then be multiplied by a constant or raised to some power. The final fitness is the sum of all the parameters raised to the power of -1:

$$f(x) = (aE_{binding}^b + cD_{complex}^d + eA_{complex}^f + gST_{complex}^h + iD_{cage}^j + kA_{cage}^l + mST_{cage}^n)^{-1} \quad (6.2)$$

### 6.2.3 Fitness function parameterisation

The fitness function parameterisation was first performed on the database of all assembled cages and their complexes. The geometry optimisation process is the bottleneck of the EA calculations in this work and the database of pre-assembled and geometry optimised cages resulting from all combinations of precursors allowed for a quick screening of a range of constants and powers for the fitness function to find the right parameters. The cage cavity and the complex cavity were excluded from the fitness function. Therefore, the  $c$ ,  $d$ ,  $i$ ,  $j$  parameters of the fitness function (eq. 6.2) were set to 0. As the result of the analysis of parameters for the entire set of assembled cages, the strain parameter was not found to be an indicative feature of the cage or complex quality. Therefore, the  $g$ ,  $h$ ,  $m$ ,  $n$  parameters of the fitness function were also set to 0. This effectively turns the feature off and it does not have an effect on the fitness of the member, nevertheless these parameters are still determined for each cage and will therefore be presented and discussed in the following results section. To further reduce the degrees of freedom for the parameterisation process, only the asymmetry of the cage in the complex is considered, therefore  $k$  and  $l$  were set to zero. The final fitness function for the fitness function parameterisation step takes the form of:

$$f(x) = (aE_{binding}^b + eA_{complex}^f)^{-1} \quad (6.3)$$

The  $a$  and  $e$  constants were screened for values between 1 to 5, in increments of 1, for all combinations. In contrast, the  $b$  and  $f$  exponents for all combinations of values in the range of

0 to 5 were considered in increments of 0.25.

For each set of  $a$ ,  $b$ ,  $f$  and  $e$  parameters, the fitness was calculated with the fitness function in equation 6.3 for each member of the population and then the population was sorted in ascending fitness value order. In the EA calculation the fitness function is being minimised, so a set of 10 candidates with the lowest fitness value was taken for each parameter set up and the solutions rated. A total rating of the top 10 candidates ( $R_{10}$ ) was calculated with a new equation based on the sum of unscaled properties for  $E_{binding}$  and  $A_{complex}$  for each member ( $i$ ) of the set:

$$R_{10} = \frac{R(E_{binding})_i + R(A_{complex})_i}{2} \quad for \quad i = 1 - 10 \quad (6.4)$$

where  $R(E_{binding})_i$  for  $i^{th}$  complex was calculated using the following formula:

$$R(E_{binding})_i = \begin{cases} 1, & \text{if } E_{binding} > 0 \text{ kJ mol}^{-1} \\ 1, & \text{if } E_{binding} < -404 \text{ kJ mol}^{-1} \\ 1 - \frac{E_{binding}}{-404 \text{ kJ mol}^{-1}}, & \text{otherwise} \end{cases} \quad (6.5)$$

A positive binding energy, *i.e.* a lack of binding affinity is penalised by adding 1 to the  $R_{10}$  value. At the same time, binding energies lower than  $-404 \text{ kJ mol}^{-1}$  also result in a +1 penalty. The reason for this is explained in the analysis of the results of the assembled cages and relates to the fact that some binding energies are unreasonable. The strongest binding energy among the assembled complexes, with the exception of the unreasonable values, was calculated to be  $\sim -404 \text{ kJ mol}^{-1}$ . Therefore, the complexes with  $E_{binding}$  between  $-404$  and  $0 \text{ kJ mol}^{-1}$  are assigned a value from 0 to 1, depending on how strong the binding affinity is, resulting in a decreasing penalty for binding energies up to  $-404 \text{ kJ mol}^{-1}$ .

The  $R(A_{complex})_i$  is calculated with the following formula:

$$R(A_{complex})_i = \frac{A_{complex}}{11.864 \text{ \AA}} \quad (6.6)$$

where the lower the asymmetry of the cage in the complex, the lower the penalty. The asymmetry parameter is here treated as a proxy for the synthesizability. This is in light of the entropy

of symmetry that might play a role in favouring structures with a higher symmetry.<sup>263</sup> The value of 11.864 is the highest asymmetry in the whole database of 3870 cages. The larger the  $R_{10}$  value, the lower the quality of the solutions. The results were then used to generate 2D heat maps (separate for each set of  $a$  and  $e$  magnitudes) that show the  $R_{10}$  value with respect to the  $b$  and  $f$  exponents. This allowed me to find the fitness function parameters  $a$ ,  $e$ ,  $b$ ,  $f$  that yield the set of ten best candidates out of the population in the most effective way.

#### 6.2.4 Evolutionary algorithm calculations

The steps described until now were used to parametrise the fitness function of the EA algorithm to run the proof of concept calculations. Five separate EA calculations were performed on the database of precursors presented in Figures A.23 - A.26. The final goal is to find the optimal POC candidates for the  $C_{60}$  encapsulation application. The calculations are limited to the consideration of [4+6] imine cages with tetrahedral topology. The fitness function for the EA calculations was constructed from the  $C_{60}$  complex binding energy, asymmetry of a cage as extracted from the  $C_{60}$  complex and asymmetry of an empty cage. The latter was added as a proxy for the shape-persistency, which I associate with the likelihood of the molecule being formed in solution and of resulting in a porous material. Lack of shape-persistence can hinder the synthesis outcome or result in the cage's collapse upon material thermal activation. The values of these parameters are normalised in the same fashion as described earlier and the fitness function's final form is:

$$f(x) = (E_{binding}^{3.25} + 0.5A_{complex}^{4.25} + 0.5A_{cage}^{4.25})^{-1} \quad (6.7)$$

The parameterisation of the fitness function (equation 6.3) was performed only for the asymmetry of the complex. As the asymmetry of the cage is a parameter of the same type, both have been included as part of the fitness function in equation 6.7. However, to retain the appropriate binding energy weighting in relation to the asymmetry aspect, both asymmetry related parameters were given half weights (constants of 0.5), so that the sum of the constants equals the binding energy feature constant. Finally, each parameter is taken to the power determined

in the fitness function parameterisation.

The size of the population is set to 20 and each EA calculation was run for 100 generations. The selection function used to choose members for the next generation was a roulette wheel, where the probability of selecting a member is proportional to its fitness value. The evolutionary algorithm steps are as follows:

1. First the initial population of 20 diverse cages is generated. A random **bb** and **lk** is chosen and a cage and the corresponding complex generated. Next, using Morgan fingerprints,<sup>280</sup> as implemented in `RDKit`, a pair of **bb** and **lk** is chosen to be as different as possible from the previously chosen one. This process is iterated until 20 cages are generated. This way a set of random but also intentionally diverse cages is generated.
2. The crossover operation is then applied to a random pair of cages that result in the exchange of building blocks that result in two offspring molecules. The crossover was performed 7 times and results in 14 offspring.
3. The mutation operation is applied 10 times within the population. Elitism was applied, meaning the cage with highest fitness in the given generation always underwent mutation. The remaining 9 candidates were chosen using roulette wheel probability. A cage is chosen at random and then its' fitness is compared to a randomly generated value between 0 and 1. If the fitness of the candidate is greater than the randomly generated number, then the cage undergoes one of the mutation functions. This is repeated until 9 cages are mutated. There were four mutations applied with equal probability; exchange of the linker to a similar one, the exchange of the building block to a similar one, the exchange of the linker to a random one and the exchange of the building block to a random one. For the similar **lks** and **bbs** Morgan fingerprints were used to assess the similarity.
4. This results in a total of 44 cages, 20 coming from the current generation, 14 from crossover and 10 from mutation. From these, 20 are chosen using the roulette wheel, to create the next generation.
5. The whole process is repeated until the 100th generation is created.



### 6.2.5 Post-processing of the results

The five EA calculations resulted in five final populations of 20 candidates each. These were then combined into a single population and the duplicates were removed. The resulting population consisted of 53 unique members. The fitness of the members of the final population was re-evaluated with the fitness function from equation 6.7 and the fitness values sorted in ascending order. The top 20 candidates with the lowest fitness values were then considered final result.

## 6.3 Results

### 6.3.1 The database of precursors

The database of precursors, tritopic aldehydes (**bbs**) and ditopic amines (**lks**), presented in Figures A.23 - A.26 in the Appendix (pp. 195-198), was first analysed. The analysis included the total number of atoms, the molecular weight of molecules, the number of rotatable bonds and the number of aromatic rings. The results are presented in Figure 6.1. The database of precursors consists of a relatively small number of **lks**, with the total number of atoms and the molecular weights skewed towards smaller values in comparison with the **bbs**. This is expected for the ditopic amine **lks** that usually contain a small number of aromatic rings. On the other hand, the tritopic aldehyde **bbs** are more evenly distributed, from smaller ones of around 25 atoms to bigger ones with 80 atoms. Therefore, from the combination of these **bbs** and **lks**, a range of sizes of cages should be expected.

### 6.3.2 The database of assembled cages

The assembly of all combinations of the **bbs** and **lks** should result in 3870 distinct cages and their  $C_{60}$  complexes. However, the assembly process failed to form 7 cages resulting in 3863 cages formed. As the failure rate is very small in comparison with the successfully assembled cages, it was not investigated further. The calculated properties of the geometry optimised cages, and the corresponding complexes, are presented in a set of graphs in Figure 6.2. The pairs of cages and their complexes were divided into three groups: the complexes that have

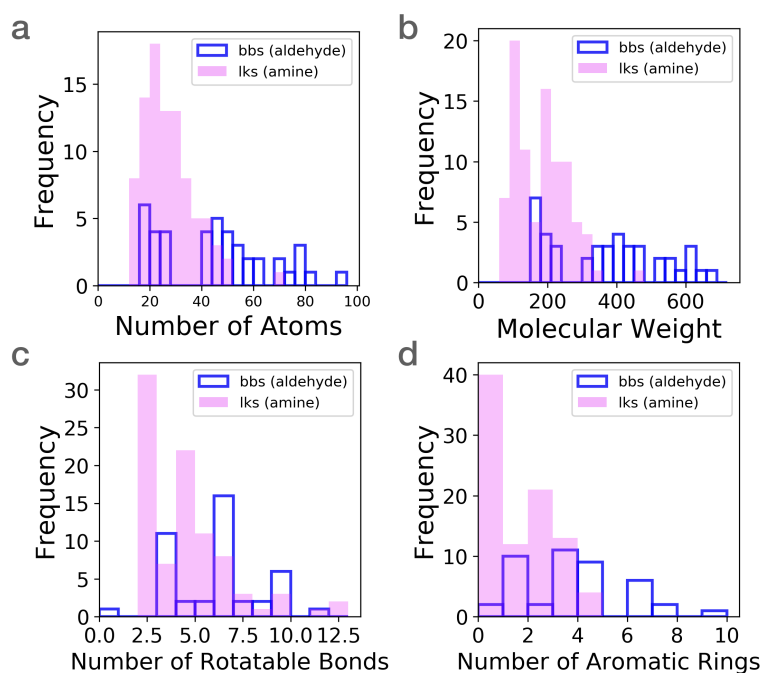


Figure 6.1: The analysis of the tritopic aldehyde **bbs** and ditopic amine **lks** from the database of precursors.

$C_{60}$  binding energies greater than  $0 \text{ kJ mol}^{-1}$  (repulsive interaction); ages with binding energies within the range of  $-404$  and  $0 \text{ kJ mol}^{-1}$ ; and the complexes with binding energies well below  $-404 \text{ kJ mol}^{-1}$ .

The last set, coloured red in the graphs, was notable as the binding energies seemed unreasonable. In Chapter 5, binding energies of  $-770$  and  $-606 \text{ kJ mol}^{-1}$  were reported for the multi-shell “hyperfullerene” complex ( $C_{60}@C_{240}$ ).<sup>261</sup> Thus, these values can be seen as a physical limit of the  $C_{60}$  interactions with a potential host. However, in the set of complexes coloured red, the binding energies are in the range of a few thousands of  $-\text{kJ mol}^{-1}$ . Additionally, these seem to aggregate around certain values and is observed for cages of larger cavities  $\sim 20 - 30 \text{ \AA}$  (the  $C_{60}$  diameter is  $\sim 10 \text{ \AA}$ ). These were later inspected and determined to have unreasonable geometries. Meaning, these cages were likely to collapse. But, their geometries seem to get stuck at as-assembled structures with an unusual orientation of hydrogens. It is believed this is a systematic error (as there is a similarity of binding energies values between groups of cages) and overall a faulty calculation. The reason these cages were not discarded is the fact that although this database is relatively small and these could be deleted manually, in future calculations on larger databases this would be impractical. Moreover, this work and the assembly

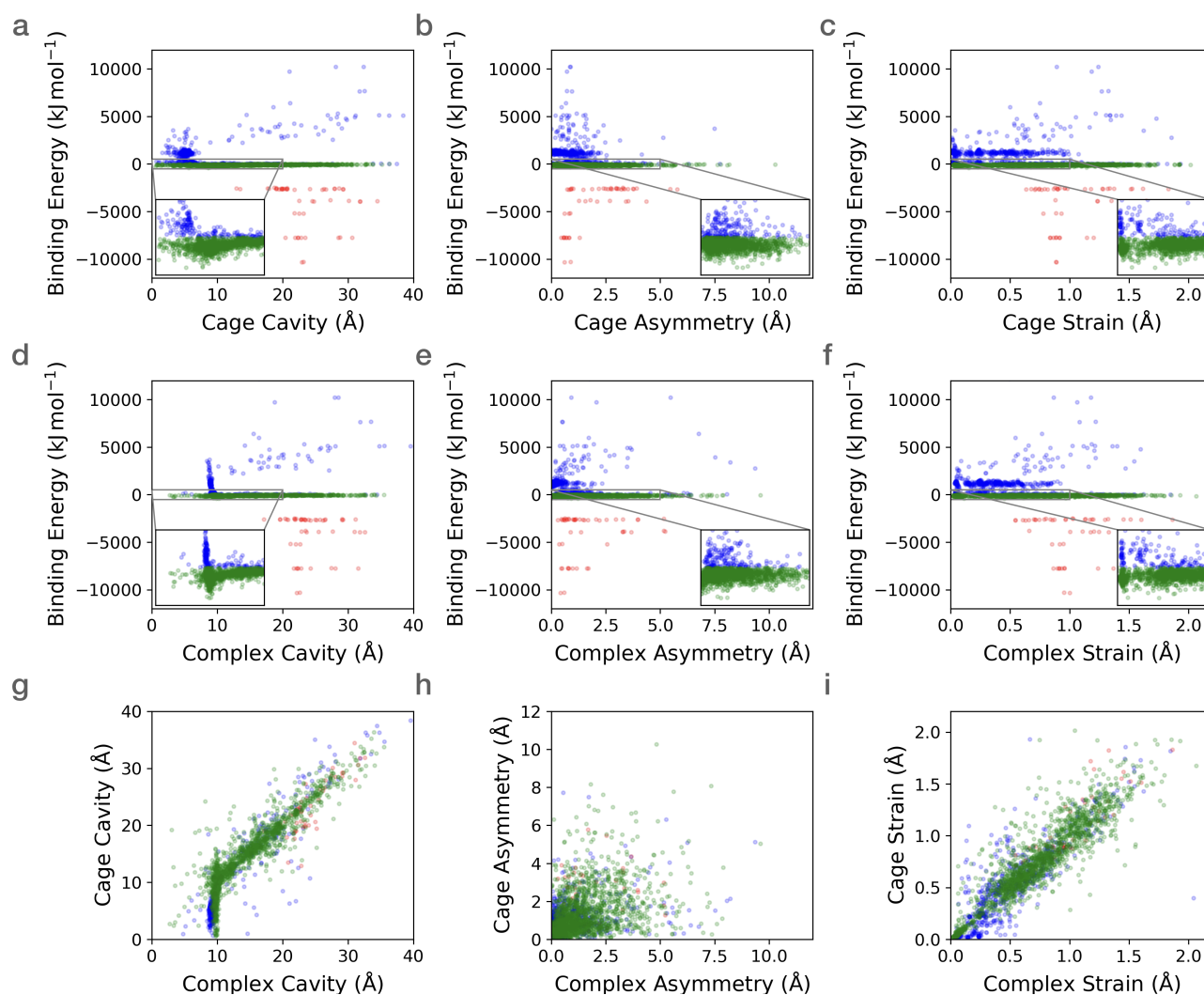


Figure 6.2: The analysis of the assembled database of cages and the corresponding  $C_{60}$  complexes. Various properties calculated as input for the fitness function include: the binding energy in the complexes, the cavity size of cage in complex, the asymmetry of cage in the complex, the strain of cage in the complex, the cavity size of an empty cage, the asymmetry of an empty cage and the strain of an empty cage. The complexes were divided into three sets. These with binding energies greater than  $0 \text{ kJ mol}^{-1}$  (coloured blue), the cages with binding energies within the range of  $-404$  and  $0 \text{ kJ mol}^{-1}$  (coloured green), and the complexes with binding energies well below  $-404 \text{ kJ mol}^{-1}$  (coloured red). On some of the plots sections are enlarged to show the important ranges in more detail.

of all combinations of cages is for the purpose of the fitness function parameterisation and EA validation. Thus, in the fitness function parameterisation process too low binding energies are penalised. The assumed cut-off is the mentioned value of  $-404 \text{ kJ mol}^{-1}$ . This is the strongest binding energy found for the complex within the range of values below that for  $C_{60}@C_{240}$ . Additionally, this value is separated by  $\sim 2000 \text{ kJ mol}^{-1}$  from the next cage that is assigned to the red set, which further supports the choice. As the EA calculations are meant to identify

the best solution for the fitness function, the parameterisation process in the next section will allow me to prevent these unreasonable complexes from propagating through the population.

The green set of complexes has binding energies between 0 and  $-404 \text{ kJ mol}^{-1}$ . These values are believed to be reasonable based on the previous work in Chapter 5 and literature values for various  $\text{C}_{60}$  complexes cited there. The trend in the graph with the binding energies correlated with the cavity diameters of cages in the complexes is especially interesting (Figure 6.2d). The blue and green sets have the cavity diameters shifted to values around  $10 \text{ \AA}$  in comparison to the values in binding energies correlated with cavity diameters of empty cages (Figure 6.2a) that are much more smaller and randomly distributed. This is an artefact of the the enforced complexation of  $\text{C}_{60}$  during the assembly, resulting in  $\text{C}_{60}$  present in cavities that are actually too small to host it. These cages are forced to expand to have larger cavities. This is also the reason for a large set of complexes with positive binding energies. The energy penalty of fitting  $\text{C}_{60}$  is far greater than the benefits of the  $\text{C}_{60}$  presence. Interestingly, this correlation is not observed in any of the plots regarding the strain values (Figures 6.2c,e,i). This can give an indication that the RMSD comparison of the building blocks in cages and complexes is not a good indication of the actual strain experienced in the molecules. In the blue set, the strain would be expected to be considerable and greater than in the green set. However, this is not observed. The green set is the target group of complexes for the fitness function, as these have reasonable binding energies in the desired range. The distribution of cages is diverse with many combinations of the properties. Therefore, the application of an EA to find the optimal candidates is justified.

### 6.3.3 The parameterisation of fitness function

Although the database initially consisted of 3863 cages, not all of the cages (or complexes) had their asymmetry determined. The application of `pywindow` to determine window diameters necessary to calculate the asymmetry for the collapsed cages will likely result in an error of the algorithm. During the fitness value calculation, if any of the parameters is missing, such cages are automatically discarded. This is acceptable as shape-persistent and symmetric molecules are targeted and the collapsed cages would be discarded eventually. Overall, the asymmetry

calculation failed for 801 empty cages, 245 complexes and in 385 pairs of both the cage and the complex (1432 in total). This results in 2432 cages that have all parameters calculated and will be used in the fitness function parameterisation. From these cages, 672 complexes have binding energies greater than  $0 \text{ kJ mol}^{-1}$ , 1709 complexes have their binding energy within the range of  $-404$  and  $0 \text{ kJ mol}^{-1}$ , and 51 complexes have binding energies well below the threshold of  $-404 \text{ kJ mol}^{-1}$ .

The parameterisation was performed on the fitness function from equation 6.3. The  $a$ ,  $e$  constants and  $b$ ,  $f$  exponents of the binding energy and asymmetry of the cage in the complex components were screened as described in the methods section. This required fitness values to be calculated for the 2432 cages and their complexes with the full set of parameters. The combination of constants and exponents resulted in a total of 9261 iterations - 21 different ratios of constants  $a$  and  $e$  and for each of these, a heat map was generated with 441 combinations of exponents  $b$  and  $f$ . For each  $a$ ,  $e$ ,  $b$ ,  $f$  parameter combination, the  $R_{10}$  score was calculated with the equation 6.4 for the set of ten cages with the highest fitness value. The  $R_{10}$  score gives the relative quality of the set of ten best cages in respect to other sets of parameters for the fitness function. The selected results are presented in Figure 6.3.

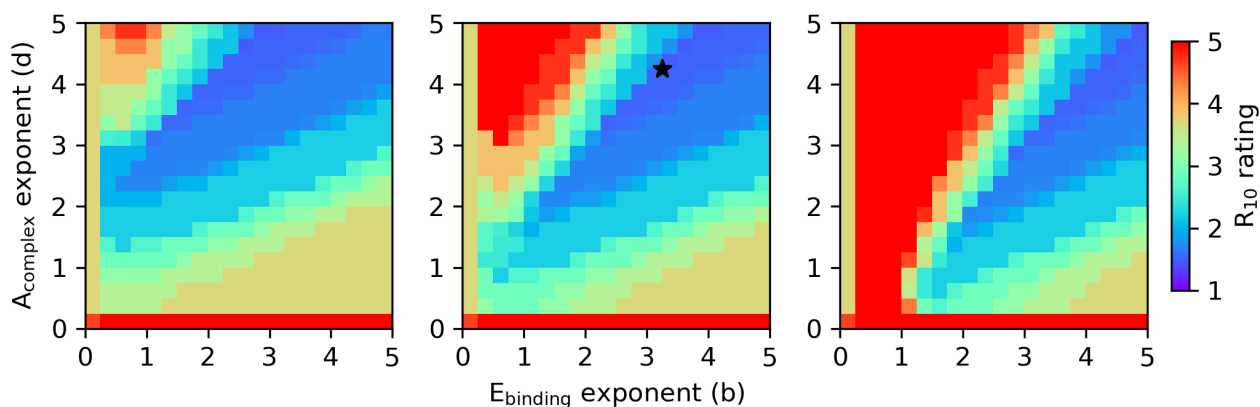


Figure 6.3: The heat maps of the fitness function parameterisation process. The plots from left to right are for the  $a$  and  $e$  constants of 5:1, 1:1 and 1:5, respectively. The  $x$  and  $y$  axes correspond to the screened  $b$  and  $f$  exponents. The heat map scale shows the  $R_{10}$  score for the ten cages with highest fitness. The lower the  $R_{10}$  score, the higher quality of the set. The  $a$ ,  $e$ ,  $b$ ,  $f$  parameter combination (1, 1, 3.25 and 4.25, respectively) that gave the lowest  $R_{10}$  score of 1.480 is marked with a star.

The results presented are for the set of constants  $a:e$  for 1:1 (middle graph), 5:1 (left hand side graph) and 1:5 (right hand side graph) ratios. This combination of  $a$  and  $e$  constants was

chosen to show the general trend observed for varied values. The heat maps are then generated representing the  $R_{10}$  score value for the combination of the  $b$ ,  $f$  exponents ( $x$  and  $y$  coordinates) for the given  $a:e$  set. The lowest  $R_{10}$  score corresponds to the optimal set of parameters. The lowest  $R_{10}$  value was 1.480 and was identified for 126 different set of parameters. The simplest set of parameters, where the sum of  $a$ ,  $e$ ,  $b$  and  $f$  was smallest, was then considered. The identified set of parameters was 1, 1, 3.25 and 4.25 for  $a$ ,  $e$ ,  $b$  and  $f$ , respectively. This calculation is marked with a black star on the middle plot in Figure 6.3.

### 6.3.4 The EA calculations

Five separate EA runs were then performed with the fitness function from equation 6.7 and the parameters from the previous section. The asymmetry of an empty cage was also added as one of the fitness parameters. This is justified as it is an equivalent type of parameter to the asymmetry of a cage in a complex. To keep the weighting of the binding energy to the asymmetry constant, in comparison with the parameterisation set up (1:1), the two asymmetry parameters were given weights of 0.5. In Figure 6.4, the evolution of the combined fitness of the given generation is plotted in respect to the 100 generations. The plots show the maximum fitness in the population, the lowest fitness and the mean fitness for the generation. All plots show that the mean fitness value quickly increases and then converges. The fitness value is with respect to a given population, therefore the scales on each of the graphs have different magnitude and can not be compared between separate EA calculations. In most cases, the convergence occurs relatively quickly after  $\sim 25$  generations. Therefore, the length of the calculations is sufficient and the optimal solutions were identified efficiently. Each generation consisted of 20 members, thus the five calculations resulted in a total of 100 candidates, although not unique.

### 6.3.5 Derived Cages

To rank the cages from all five EA calculations, the results were combined and reweighed with respect to the same fitness function from equation 6.7. The combined results consisted of 53 unique cages (duplicates were discarded) and their complexes. The 20 cages with the lowest fitness value were considered final results and are presented in Figure 6.5.

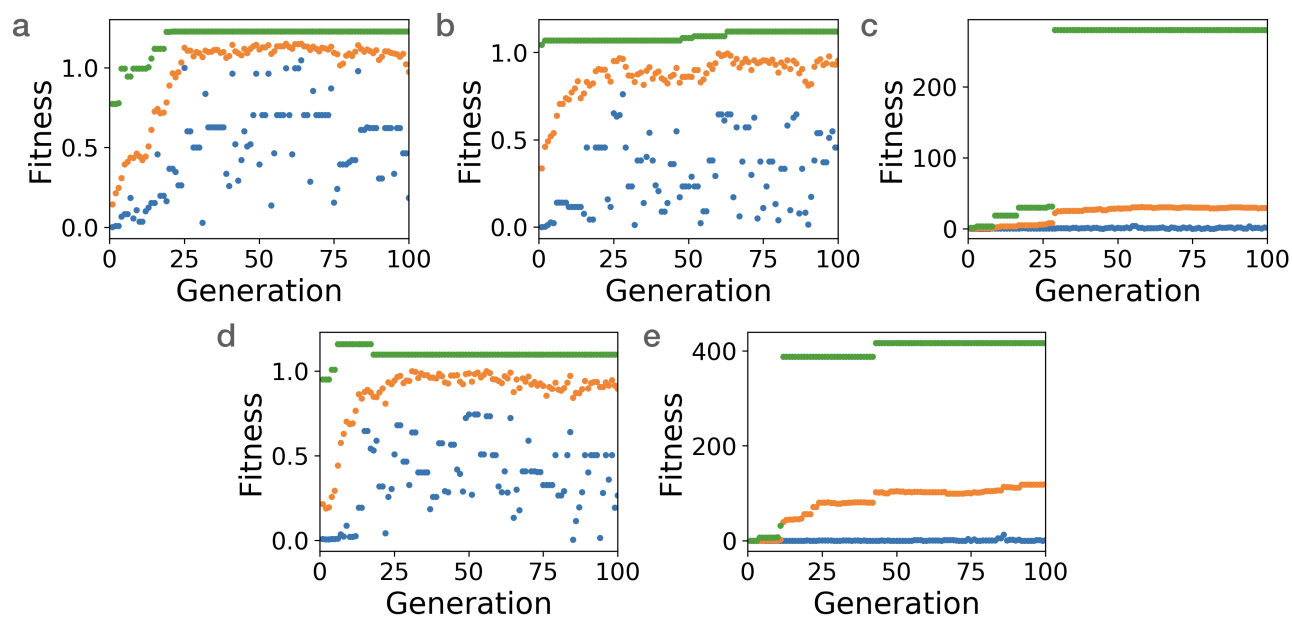


Figure 6.4: The evolution of the fitness values in the five EA calculations. The lowest fitness value in the (blue), the highest fitness value (green) in a given generation of 20 members are presented. The average fitness value for the current generation is in orange. The plots are for the duration of 100 generations.

In Table 6.1, the **bbs** and **lks** that the resulting cages were assembled from, with an index for easy reference to these in Figures A.23 - A.26, are listed. In addition, each EA calculation that identified a given cage is marked with a tick sign. The fact that many of these cages were identified multiple times shows the effectiveness of the constructed fitness function and that the screening of the databases is quick and broad. The evolution of the structures with the specifically parametrised fitness function allow the structures to converge to similar structures. The **bbs** in the resulting cages share similar features. They are planar, have a small number of rotatable bonds, and have a high number of aromatic rings. They are also very similar in size. The **bb38** has a spherical diameter of 16.1 Å, and the **bb15** and **bb13** have diameters of 16.8 and 17.0 Å, whereas the **bb17** and **bb16** are slightly smaller and have diameters of 14.1 and 14.4 Å, respectively. The **lk38** in **C1** is the most distinct from the set of linkers in the resulting cages, as the separation of the nitrogens between neighbouring imine bonds is 6.9 Å. All the other **lks** have the amine functionality on neighbouring carbons, resulting in the spacing between nitrogens in the close imine bond pairs in range of 3.0 - 3.4 Å. However, the larger linker in **C1** does not result in a larger cavity diameter in comparison with the rest of the cages (values reported in the next section).

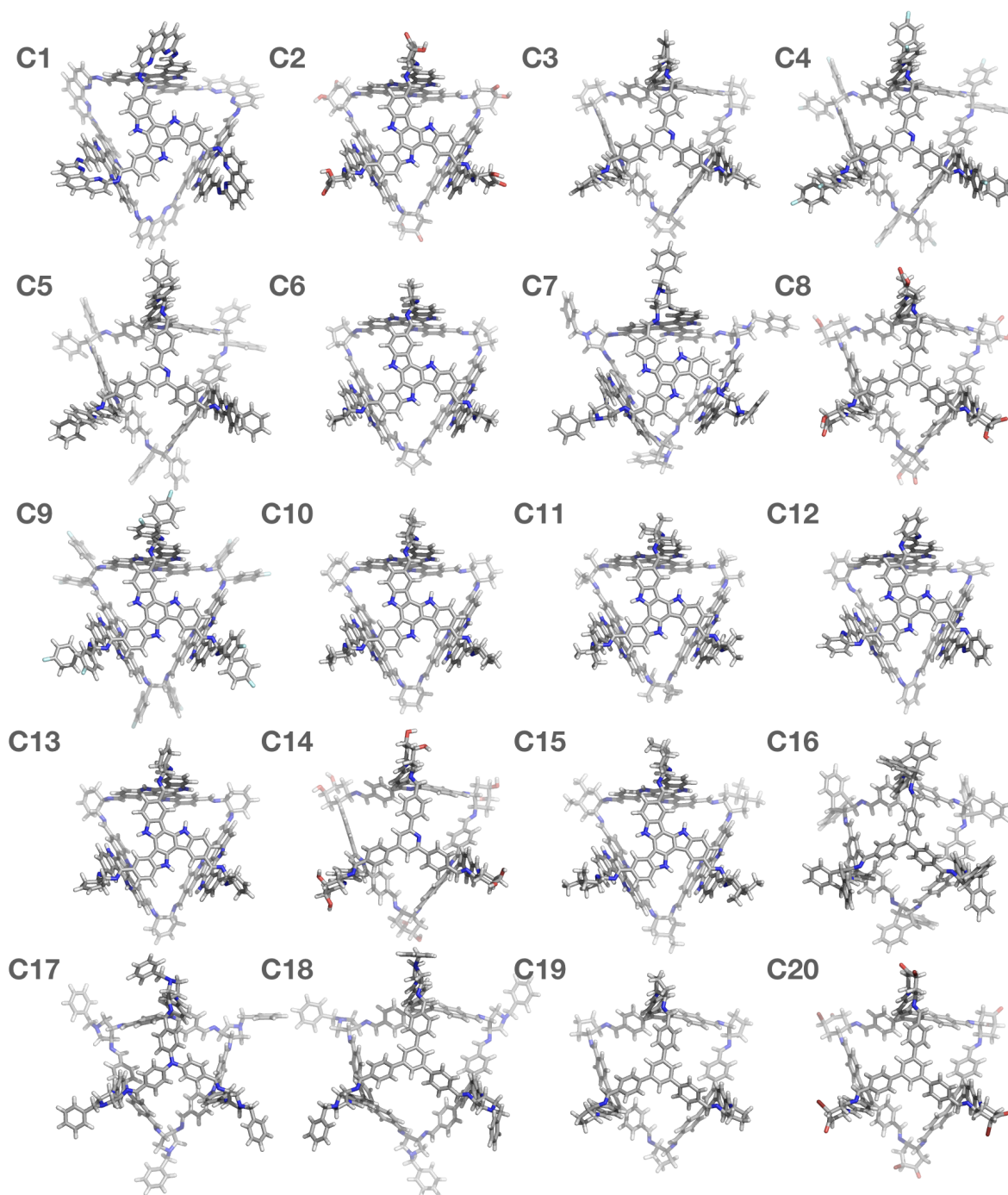


Figure 6.5: The twenty cages with highest fitness value from the combined results of the five EA calculations sorted in descending fitness value order from top left to bottom right. The geometry optimised structures of the empty cages are presented.

In Table 6.2, the re-scaled fitness value for the combined results of the five EA calculations and the unscaled parameters are presented. The cages have relatively high magnitude of binding energies, between  $-160$  and  $-270$   $\text{kJ mol}^{-1}$ . The asymmetry for both the empty cage and the cage in the complex are also in the lower range of values present in the database. These parameters were part of the fitness function. However, all cages seem naturally converged to structures



Table 6.1: The aldehyde **bbs** and amine **lks** used to assembly the final population of cages. The **bbs** and **lks** index numbers correspond to Figures A.23 - A.26 in the Appendix. The specific EA calculations in which each cage was a member of the 100<sup>th</sup> generation are marked with a tick sign.

Cage	bb	lk	EA calculation				
			1	2	3	4	5
<b>C1</b>	38	38	✓				✓
<b>C2</b>	38	19	✓	✓			✓
<b>C3</b>	15	31			✓	✓	
<b>C4</b>	15	62	✓				
<b>C5</b>	15	61			✓		
<b>C6</b>	38	24	✓	✓	✓	✓	✓
<b>C7</b>	38	33		✓	✓	✓	✓
<b>C8</b>	13	19	✓		✓		✓
<b>C9</b>	38	62	✓				✓
<b>C10</b>	38	25	✓	✓	✓	✓	✓
<b>C11</b>	38	16	✓		✓	✓	✓
<b>C12</b>	38	58	✓		✓	✓	✓
<b>C13</b>	38	31		✓	✓	✓	
<b>C14</b>	15	20			✓	✓	
<b>C15</b>	38	17	✓	✓	✓	✓	✓
<b>C16</b>	17	72		✓			
<b>C17</b>	16	33		✓			
<b>C18</b>	13	32		✓			✓
<b>C19</b>	13	24					✓
<b>C20</b>	13	22			✓		

with cavities close in size to the C<sub>60</sub> diameter ( $\sim 10$  Å), between 9.7 and 10.6 Å. But, the cavity diameter of an empty cage and cage in the complex were not part of the fitness function. This shows how the binding energy was a good choice for a parameter that would also affect other features such as the cavity size. In Figure 6.2, the distribution of similar binding energies is observed for much broader set of cavity sizes. Possibly, the binding energy and asymmetry also allowed to cavities to converge to the optimal solutions as a result.

The results are especially promising, as some of the building blocks that repeat in optimal solutions were used in the past to synthesise cages. Ding *et al.* in 2015 synthesised a [4+6] triazine cage with cyclohexane.<sup>86</sup> The triazine building block is similar to the **bb15** in cages **C3**, **C4**, **C5**, **C14** that contains single nitrogen substitution in the central heteroatomic benzene ring in comparison with three in triazine. The **bb16** in cage **C17** was previously used to synthesise

Table 6.2: The fitness values and the unscaled properties of the final cages.

Cage	$f(x)$	$E_{binding}$	$D_{complex}$	$A_{complex}$	$ST_{complex}$	$D_{cage}$	$A_{cage}$	$ST_{cage}$
<b>C1</b>	15.732	-265.5	10.4	0.034	0.09	11.6	0.132	0.07
<b>C2</b>	14.047	-268.7	10.2	0.046	0.33	10.3	0.142	0.38
<b>C3</b>	4.136	-227.6	10.2	0.059	0.47	10.5	0.071	0.41
<b>C4</b>	3.683	-223.6	10.2	0.018	0.73	10.5	0.057	0.62
<b>C5</b>	3.180	-219.6	10.2	0.021	0.84	10.4	0.051	0.85
<b>C6</b>	2.973	-217.7	10.3	0.011	0.23	11.2	0.022	0.22
<b>C7</b>	2.435	-223.7	10.1	0.049	0.39	9.9	0.184	0.47
<b>C8</b>	1.417	-194.8	10.4	0.031	0.76	10.8	0.094	0.74
<b>C9</b>	1.361	-193.0	10.1	0.025	0.42	10.0	0.034	0.24
<b>C10</b>	1.211	-188.7	10.1	0.026	0.24	10.2	0.023	0.24
<b>C11</b>	1.103	-185.2	10.1	0.023	0.07	10.2	0.032	0.07
<b>C12</b>	1.074	-204.2	9.9	0.082	0.04	9.3	0.233	0.04
<b>C13</b>	1.042	-190.7	10.1	0.116	0.08	10.0	0.133	0.09
<b>C14</b>	1.025	-218.5	10.2	0.050	0.87	10.7	0.265	0.91
<b>C15</b>	0.894	-200.5	10.1	0.155	0.36	10.2	0.139	0.38
<b>C16</b>	0.664	-164.7	9.9	0.064	0.53	9.4	0.036	0.56
<b>C17</b>	0.640	-169.4	9.7	0.128	0.74	8.9	0.059	0.53
<b>C18</b>	0.634	-172.9	10.6	0.040	0.96	11.1	0.227	1.28
<b>C19</b>	0.630	-163.6	10.6	0.093	0.68	11.5	0.039	0.68
<b>C20</b>	0.628	-195.5	10.4	0.179	0.58	10.8	0.032	0.71

a [4+6] **CC5** cage with cyclopentane.<sup>226</sup> Here, the **lk33** is a derivative of a cyclopentane. Most recently, a truxene building block, structurally similar to **bb38** in **C1** and nine other cages, was used to synthesise a [4+6] cage with ethylenediamine.<sup>281</sup> The results were compared to the whole database in Figure 6.6. The identified solutions are highly localised, especially for the features that were part of the fitness function. This is somewhat equivalent to finding the global minimum on the chemical hyperspace.

We can summarise that the optimal cavity size for  $C_{60}$  encapsulation is in the range of 9.3 - 11.6 Å. The cages studied in the previous chapter were of similar dimensions. Good candidates would also be highly symmetric and shape-persistent, meaning they do not collapse upon guest or solvent molecules evacuation, with relatively rigid building blocks. The considered tritopic aldehyde **bbs** should be roughly  $\sim 16$  Å in diameter and the ditopic amine **lks** should have the amine functionality on neighbouring carbon atoms. Many of these could be simplified towards alternatives that were successfully used to synthesise cages in the past. However, the best

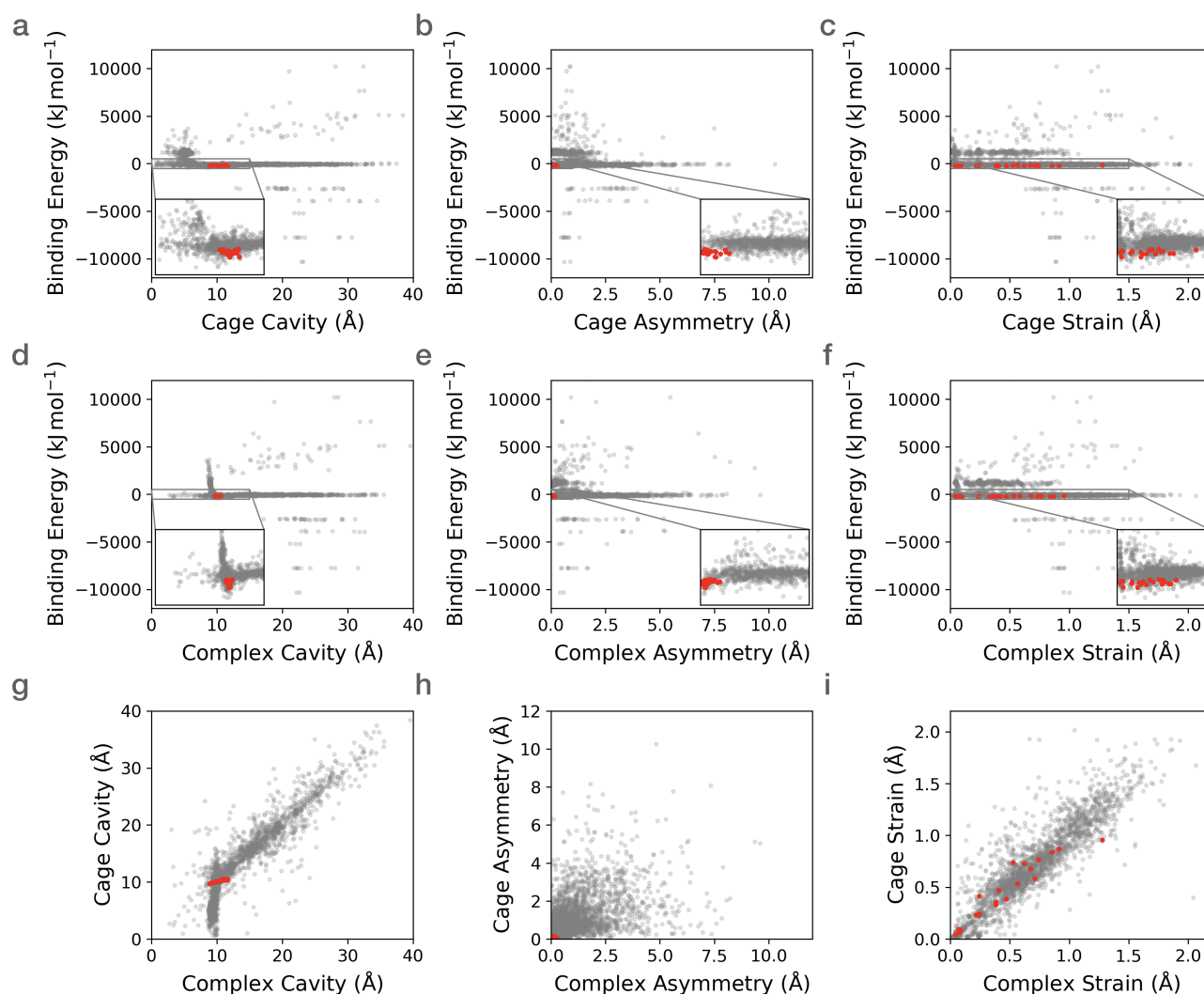


Figure 6.6: The features of the final cages (coloured red) in comparison to the features of the database of pre-screened cages (coloured grey).

molecule, **C1**, has the amine functionality in greater separation (not on neighbouring carbon atoms), thus larger **lks** should also be considered. To my knowledge there are currently no studies in the literature of  $\text{C}_{60}$  encapsulation in [4+6] imine cages. However, the experimental examples of cages, listed in the previous section, are structurally similar to the presented set of cages here.

## 6.4 Discussion

In this proof-of-concept work, all combinations of **bbs** and **lks** were assembled into organic cages. This was possible as a result of a relatively small number of considered precursors. The taken steps were necessary to first understand the different types of resulting  $\text{C}_{60}$ @POC com-

plexes and identify the key features for the construction of the fitness function. The assembly of a full database allowed me to parametrise the behaviour of the fitness function and to follow expected behaviour, such as excluding candidates with unreasonable geometries and binding energies. These steps can be considered when determining if the whole database was screened with reasonable computational resources for future EA calculations. However, if we consider a hypothetical database of 30000 **lks** and 10000 **bbs**, and if we extend the possible topologies ( $[2+3]$ ,  $[4+6]$ ,  $[8+12]$ ), the resulting combination of all possibilities would reach 900 million. Therefore, this work is a proof-of-concept that the EA calculations and tailored fitness function can identify optimal solutions for applications. From the presented results, the following summary of steps taken and methodology established should be considered in future work:

- From a large database of precursors a smaller subset should be extracted that will result in few thousands combinations of diverse solutions and will allow the analysis of the database of candidates and the fitness function parameterisation. The parametrised fitness function should be then used in the EA calculations including all precursors. In this work the initial database of  $\sim 4000$  solutions provided  $\sim 2500$  candidates for fitness function parameterisation with especially promising results. However, the size of the potential training database size might vary based on the number of precursors, their diversity and problem studied.
- The parameterisation of the fitness function should include a small number of parameters to allow easy visualisation of the results and correlation of the resulting solutions of the fitness function to these parameters. In this work, two parameters, the binding energy and asymmetry of a cage in a complex, were used. It was necessary to find a total of four related parameters, the constants and exponents that can control the behaviour of the fitness function. After parameterisation, parameters of equivalent type can be added to the fitness function, as was done in this work for the asymmetry of an empty cage.
- Many short EA calculations should be prioritised over a few long calculations. This is especially true for larger databases of precursors to efficiently sample relatively distant points on the chemical hyperspace. Additionally, exit functions can be used, rather than

a constant length of calculations, such as the termination of the calculation when convergence criteria are satisfied.

- In this work, the relative orientation of the guest in the cage cavity was not considered and was initialised as random. This is in continuation of the conclusions drawn from the previous Chapter 5. However, this is only possible for highly symmetrical guest molecules. Thus, this should be considered for guest molecules of low symmetry.

## 6.5 Conclusions

In this chapter, I have showed a function-led computational design of new potential  $C_{60}$  fullerene encapsulants with a novel evolutionary algorithm methodology. The whole process from the choice of the database of precursors, the assembly of the database for parameterisation process, fitness function construction and the analysis of the results in respect to the pre-screened candidates provided insights into each of these steps. The presented methodology can be used in place of experimental serendipitous discovery, or for the opposite, to facilitate and improve rational design of new functional materials by providing insightful feedback. The EA algorithm and the constructed fitness function were found to efficiently identify promising candidates for the experimental consideration to find new  $C_{60}$  fullerene encapsulants.

Design principles can be formulated from the results. The aldehyde building blocks should be fairly planar with a circular diameter in range of 14.1 - 17.0 Å. The amine linkers with the amine functional groups on the neighbouring carbons result in the most promising cages. However, larger linkers such as the one in **C1** should also be considered. In all cases the building blocks have a small number of rotatable bonds, and high number of aromatic rings. The combination of the building block and linker should result in a cavity size of approximately 10 Å in diameter. Overall, a good cage for  $C_{60}$  encapsulation application should be highly symmetric and shape-persistent. The results have been shared with our experimental collaborators and await synthetic realisation. We are currently considering combining other properties such as the shape of the cavity or the window cross-section to target guest size and shape complementarity. This could possibly allow us to focus on the possibility of guest diffusion.

State-of-the art materials design studies mostly utilise large databases of synthetically realised precursors. However, in the future, it could be possible to replace databases of precursors with algorithms that allow the **bbs** and **lks** to evolve in line with the evolution of the fitness function solution. However, the chemical stability, synthesisability and toxicity of the precursors would necessarily become a part of the fitness function to introduce a positive bias towards structures that are physically realisable. This approach would require resources of unprecedented scale, although it seems achievable in the foreseeable future with ever growing computational power.

# Chapter 7

## Summary and Perspectives

The goal of this thesis was to present advances in the screening of porous organic molecules for useful properties. Developed software and methodologies were used in combination with established computational chemistry techniques. Both the experimentally reported porous organic cages and belt-like molecules, *i.e.* molecular pores, as well as computationally generated hypothetical porous organic cages, were a subject of this thesis. I have identified new promising materials for Xe/Kr separation application and possible C<sub>60</sub> encapsulants. I have shown that the properties of single molecules can be used as a proxy to the bulk and in-solution properties. This approach will allow one to study the diffusion of gases and molecules in porous materials or the application of molecular pores in-solution at a fraction of the cost of the currently applied methodologies for the solid state.

The developed `pywindow` software was initially an attempt to find a more robust and transferable methodology to calculate window diameters than the circumcircle method that had previously been reported. The circumcircle method required user input, thus it was not a viable option for this application. As a result, an automated way of calculating window diameters was developed. Additionally, this software was extended to calculate a variety of structural properties and to study single molecules in the bulk from molecular dynamics trajectories. The latter was possible by introducing a custom unit cell rebuilding algorithm that reconstructs molecules passing through the periodic boundary allowing one to analyse them for structural properties. The `pywindow` software was then implemented in the evolutionary algorithm developed in our group. It allowed quick structural property screening. This would not be possible

without a reliable and automated tool to quickly calculate window diameters to use in the assessment of cage symmetry. The other software examples for the structural analysis of porous materials, I discussed in Chapter 2, work only with periodic systems. Therefore, `pywindow` proved to be a robust tool that allows further advances in computational materials design. In the future, I plan to extend `pywindow` functionality to determine further structural properties of molecular pores. The windows can be analysed not only for their diameters but also for their shape. This should allow one to study the complementarity of the windows with the possible size and shape of guests and to predict the diffusion mechanisms. Similarly, the intrinsic cavities can be analysed for their 3D shape and the total volume, rather than spherical approximation. This should result in a better correlation of the pore size and shape with the guest's binding affinities. In the near future, shape descriptors such as asphericity, acylindricity and relative shape anisotropy for molecular pores will be implemented as well.

In the screening of molecular pores for Xe/Kr separation, I have identified an existing material that was not previously considered for this application. The experimental Xe/Kr selectivity of Noria is comparable with the best performing porous materials, such as the porous organic cage **CC3**, as determined by our collaborators in the University of Liverpool. During the course of that study, another group also suggested Noria for Xe/Kr separation, supporting our findings. I have shown how the structural information, the binding energies, the window size fluctuations and the barriers for Xe and Kr diffusion, calculated for single molecules, can be used to identify materials with excellent Xe/Kr separation performance in the bulk. This methodology can be applied in the future to screen existing molecular pores for other properties and applications.

The following study of favourable C<sub>60</sub> fullerene encapsulants had a twofold outcome. Firstly, the studied porous organic cages have been successfully realised synthetically and had big enough cavities for C<sub>60</sub> encapsulation. The templating effect the C<sub>60</sub> has on cage formation has been shown experimentally, but only one of the cages has formed a possible C<sub>60</sub> complex. I used the pore limiting envelope analysis to propose the likely route of the C<sub>60</sub> into the pore. From the pore limiting envelopes, the diffusion through windows of already formed cage seems unlikely and an alternative mechanism of the diffusion of C<sub>60</sub> into the pore during cage formation was



proposed instead. However, to support these claims more experimental testing is necessary. Secondly, I analysed the binding energy changes with respect to the relative host-guest orientation in the  $C_{60}$  complex. As the binding energy changes were relatively small, my conclusion was that the relative host-guest orientation did not need to be considered as an important factor in the the following chapter. This significantly reduced computational costs and thus the time necessary to perform the evolutionary algorithm calculations was shortened to acceptable lengths. Also, the DFT and force field geometry optimisations were compared. The latter shows a higher freedom of guest rotation in the cage cavity, increasing the likelihood of finding local minima than with DFT. Although the OPLS3 force field was shown to mostly disagree with the DFT derived binding energies for the particular complexes, the magnitudes of binding energies in both methods are comparable. Thus, this has justified the use of force field binding energies in the following study, that was required due to the computational expense of DFT. The analysis of the binding energies, both the ones reported in the literature and determined by myself for the studied  $C_{60}$  complex, allowed me to identify the physical limit to the binding energy interaction in these complexes. This knowledge was used in the function-led evolutionary algorithm-assisted design of porous organic cages to identify unreasonable complexes.

In the last chapter I showed the development of a novel approach to design molecules for applications. From the databases of precursors, a set of hypothetical porous organic cages was developed with the desired features of  $C_{60}$  encapsulants. The STK software developed in our group allowed for evolutionary algorithm-assisted function-led development of novel cages. I have developed a methodology to parameterise the constructed fitness function. As a result, 20 promising cages were identified out of the combination of building blocks and linkers. The tailored fitness function prevented the complexes with unreasonable magnitudes of binding energies to enter the final population. The proposed cages were structurally very similar. This proves that this approach can successfully identify good solutions from a diverse set. Based on the analysis of the building blocks and linkers, the design principles that can be used in future rational synthesis of  $C_{60}$  encapsulants were formulated. Several of these cages are structurally similar to existing cages in the literature. Although this work was performed on a small database, this approach can be easily extended to much larger databases where the

use of the evolutionary algorithm is especially justified. The developed methodology for the fitness function parameterisation can be used to construct tailored fitness functions for other applications. In the future, the parameterisation of the fitness function could be automated with machine learning approaches. From the database of precursors, a small set was selected to test and fully tailor the fitness function. Furthermore, the precursors could also be optimised during the evolutionary algorithm to find the ones that are best suited for the application, but missing from the database. I believe this work has shown how a selection of the computational methods available and the custom developed software in our group, `pywindow` by myself and `STK` in collaboration, can advance the design of novel porous molecular materials.

## Bibliography

- [1] M. Thommes, K. Kaneko, A. V. Neimark, J. P. Olivier, F. Rodriguez-Reinoso, J. Rouquerol and K. S. W. Sing, *Pure Appl. Chem.*, 2015, **87**, 1051–1069.
- [2] F. Schuth, K. S. W. Sing and J. Weitkamp, *Handbook of Porous Solids*, Wiley VCH, Heidelberg, 2002.
- [3] C. S. Cundy and P. A. Cox, *Micropor. Mesopor. Mat.*, 2005, **82**, 1–78.
- [4] C. S. Cundy and P. A. Cox, *Chem. Rev.*, 2003, **103**, 663–701.
- [5] D. S. Sholl and R. P. Lively, *Nature*, 2016, **532**, 435–437.
- [6] O. M. Yaghi, M. O’Keeffe, N. W. Ockwig, H. K. Chae, M. Eddaoudi and J. Kim, *Nature*, 2003, **423**, 705–714.
- [7] X. Feng, X. Ding and D. Jiang, *Chem. Soc. Rev.*, 2012, **41**, 6010–6022.
- [8] J.-P. Zhang and X.-M. Chen, *Chem. Commun.*, 2006, **0**, 1689–1699.
- [9] S. Kitagawa, *Acc. Chem. Res.*, 2017, **50**, 514–516.
- [10] R. Banerjee, H. Furukawa, D. Britt, C. Knobler, M. O’Keeffe and O. M. Yaghi, *J. Am. Chem. Soc.*, 2009, **131**, 3875–3877.
- [11] J.-R. Li, R. J. Kuppler and H.-C. Zhou, *Chem. Soc. Rev.*, 2009, **38**, 1477–1504.
- [12] H. Furukawa, K. E. Cordova, M. O’Keeffe and O. M. Yaghi, *Science*, 2013, **341**, 1230444.
- [13] N. Notman, *MOFs find a use*, *Chemistry World*, 2017.
- [14] *Oak Ridge National Laboratory*, <http://www.ornl.gov>.
- [15] T. Carey, A. Corma, F. Rey, C. C. Tang, J. A. Hriljac and P. A. Anderson, *Chem. Commun.*, 2012, **48**, 5829–5831.
- [16] A. P. Côté, A. I. Benin, N. W. Ockwig, M. O’Keeffe, A. J. Matzger and O. M. Yaghi, *Science*, 2005, **310**, 1166–1170.

- [17] N. Lock, Y. Wu, M. Christensen, L. J. Cameron, V. K. Peterson, A. J. Bridgeman, C. J. Kepert and B. B. Iversen, *J. Phys. Chem. C*, 2010, **114**, 16181–16186.
- [18] K. S. Park, Z. Ni, A. P. Côté, J. Y. Choi, R. Huang, F. J. Uribe-Romo, H. K. Chae, M. O’Keeffe and O. M. Yaghi, *Proc. Natl. Acad. Sci. U. S. A.*, 2006, **103**, 10186–10191.
- [19] 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. Bacsá, A. M. Z. Slawin, A. Steiner and A. I. Cooper, *Nat. Mater.*, 2009, **8**, 973–978.
- [20] J. D. Evans, K. E. Jelfs, G. M. Day and C. J. Doonan, *Chem. Soc. Rev.*, 2017, **46**, 3286–3301.
- [21] S. A. Allison and R. M. Barrer, *Inorg. Phys. Theor.*, 1969, **0**, 1717–1723.
- [22] R. M. Barrer and V. H. Shanson, *J. Chem. Soc., Chem. Commun.*, 1976, **0**, 333–334.
- [23] D. D. MacNicol and F. B. Wilson, *J. Chem. Soc. D*, 1971, **0**, 786–787.
- [24] J. Tian, P. K. Thallapally and B. P. McGrail, *CrystEngComm*, 2012, **14**, 1909–1919.
- [25] T. Hasell and A. I. Cooper, *Nat. Rev. Mater.*, 2016, **1**, 16053.
- [26] C. R. Groom, I. J. Bruno, M. P. Lightfoot and S. C. Ward, *Acta Crystallogr. Sect. B*, 2016, **72**, 171–179.
- [27] J. D. Evans, D. M. Huang, M. Haranczyk, A. W. Thornton, C. J. Sumby and C. J. Doonan, *CrystEngComm*, 2016, **18**, 4133–4141.
- [28] P. Z. Moghadam, A. Li, S. B. Wiggin, A. Tao, A. G. P. Maloney, P. A. Wood, S. C. Ward and D. Fairen-Jimenez, *Chem. Mater.*, 2017, **29**, 2618–2625.
- [29] T. M. Long and T. M. Swager, *Adv. Mater.*, 2001, **13**, 601–604.
- [30] M. Mastalerz, *Chem. Eur. J.*, 2012, **18**, 10082–10091.
- [31] M. Mastalerz and I. M. Oppel, *Angew. Chem. Int. Ed.*, 2012, **51**, 5252–5255.

- [32] V. Böhmer, *Angew. Chem. Int. Ed.*, 1995, **34**, 713–745.
- [33] J. L. Atwood, L. J. Barbour, A. Jerga and B. L. Schottel, *Science*, 2002, **298**, 1000–1002.
- [34] C. M. Kane, O. Ugono, L. J. Barbour and K. T. Holman, *Chem. Mater.*, 2015, **27**, 7337–7354.
- [35] W. A. Freeman, W. L. Mock and N. Y. Shih, *J. Am. Chem. Soc.*, 1981, **103**, 7367–7368.
- [36] E. Masson, X. Ling, R. Joseph, L. Kyeremeh-Mensah and X. Lu, *RSC Adv.*, 2012, **2**, 1213–1247.
- [37] G. Crini, *Chem. Rev.*, 2014, **114**, 10940–10975.
- [38] M. J. Bojdys, M. E. Briggs, J. T. A. Jones, D. J. Adams, S. Y. Chong, M. Schmidtman and A. I. Cooper, *J. Am. Chem. Soc.*, 2011, **133**, 16566–16571.
- [39] M. Mastalerz, M. W. Schneider, I. M. Oppel and O. Presly, *Angew. Chem. Int. Ed.*, 2011, **50**, 1046–1051.
- [40] T. Hasell, J. L. Culshaw, S. Y. Chong, M. Schmidtman, M. A. Little, K. E. Jelfs, E. O. Pyzer-knapp, H. Shepherd, D. J. Adams, G. M. Day and A. I. Cooper, *J. Am. Chem. Soc.*, 2014, **136**, 1438–1448.
- [41] T. Guo, A. Bedane, Y. Pan, B. Shirani, H. Xiao and M. Eić, *Energy Fuels*, 2017, **31**, 4186–4192.
- [42] T.-x. Guo, A. H. Bedane, Y. Pan, H. Xiao and M. Eić, *Mater. Lett.*, 2017, **189**, 114–117.
- [43] E. Christoforides, D. Mentzafos and K. Bethanis, *J. Incl. Phenom. Macrocycl. Chem.*, 2015, **81**, 193–203.
- [44] D. Bardelang, K. A. Udachin, D. M. Leek, J. C. Margeson, G. Chan, C. I. Ratcliffe and J. A. Ripmeester, *Cryst. Growth Des.*, 2011, **11**, 5598–5614.
- [45] J. Tian, P. K. Thallapally, S. J. Dalgarno, P. B. McGrail and J. L. Atwood, *Angew. Chem. Int. Ed.*, 2009, **48**, 5492–5495.

- [46] S. Lim, H. Kim, N. Selvapalam, K. J. Kim, S. J. Cho, G. Seo and K. Kim, *Angew. Chem. Int. Ed.*, 2008, **47**, 3352–3355.
- [47] H. Kudo, R. Hayashi, K. Mitani, T. Yokozawa, N. C. Kasuga and T. Nishikubo, *Angew. Chem. Int. Ed.*, 2006, **45**, 7948–7952.
- [48] H. Kudo, Y. Suyama, H. Oizumi, T. Itani and T. Nishikubo, *J. Mater. Chem.*, 2010, **20**, 4445–4450.
- [49] M. L. C. Quan and D. J. Cram, *J. Am. Chem. Soc.*, 1991, **113**, 2754–2755.
- [50] B. S. Park, C. B. Knobler, J. Eid, Clark N., R. Warmuth and D. J. Cram, *Chem. Commun.*, 1998, **113**, 55–56.
- [51] X. Liu, Y. Liu, G. Li and R. Warmuth, *Angew. Chem. Int. Ed.*, 2006, **45**, 901–904.
- [52] P. Skowronek and J. Gawroński, *Org. Lett.*, 2008, **10**, 4755–4758.
- [53] M. Mastalerz, *Chem. Commun.*, 2008, **0**, 4756–4758.
- [54] G. Zhang and M. Mastalerz, *Chem. Soc. Rev.*, 2014, **43**, 1934–1947.
- [55] J. D. Evans, C. J. Sumbly and C. J. Doonan, *Chem. Lett.*, 2015, **44**, 582–588.
- [56] Y. Jin, C. Yu, R. J. Denman and W. Zhang, *Chem. Soc. Rev.*, 2013, **42**, 6634–6654.
- [57] Y. Jin, Q. Wang, P. Taynton and W. Zhang, *Acc. Chem. Res.*, 2014, **47**, 1575–1586.
- [58] Y. C. Horng, T. L. Lin, C. Y. Tu, T. J. Sung, C. C. Hsieh, C. H. Hu, H. M. Lee and T. S. Kuo, *Eur. J. Org. Chem.*, 2009, 1511–1514.
- [59] A. Avellaneda, P. Valente, A. Burgun, J. D. Evans, A. W. Markwell-Heys, D. Rankine, D. J. Nielsen, M. R. Hill, C. J. Sumbly and C. J. Doonan, *Angew. Chem. Int. Ed.*, 2013, **52**, 3746–3749.
- [60] K. Matsui, Y. Segawa and K. Itami, *J. Am. Chem. Soc.*, 2014, **136**, 16452–16458.
- [61] J. D. Evans, D. M. Huang, M. R. Hill, C. J. Sumbly, A. W. Thornton and C. J. Doonan, *J. Phys. Chem. C*, 2014, **118**, 1523–1529.

- [62] H. Mao and S. Zhang, *J. Colloid Interface Sci.*, 2017, **490**, 29–36.
- [63] Y. Jin, B. A. Voss, R. D. Noble and W. Zhang, *Angew. Chem. Int. Ed.*, 2010, **49**, 6348–6351.
- [64] M. Liu, M. A. Little, K. E. Jelfs, J. T. A. Jones, M. Schmidtman, S. Y. Chong, T. Hasell and A. I. Cooper, *J. Am. Chem. Soc.*, 2014, **136**, 7583–7586.
- [65] T. Hasell, M. Schmidtman, C. A. Stone, M. W. Smith and A. I. Cooper, *Chem. Commun.*, 2012, **48**, 4689–4691.
- [66] M. W. Schneider, I. M. Oppel, A. Griffin and M. Mastalerz, *Angew. Chem. Int. Ed.*, 2013, **52**, 3611–3615.
- [67] K. Acharyya and P. S. Mukherjee, *Chem. Eur. J.*, 2015, **21**, 6823–6831.
- [68] S. Jiang, K. E. Jelfs, D. Holden, T. Hasell, S. Y. Chong, M. Haranczyk, A. Trewin and A. I. Cooper, *J. Am. Chem. Soc.*, 2013, **135**, 17818–17830.
- [69] S. L. Price, *Acta Crystallogr. Sect. B*, 2013, **69**, 313–328.
- [70] J. T. A. Jones, D. Holden, T. Mitra, T. Hasell, D. J. Adams, K. E. Jelfs, A. Trewin, D. J. Willock, G. M. Day, J. Bacsá, A. Steiner and A. I. Cooper, *Angew. Chem.*, 2011, **123**, 775–779.
- [71] T. Hasell, S. Y. Chong, M. Schmidtman, D. J. Adams and A. I. Cooper, *Angew. Chem. Int. Ed.*, 2012, **51**, 7154–7157.
- [72] J. Liu, P. K. Thallapally and D. Strachan, *Langmuir*, 2012, **28**, 11584–11589.
- [73] 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 and A. I. Cooper, *Nat. Mater.*, 2014, **13**, 954–960.
- [74] B. J. Sikora, C. E. Wilmer, M. L. Greenfield and R. Q. Snurr, *Chem. Sci.*, 2012, **3**, 2217–2223.

- [75] T. Mitra, K. E. Jelfs, M. Schmidtman, A. Ahmed, S. Y. Chong, D. J. Adams and A. I. Cooper, *Nat. Chem.*, 2013, **5**, 276–281.
- [76] N. O'Reilly, N. Giri and S. L. James, *Chem. Eur. J.*, 2007, **13**, 3020–3025.
- [77] N. Giri, C. E. Davidson, G. Melaugh, M. G. Del Pópolo, J. T. A. Jones, T. Hasell, A. I. Cooper, P. N. Horton, M. B. Hursthouse and S. L. James, *Chem. Sci.*, 2012, **3**, 2153–2157.
- [78] N. Giri, M. G. Del Pópolo, G. Melaugh, R. L. Greenaway, K. Rätzke, T. Koschine, L. Pison, M. F. Gomes, A. I. Cooper and S. L. James, *Nature*, 2015, **527**, 216–220.
- [79] F. Zhang, F. Yang, J. Huang, B. G. Sumpter and R. Qiao, *J. Phys. Chem. B*, 2016, **120**, 7195–7200.
- [80] R. L. Greenaway, D. Holden, E. G. B. Eden, A. Stephenson, C. W. Yong, M. J. Bennison, T. Hasell, M. E. Briggs, S. L. James and A. I. Cooper, *Chem. Sci.*, 2017, **8**, 2640–2651.
- [81] M. W. Schneider, H.-J. Siegfried Hauswald, R. Stoll and M. Mastalerz, *Chem. Commun.*, 2012, **48**, 9861–9863.
- [82] S. M. Elbert, F. Rominger and M. Mastalerz, *Chem. Eur. J.*, 2014, **20**, 1–15.
- [83] Y. Jin, B. A. Voss, A. Jin, H. Long, R. D. Noble and W. Zhang, *J. Am. Chem. Soc.*, 2011, **133**, 6650–6658.
- [84] Y. Jin, A. Jin, R. McCaffrey, H. Long and W. Zhang, *J. Org. Chem.*, 2012, **77**, 7392–7400.
- [85] S. Jiang, J. Bacsa, X. Wu, J. T. A. Jones, R. Dawson, A. Trewin, D. J. Adams and A. I. Cooper, *Chem. Commun.*, 2011, **47**, 8919–8921.
- [86] H. Ding, Y. Yang, B. Li, F. Pan, G. Zhu, M. Zeller, D. Yuan and C. Wang, *Chem. Commun.*, 2015, **51**, 1976–1979.
- [87] M. A. Little, M. E. Briggs, J. T. A. Jones, M. Schmidtman, T. Hasell, S. Y. Chong, K. E. Jelfs, L. Chen and A. I. Cooper, *Nat. Chem.*, 2015, **7**, 153–159.
- [88] T. Hasell and A. I. Cooper, *Nat. Rev. Mater.*, 2016, **1**, 16053.



- [89] R. S. Patil, D. Banerjee, C. M. Simon, J. L. Atwood and P. K. Thallapally, *Chem. Eur. J.*, 2016, **22**, 12618–12623.
- [90] M. Miklitz, S. Jiang, R. Clowes, M. E. Briggs, A. I. Cooper and K. E. Jelfs, *J. Phys. Chem. C*, 2017, **121**, 15211–15222.
- [91] T. Hasell, M. Miklitz, A. Stephenson, M. A. Little, S. Y. Chong, R. Clowes, L. Chen, D. Holden, G. A. Tribello, K. E. Jelfs and A. I. Cooper, *J. Am. Chem. Soc.*, 2016, **138**, 1653–1659.
- [92] D. Banerjee, A. J. Cairns, J. Liu, R. K. Motkuri, S. K. Nune, C. A. Fernandez, R. Krishna, D. M. Strachan and P. K. Thallapally, *Acc. Chem. Res.*, 2015, **48**, 211–219.
- [93] C. M. Simon, R. Mercado, S. K. Schnell, B. Smit and M. Haranczyk, *Chem. Mater.*, 2015, **27**, 4459–4475.
- [94] M. Kitchin, K. Konstas, M. L. Czyz, P. Valente, C. J. Sumby, M. R. Hill, A. Polyzos and C. J. Doonan, *Chem. Commun.*, 2015, **51**, 14231–14234.
- [95] M. E. Briggs and A. I. Cooper, *Chem. Mater.*, 2017, **29**, 149–157.
- [96] T. Hasell, S. Y. Chong, K. E. Jelfs, D. J. Adams and A. I. Cooper, *J. Am. Chem. Soc.*, 2012, **134**, 588–598.
- [97] A. F. Bushell, P. M. Budd, M. P. Attfield, J. T. A. Jones, T. Hasell, A. I. Cooper, P. Bernardo, F. Bazzarelli, G. Clarizia and J. C. Jansen, *Angew. Chem.*, 2013, **125**, 1291–1294.
- [98] Q. Song, S. Jiang, T. Hasell, M. Liu, S. Sun, A. K. Cheetham, E. Sivaniah and A. I. Cooper, *Adv. Mater.*, 2016, **28**, 2629–2637.
- [99] M. A. Zwijnenburg, E. Berardo, W. J. Peveler and K. E. Jelfs, *J. Phys. Chem. B*, 2016, **120**, 5063–5072.
- [100] T. Uemura, R. Nakanishi, S. Mochizuki, S. Kitagawa and M. Mizuno, *Angew. Chem. Int. Ed.*, 2016, **55**, 6443–6447.

- [101] J. Kong, M. Zhang, A.-H. Duan, J.-H. Zhang, R. Yang and L.-M. Yuan, *J. Sep. Sci.*, 2015, **38**, 556–561.
- [102] A. Kewley, A. Stephenson, L. Chen, M. E. Briggs, T. Hasell and A. I. Cooper, *Chem. Mater.*, 2015, **27**, 3207–3210.
- [103] X. Kong and J. Jiang, *J. Phys. Chem. C*, 2018, **122**, 1732–1740.
- [104] K. Acharyya and P. S. Mukherjee, *Chem. Commun.*, 2014, **50**, 15788–15791.
- [105] M. Brutschy, M. W. Schneider, M. Mastalerz and S. R. Waldvogel, *Adv. Mater.*, 2012, **24**, 6049–6052.
- [106] M. Brutschy, M. W. Schneider, M. Mastalerz and S. R. Waldvogel, *Chem. Commun.*, 2013, **49**, 8398–8400.
- [107] M. Liu, L. Chen, S. Lewis, S. Y. Chong, M. A. Little, T. Hasell, I. M. Aldous, C. M. Brown, M. W. Smith, C. A. Morrison, L. J. Hardwick and A. I. Cooper, *Nat. Commun.*, 2016, **7**, 1–9.
- [108] R. Gaillac, P. Pullumbi, K. A. Beyer, K. W. Chapman, D. A. Keen, T. D. Bennett and F.-X. Coudert, *Nat. Mater.*, 2017, **16**, 1149–1155.
- [109] C. J. J. D. Ouden, B. Smit, A. F. H. Wielers, R. A. Jackson and A. K. Nowak, *Mol. Simulat.*, 1989, **4**, 121–136.
- [110] E. C. D. Lara, R. Kahn and A. M. Goulay, *J. Chem. Phys.*, 1989, **90**, 7482–7491.
- [111] T. Düren, Y.-S. Bae and R. Q. Snurr, *Chem. Soc. Rev.*, 2009, **38**, 1237–1247.
- [112] S. Roy, C. B. George and M. A. Ratner, *J. Phys. Chem. C*, 2012, **116**, 23494–23502.
- [113] T. Düren, L. Sarkisov, O. M. Yaghi and R. Q. Snurr, *Langmuir*, 2004, **20**, 2683–2689.
- [114] A. M. Banu, D. Friedrich, S. Brandani and T. Düren, *Ind. Eng. Chem. Res.*, 2013, **52**, 9946–9957.

- [115] C. E. Wilmer, M. Leaf, C. Y. Lee, O. K. Farha, B. G. Hauser, J. T. Hupp and R. Q. Snurr, *Nat. Chem.*, 2012, **4**, 83–89.
- [116] E. Haldoupis, S. Nair and D. S. Sholl, *Phys. Chem. Chem. Phys.*, 2011, **13**, 5053–5060.
- [117] Y. Long, J. C. Palmer, B. Coasne, M. Śliwinska-Bartkowiak, G. Jackson, E. A. Müller and K. E. Gubbins, *J. Chem. Phys.*, 2013, **139**, 144701.
- [118] F. J. A. L. Cruz, E. A. Müller and J. P. B. Mota, *RSC Adv.*, 2011, **1**, 270–281.
- [119] K. V. Kumar, E. A. Müller and F. Rodríguez-Reinoso, *J. Phys. Chem. C*, 2012, **116**, 11820–11829.
- [120] S. Mecozzi and J. Rebek, *Chem. Eur. J.*, 1998, **4**, 1016–1022.
- [121] K. E. Jelfs and A. I. Cooper, *Curr. Opin. Solid State Mater. Sci.*, 2013, **17**, 19–30.
- [122] I. G. Garcia, M. Bernabei, R. P. Soto and M. Haranczyk, *Cryst. Growth Des.*, 2017, **17**, 5614–5619.
- [123] K. E. Jelfs, E. G. B. Eden, J. L. Culshaw, S. Shakespeare, E. O. Pyzer-Knapp, H. P. G. Thompson, J. Bacsá, G. M. Day, D. J. Adams and A. I. Cooper, *J. Am. Chem. Soc.*, 2013, **135**, 9307–9310.
- [124] E. O. Pyzer-Knapp, H. P. G. Thompson, F. Schiffmann, K. E. Jelfs, S. Y. Chong, M. A. Little, A. I. Cooper and G. M. Day, *Chem. Sci.*, 2014, **5**, 2235–2245.
- [125] J. D. Evans, D. M. Huang, M. R. Hill, C. J. Sumby, D. S. Sholl, A. W. Thornton and C. J. Doonan, *J. Phys. Chem. C*, 2015, **119**, 7746–7754.
- [126] B. J. Sikora, C. E. Wilmer, M. L. Greenfield and R. Q. Snurr, *Chem. Sci.*, 2012, **3**, 2217–2223.
- [127] E. R. Cooper, C. D. Andrews, P. S. Wheatley, P. B. Webb, P. Wormald and R. E. Morris, *Nature*, 2004, **430**, 1012–1016.
- [128] D. Banerjee, Z. Zhang, A. M. Plonka, J. Li and J. B. Parise, *Cryst. Growth Des.*, 2012, **12**, 2162–2165.

- [129] A. J. Fletcher, K. M. Thomas and M. J. Rosseinsky, *J. Solid State Chem.*, 2005, **178**, 2491–2510.
- [130] T. J. Vlugt and M. Schenk, *J. Phys. Chem. B*, 2002, **106**, 12757–12763.
- [131] D. Fairen-Jimenez, R. Galvelis, A. Torrisi, A. D. Gellan, M. T. Wharmby, P. A. Wright, C. Mellot-Draznieks and T. Düren, *Dalton Trans.*, 2012, **41**, 10752–10762.
- [132] D. Fairen-Jimenez, S. A. Moggach, M. T. Wharmby, P. A. Wright, S. Parsons and T. Düren, *J. Am. Chem. Soc.*, 2011, **133**, 8900–8902.
- [133] L. Chen, J. P. S. Mowat, D. Fairen-Jimenez, C. A. Morrison, S. P. Thompson, P. A. Wright and T. Düren, *J. Am. Chem. Soc.*, 2013, **135**, 15763–15773.
- [134] K. Theisen, B. Smit and M. Haranczyk, *J. Chem. Inf. Model.*, 2010, **50**, 461–469.
- [135] M. Witman, S. Ling, S. Jawahery, P. G. Boyd, M. Haranczyk, B. Slater and B. Smit, *J. Am. Chem. Soc.*, 2017, **139**, 5547–5557.
- [136] R. Manurung, D. Holden, M. Miklitz, L. Chen, T. Hasell, S. Y. Chong, M. Haranczyk, A. I. Cooper and K. E. Jelfs, *J. Phys. Chem. C*, 2015, **119**, 22577–22586.
- [137] L. J. Barbour, *Chem. Commun.*, 2006, **0**, 1163–1168.
- [138] D. Holden, S. Y. Chong, L. Chen, K. E. Jelfs, T. Hasell and A. I. Cooper, *Chem. Sci.*, 2016, **7**, 4875–4879.
- [139] I. Langmuir, *J. Am. Chem. Soc.*, 1918, **40**, 1361–1403.
- [140] S. Brunauer, P. H. Emmett and E. Teller, *J. Am. Chem. Soc.*, 1938, **60**, 309–319.
- [141] D. A. H. Hanaor, M. Ghadiri, W. Chrzanowski and Y. Gan, *Langmuir*, 2014, **30**, 15143–15152.
- [142] G. Zhang, O. Presly, F. White, I. M. Oppel and M. Mastalerz, *Angew. Chem. Int. Ed.*, 2014, **53**, 1516–1520.
- [143] E. Haldoupis, S. Nair and D. S. Sholl, *J. Am. Chem. Soc.*, 2010, **132**, 7528–7539.

- [144] S. Builes, S. I. Sandler and R. Xiong, *Langmuir*, 2013, **29**, 10416–10422.
- [145] R. P. Lively and M. J. Realff, *AIChE J.*, 2016, **62**, 3699–3705.
- [146] S. Horike, S. Shimomura and S. Kitagawa, *Nat. Chem.*, 2009, **1**, 695–704.
- [147] M. V. Parkes, H. Demir, S. L. Teich-Mcgoldrick, D. S. Sholl, J. A. Greathouse and M. D. Allendorf, *Micropor. Mesopor. Mat.*, 2014, **194**, 190–199.
- [148] S. Yang, J. Sun, A. J. Ramirez-Cuesta, S. K. Callear, W. I. F. David, D. P. Anderson, R. Newby, A. J. Blake, J. E. Parker, C. C. Tang and M. Schröder, *Nat. Chem.*, 2012, **4**, 887–894.
- [149] 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 and B. Smit, *Energy Environ. Sci.*, 2015, **8**, 1190–1199.
- [150] A. W. Thornton, D. A. Winkler, M. S. Liu, M. Haranczyk and D. F. Kennedy, *RSC Adv.*, 2015, **5**, 44361–44370.
- [151] Y. Bao, R. L. Martin, C. M. Simon, M. Haranczyk, B. Smit and M. W. Deem, *J. Phys. Chem. C*, 2015, **119**, 186–195.
- [152] T. F. Willems, C. H. Rycroft, M. Kazi, J. C. Meza and M. Haranczyk, *Micropor. Mesopor. Mat.*, 2012, **149**, 134–141.
- [153] M. Pinheiro, R. L. Martin, C. H. Rycroft, A. Jones, E. Iglesia and M. Haranczyk, *J. Mol. Graphics Modell.*, 2013, **44**, 208–19.
- [154] O. S. Smart, J. G. Neduvellil, X. Wang, B. A. Wallace and M. S. P. Sansomt, *J. Mol. Graphics*, 1996, **7855**, 354–360.
- [155] *Surface area program*, [http://people.bath.ac.uk/td222/research/surface\\_area/index.html](http://people.bath.ac.uk/td222/research/surface_area/index.html).
- [156] E. L. First, C. E. Gounaris, J. Wei and C. A. Floudas, *Phys. Chem. Chem. Phys.*, 2011, **13**, 17339.

- [157] E. L. First and C. A. Floudas, *Micropor. Mesopor. Mat.*, 2013, **165**, 32–39.
- [158] L. Sarkisov and A. Harrison, *Mol. Simulat.*, 2011, **37**, 1248–1257.
- [159] I. Heimbach, F. Rhiem, F. Beule, D. Knodt, J. Heinen and R. O. Jones, *J. Comput. Chem.*, 2017, **38**, 389–394.
- [160] D. Holden, K. E. Jelfs, A. I. Cooper, A. Trewin and D. J. Willock, *J. Phys. Chem. C*, 2012, **116**, 16639–16651.
- [161] S. Kim, P. A. Thiessen, E. E. Bolton, J. Chen, G. Fu, A. Gindulyte, L. Han, J. He, S. He, B. A. Shoemaker, J. Wang, B. Yu, J. Zhang and S. H. Bryant, *Nucleic Acids Res.*, 2016, **44**, D1202–D1213.
- [162] S. L. Mayo, B. D. Olafson and W. A. Goddard, *J. Phys. Chem.*, 1990, **94**, 8897–8909.
- [163] A. K. Rappé, C. J. Casewit, K. S. Colwell, W. A. Goddard and W. M. Skiff, *J. Am. Chem. Soc.*, 1992, **114**, 10024–10035.
- [164] W. L. Jorgensen, J. D. Madura and C. J. Swenson, *J. Am. Chem. Soc.*, 1984, **106**, 6638–6646.
- [165] M. G. Martin and J. I. Siepmann, *J. Phys. Chem. B*, 1998, **102**, 2569–2577.
- [166] A. D. MacKerell, D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. Reiher, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiórkiewicz-Kuczera, D. Yin and M. Karplus, *J. Phys. Chem. B*, 1998, **102**, 3586–3616.
- [167] J. L. Banks, H. S. Beard, Y. Cao, A. E. Cho, W. Damm, R. Farid, A. K. Felts, T. A. Halgren, D. T. Mainz, J. R. Maple, R. Murphy, D. M. Philipp, M. P. Repasky, L. Y. Zhang, B. J. Berne, R. A. Friesner, E. Gallicchio and R. M. Levy, *J. Comput. Chem.*, 2005, **26**, 1752–1780.

- [168] E. Harder, W. Damm, J. Maple, C. Wu, M. Reboul, J. Y. Xiang, L. Wang, D. Lupyán, M. K. Dahlgren, J. L. Knight, J. W. Kaus, D. S. Cerutti, G. Krilov, W. L. Jorgensen, R. Abel and R. A. Friesner, *J. Chem. Theory Comput.*, 2016, **12**, 281–296.
- [169] W. Smith, C. W. Yong and P. M. Rodger, *Mol. Simulat.*, 2002, **28**, 385–471.
- [170] D. Frenkel and B. Smit, *Understanding Molecular Simulation: From Algorithms to Applications*, Elsevier Science, 2001.
- [171] D. J. Evans and O. P. Morriss, *Comput. Phys. Rep.*, 1984, **1**, 297–343.
- [172] H. J. C. Berendsen, J. P. M. Postma, W. F. van Gunsteren, A. DiNola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.
- [173] W. G. Hoover, *Phys. Rev. A*, 1985, **31**, 1695–1697.
- [174] W. G. Hoover, *Phys. Rev. A*, 1986, **34**, 2499–2500.
- [175] A. Laio and F. L. Gervasio, *Rep. Prog. Phys.*, 2008, **71**, 126601.
- [176] A. Barducci, M. Bonomi and M. Parrinello, *WIREs Comput. Mol. Sci.*, 2011, **1**, 826–843.
- [177] C. J. Cramer, *Essentials of Computational Chemistry: Theories and Models*, John Wiley & Sons Ltd., Chichester, 2nd edn., 2004.
- [178] W. Koch and M. C. Holthausen, *A Chemist's Guide to Density Functional Theory*, WILEY-VCH Verlag GmbH, Weinheim, 2nd edn., 2001.
- [179] J. P. J. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.
- [180] A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098–3100.
- [181] J. VandeVondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing and J. Hutter, *Comput. Phys. Commun.*, 2005, **167**, 103–128.
- [182] G. Maurin, in *Adsorption by Powders and Porous Solids*, ed. F. Rouquerol, J. Rouquerol, K. S. W. Sing, P. Llewellyn and G. Maurin, Academic Press, Oxford, 2nd edn., 2014, pp. 191–235.

- [183] P. Ryan, O. K. Farha, L. J. Broadbelt and R. Q. Snurr, *AIChE J.*, 2011, **57**, 1759–1766.
- [184] L. Chen, C. A. Morrison and T. Düren, *J. Phys. Chem. C*, 2012, **116**, 18899–18909.
- [185] T. Van Heest, S. L. Teich-McGoldrick, J. A. Greathouse, M. D. Allendorf and D. S. Sholl, *J. Phys. Chem. C*, 2012, **116**, 13183–13195.
- [186] K. V. Lawler, A. Sharma, B. Alagappan and P. M. Forster, *Micropor. Mesopor. Mat.*, 2016, **222**, 104–112.
- [187] A. L. Myers and J. M. Prausnitz, *AIChE J.*, 1965, **11**, 121–127.
- [188] C. M. Simon, B. Smit and M. Haranczyk, *Comput. Phys. Commun.*, 2016, **200**, 364–380.
- [189] N. F. Cessford, N. A. Seaton and T. Düren, *Ind. Eng. Chem. Res.*, 2012, **51**, 4911–4921.
- [190] D. B. Hibbert, *Chemom. Intell. Lab. Syst.*, 1989, **19**, 1–9.
- [191] C. B. Lucasius and G. Kateman, *Chemom. Intell. Lab. Syst.*, 1993, **19**, 1–33.
- [192] T. C. Le and D. A. Winkler, *Chem. Rev.*, 2016, **116**, 6107–6132.
- [193] E. W. Lamelijer, J. N. Kok, T. Bäck and A. P. Ijzerman, *J. Chem. Inf. Model.*, 2006, **46**, 545–552.
- [194] A. R. Oganov and C. W. Glass, *J. Chem. Phys.*, 2006, **124**, 244704.
- [195] Q. Zhu, A. R. Oganov, C. W. Glass and H. T. Stokes, *Acta Crystallogr. Sect. B*, 2012, **68**, 215–226.
- [196] A. R. Oganov, A. O. Lyakhov and M. Valle, *Acc. Chem. Res.*, 2011, **44**, 227–237.
- [197] V. Venkatasubramanian, K. Chan and J. M. Caruthers, *J. Chem. Inf. Comput. Sci.*, 1995, **35**, 188–195.
- [198] R. H. Herring and M. R. Eden, *Comput. Chem. Eng.*, 2015, **83**, 267–277.
- [199] M. A. Little, S. Y. Chong, M. Schmidtman, T. Hasell and A. I. Cooper, *Chem. Commun.*, 2014, **50**, 9465–9468.



- [200] Python Software Foundation. *Python Language Reference, version 3*. Available at , <http://www.python.org>.
- [201] Wikipedia: *Object-oriented programming*, [https://en.wikipedia.org/wiki/Object-oriented\\_programming](https://en.wikipedia.org/wiki/Object-oriented_programming).
- [202] F. Pérez and B. E. Granger, *Computing in Science and Engineering*, 2007, **9**, 21–29.
- [203] *Project Jupyter*, <http://jupyter.org/>.
- [204] CSD, *CSD Elemental Radii*, [https://www.ccdc.cam.ac.uk/support-and-resources/ccdcresources/Elemental{\\\_}Radii.xlsx](https://www.ccdc.cam.ac.uk/support-and-resources/ccdcresources/Elemental{\_}Radii.xlsx).
- [205] H. Vogel, *Math. Biosci.*, 1979, **44**, 179–189.
- [206] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot and E. Duchesnay, *J. Mach. Learn. Res.*, 2011, **12**, 2825–2830.
- [207] E. Jones, T. Oliphant and P. Peterson, *SciPy: Open Source Scientific Tools for Python*, 2001–, <http://www.scipy.org/>.
- [208] *Accelrys Software Inc., Discovery Studio Modeling Environment, Release 7.0* , San Diego: *Accelrys Software Inc.*, 2007.
- [209] W. L. Jorgensen, D. S. Maxwell and J. Tirado-Rives, *J. Am. Chem. Soc.*, 1996, **118**, 11225–11236.
- [210] C. W. Yong, *J. Chem. Inf. Model.*, 2016, **56**, 1405–1409.
- [211] W. F. V. Gunsteren and H. J. C. Berendsen, *Mol. Simulat.*, 1988, **1**, 173–185.
- [212] L. Brancalion and H. Moseley, *Lasers Med. Sci.*, 2002, **17**, 173–186.
- [213] N. P. Franks, R. Dickinson, S. L. M. de Sousa, A. C. Hall and W. R. Lieb, *Nature*, 1998, **396**, 324.

- [214] A. Bifone, Y. Q. Song, R. Seydoux, R. E. Taylor, B. M. Goodson, T. Pietrass, T. F. Budinger, G. Navon and A. Pines, *Proc. Natl. Acad. Sci. U. S. A.*, 1996, **93**, 12932–12936.
- [215] M. Drescher, M. Hentschel, R. Kienberger, G. Tempea, C. Spielmann, G. A. Reider, P. B. Corkum and F. Krausz, *Science*, 2001, **291**, 1923–1927.
- [216] F. G. Kerry, *Industrial Gas Handbook: Gas Separation and Purification*, CRC Press, Boca Raton, FL, 2007.
- [217] R. E. Bazan, M. Bastos-Neto, A. Moeller, F. Dreisbach and R. Staudt, *Adsorption*, 2011, **17**, 371–383.
- [218] A. Soleimani Dorcheh, D. Denysenko, D. Volkmer, W. Donner and M. Hirscher, *Micropor. Mesopor. Mat.*, 2012, **162**, 64–68.
- [219] S. Pan, S. Mandal and P. K. Chattaraj, *J. Phys. Chem. B*, 2015, **119**, 10962–10974.
- [220] W. Saenger and M. Noltemeyer, *Angew. Chem. Int. Ed.*, 1974, **13**, 552–553.
- [221] A. I. Joseph, S. H. Lapidus, C. M. Kane and K. T. Holman, *Angew. Chem. Int. Ed.*, 2015, **54**, 1471–1475.
- [222] J. Kim, I. S. Jung, S. Y. Kim, E. Lee, J. K. Kang, S. Sakamoto, K. Yamaguchi and K. Kim, *J. Am. Chem. Soc.*, 2000, **122**, 540–541.
- [223] J. Zhang, H. Ren and L. Liu, *Chem. Lett.*, 2010, **39**, 1016–1017.
- [224] J. L. Culshaw, G. Cheng, M. Schmidtman, T. Hasell, M. Liu, D. J. Adams and A. I. Cooper, *J. Am. Chem. Soc.*, 2013, **135**, 10007–10010.
- [225] T. Mitra, X. Wu, R. Clowes, J. T. A. Jones, K. E. Jelfs, D. J. Adams, A. Trewin, J. Bacsá, A. Steiner and A. I. Cooper, *Chem. Eur. J.*, 2011, **17**, 10235–10240.
- [226] J. T. A. Jones, T. Hasell, X. Wu, J. Bacsá, K. E. Jelfs, M. Schmidtman, S. Y. Chong, D. J. Adams, A. Trewin, F. Schiffman, F. Cora, B. Slater, A. Steiner, G. M. Day and A. I. Cooper, *Nature*, 2011, **474**, 367–371.

- [227] K. Harata, *Bull. Chem. Soc. Jpn.*, 1977, **50**, 1416–1424.
- [228] T. Steiner and W. Saenger, *Acta Crystallogr. Sect. B*, 1998, **54**, 450–455.
- [229] T. Brotin and J. P. Dutasta, *Chem. Rev.*, 2009, **109**, 88–130.
- [230] O. Taratula, P. A. Hill, N. S. Khan, P. J. Carroll and I. J. Dmochowski, *Nat. Commun.*, 2010, **1**, 148.
- [231] A. I. Joseph, G. El-Ayle, C. Boutin, E. L. Once, P. Berthault and K. T. Holman, *Chem. Commun.*, 2014, **50**, 15905–15908.
- [232] P. Skowronek, B. Warżajtis, U. Rychlewska and J. Gawroński, *Chem. Commun.*, 2013, **49**, 2524–6.
- [233] K. Ono, K. Johmoto, N. Yasuda, H. Uekusa, S. Fujii, M. Kiguchi and N. Iwasawa, *J. Am. Chem. Soc.*, 2015, **137**, 7015–7018.
- [234] M. W. Schneider, I. M. Oppel, H. Ott, L. G. Lechner, H.-J. S. Hauswald, R. Stoll and M. Mastalerz, *Chem. Eur. J.*, 2012, **18**, 836–47.
- [235] M. W. Schneider, I. M. Oppel and M. Mastalerz, *Chem. Eur. J.*, 2012, **18**, 4156–4160.
- [236] R. Warmuth, E. F. Maverick, C. B. Knobler and D. J. Cram, *J. Org. Chem.*, 2003, **68**, 2077–2088.
- [237] C. Givélet, J. Sun, D. Xu, T. J. Emge, A. Dhokte and R. Warmuth, *Chem. Commun.*, 2011, **47**, 4511–4513.
- [238] *CP2K*, <https://www.cp2k.org/>.
- [239] G. Lippert, J. Hutter and M. Parrinello, *Mol. Phys.*, 1997, **92**, 477–487.
- [240] S. Goedecker, M. Teter and J. Hutter, *Phys. Rev. B*, 1996, **54**, 1703–1710.
- [241] S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.
- [242] J. VandeVondele and J. Hutter, *J. Chem. Phys.*, 2007, **127**, 114105.

- [243] S. Simon, M. Duran and J. J. Dannenberg, *J. Chem. Phys.*, 1996, **105**, 11024–11031.
- [244] T. Risthaus and S. Grimme, *J. Chem. Theory Comput.*, 2013, **9**, 1580–1591.
- [245] G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni and G. Bussi, *Comput. Phys. Commun.*, 2014, **185**, 604–613.
- [246] J. S. Camp and D. S. Sholl, *J. Phys. Chem. C*, 2016, **120**, 1110–1120.
- [247] D. Banerjee, C. M. Simon, A. M. Plonka, R. K. Motkuri, J. Liu, X. Chen, B. Smit, J. B. Parise, M. Haranczyk and P. K. Thallapally, *Nat. Commun.*, 2016, **7**, ncomms11831.
- [248] H. W. Kroto, J. R. Heath, S. C. O'Brien, R. F. Curl and R. E. Smalley, *Nature*, 1985, **318**, 162–163.
- [249] V. Biju, *Chem. Soc. Rev.*, 2014, **43**, 744–764.
- [250] A. Rodríguez-Forteza, N. Alegret and J. M. Poblet, *Chem. Rev.*, 2013, **9**, 907–924.
- [251] T. Liu and A. Troisi, *Adv. Mater.*, 2013, **25**, 1038–1041.
- [252] A. Y. Ganin, Y. Takabayashi, P. Jeglič, D. Arčon, A. Potočnik, P. J. Baker, Y. Ohishi, M. T. McDonald, M. D. Tzirakis, A. McLennan, G. R. Darling, M. Takata, M. J. Rosseinsky and K. Prassides, *Nature*, 2010, **466**, 221–225.
- [253] C. García-Simón, M. Garcia-Borràs, L. Gómez, T. Parella, S. Osuna, J. Juanhuix, I. Imaz, D. MasPOCH, M. Costas and X. Ribas, *Nat. Commun.*, 2014, **5**, 5557.
- [254] F. J. Rizzuto, D. M. Wood, T. K. Ronson and J. R. Nitschke, *J. Am. Chem. Soc.*, 2017, **139**, 11008–11011.
- [255] W. Brenner, T. K. Ronson and J. R. Nitschke, *J. Am. Chem. Soc.*, 2017, **139**, 75–78.
- [256] A. R. Mulholland, C. P. Woodward and S. J. Langford, *Chem. Commun.*, 2011, **47**, 1494–1496.
- [257] A. Sygula, F. R. Fronczek, R. Sygula, P. W. Rabideau and M. M. Olmstead, *J. Am. Chem. Soc.*, 2007, **129**, 3842–3843.

- [258] D. Canevet, E. M. Pérez and N. Martín, *Angew. Chem. Int. Ed.*, 2011, **50**, 9248–9259.
- [259] Q. Wang, C. Zhang, B. C. Noll, H. Long, Y. Jin and W. Zhang, *Angew. Chem. Int. Ed.*, 2014, 10663–10667.
- [260] C. Zhang, Q. Wang, H. Long and W. Zhang, *J. Am. Chem. Soc.*, 2011, **133**, 20995–21001.
- [261] S. Grimme, C. Mück-Lichtenfeld and J. Antony, *J. Phys. Chem. C*, 2007, **111**, 11199–11207.
- [262] C. Mück-Lichtenfeld, S. Grimme, L. Kobryn and A. Sygula, *Phys. Chem. Chem. Phys.*, 2010, **12**, 7091–7097.
- [263] V. Santolini, M. Miklitz, E. Berardo and K. E. Jelfs, *Nanoscale*, 2017, **9**, 5280–5298.
- [264] J.-L. Brédas, K. Persson and R. Seshadri, *Chem. Mater.*, 2017, **29**, 2399–2401.
- [265] T. Kalil and C. Wadia, *Materials Genome Initiative for Global Competitiveness*, June, 2011.
- [266] R. L. Martin, C. M. Simon, B. Smit and M. Haranczyk, *J. Am. Chem. Soc.*, 2014, **136**, 5006–5022.
- [267] G. Mondragón-Solórzano, R. Sierra-Álvarez, E. López-Honorato and J. Barroso-Flores, *J. Inclusion Phenom. Macrocyclic Chem.*, 2016, **85**, 169–174.
- [268] M. Witman, S. Ling, S. Anderson, L. Tong, K. C. Stylianou, B. Slater, B. Smit and M. Haranczyk, *Chem. Sci.*, 2016, **7**, 6263–6272.
- [269] C. L. Hobday, T. D. Bennett, D. Fairen-Jimenez, A. J. Graham, C. A. Morrison, D. R. Allan, T. Düren and S. A. Moggach, *J. Am. Chem. Soc.*, 2018, **140**, 382–387.
- [270] A. G. Slater and A. I. Cooper, *Science*, 2015, **348**, aaa8075.
- [271] K. E. Jelfs, X. Wu, M. Schmidtman, J. T. A. Jones, J. E. Warren, D. J. Adams and A. I. Cooper, *Angew. Chem. Int. Ed.*, 2011, **50**, 10653–10656.
- [272] T. Fink, H. Bruggesser and J. L. Reymond, *Angew. Chem. Int. Ed.*, 2005, **44**, 1504–1508.

- [273] L. Ruddigkeit, R. Van Deursen, L. C. Blum and J. L. Reymond, *J. Chem. Inf. Model.*, 2012, **52**, 2864–2875.
- [274] R. Van Deursen and J. L. Reymond, *ChemMedChem*, 2007, **2**, 636–640.
- [275] T. E. B. Davies, M. M. Popa and C. V. Ciobanu, *AIP Conf. Proc.*, 2007, **963**, 168–175.
- [276] K. Huwig, C. Fan and M. Springborg, *J. Chem. Phys.*, 2017, **147**, 234105.
- [277] D. Holden, K. E. Jelfs, A. Trewin, D. J. Willock, M. Haranczyk and A. I. Cooper, *J. Phys. Chem. C*, 2014, **118**, 12734–12743.
- [278] M. J. Lennox and T. Düren, *J. Phys. Chem. C*, 2016, **120**, 18651–18658.
- [279] *RDKit: Open-source cheminformatics*, <http://www.rdkit.org>.
- [280] H. L. Morgan, *J. Chem. Doc.*, 1965, **5**, 107–113.
- [281] Y. Wang, H. Fang, I. Tranca, H. Qu, X. Wang, A. J. Markvoort, Z. Tian and X. Cao, *Nat. Commun.*, 2018, **9**, 488–496.

# Appendix

## A.1 The *pywindow* software window analysis schematics

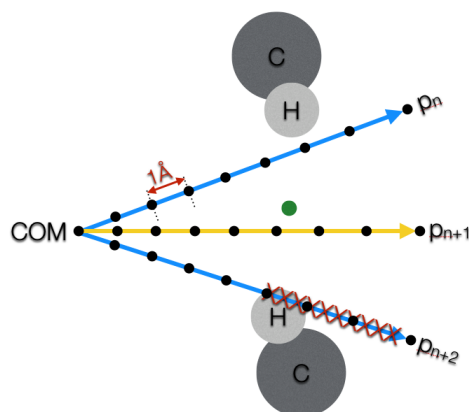


Figure A.1: Process of finding windows in molecular pores. From the centre of mass of a molecule (COM) the projected vectors are sampled in increments of  $1 \text{ \AA}$  along the vectors path (black dots). At each such point the distance to the closest atom (corrected for the appropriate van der Waals radius) is calculated. If vector passes through an atom, such as the  $p_{n+2}$  vector, it is discarded. From the resulting vectors, the one that passes closest to the centre of the window marked with a green dot (vector in yellow) is then analysed as shown in Figure A.2.

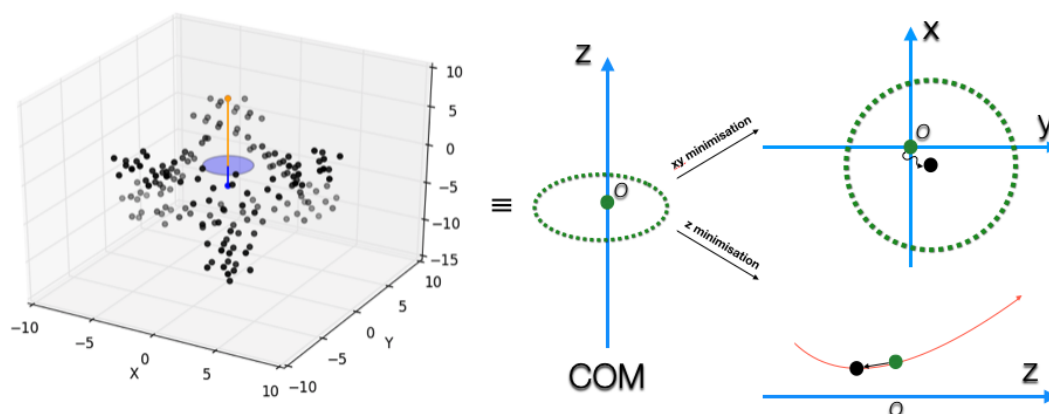


Figure A.2: In this schematic the process of finding the window centre is presented. The coordinates of the molecule are translated and rotated so that the vector that passes closest to the window centre (in yellow in Figure A.1) starts at the origin of the Cartesian system and is aligned to the Z axis. The window plane is then shifted to the origin. The  $x$  and  $y$  coordinates of the window centre are optimised to find the largest window diameter. The  $z$  coordinate is also minimised to find the necking of the window channel and true window centre.



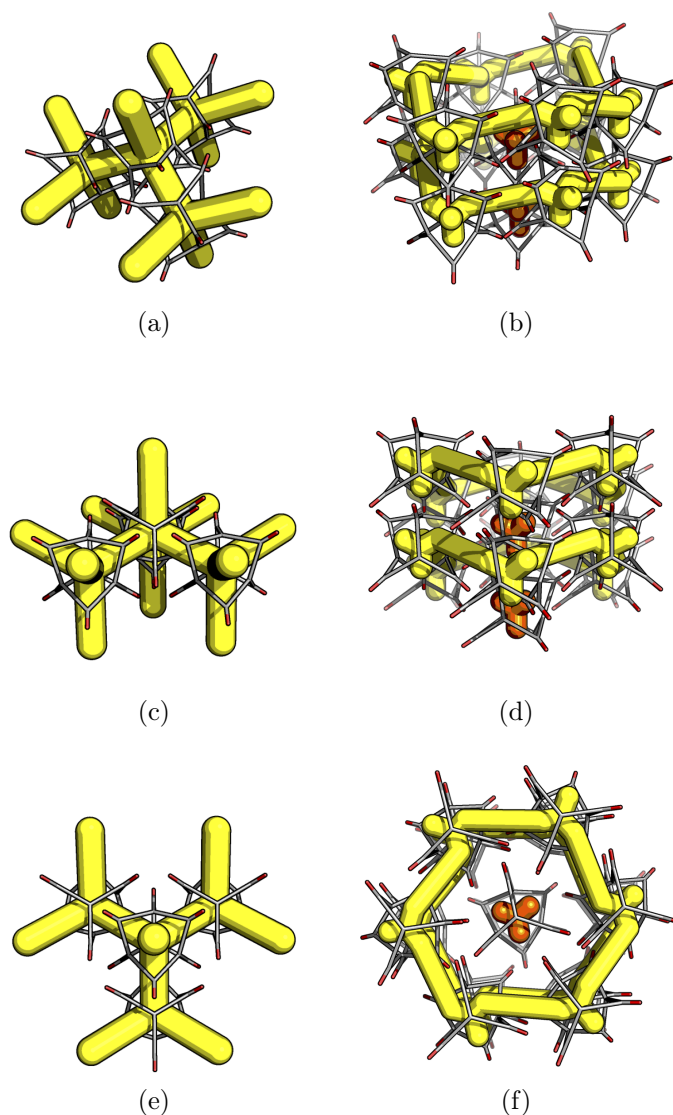
A.2 Solid state structures of *CC3* polymorphs

Figure A.3: The subfigures *a*, *c* and *e* show  $\text{CC3}\alpha\text{-}R$  and  $\text{CC3}\alpha\text{-}R,S$  crystal packing from different perspectives, with the diamondoid shaped pore network shown as yellow tubes and  $\text{CC3}$  replaced with the simplified wireframe model. The subfigures *b*, *d*, *f* show  $\text{CC3}\beta\text{-}R$  crystal packing with the honey-comb shaped interconnected porous network considered as ‘accessible’ showed as yellow tubes and the intrinsic excluded intrinsic voids in orange that are considered as ‘non-accessible’.

### A.3 DFT binding energies for CDB26

Table A.1: The binding energies for Kr ( $E_{b,Kr}$ ) and Xe ( $E_{b,Xe}$ ) and relative binding ( $E_{b,rel}$ ) calculated using DFT for CDB26 as described in the calculating binding energies methods section of the main paper. The table is sorted in order of ascending relative binding energy.

<i>CDB26</i>	$E_{b,Kr}$ ( $kJ\ mol^{-1}$ )	$E_{b,Xe}$ ( $kJ\ mol^{-1}$ )	$E_{b,rel} = E_{b,Kr} - E_{b,Xe}$ ( $kJ\ mol^{-1}$ )
<b>WC3</b>	-25.7	-23.4	-2.3
<b>RCC1b</b>	-21.5	-23.9	2.4
<b>MC3</b>	-12.6	-15.2	2.6
<b>CD3</b>	-11.2	-14.3	3.0
<b>MC7</b>	-14.7	-17.9	3.2
<b>MC4</b>	-11.9	-15.9	4.0
<b>CP3</b>	-29.1	-33.9	4.7
<b>HC1</b>	-11.2	-16.6	5.4
<b>WC2</b>	-27.7	-33.2	5.5
<b>CB7</b>	-17.7	-23.5	5.8
<b>RCC1c</b>	-25.1	-31.1	6.0
<b>CC10</b>	-23.4	-29.4	6.0
<b>RCC1d</b>	-24.8	-31.2	6.4
<b>RCC1a</b>	-21.6	-28.6	7.0
<b>CD2</b>	-13.1	-20.5	7.5
<b>CC9</b>	-21.9	-29.4	7.5
<b>CC1</b>	-19.5	-27.0	7.6
<b>CC2</b>	-20.2	-27.8	7.6
<b>NC2</b>	-24.3	-31.9	7.6
<b>CC4</b>	-18.3	-26.0	7.7
<b>CC3</b>	-20.3	-28.1	7.7
<b>CP1</b>	-30.2	-39.0	8.8
<b>RCC3b</b>	-22.5	-31.9	9.4
<b>WC4</b>	-27.7	-37.7	10.0
<b>CB6</b>	-23.1	-33.5	10.4
<b>NC1</b>	-17.9	-31.5	13.6

## DFT binding energies compared to void diameters

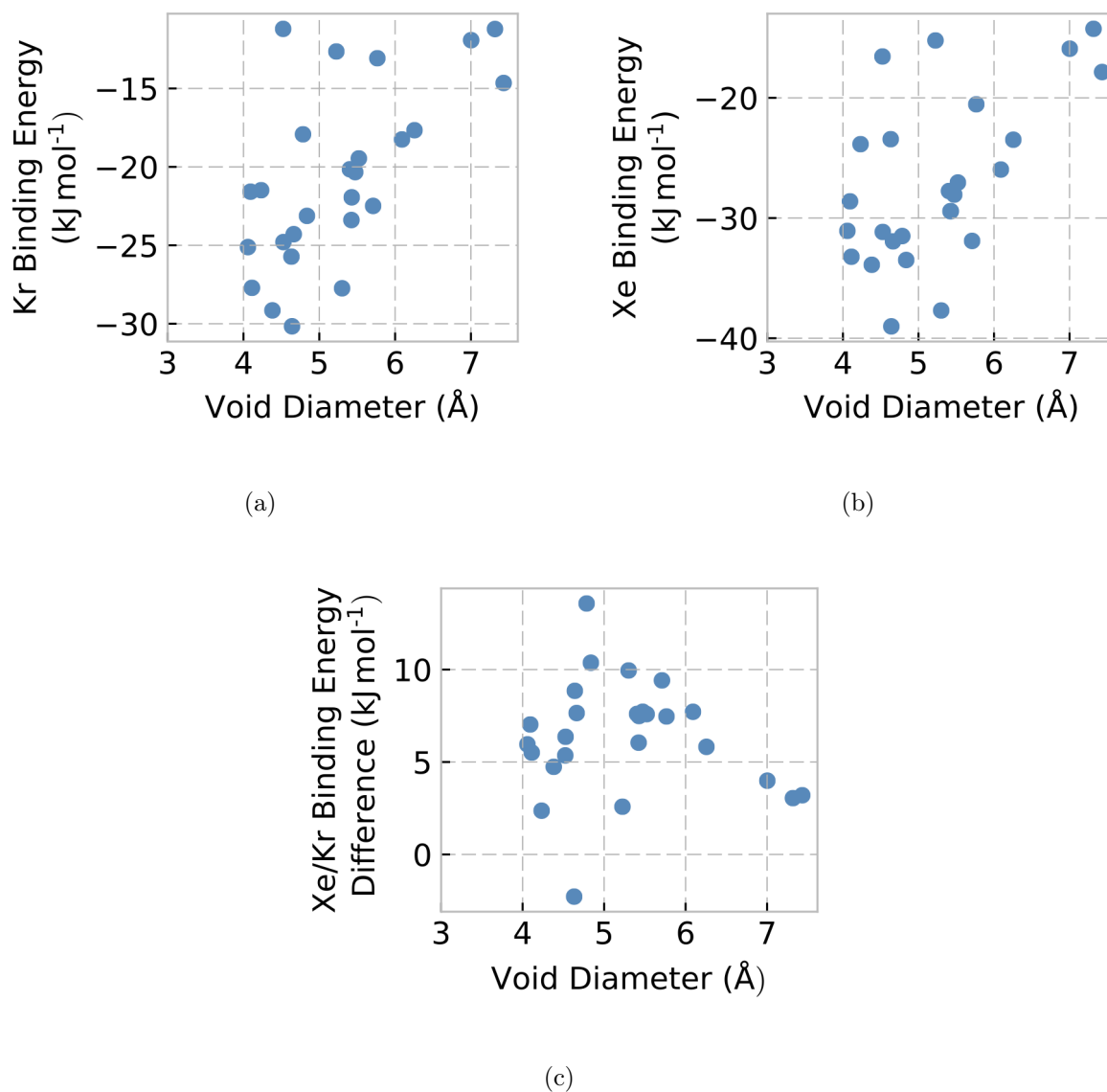


Figure A.4: Scatter plots of the measured void diameters (X-axis) against the calculated (Y-axis) a) DFT Kr binding energies b) DFT Xe binding energies c) Binding energy differences between the host-guest complex with Kr and Xe.

## A.4 Structural overlays of NC1 and NC2

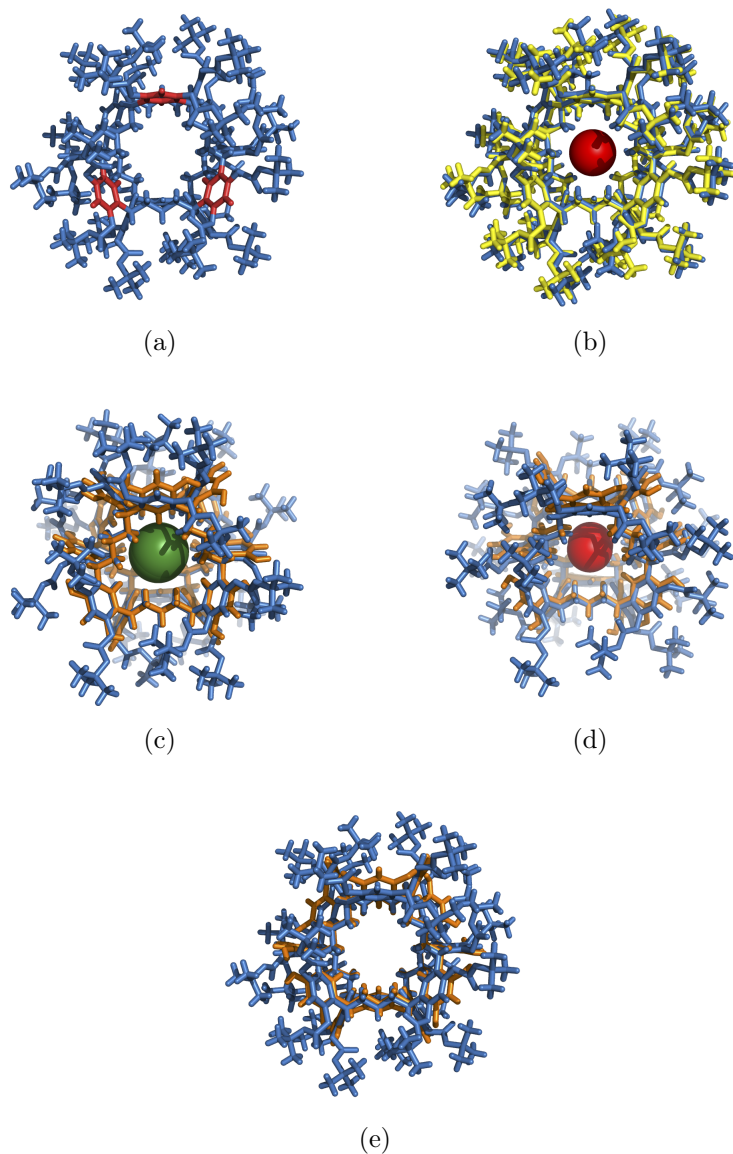


Figure A.5: a) The structure of DFT optimised **NC1** in blue with the benzene rings that are used to compare the structural differences to **NC2** in red; b) the overlay of the empty DFT optimised **NC1** in blue and **NC1-Kr** in yellow, with Kr colored red; c) DFT optimised **NC1-Xe** in blue with Xe in green and **NC2-Xe** in orange; d) DFT optimised **NC1-Kr** in blue and **NC2-Kr** in orange; e) the overlay of DFT optimised empty structures of **NC1** in blue and **NC2** in orange.

## A.5 Experimental and DFT void diameters for *CDB12*

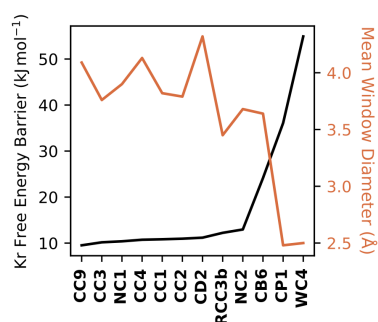
Table A.2: The void diameters for molecules calculated for: as extracted from X-ray crystal structure ( $d_{void,cryst.}$ ), for the DFT-D3 geometry-optimised empty molecule ( $d_{void,DFT-empty}$ ), for the DFT-D3 geometry-optimised complex with Kr ( $d_{void,DFT-Kr}$ ) and for the DFT-D3 geometry-optimised complex with Xe ( $d_{void,DFT-Xe}$ ).

<i>CDB12</i>	$d_{void,cryst.}$ (Å)	$d_{void,DFT-empty}$ (Å)	$d_{void,DFT-Kr}$ (Å)	$d_{void,DFT-Xe}$ (Å)
<b>CB6</b>	4.84	5.15	5.12	5.15
<b>CC1</b>	5.23	5.60	5.48	5.48
<b>CC2</b>	5.31	5.50	5.38	5.38
<b>CC3</b>	5.40	5.54	5.52	5.53
<b>RCC3b</b>	5.74	5.61	5.61	5.61
<b>CC4</b>	5.26	5.78	5.69	5.66
<b>CC9</b>	4.93	5.26	5.19	5.20
<b>CD2</b>	5.60	5.67	5.68	5.56
<b>CP1</b>	4.64	4.71	4.70	4.71
<b>NC1</b>	5.14	4.87	4.88	4.88
<b>NC2</b>	5.04	4.55	4.52	4.61
<b>WC4</b>	5.30	5.33	5.30	5.32

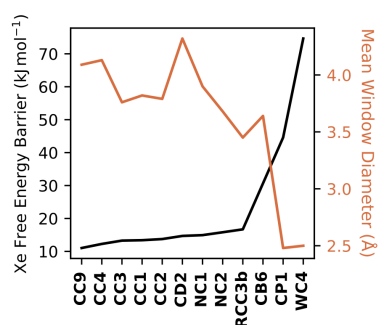
## A.6 Free energy barriers for *CDB12*

Table A.3: The free energy barriers for a Kr molecule ( $FEB_{Kr}$ ) or a Xe molecule ( $FEB_{Xe}$ ) diffusing through windows calculated using metadynamics simulations. The relative difference and the production run simulation times for Kr and Xe are also provided.

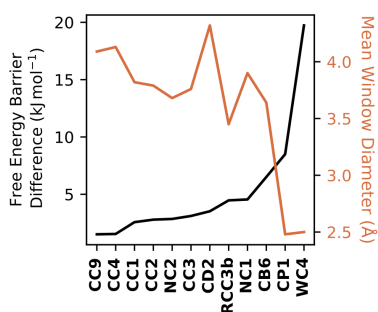
<i>CDB12</i>	$FEB_{Kr}$ (kJ mol <sup>-1</sup> )	$FEB_{Xe}$ (kJ mol <sup>-1</sup> )	<i>difference</i> (kJ mol <sup>-1</sup> )	<i>simulation time</i> Kr / Xe (ns)
<b>CB6</b>	24.13 ± 0.04	30.64 ± 0.05	6.51	300 / 300
<b>CC1</b>	10.83 ± 0.03	13.40 ± 0.03	2.57	300 / 300
<b>CC2</b>	10.96 ± 0.02	13.75 ± 0.04	2.79	300 / 300
<b>CC3</b>	10.17 ± 0.02	13.28 ± 0.03	3.11	300 / 300
<b>RCC3b</b>	12.23 ± 0.03	16.70 ± 0.04	4.47	300 / 300
<b>CC4</b>	10.73 ± 0.02	12.27 ± 0.02	1.54	300 / 300
<b>CC9</b>	9.52 ± 0.02	11.03 ± 0.03	1.51	300 / 300
<b>CD2</b>	11.18 ± 0.02	14.70 ± 0.03	3.52	400 / 400
<b>CP1</b>	36.08 ± 0.04	44.58 ± 0.07	8.50	400 / 400
<b>NC1</b>	10.39 ± 0.04	14.94 ± 0.03	4.55	700 / 1100
<b>NC2</b>	12.96 ± 0.02	15.81 ± 0.02	2.85	800 / 1100
<b>WC4</b>	54.91 ± 0.16	74.64 ± 0.18	19.73	300 / 400

Free energy barriers *vs.* void diameters

(a)



(b)



(c)

Figure A.6: a) Kr free energy barriers (left-hand side Y-axis) plotted against the mean window diameter (right-hand side Y-axis in orange) for *CDB12* b) Xe free energy barriers (left-hand side Y-axis) plotted against the mean window diameter (right-hand side Y-axis in orange) for *CDB12* c) free energy barriers difference (left-hand side Y-axis) plotted against the mean window diameter (right-hand side Y-axis in orange) for *CDB12*.

## A.7 Experimental details for Chapter 4

The whole of the experimental work presented in this section was performed by Dr. Shan Jiang, Mr. Rob Clowes and Dr. Michael Briggs, from the University of Liverpool.

### Synthesis of **aNC1**, **aNC2** and **aCC3-R**

**NC2** (Noria) and **NC1** (Noria-BOC) were synthesized according to the procedures reported by Kudo et al.<sup>47</sup> The recorded <sup>1</sup>H NMR spectra for both molecules was consistent with the literature. The as-synthesised **NC1** and **NC2** were observed by PXRD to be amorphous and poorly crystalline, respectively. **CC3-R** (100 mg) was dissolved in a mixture of DCM (10 mL) and methanol (2 mL) and the solution was frozen in liquid nitrogen. The solvent was then removed via freeze drying to afford **aCC3-R** as an amorphous solid.<sup>96</sup>

### Experimental gas adsorption measurements

Kr and Xe adsorption and desorption isotherms were measured at 298 K up to 1 bar using a Micromeritics 2020 volumetric adsorption analyser. Powder samples were degassed offline at 90 °C under dynamic vacuum ( $10^{-5}$  bar) before analysis.



## Experimental analysis of aNC1, aNC2 and aCC3

### NMR spectroscopy

Solution  $^1\text{H}$  NMR spectra were recorded at 400.13 MHz using a Bruker Avance 400 NMR spectrometer.

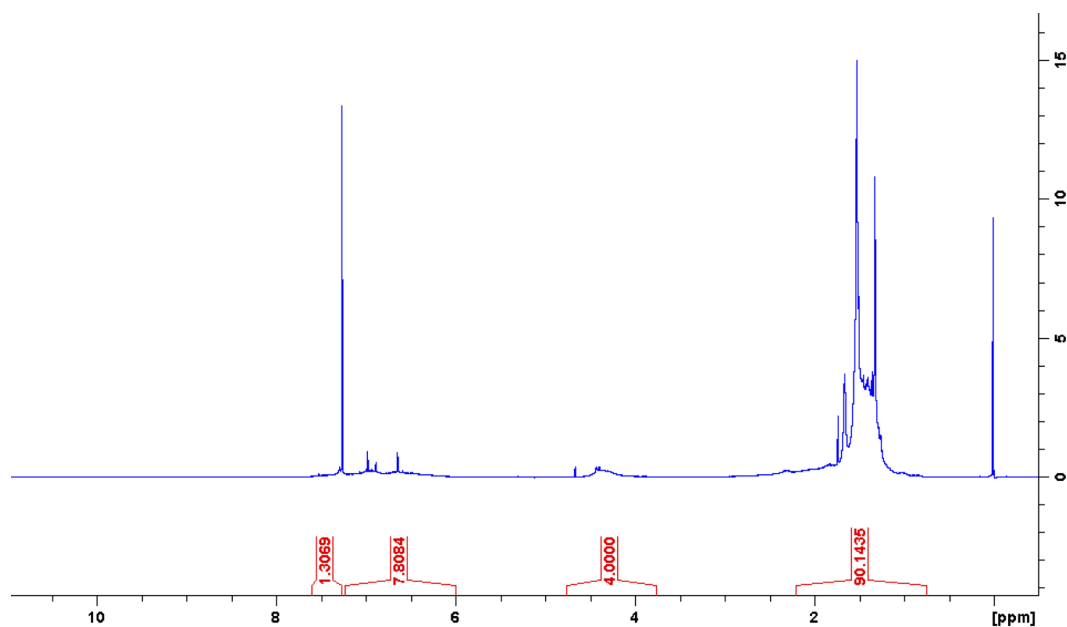


Figure A.7:  $^1\text{H}$  NMR ( $\text{CDCl}_3$ ) of NC1 after vacuum drying at 90 °C.

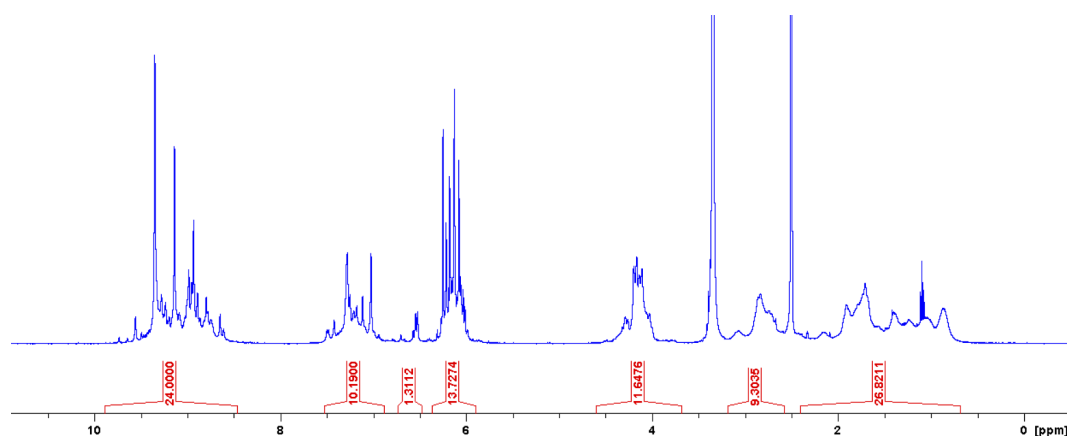


Figure A.8:  $^1\text{H}$  NMR (DMSO) of NC2 after vacuum drying at 90 °C.

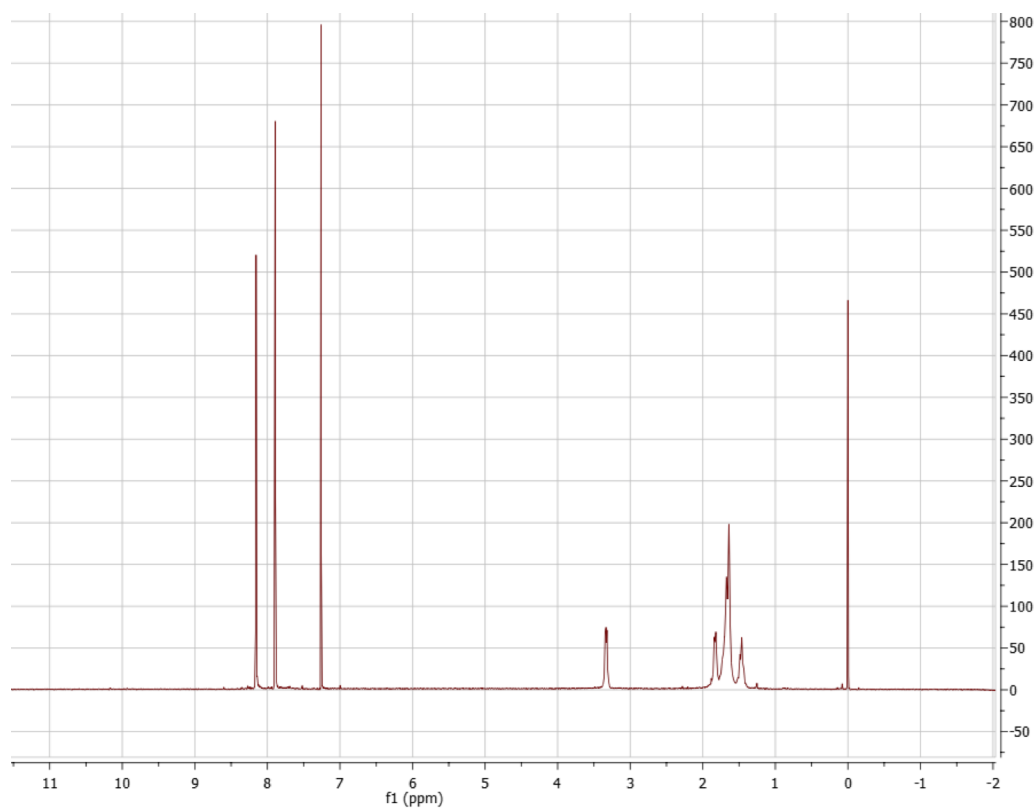


Figure A.9:  $^1\text{H}$  NMR ( $\text{CDCl}_3$ ) of **aCC3** after freeze drying.

## PXRD data

Powder X-ray diffraction (PXRD) data were collected in transmission mode on loose powder samples held on thin Mylar film in aluminum well plates on a Panalytical X'Pert PRO MPD equipped with a high throughput screening (HTS) XYZ stage, X-ray focusing mirror, and PIXcel detector, using Cu K $\alpha$  radiation.

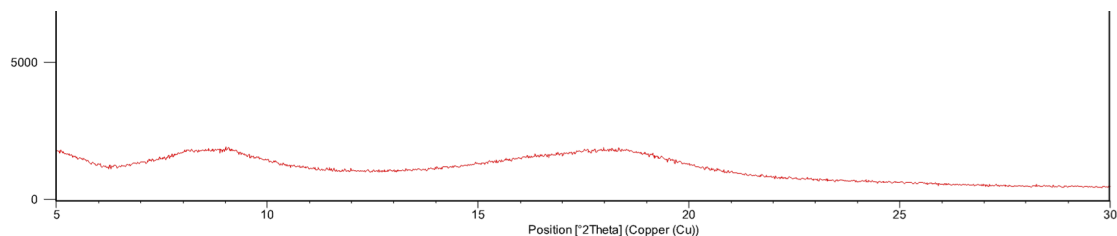


Figure A.10: PXRD of **NC1** after vacuum drying at 90 °C.

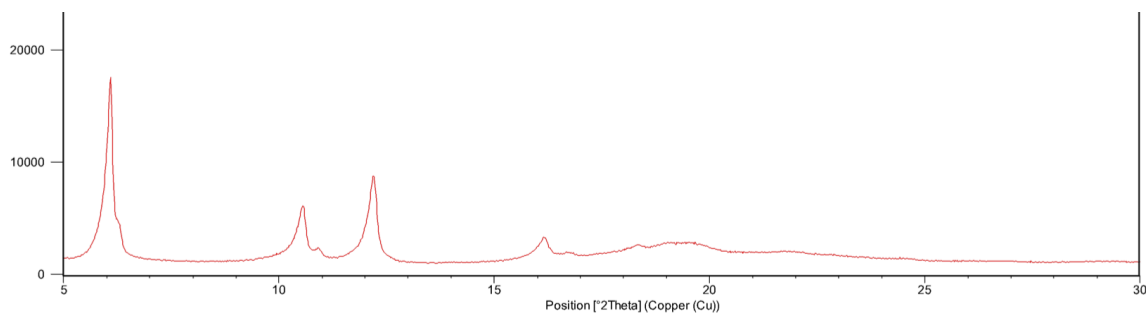


Figure A.11: PXRD of **NC2** after vacuum drying at 90 °C.

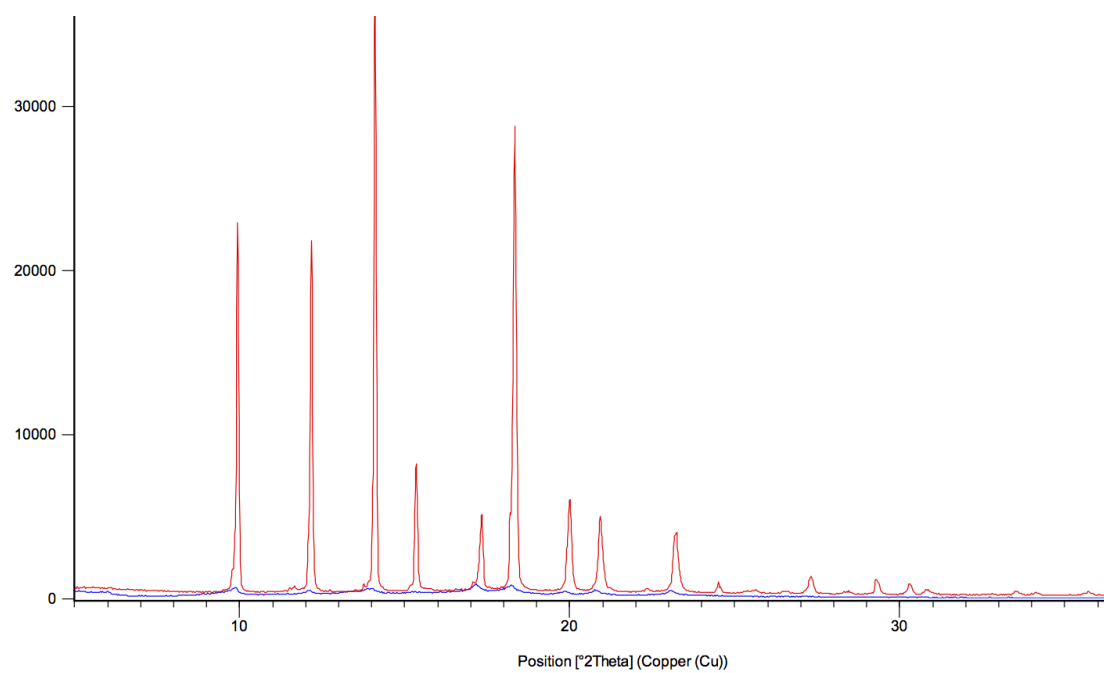


Figure A.12: PXRD of **CC3 $\alpha$**  (red) isolated from synthesis and **aCC3** (blue) isolated after freeze drying.

## Gas adsorption measurements

Kr and Xe adsorption and desorption isotherms were measured at 298 K up to 1 bar using a Micromeritics 2020 volumetric adsorption analyser. Powder samples were degassed offline at 90 °C under dynamic vacuum ( $10^{-5}$  bar) before analysis.

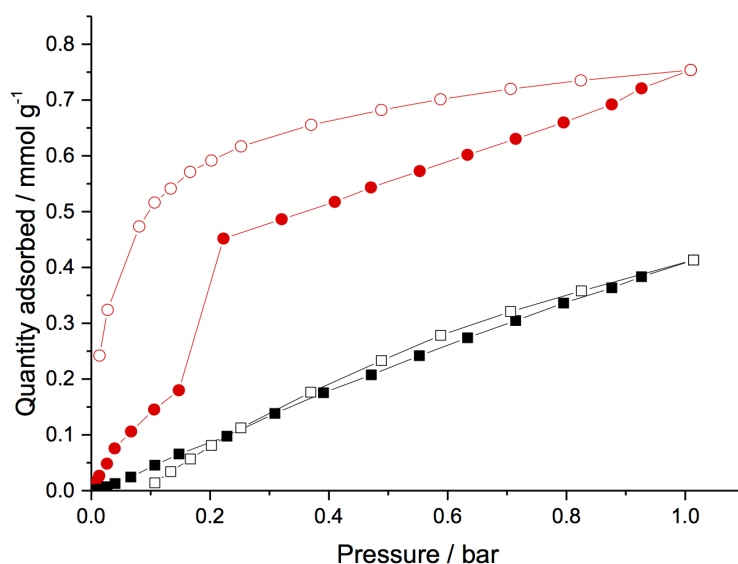


Figure A.13: Gas sorption isotherms for the uptake of Kr (black squares) and Xe (red circles) in **aNC1**. Closed and open symbols represent adsorption and desorption, respectively.

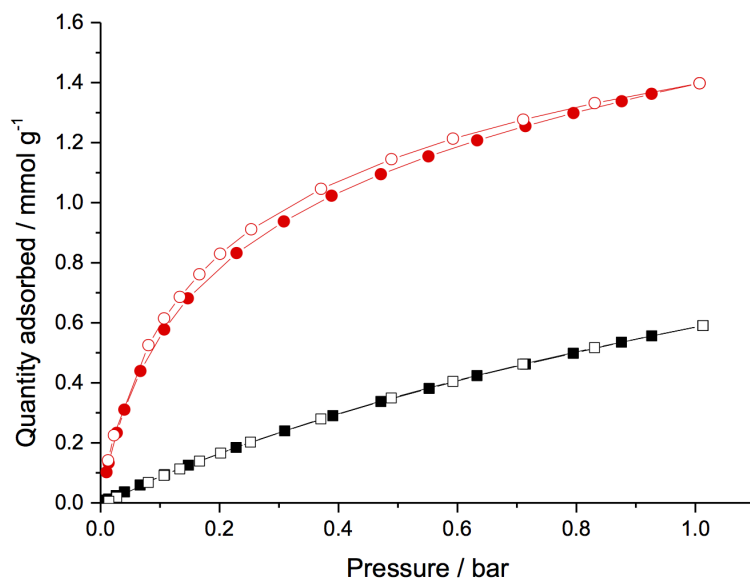


Figure A.14: Gas sorption isotherms for the uptake of Kr (black squares) and Xe (red circles) in **aNC2**. Closed and open symbols represent adsorption and desorption, respectively.

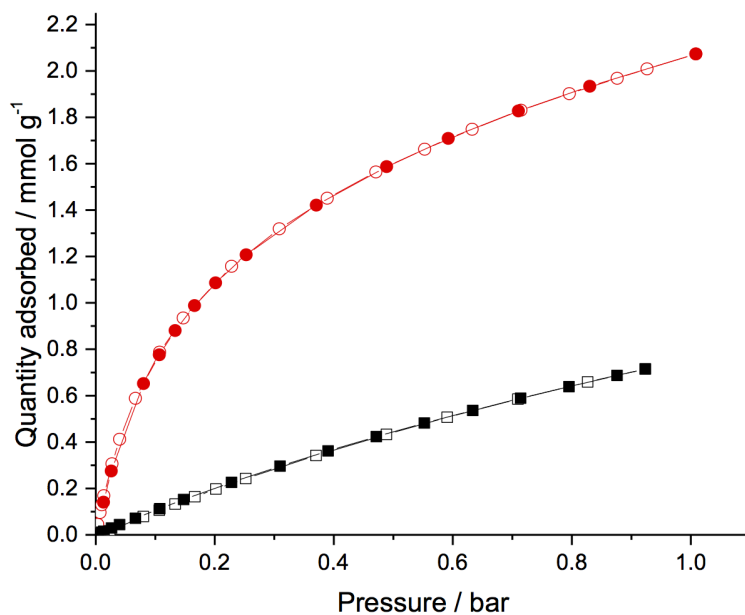


Figure A.15: Gas sorption isotherms for the uptake of Kr (black squares) and Xe (red circles) in **aCC3**. Closed and open symbols represent adsorption and desorption, respectively.

## pyIAST adsorption isotherms fitting

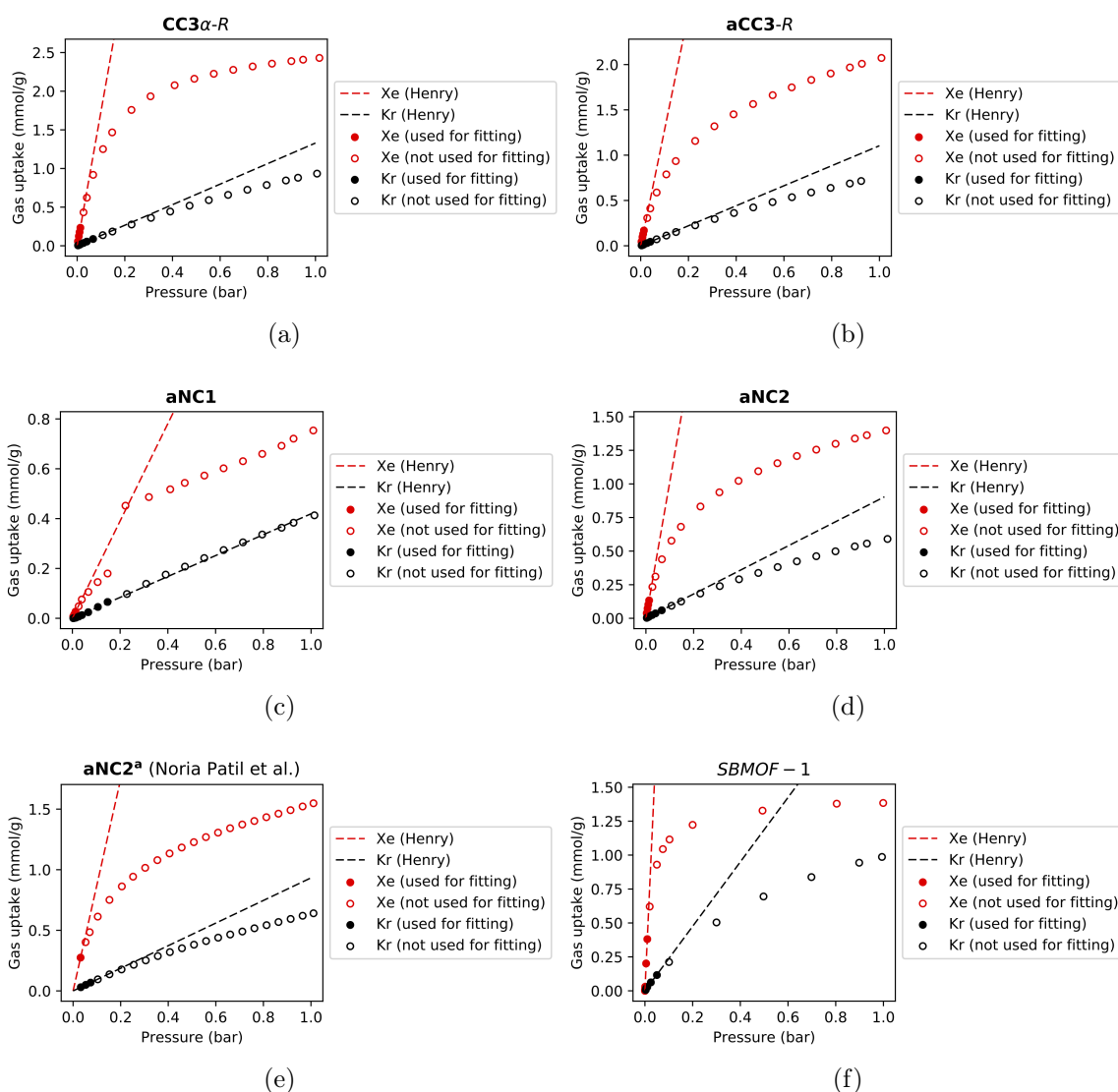


Figure A.16: Points in Henry's region of the linear adsorption increase phase used for the fitting and calculating the Henry's constants. The Kr and Xe adsorption isotherms data were taken from the literature for **CC3 $\alpha$ -R**<sup>73</sup> (plot a), **aNC2<sup>a</sup>** (Noria molecule as synthesised by Patil *et al.*<sup>89</sup> in plot e) and **SBMOF-1**<sup>247</sup> (plot f). Single component adsorption isotherms data for **aCC3-R** (plot b), **aNC1** (plot c) and **aNC2** (plot d) are from this work.

## A.8 Binding energies and cavity diameters for $C_{60}@POC$ complexes

Table A.4: DFT and OPLS3 derived binding energies for  $C_{60}@POC$  complexes.

POC	DFT ( $kJ mol^{-1}$ )	Dispersion		OPLS3 ( $kJ mol^{-1}$ )	van der Waals contribution ( $kJ mol^{-1}$ )
		Correction ( $kJ mol^{-1}$ )			
<b>A1</b>	-96.3	-96.4		-138.1	-173.9
<b>A2</b>	-100.0	-86.0		-76.1	-76.8
<b>A3</b>	-149.6	-145.8		-172.8	-171.0
<b>A4</b>	-169.6	-150.1		-159.6	-166.2
<b>B1</b>	-108.8	-101.8		-140.2	-174.6
<b>B2</b>	-97.0	-85.2		-86.6	-81.8
<b>B3</b>	-145.0	-142.7		-169.2	-171.8
<b>B4</b>	-166.3	-151.3		-162.1	-164.9

Table A.5: DFT and OPLS3 calculated  $D_{void\_opt}$  for empty cages and  $C_{60}@POC$  complexes.

POC	DFT		OPLS3	
	Empty Cage ( $\text{\AA}$ )	Complex ( $\text{\AA}$ )	Empty Cage ( $\text{\AA}$ )	Complex ( $\text{\AA}$ )
<b>A1</b>	6.32	9.54	6.20	9.01
<b>A2</b>	10.10	9.83	10.06	9.93
<b>A3</b>	9.74	9.89	9.46	9.97
<b>A4</b>	8.58	9.70	8.61	10.03
<b>B1</b>	6.31	9.63	6.17	9.80
<b>B2</b>	10.63	11.00	10.00	10.25
<b>B3</b>	9.39	9.79	9.24	9.92
<b>B4</b>	9.03	9.65	8.86	10.12



## A.9 The NMR spectra for B-series C<sub>60</sub>@POC complexes

The whole of the experimental work presented in this section was performed by Dr. Rebecca Greenaway, from the University of Liverpool.

No cages were found to form in the absence of the C<sub>60</sub> at the 0.0014 M aldehydes concentration in respect to the 2,4,6-Triethylbenzene-1,3,5-triyl)trimethanamine (0.020 mmol mL<sup>-1</sup>). However, these cages were found to form in the presence of one molar equivalent of C<sub>60</sub> relative to the potential formed cages (see Figures A.17-A.20). This indicates a templating effect of C<sub>60</sub> on the cages formation.

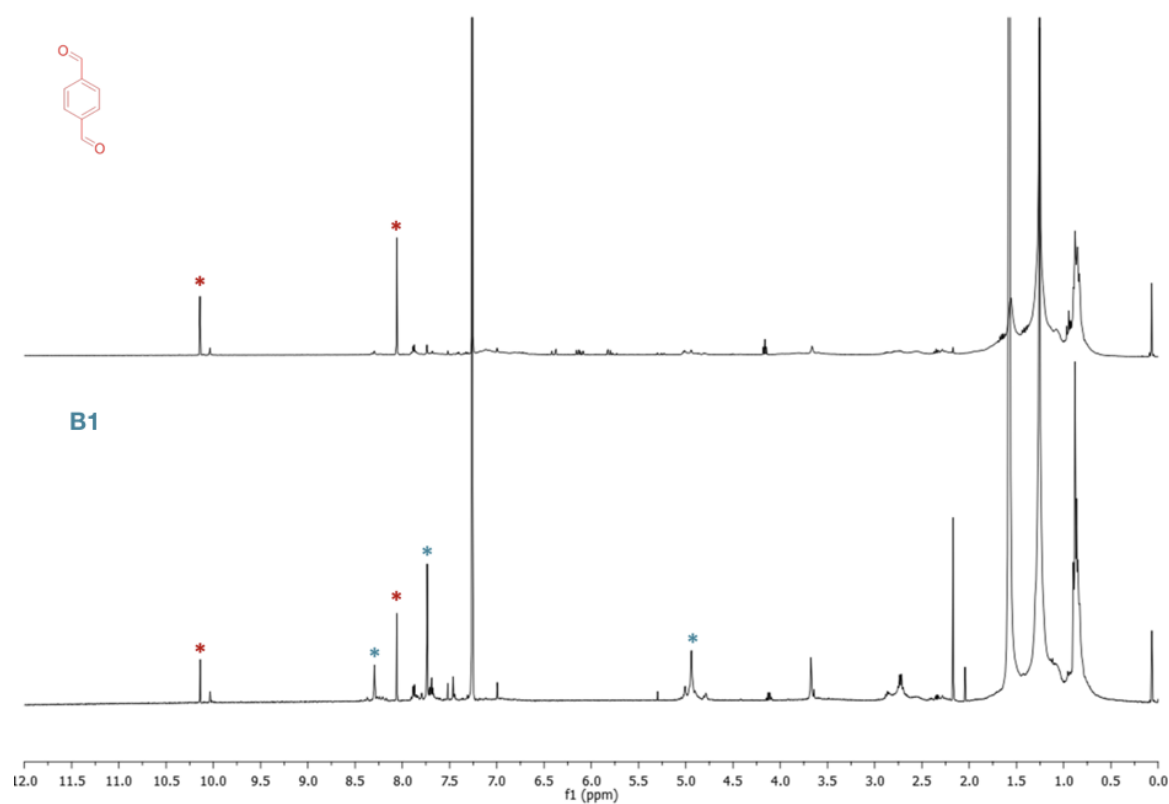


Figure A.17: Stacked <sup>1</sup>H NMR spectra for the attempted formation of **B1**, without C<sub>60</sub> (top spectra, aldehyde starting material \*) showing no conversion to cage, and in the presence of C<sub>60</sub> (bottom spectra, aldehyde starting material \*, **B1** [4+6] cage \*), both carried out at the same concentration (0.0014 M with respect to triamine), indicating C<sub>60</sub> has a templating effect during the cage formation.

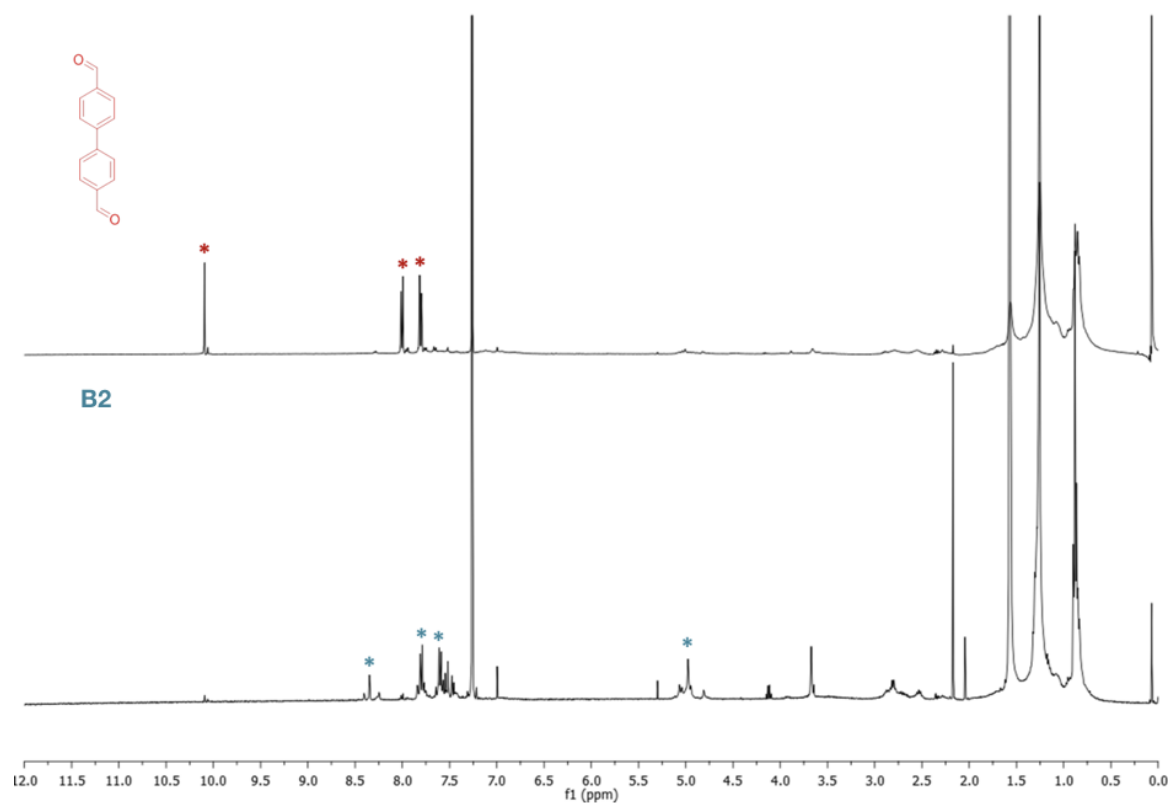


Figure A.18: Stacked  $^1\text{H}$  NMR spectra for the attempted formation of **B2**, without  $C_{60}$  (top spectra, aldehyde starting material \*) showing no conversion to cage, and in the presence of  $C_{60}$  (bottom spectra, aldehyde starting material \*, **B2** [4+6] cage \*), both carried out at the same concentration (0.0014 M with respect to triamine), indicating  $C_{60}$  has a templating effect during the cage formation.

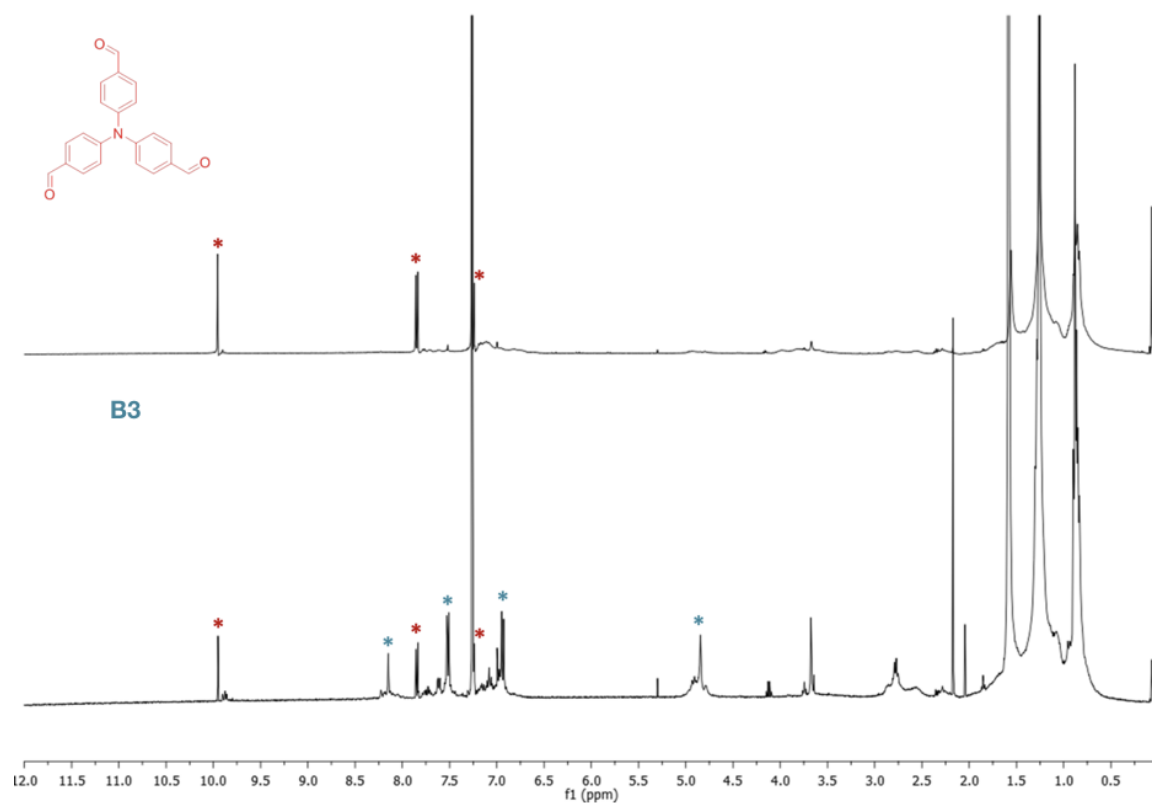


Figure A.19: Stacked <sup>1</sup>H NMR spectra for the attempted formation of **B3**, without C<sub>60</sub> (top spectra, aldehyde starting material \*) showing no conversion to cage, and in the presence of C<sub>60</sub> (bottom spectra, aldehyde starting material \*, **B3** [4+4] cage \*), both carried out at the same concentration (0.0014 M with respect to triamine), indicating C<sub>60</sub> has a templating effect during the cage formation.

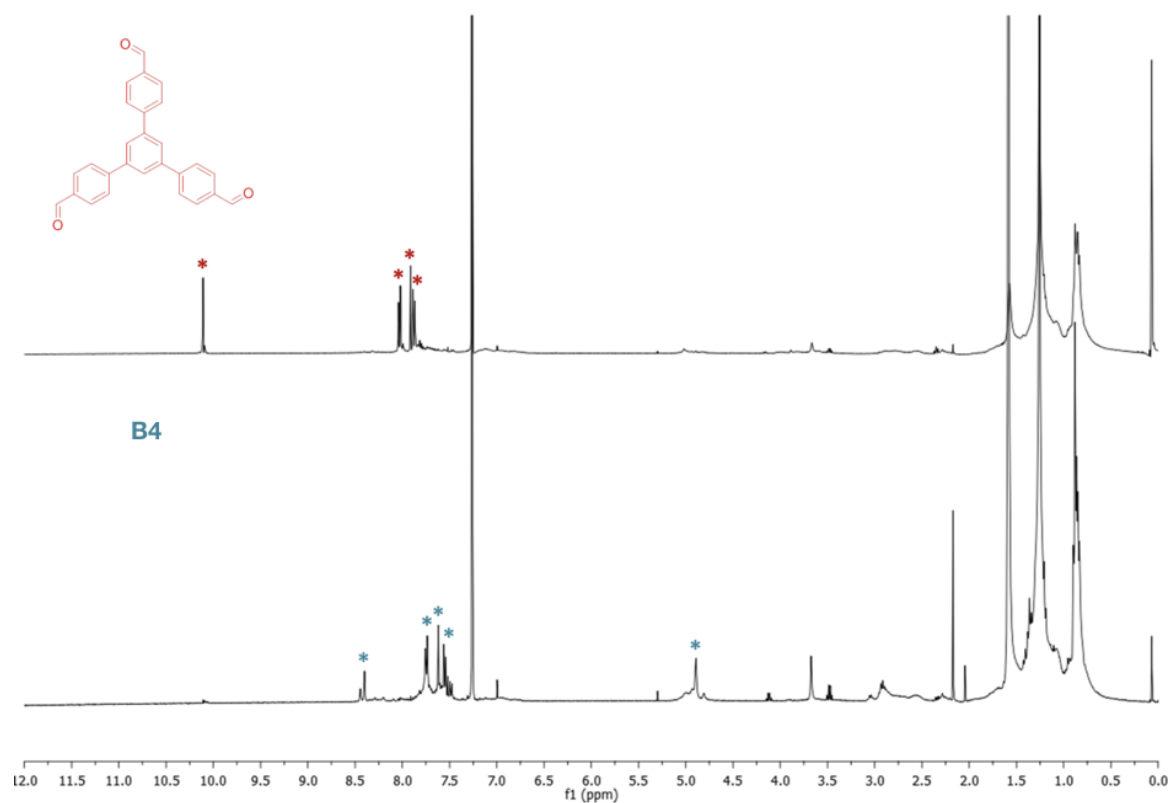


Figure A.20: Stacked  $^1\text{H}$  NMR spectra for the attempted formation of **B4**, without  $C_{60}$  (top spectra, aldehyde starting material  $*$ ) showing no conversion to cage, and in the presence of  $C_{60}$  (bottom spectra, aldehyde starting material  $*$ , **B4** [4+4] cage  $*$ ), both carried out at the same concentration (0.0014 M with respect to triamine), indicating  $C_{60}$  has a templating effect during the cage formation.

These cages were found previously to form at higher relative aldehyde concentration (0.0046 M) relative to the 2,4,6-Triethylbenzene-1,3,5-triyl)trimethanamine (0.020 mmol mL<sup>-1</sup>). In the repeated reactions of **B1**, **B2**, and **B3**, cages formed both in the absence and presence of  $C_{60}$  (see Figure A.19). This suggests that the inclusion of  $C_{60}$  in the dilute reactions had a templating effect on cage formation, but is not irreversibly bound inside the cage cavity.

However, in the repeated reaction of **B4**, there appears to be the presence of 2 distinct aromatic imine peaks at 8.44 and 8.40 ppm in a 1:1.5 ratio respectively, and an additional aromatic doublet clearly visible at 7.49 ppm. This is indicative of the presence of another cage species from the reaction containing  $C_{60}$ , different to that observed in the reaction without  $C_{60}$ , which may be of a 'bound' ( $C_{60}$ @**B4** complex) and 'unbound' cage species respectively (see Figure A.20)

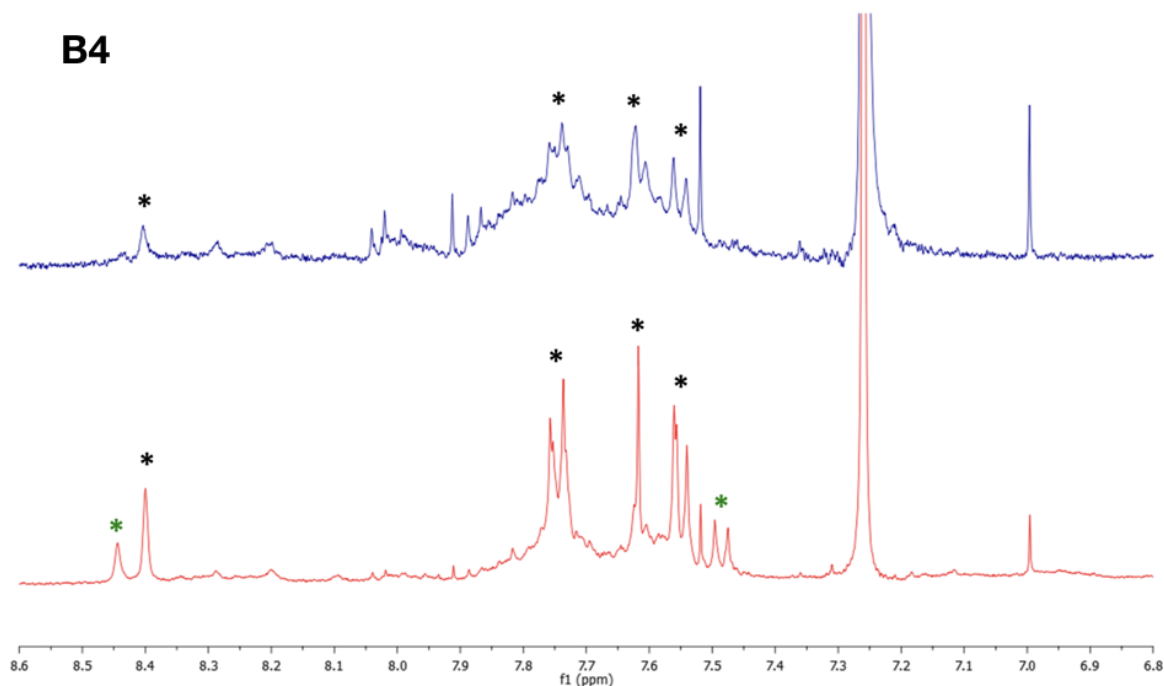


Figure A.21: Stacked <sup>1</sup>H NMR spectra of the aromatic regions for cage **B4** formed in the absence (top spectra, blue) and presence (bottom spectra, red) of  $C_{60}$ . Additional peaks at 8.44 ppm and 7.49 ppm (\*) indicate the presence of another cage species from the reaction containing  $C_{60}$ , different to that observed in the reaction without  $C_{60}$  (\*) which may be indicative of a 'bound' ( $C_{60}$ @**B4** complex) and 'unbound' cage species respectively.

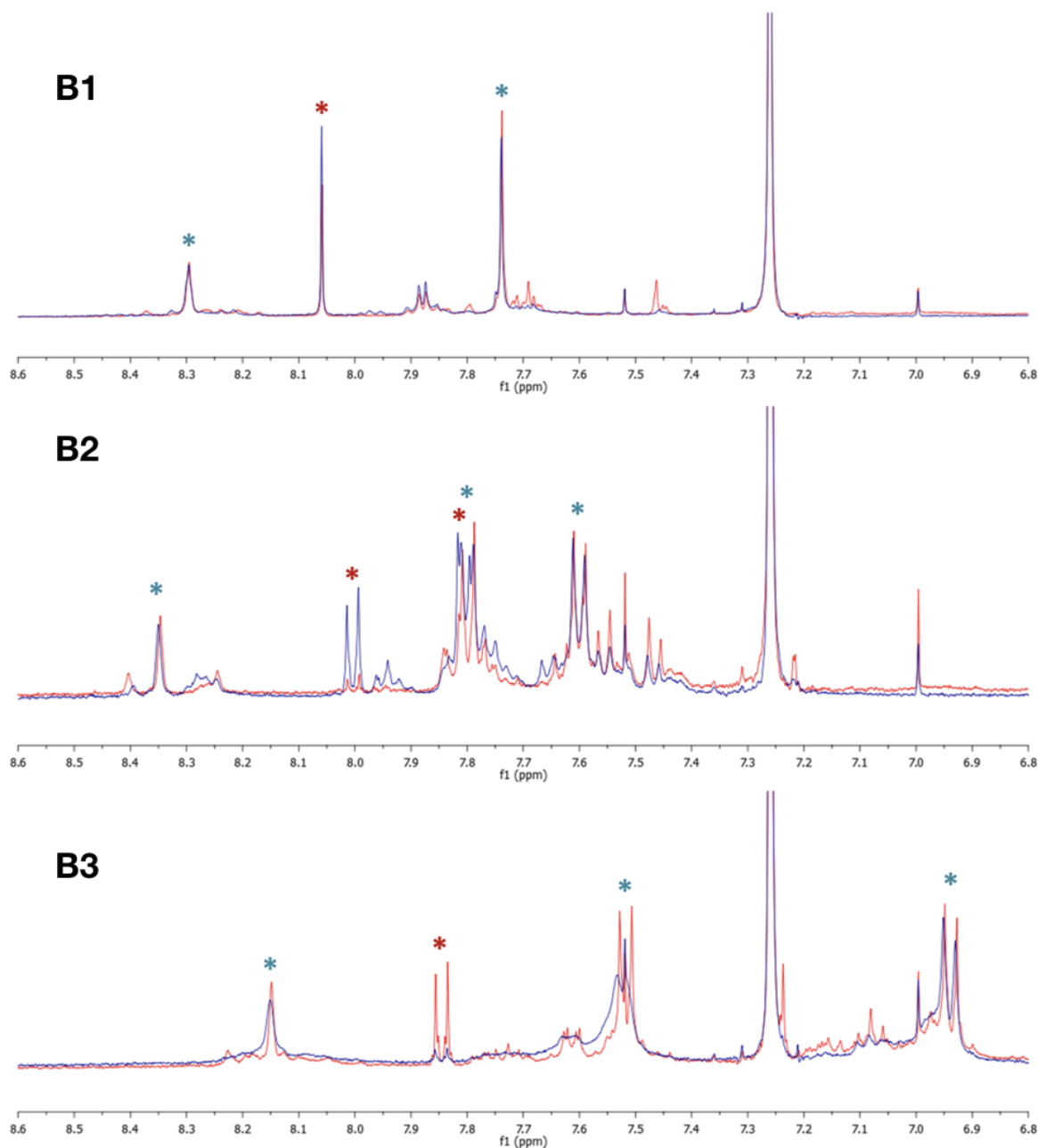


Figure A.22: Stacked overlaid  $^1\text{H}$  NMR spectra of the aromatic regions for cages **B1** (top spectra), **B2** (middle spectra) and **B3** (bottom spectra) formed in the absence (blue spectra), and presence (red spectra), of  $C_{60}$ . No clear peak shifts are observed between each pair of spectra, suggesting the  $C_{60}$  had a templating effect on cage formation but is not irreversibly bound inside the cage cavity.

## A.10 The trialdehyde building blocks and diamine linkers

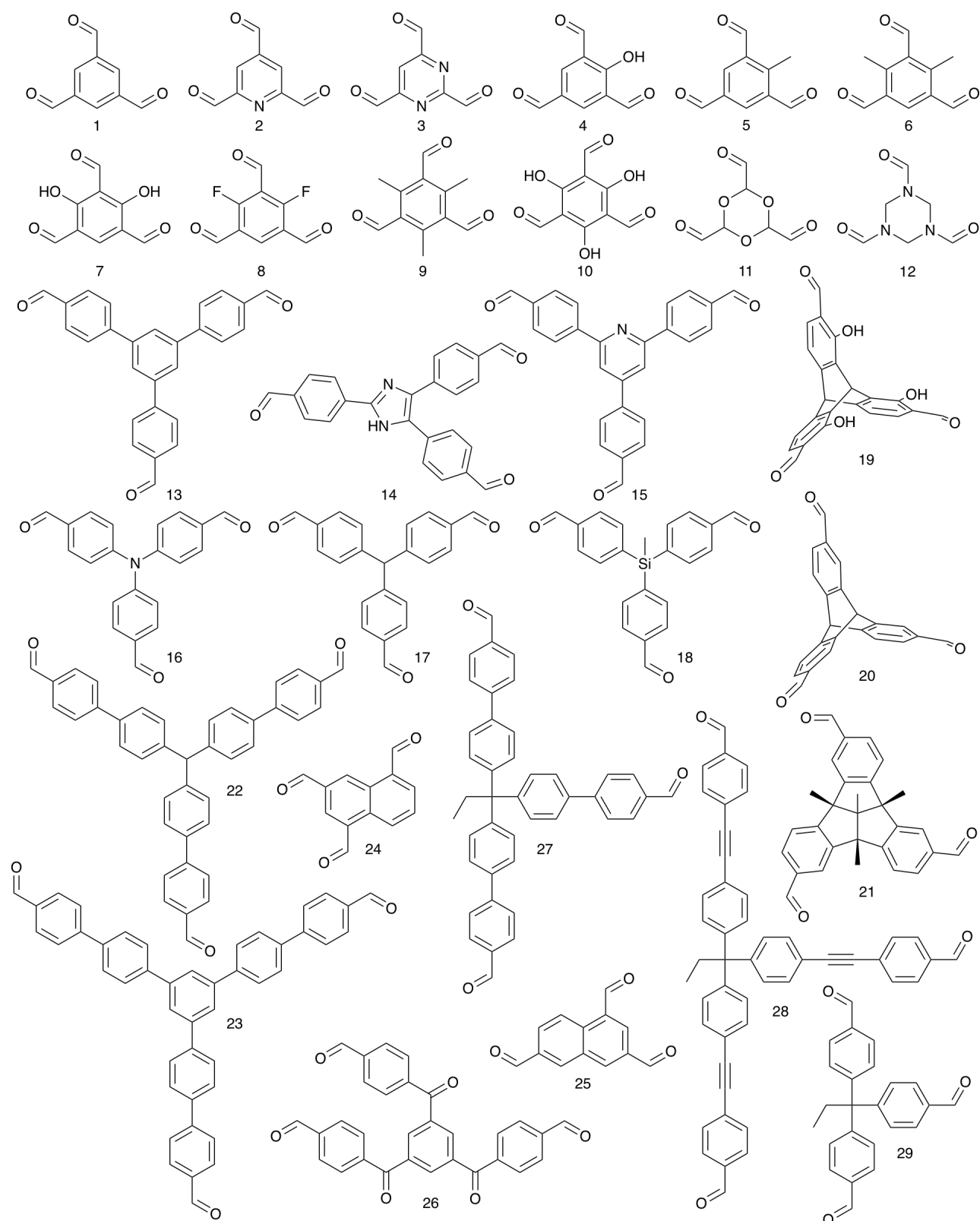


Figure A.23: The trialdehyde building blocks used in the assembly of POCs - part 1.

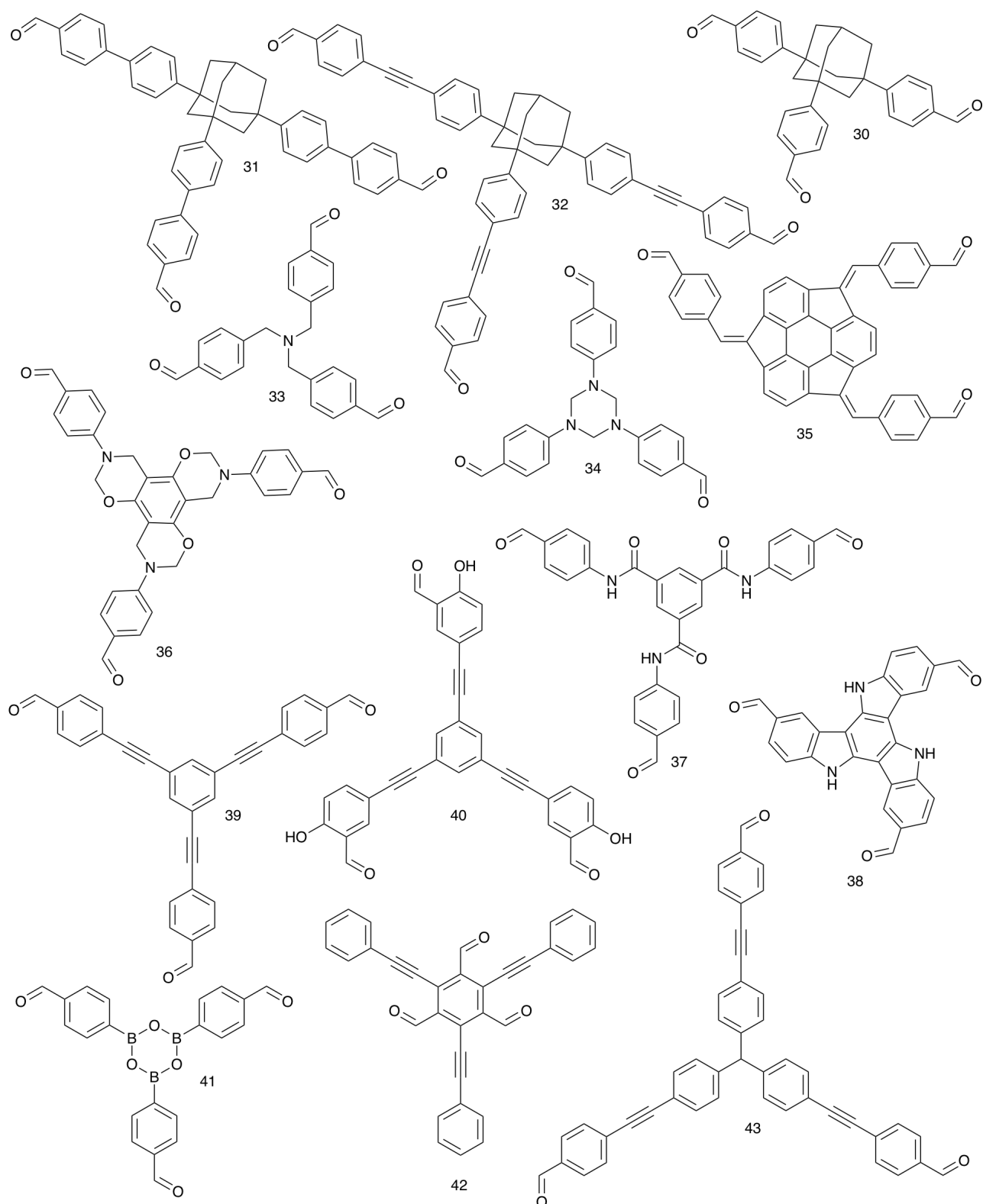


Figure A.24: The trialdehyde building blocks used in the assembly of POCs - part 2.



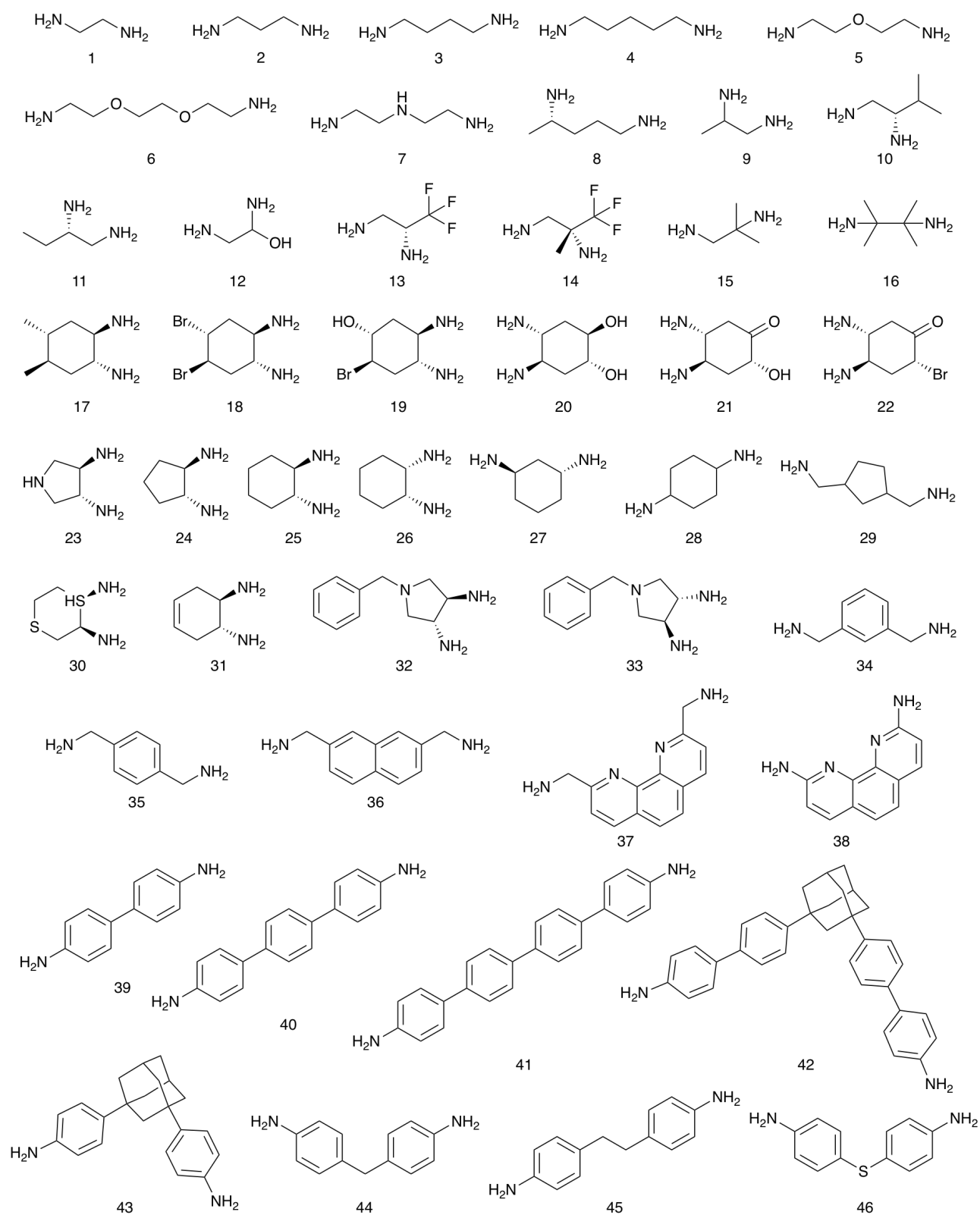


Figure A.25: The diamine linkers used in the assembly of POCs - part 1.

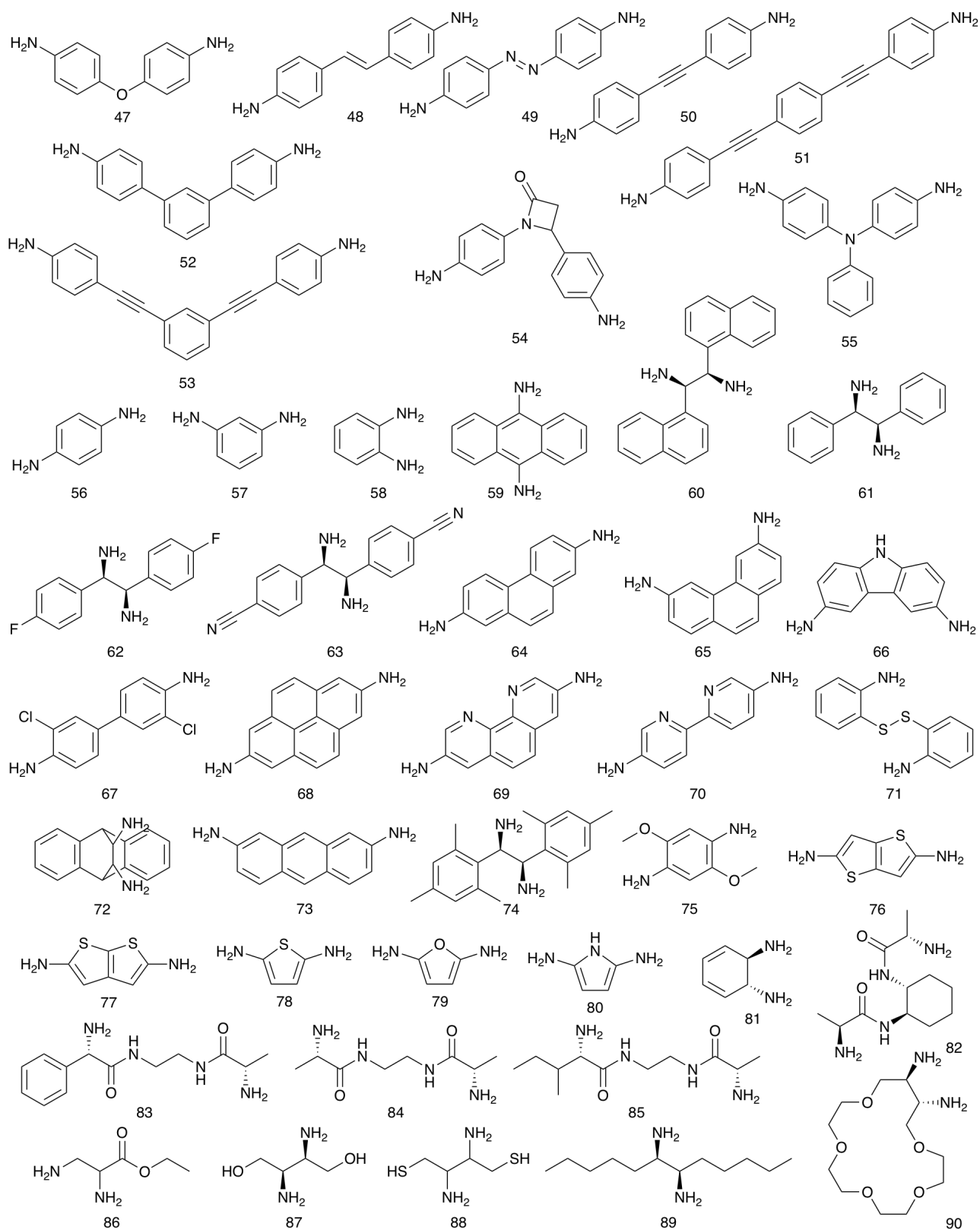


Figure A.26: The diamine linkers used in the assembly of POCs - part 2.