Unravelling the Origin and Influence of Point Defects in UiO-66 Through Computer Simulation

Martin Hutereau

A thesis submitted to University College London in partial fulfilment of the degree of Doctor of Philosophy

Supervised by Professor Ben Slater

March 2023

Department of Chemistry

20 Gordon Street

London

WC1H 0AJ

United Kingdom

Declaration of own work

I, Martin Hutereau, confirm that the work presented in my thesis is my own. Where information has been derived from other sources, I confirm that this has been indicated in the thesis.

Abstract

Density functional theory has been used to study point defects in the metal-organic framework UiO-66. The thermodynamics of building block vacancies has been surveyed in order to uncover the driving forces behind the formation of ordered nanoregions. While trends found with most charge-compensating species are consistent with experiment, the energetics of defects and domain interfaces suggest that these features should be well-dispersed rather than spatially aggregated and correlated. Examining the inorganic nodes in vacuum revealed that some linkers would be missing prior to framework assembly; this implies a thermodynamic impetus in addition to the kinetic factors conventionally used to explain defectivity. A number of experiments have been carried out using probe molecules to investigate the impact that these features have on acidic properties. In this way, the potential for tuning the strength of native Brønsted acid sites has been ruled out, but it was confirmed that these can be supplemented using charge-compensation schemes involving water. It was also affirmed that the Lewis acidity of zirconium is not strongly affected by nearby defect-terminating species, though these remain important in catalysis. It was found that in the cyclisation of citronellal, the local structure at vacancies affects how readily catalytic sites can be accessed, with steric hindrance a key factor. Finally, the heterometallic Zr-Ce framework was scrutinised in order to rationalise the distribution of cerium. The periodic material did not exhibit any preferences for the Zr_5Ce_1 building blocks seen experimentally: homometallic bonds are instead preferred. The inclusion of defects in the structure further revealed that higher framework flexibility allows cerium to be better accommodated, which has implications for synthesis. The possible role of Ce^{3+} was additionally considered: this species was found to be thermodynamically unavailable for the formation of building blocks during synthesis, providing a clue towards the prevalence of Zr_5Ce_1 over other mixed clusters.

Impact Statement

UiO-66 is an extensively researched member of an established class of porous materials known as metal-organic frameworks (MOF). It is unique among its peers in that it is able to host a remarkably high concentration of vacancies (defects) without severely compromising its stability. This has been exploited in past research to tune structural and chemical properties for the purposes of improving performance in processes of commercial and industrial interest. While this defect behaviour has been thoroughly documented from an experimental perspective, much remains unknown regarding how these features form and the influence they exert on the material's chemistry. The aim of this thesis was to further this understanding by using computational simulations to derive insights with atomic-level detail.

One outcome of this work is that a wide range of defect types have been evaluated on an equal basis so as to allow for fair comparison. This makes it easier to understand how the structures of vacancies will change as a result of the choice of reactants and post-synthetic processes. In parallel, the impact that these features have on catalytically important Brønsted (able to give away or receive a proton) and Lewis (able to give away or receive a pair of electrons) acid properties is now better known. With this knowledge, it will be possible to make more informed decisions on the synthesis and subsequent treatment of the material, such that it can be better optimised for a given application. Another result is that thermodynamics is now thought to play a role in the formation of defects at the solution stage of making UiO-66; this should direct future work towards this under-researched aspect of the synthesis. The way in which cerium is dispersed in the mixed framework was also investigated. A key conclusion of this is that the arrangements of cerium and vacancies are interlinked, providing both a design criterion for improving this material, as well as an interesting direction for future research. The choice of solvent and reactants was also shown to be important in this regard; a more

thorough investigation of these factors would enable better control over the uptake and distribution in UiO-66 of not just cerium, but potentially other problematic metals such as titanium.

Finally, the study carried out here has implications beyond UiO-66. A number of other MOFs share similar structural motifs, for which the insights found here may also be useful in designing better frameworks. To a lesser extent, this will also apply to other inherently defective MOFs with different chemistry, such as zeolitic imidazole frameworks.

Acknowledgements

I would firstly like to acknowledge UCL and EPSRC for funding this project and providing me with a positive research environment. My work would also not have been possible without the computational resources made available by the Materials Chemistry Consortium.

I would like to thank my supervisor, professor Ben Slater, for his invaluable guidance and support throughout the course of my doctorate. I am thankful to professor Angelos Michaelides for the advice he has given me as secondary supervisor. Recognition is also due to Zhao Baiwen and professor Richard Walton for our exchanges on cerium UiO-66; and members of the CMM in Ghent for interesting discussions on all things QuickFF. I would also like to thank the other members of the Slater group, particularly Tom and Dan, for making the office a fun and stimulating place to be in.

I want to thank my family for their unwavering support throughout my doctorate. I can hardly see having made it to the end without the encouragement from especially my parents and siblings, but also grandparents, uncles, aunt, and nephews. My friends have also been a constant source of support, and have believed in me from the beginning. I am grateful for all the interest in my work, no matter how alien the subject matter may at times have seemed.

A very special mention is reserved for my dear wife, who has lived through this alongside me. It would not have been the same without you, and I have no regrets over the path we have taken together. Thank you for being by my side.

And finally, I credit my grandfather for nurturing my interest in science from a young age. Had he been here today, I know he would have been tremendously proud and excited. *Merci, grand-père.*

Publication strategy

The work contained in this thesis has not yet been published. However, it is expected that a number of publications will arise from the different results chapters. This short section describes what tasks remain to be done and how these articles will be organised.

- **Chapter 3** focuses on the structure of defects in UiO-66 and correlated defective nanoregions. The article detailing the results from this chapter will discuss the thermodynamics of linker and cluster vacancies, domain interfaces, and building blocks. Existing data has already been assembled to form the bulk of a publication, but on-going work with force-field methods needs to be completed.

- **Chapter 4** presents two sets of results, on Brønsted and Lewis acidity, that may not necessarily belong in the same article. The latter, which involves the cyclisation of citronellal to isopulegol, has more bearing on tuning UiO-66 to improve catalyst design and so will be prioritised. All necessary experiments have been carried out, but more analysis of transition states will benefit our understanding of the controlling factors in this catalysis

- **Chapter 5** details work on Ce-Zr-UiO-66 and the distribution of cerium in this material. Collaboration with experimentalists and some related simulations are still on-going, and these should enrich the discussion on inorganic building blocks in this framework. It is expected that the existing results, along with an array of experimental data, will be presented together to detail the distribution of cerium and its implications on redox catalysis.

In addition to these planned works, the knowledge and experience of metal-organic frameworks gained during this thesis has already been put to use to contribute to the following review article:

- I.G. Clayson, D. Hewitt, M. Hutereau, T. Pope and B. Slater, *Adv. Mater.*, 2020, **32**, 2002780

Contents

Glossary **16**

1 Introduction **20**

1.1 Thesis overview 20

1.2 Metal-Organic frameworks and UiO-66 20

1.3 Defect fundamentals 26

1.3.1 Linker vacancies 26

1.3.2 Cluster vacancies and correlation 31

1.4 Defect tuning 35

1.5 Metal-mixing 37

1.6 Impact on material properties 40

1.6.1 Stability 40

1.6.2 Pore structure 42

1.6.3 Catalytic potential 44

2 Theoretical Background and Methods **59**

2.1 Hartree-Fock Theory 59

2.2 Density Functional Theory 66

2.2.1 Open-shell systems 71

2.2.2 Dispersion Interactions 73

2.2.3 Periodic calculations 75

2.2.4 Basis sets 77

2.2.5 Forces at the quantum mechanical level 80

2.3 Molecular Dynamics 81

2.3.1 Ensembles and Thermostats 83

2.4 Vibrational Analysis 85

2.5 Transition States 87

2.5.1 Elastic Band Method 90

2.5.2 The Dimer Method 93

2.6 Software details 96

3 Defects and Correlated Nanodomains 108

3.1 Motivation . . . 108

3.2 Approach . . . 114

3.2.1 Defining relative energies of defects . . . 114

3.2.2 Free energies . . . 117

3.2.3 Evaluating linker strain . . . 121

3.2.4 Constructing defects . . . 123

3.2.5 Interfacial energies . . . 126

3.2.6 Specific computational details . . . 127

3.3 Results and discussion . . . 128

3.3.1 Periodic defect structures . . . 128

3.3.2 Defect free energies . . . 132

3.3.3 Nanodomain interfaces . . . 137

3.3.4 Enumeration of defects around building blocks . . . 140

3.4 Conclusions . . . 147

4 The Influence of Defects on Acidic Properties 154

4.1 Motivation . . . 154

4.2 Approach . . . 158

4.2.1 Quantifying acidity . . . 158

4.2.2 Functional benchmarking . . . 160

4.2.3 Energy barriers . . . 164

4.2.4 Stereoisomers . . . 167

4.2.5 Specific computational details . . . 168

4.3 Results and discussion . . . 170

4.3.1 Adsorption of ammonia and pyridine . . . 170

4.3.2 Carbon monoxide frequencies . . . 180

4.3.3 Cyclisation of citronellal . . . 184

4.4 Conclusions . . . 195

5 Mixed Ce-Zr UiO-66 202

5.1 Motivation . . . 202

5.2 Approach . . . 206

5.2.1 Modelling cerium 206
5.2.2 Computational details 210
5.3 Results and discussion 211
5.3.1 Isolated Ce-Zr building blocks 211
5.3.2 Metal-mixing in the unit cell 213
5.3.3 Building blocks containing Ce^{3+} 218
5.4 Conclusions 228

6 Conclusions and Outlook **235**

Appendices **251**
A Chemical equations for defects 251
A.1 FA - formic acid 251
A.2 AA - acetic acid 251
A.3 TFAA - trifluoroacetic acid 251
A.4 BA - benzoic acid 251
A.5 $2\,H_2O$–OH 251
A.6 H_2O–OH 252
A.7 Cl–OH 252
A.8 FA–H_2O 252
A.9 Cl 252
A.10 OH 252
A.11 **e** 252
B Using Python to construct defects 253
C Supercell schematics 264
C.1 **fcu-reo** supercells 264
C.2 **fcu-bcu** supercells 264
D CO adsorption geometric data 266
E Cyclisation of citronellal transition states 267
F Equations for systems containing Ce^{3+} 270
F.1 Uncapped 270
F.2 FA 270
F.3 BA 271

F.4 DMF 271

List of Figures

1 Atom colour legend 15
2 Archetypal MOF examples 22
3 Schematic depiction of UiO-66 24
4 Building blocks of UiO-66 25
5 UiO-66 defect terminations 28
6 Correlated nanodomain schematics 34
7 Hartree-Fock self-consistency flowchart 65
8 Density functional theory self-consistency flowchart . . . 68
9 Schematic of a model potential energy surface 89
10 Schematic illustration of the elastic band method 92
11 Schematic of the dimer method construction 94
12 Electron microscopy images of defective regions 109
13 Missing linker defects in UiO-66 110
14 **bcu** inelastic neutron scattering fingerprints 112
15 Formate missing linker defect scheme 115
16 Formate missing cluster defect scheme 116
17 Schematics of mobile atoms for vibration 119
18 Evaluating linker strain 122
19 Correlation between inorganic nodes 124
20 Schematic of interfaces between the **fcu** and **reo** topologies 126
21 Breakdown of contributions to defect free energies 135
22 **fcu**-**bcu** interfacial energies 139
23 Defect enumeration with formic acid 141
24 Aperiodic building block schematics 143
25 Defect energies of isolated clusters 144
26 Dangling linker geometry in isolated clusters 145
27 Acidic environments in defective UiO-66 155
28 Zirconium decoordination events 157
29 Benchmarking structures for probe adsorption 160
30 Functional benchmarking for Brønsted acidity 161
31 Functional benchmarking for CO frequencies 162

32 Mobile atoms for CO frequency calculations 163
33 Collective variables for dangling step 165
34 Collective variables used for cyclisation step 166
35 Mobile atoms during vibrational analysis with citronellal 167
36 Stereoisomers of citronellal and isopulegol 168
37 Adsorption sites in defective UiO-66 171
38 Ammonia binding energies 172
39 Irregular binding geometries of ammonia 174
40 Secondary hydrogen bonds between ammonia and the framework 175
41 Pyridine binding energies 176
42 π-π interactions with pyridine 178
43 CO stretching frequencies 180
44 CO adsorption geometric parameters 181
45 CO frequency-geometry correlation 183
46 Conversion of citronellal to isopulegol 186
47 Opening up the coordination sphere of Zr 187
48 Barriers for accessing Lewis acid sites 189
49 Citronellal cyclisation energy barriers 192
50 Dispersion interactions during citronellal cyclisation . . . 194
51 Thermal stability of Ce-Zr UiO-66 203
52 XANES spectrum of Ce-Zr UiO-66 205
53 DFT+U benchmarking - Ce^{4+} 208
54 Aperiodic Ce-Zr building block relative energies 212
55 Periodic Ce-Zr UiO-66 relative energies 214
56 Linker distortions in Zr-Ce-UiO-66 215
57 Defective periodic Ce-Zr UiO-66 relative energies 216
58 Favourable defective Ce-Zr unit cell schematics 218
59 Local structure around Ce^{3+} 220
60 Spin density around Ce^{3+} 223
61 $5 \times 5 \times 5$ supercell schematics of UiO-66 238
62 UiO-66 building block oligomer 240

63 Alternate Zr and Ce building blocks 246

List of Tables

1 Energy cost of the **reo** defect 33
2 Thermodynamic ensembles used in molecular dynamics . 83
3 PHVA defect free energy benchmark 120
4 Energy cost of rotating SBUs in UiO-66 125
5 UiO-66 DFT defect energies 129
6 Dispersion contributions to defect energies 131
7 Free energies of linker and cluster defects 133
8 **fcu-reo** interfacial energies 137
9 Dispersion components of isolated cluster defect energies 146
10 CO mobile atoms benchmark 164
11 CO frequencies on coordinatively unsaturated Zr 188
12 Free energy barriers for accessing Lewis acid sites 190
13 Free energy barriers for citronellal cyclisation 193
14 DFT+U benchmarking - Ce^{3+} 209
15 Relative energies of Ce^{3+} reference states 221
16 Relative energies of building blocks containing Ce^{3+} . . . 224
17 Redox energies of cerium clusters 226

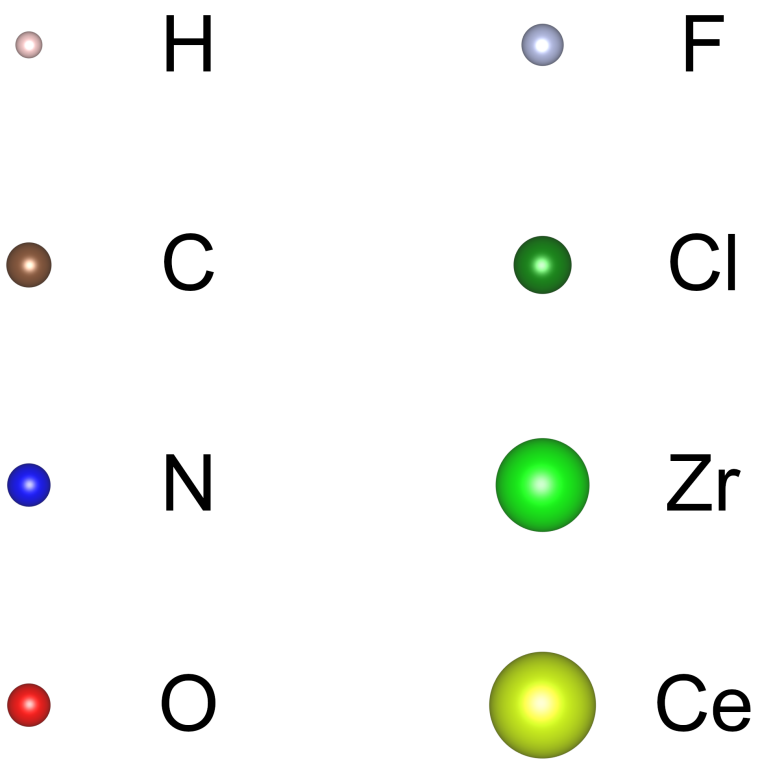


Figure 1: Except where indicated, atoms in diagrams contained within this work follow this colouring scheme. The relative sizes used do not follow any particular rules.

Glossary

bcu A topology with 8-connected nodes arranged in a body-centred cubic manner.

fcu A topology with 12-connected nodes arranged in a face-cubic centred manner; adopted by UiO-66.

reo A topology with 8-connected nodes and 2-connected linkers, typified by ReO_3.

scu A topology with half of nodes 4-connected, and the other half 8-connected.

AA Acetic acid.

ADMM Auxiliary density matrix method.

AIMD *Ab initio* molecular dynamics.

BA Benzoic acid.

BDC Benzene dicarboxylate: the linker building block in UiO-66. Also called terephthalic acid.

BET Brunauer–Emmett–Tell (surface area).

BLYP Becke–Lee–Yang–Parr (density functional).

bnn A topology with 5-connected nodes and nominally 2-connected linkers.

BO Born–Oppenheimer (approximation).

bpq A topology with 6-connected nodes, arranged in 1-dimensional chains, and 4-connected linkers.

CAN Cerium(IV) ammonium nitrate.

CI-NEB Climbing-image nudged elastic band.

CSVR Canonical sampling through velocity rescaling (thermostat).

DFT Density functional theory.

DMF N,N-Dimethyl formamide.

EDX Energy-dispersive X-ray (spectroscopy).

EXAFS Extended X-ray absorption fine structure.

FA Formic acid.

GGA Generalised gradient approximation.

GPU Graphical processing unit.

GTH Goedecker–Tetter–Hutter (pseudopotential).

HF Hartree-Fock.

HKUST-1 Hong Kong University of Science and Technology 1, a MOF.

HOMO Highest occupied molecular orbital.

HP Hierarchically porous.

HRTEM High-resolution transmission electron microscopy.

INS Inelastic neutron scattering.

IR Infrared (spectroscopy).

IT-NEB Improved-tangent nudged elastic band.

KS Kohn-Sham.

LCMT Ligand-to-metal charge transfer.

LDA Local density approximation.

LUMO Lowest occupied molecular orbital.

MD Molecular dynamics.

MIL-53 Matériaux de l'Institut Lavoisier 53, a MOF.

MOF Metal-Organic Framework.

MOF-74 Metal-organic framework 74, a MOF.

MOF-808 Metal-Organic Framework 808, a MOF.

multivariate A multivariate MOF contains more than one type of linker or SBU, though this usually refers to the former.

NEB Nudged elastic band.

NMR Nuclear magnetic resonance.

NU-1000 Northwestern University 1000, a MOF.

PBE Perdew–Burke–Ernzerhof (density functional).

PXRD Powder X-ray diffraction.

QMMM Quantum mechanics/molecular mechanics, a simulation technique in which different parts of the system are treated at quantum and force-field levels.

QuickFF Quick Force Field.

SBU Secondary Building Unit: the inorganic nodes in MOFs. Also called Zr brick, Zr cluster, inorganic cornerstone.

SED Scanning electron diffraction.

sod A topology with 4-connected nodes and 2-connected linkers; typified by sodalite.

tbo A topology with 4-connected nodes and 3-connected linkers; typified by twisted boracite.

TCM Three-cluster model; a model used to rationalise the trend in thermal stability in Ce-Zr UiO-66. It proposes that the framework is built almost exlusively from Zr_6,Ce_6, and Zr_5Ce_1 SBUs.

TEMPO (2,2,6,6-Tetramethylpiperidin-1-yl)oxyl.

TFAA Trifluoroacetic acid.

TGA Thermogravimetric analysis.

UiO-66 Universitet i Oslo-66, the titular MOF of this thesis.

XANES X-Ray absorption near edge structure.

XRD X-ray diffraction.

ZIF-8 Zeolitic imidazole framework 8, a MOF.

1 Introduction

1.1 Thesis overview

This doctoral thesis focuses on a metal-organic framework (MOF) known as UiO-66 and its defect behaviour. The aim of this work has been to use computational methods to build upon the current understanding of vacancies for this material. This document is organised into 6 chapters which will be briefly described here. The remainder of **Chapter 1** broadly introduces MOFs and UiO-66. The existing literature on this material and its defects is presented in order to contextualise and motivate this research. **Chapter 2** covers the general computational techniques used here to simulate UiO-66 and its building blocks as well as details on the software packages with which these were applied. **Chapter 3** focuses on the thermodynamics of the different vacancies found in this material in an attempt to explain the existence of experimentally-reported defective nanoregions. Both the periodic framework and its aperiodic building blocks are considered in this context. **Chapter 4** explores the relationship between UiO-66's acid sites and vacancies. This includes modelling a catalytic reaction within different defective structures to evaluate the impact on activity. **Chapter 5** examines cerium-substituted versions of UiO-66 to rationalise unusual experimentally-reported trends in metal distributions. As part of this study, the redox chemistry of cerium and the possible role of Ce^{3+} are taken into account. Finally, **Chapter 6** broadly discusses the key findings and conclusions of the current work. Open questions which remain unanswered are highlighted and possible directions for further research are proposed.

1.2 Metal-Organic frameworks and UiO-66

Metal-organic frameworks (MOFs) are a class of nano-porous materials comprising inorganic clusters connected together by organic ligands.[1,2] The first synthesis of a MOF dates back to 1994: Yaghi *et al.* described a 3D open framework of tetrahedral Cu(I) centres connected together by

4,4-bypiridine ligands.[3] Although similar systems have continued to be studied, often under the umbrella of similar structural archetypes such as zeolites and coordination polymers, it wasn't until 2004 that the current broad definition of a MOF was formalised by Rowsell and Yaghi.[4] Thereafter, interest in these systems continued to grow, involving an ever-increasing number of building blocks and synthesis techniques to generate more intricate and exotic frameworks. These have been almost exclusively synthetic; only a handful of naturally-occurring minerals have been characterised as analogous to MOFs.[5] Today, these materials represent an important growing field of the order of more than 10000 publications every year.[6]

MOF structure can be minimally described by 3 pieces of information: the namesake organic and inorganic constituents, and the manner in which these interconnect. The first of these, sometimes referred to as the 'linker', can be broken down into a backbone and functional end groups. The backbone is primarily carbon-based, but may be appended with functional side groups. The ends typically contain oxygen, nitrogen, or phosphorus atoms which form strong bonds with metals, most often in the form of carboxylate,[12] imide,[13] sulphonate, or phosphonate groups.[14] The inorganic node, often called secondary building unit (SBU), is usually built from transition metal ions,[9,15,16] but can also contain alkali and alkali earth metals,[17,18] as well as p-block metals such as aluminium or lead.[19,20] These also range from single-atom to multi-atomic, mixed-metal clusters with bridging oxygen and hydroxy ligands. Finally, there is significant diversity in how these blocks can be combined together into networks, or reticules. The ranges of linker topicities (from ditopic to hexatopic) and cluster connectivities (up to 12) lead to a large number of 2D and 3D structures that can be described by a topological net, detailed by a 3 or 4-letter code.[21] Some well-known frameworks are shown in **Figure 2** to illustrate different structures that can be constructed by combining together various SBUs and linkers.

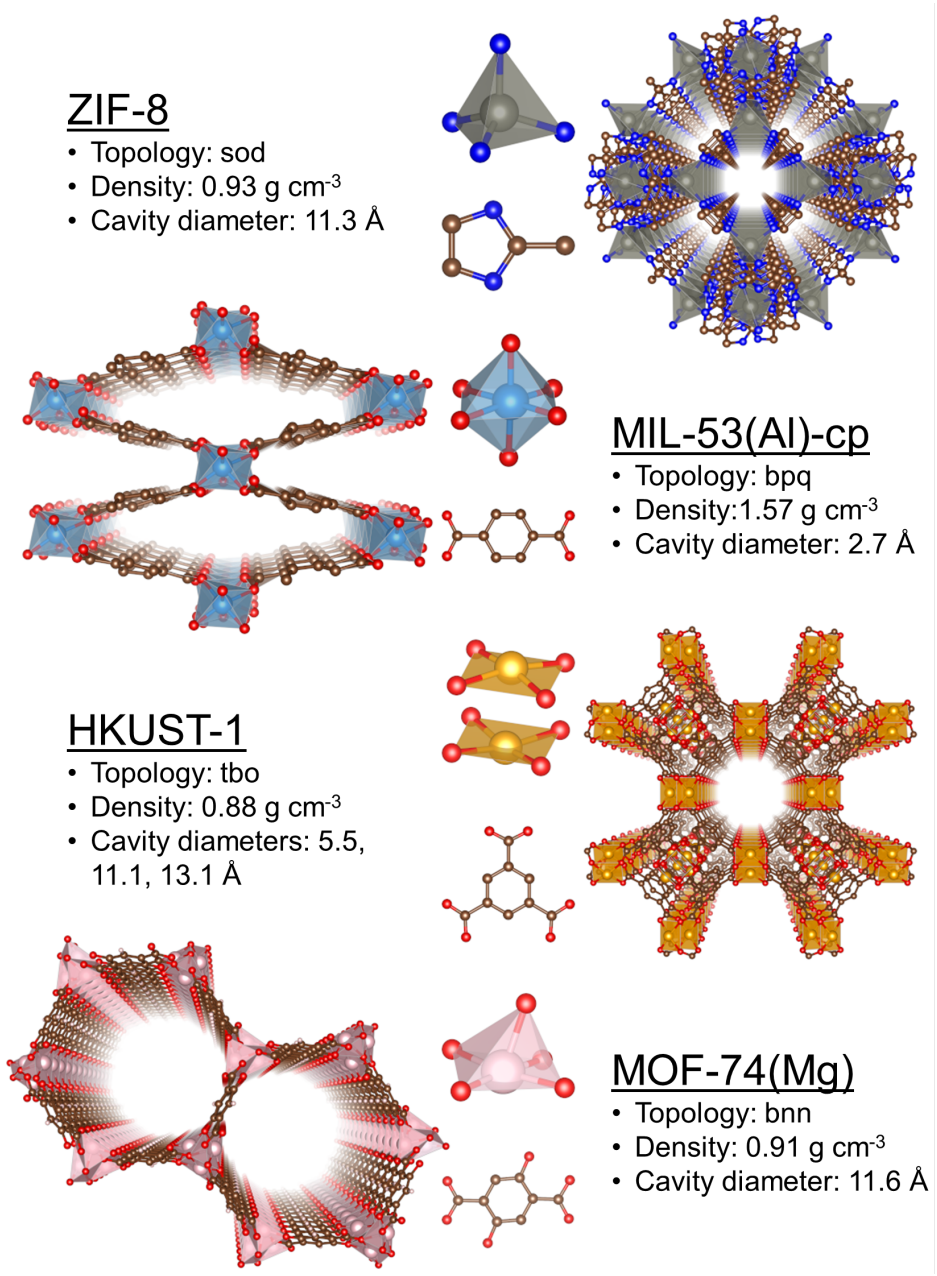


Figure 2: Schematic depictions of some archetypal MOFs intended to represent a range of different SBUs, linkers, and topologies. The data shown has been derived from structure files taken from references [7–10]; cavity diameters have been estimated via pore size distribution calculations carried out in Zeo++.[11] Note that the structure of MIL-53(Al) given here is the closed-pore (cp) form. All solvent molecules have been removed and some protons have been omitted from the illustrations. The additional colours used here correspond to Zn (grey), Al (light blue), Cu (orange), and Mg (pink).

The myriad available building blocks, as well as the manner in which they can be meshed together, allow for an incredible range of possible frameworks. This represents equally rich physical and chemical properties, for example: porosity can be modified by incrementing the linker size;[22] mechanical stability can be maximised using highly-connected nodes;[23] affinity towards certain adsorbates can be optimised by using functionalised linkers;[24,25] and catalytic activity can be altered through the choice of metal.[26] This diversity is epitomised by the number of realised and hypothesised MOF structures, available in databases such as the Cambridge Structural Database MOF subset, CoreMOF, hMOF, TOBACCO, and TABASCO.[27–31] However, even this fails to capture the chemical space encompassed by these frameworks, as synthetic and post-synthetic techniques can be further used to generate hierarchical pores and complex domain structures, tune the presence of defects, introduce additional functional groups, and even graft metal oxide nanoparticles.[32–36]

It is this range of properties and the scope for modification that has attracted so much attention towards MOFs, as reflected by the breadth of possible applications. A wide range of porosities and 3-dimensional channel systems with varying window sizes has led to interest for gas storage, greenhouse gas capture, water harvesting, and molecular separation processes.[37–40] The abundance of well-defined and tunable metal sites with variable coordination allow for a range of catalytic behaviour, including Lewis and Brønsted acidity, photocatalysis, redox, and the breakdown of nerve gas simulants.[41–45] Careful control over electronic structure has also led to the development of frameworks with magnetic properties or with uses in batteries and as semiconductors.[46–48] The low cytotoxicity of certain MOF building blocks has also driven research on bio-compatible frameworks (including a new family termed 'bio-MOFs') for targeted drug delivery, sensors and imaging.[49–51] Although the commercial use of MOFs is currently limited to niche applications, they remain an incredibly versatile and promising archetype which will continue to grow.

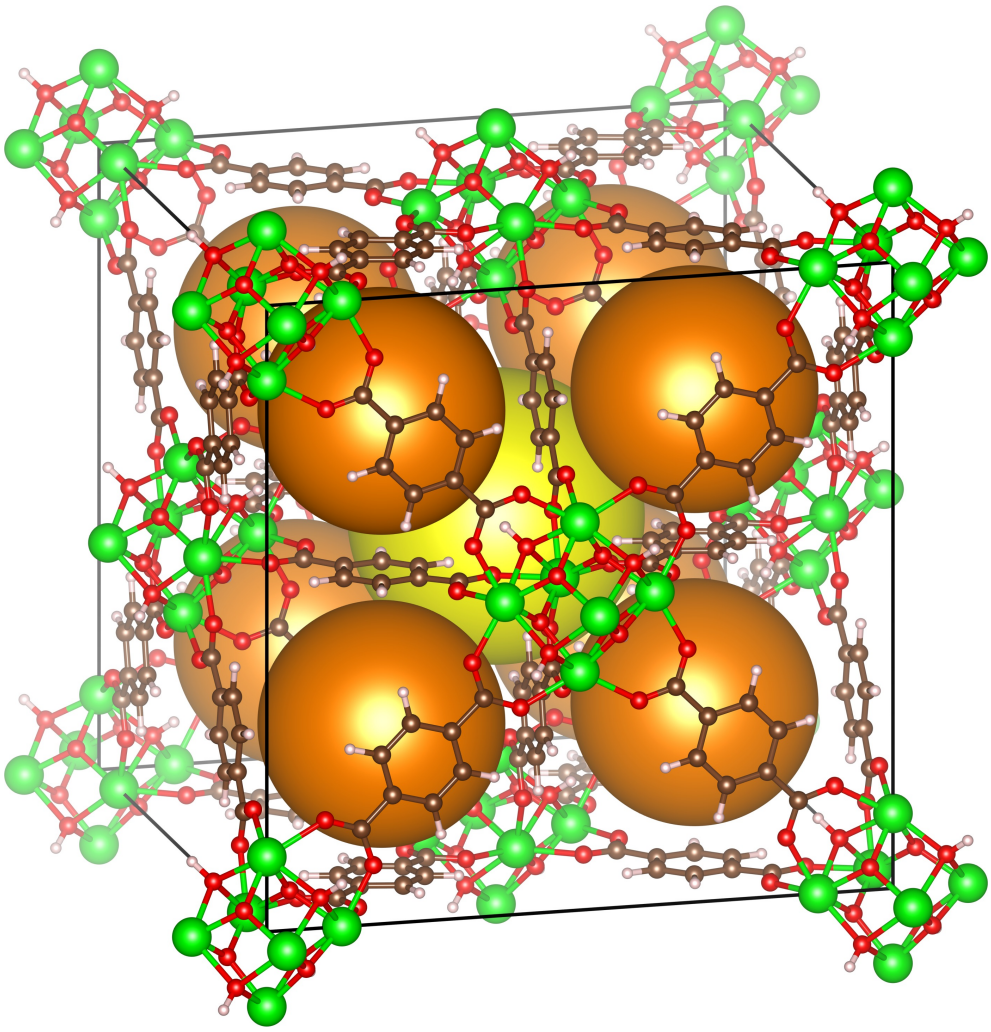


Figure 3: The conventional unit cell of pristine UiO-66. Zr is held together into $Zr_6O_4(OH)_4$ octahedra by bridging oxygen and hydroxyl groups. Each of these is connected to 12 other SBUs via BDC linkers distributed in planes parallel to (100), (010), and (001); the cell contains in total 4 of the inorganic and 24 of the organic building blocks, with an overall composition $(Zr_6O_4(OH)_4)_4(C_8O_4H_4)_{24}$. It is easy to see the **fcu** topology from the face-centred distribution of Zr nodes. The central octahedral and secondary tetrahedral cavities are shown respectively as yellow and orange spheres.

UiO-66 is a MOF built from octahedral $Zr_6O_4(OH)_4$ SBUs and ditopic 1,4-benzene dicarboxylate (BDC) ligands, first synthesised in 2008 by Cavka *et al.*[52] A schematic representation of it can be seen in **Figure 3**. The individual building blocks are additionally detailed in **Figure 4**. Each inorganic node is connected to 12 others, an unusually high number, giving an overall topology of **fcu**. This redundancy in connectivity, combined with strong Zr–O bonds, leads to exceptional stability: thermal degradation does not begin until ∼750 K; amorphisation requires a pressure exceeding 1.3 GPa;[53] and the material is resilient towards aqueous conditions over a pH range of 1-8.5.[54,55] The moderate linker size and network connectivity result in tetrahedral and octahedral cavities that offer good porosity, with Brunauer–Emmett–Tell (BET) surface areas over a range of 1100-1500 m^2g^{-1}.[56] UiO-66 is also a platform for other $Zr_6O_4(OH)_4$ MOFs; isoreticles are available via linker expansion and varied topologies can be reached by altering linker topicity (in frameworks such as MOF-808 and NU-1000).[57,58]

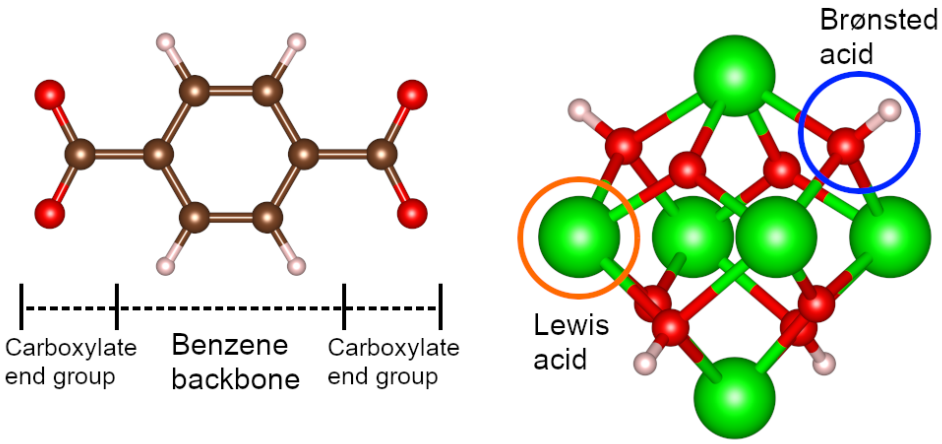


Figure 4: The building blocks of UiO-66. Left: the BDC linker ligand (also known as terephthalate) with its end groups and backbone labelled. Right: the $Zr_6O_4(OH_4)$ SBU with some of its μ_3 hydroxyl Brønsted and Zr^{4+} Lewis acidic groups labelled. Note the arrangement of protons shown here which maximises their separations; this is the global minimum for this type of inorganic building block.[59]

Much of the interest towards UiO-66 has been due to its stability. Chemical and thermal robustness mean linkers and SBUs can easily be functionalised post-synthetically, making the framework an ideal platform onto which further features can be built. This durability also means the Lewis and Brønsted acidic groups on the SBUs can be effectively exploited as catalytic sites for various reactions. Additionally, this MOF exhibits exotic defect behaviour through linker and node vacancies, which can alter its chemical and mechanical properties. These defects have been the subject of extensive research, both from a fundamental perspective and as a means of tuning the material for specific applications. As a result of these characteristics, UiO-66 has been one of the most thoroughly studied frameworks, to the extent that it has been the singular subject of entire reviews.[41,60,61]

1.3 Defect fundamentals

1.3.1 Linker vacancies

The presence of defects in UiO-66 was first noted in 2011 as part of detailed structural characterisation work by Valenzano *et al.*[62] Thermogravimetric analysis (TGA) for this framework revealed a mass loss corresponding to a sub-stoichiometric amount of BDC. This discrepancy with respect to the perfect material was attributed to vacancies at linker positions. A later neutron diffraction experiment in 2013 by Wu *et al.* confirmed that the structure is best described by a fractional occupancy of ~ 0.9 on linker sites, implying both vacancies and some degree of disorder about them.[63] These 2 studies initially established the presence of missing linker defects in UiO-66 and confirmed the suitability of TGA and neutron diffraction for quantifying their concentration. Although other techniques have been applied to identify and measure missing linkers in this material, TGA remains the most common due to its ease of application.[36]

While the presence of defects in UiO-66 was quickly acknowledged and the measurement of missing linkers became commonplace, little was known regarding atomic structure at these sites. Introducing a vacancy in the material leaves the framework with a +2 charge, necessitating some charge-compensating species. Valenzano *et al.*'s initial study proposed that non-native hydroxide anions played this role.[62] Subsequent XRD work by Øien *et al.* and Trickett *et al.* narrowed this down to a cluster of 2 water molecules and a hydroxyl species.[64,65] Trickett's motif had the hydroxyl lying directly over a μ_3 hydroxyl on the node (see **Figure 5** panel a). Subsequent potentiometric acid-base titrations by Klet *et al.*, which revealed 3 distinct equivalence points, seemed to confirm these findings.[66] Nevertheless, the likely instability of this arrangement prompted Ling and Slater to reconstruct this defect and simulate it dynamically using *ab initio* molecular dynamics (AIMD) at the hybrid density functional theory (DFT) level.[67] They found that protons dynamically exchange between the capping species at the defect, but that the hydroxyl is ultimately bonded directly to zirconium (**Figure 5** panel b). At sufficiently elevated temperatures, one of the water molecules diffuses away into the cavity (leaving behind the geometry seen in **Figure 5** panel c). Detailed DFT work by Vandichel *et al.* confirmed this new geometry to be the most favourable and also linked this water motif to dehydration processes seen experimentally.[68]

Another possibility for the capping species is a deprotonated monocarboxylic acid. In Wu *et al.*'s neutron diffraction study, inelastic neutron scattering peaks corresponding to tunneling between methyl rotational ground states were observed, even after the framework was evacuated of loosely-bound molecules.[63] It was postulated that the acetic acid (AA) additive used during synthesis was playing a structural role in these defects. Vermoortele *et al.* examined similar samples made using trifluoroacetic acid (TFAA) and HCl instead. Energy-dispersive X-ray (EDX) measurements revealed that when only HCl was used, Cl^- was present

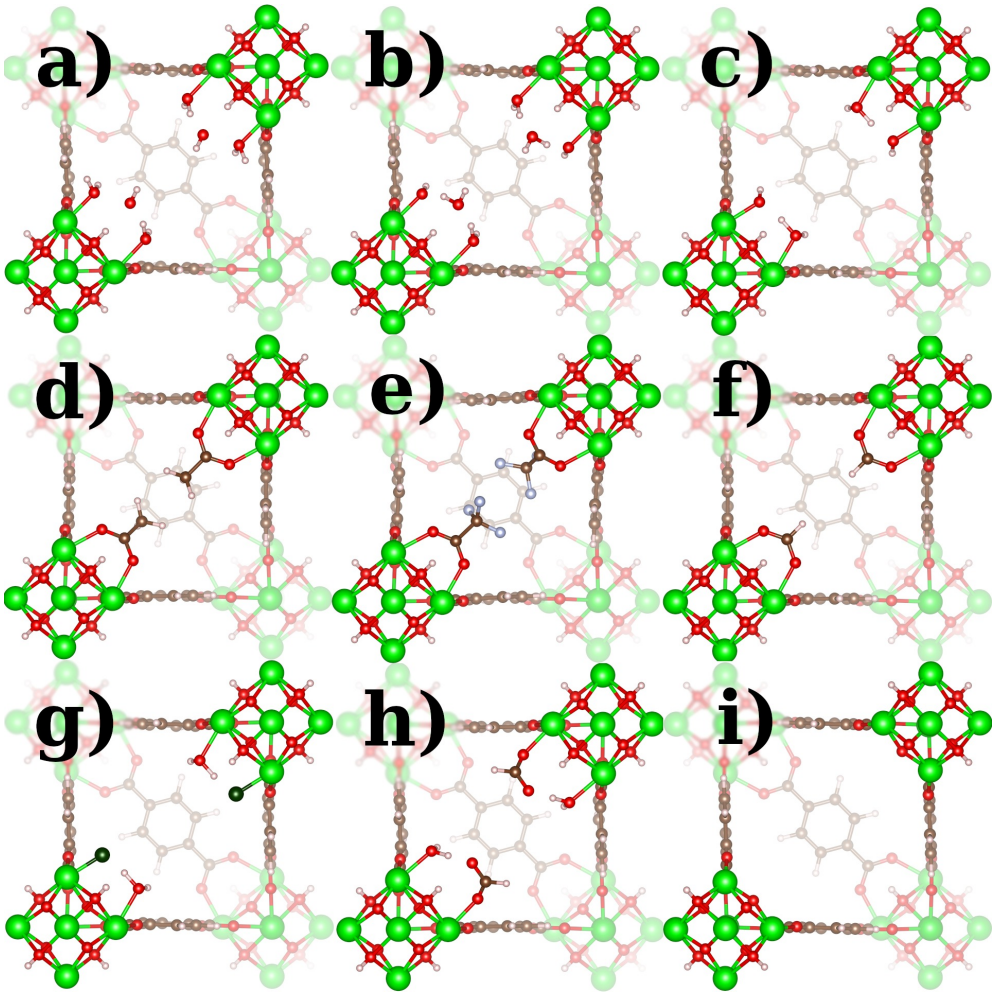


Figure 5: Different charge-compensation schemes for missing linker defects. a): The proposed structure of Trickett *et al.* involving an additional hydroxide anion sitting atop the native μ_3-OH group and flanked by two further water molecules.[65] b), c): The geometries containing a hydroxide anion bound to Zr and water molecules found by Ling and Slater during low and high temperature AIMD simulations, respectively.[67] d), e), f): replacement of the dicarboxylate linker respectively by two monocarboxylate acetate, trifluoroacetate, or formate anions. g): water and a chloride anion, connected via a hydrogen bond, bound to adjacent zirconium atoms. f): The structure proposed by Tan *et al.* for a formate-water pair.[69] g): the unterminated **e** defect in which protons have been removed from both nearby μ_3 bridging hydroxyl groups.

in the framework in significant quantities, whereas when both acids were used, ^{19}F nuclear magnetic resonance (NMR) pointed exclusively towards TFAA being present.[70] Their supporting DFT calculations on the dehydroxylation of SBUs reproduced the same temperature behaviour for clusters with TFAA attached that are seen experimentally. Further DFT calculations from Vandichel *et al.* confirmed that, of the possible combinations of Cl^- and TFAA, a pure TFAA motif is the most stable.[71] Other monocarboxylic acids have been studied by Shearer *et al.*, including formic acid (FA), for charge compensation.[72] The authors made use of ^{1}H NMR on samples digested in NaOH to confirm for each acid used that they were present in the framework post-synthesis. To this end, the activation procedure was carefully controlled to remove all solvent and loosely-bound species. It was further noted that FA was found in all samples, even when this was not used as an additive; this was ascribed to the hydrolysis of the N,N-dimethyl formamide (DMF) solvent used in the synthesis, which yields FA. The cappings for AA, TFAA, and FA are shown in **Figure 5** panels d, e, and f.

One of the most comprehensive analyses of the capping species has been the comparative work of Bristow *et al.*[73] A force field parametrised using DFT data was used to compute free energy differences of different defect motifs (combinations of Cl^-, AA, H_2O, OH^-, and DMF) relative to the pristine material. Some solvent effects were also taken into account via implicit solvation and limited binding energy calculations. AA proved the most stable, with a free energy cost of ∼75 kJ mol^{-1}, followed by Cl^--H_2O (**Figure 5** panel g) at ∼100 kJ mol^{-1}. This appears to agree with the aforementioned work regarding monocarboxylic acids. Another complete study has been that of Wei *et al.*, which compared experimental infrared (IR) spectra with frequencies simulated via DFT across samples synthesised with a range of monocarboxylic acids.[74] Good agreement was seen for all species except acetic acid at low concentration; in this case the high-temperature water-hydroxyl motif of Ling and Slater correlated

well with the observed frequencies, in apparent contradiction with the results of Bristow *et al.*[67,73] This disagreement exemplifies the lack of consistency for the capping species between different syntheses. However, conditions are rarely uniform and will differ between experiments: molar ratios for the acidic modulator and BDC; crystallisation time and temperature; and the post-synthetic washing and activation procedures. Variation in these factors can lead to large differences in defect type and concentration from one study to the next.[75] Therefore, the species ultimately present at defect sites may depend more strongly on the synthesis and subsequent treatments than thermodynamic stability predicted by atomistic simulations.

The most recent proposed structures involve both water and monocarboxylic acids. Tan *et al.* have considered a geometry in which FA or AA is bound in unidentate fashion to one zirconium while the other is capped by water, with a hydrogen bond holding both species together (**Figure 5** panel h).[69] The authors argue this helps to explain the finer structure found in IR spectra around 1000-2000 cm^{-1}. These features disappear upon activation at 430 K and reappear upon exposure to ambient moisture, suggesting the loss of and subsequent return of the capping water. Supporting DFT calculations on bidentate monocarboxylic acid and their suggested structure show the latter is more stable by $\sim$ 30 kJ mol^{-1}; the forward energy barrier of only 17 kJ mol^{-1} and larger reverse barrier of 50 kJ mol^{-1} are consistent with the observed temperature-dependent IR behaviour.

For completeness, a further capping geometry will be mentioned, sometimes referred to as **e**. For this defect, the linker is removed and charge compensation is achieved by the removal of 1 proton from each connected Zr cluster (**Figure 5** panel i). This model has been used mainly where highly activated processes are being studied, such as photocatalysis.[76] This is therefore not representative of the equilibrium structure

of UiO-66, for which monocarboxylic acids and water-hydroxyl motifs are more suitable. Indeed, Vandichel *et al.* have shown via DFT that this structure has a free energy ∼60 kJ mol^{-1} higher than water, chloride or TFAA-based cappings.[71] Subsequent exposure to moisture from the solvent, impurities, or the environment progressively refills the Zr coordination sphere back to saturation.[77–79]

The studies discussed in this section suggest that the precise atomic structure of defects will depend closely on synthesis, activation, and operating conditions. Monocarboxylic acids appear to be the preferred cap, with formic acid (as a byproduct of DMF hydrolysis) an almost omnipresent candidate. At low water content, one expects these to be primarily bidentate, whereas the unidentate may be preferred when sufficient moisture is present. This initial species is controlled by the synthetic parameters. Subsequently, the activation step may remove water (converting unidentate caps to bidentate ones) at intermediate temperature, or even the monocarboxylates themselves at sufficiently high temperatures (supposedly generating **e** defects). When the sample is cooled down and re-exposed to air, these sites may be repopulated by water, thereby regenerating H_2O-OH or unidentate monocarboxylate-water motifs. It appears clear then that special care must be taken when computationally modelling defects to consider the most appropriate and relevant charge-compensation schemes based on the criteria discussed above: synthesis, activation, and operating conditions. Fortunately, DFT is capable of describing these structures appropriately and simulated IR frequencies can allow for convenient corroboration with experiment.

1.3.2 Cluster vacancies and correlation

The defect behaviour of UiO-66 becomes even more complex when consideration is given to cluster vacancies. By using very high FA : BDC ratios, Cliffe *et al.* have made samples of Hf-UiO-66 which exhibited symmetry-forbidden peaks in their PXRD patterns (and which are absent in the

'pristine' material).[80] The authors showed that this was consistent with a structural model in which a single zirconium node (and its 12 adjacent linkers) has been removed from the unit cell, leaving a cluster vacancy. As this changes the framework topology from **fcu** to **reo**, this is often referred to as the **reo** defect or structure. These are shown in **Figure 6** panels a and c. They generated a supercell in which SBUs were randomly removed (at most 1 per unit cell) until the experimentally measured missing linker content was reached; the same symmetry-forbidden reflections were found in generated patterns only for arrangements where missing clusters were aggregated into nanodomains of adjacent unit cells. Supporting DFT calculations revealed the **reo** cell parameters differed from the **fcu** ones by only ∼0.05 %, explaining the apparent miscibility between structures. Later work by Atzori *et al.* on the Zr material yielded good agreement between the experimental and DFT-simulated IR spectra of the **reo** unit cell, suggesting the missing cluster model is appropriate.[81]

The most striking aspect of these defects is their tendency to form nanodomains. This has driven further work seeking to uncover the driving force and mechanism through which this happens, as this understanding could allow for control over domain structure. Part of Bristow *et al.* 's study of capping species examined correlation between missing linker defects.[73] It was found that removing linkers from the same tetrahedral cage is favourable over distributing vacancies by a free energy of ∼25 kJ mol^{-1}. Moreover, the overall free energy cost of creating the defects decreases as more are introduced. Although both these findings suggest a significant thermodynamic component to defect correlation, the **reo**-type structure was only addressed in a follow-up study. Svane *et al.* carried out detailed DFT calculations to evaluate some informative metrics: the energy difference between distributed and aggregated missing linkers, and the energy required to create a **reo** defect from a linker-deficient unit cell.[82] The results are shown in **Table 1**.

Table 1: Energy cost of the **reo** defect

Capping	**e**	Cl–H_2O	FA	AA	TFAA
ΔE^6	16.4	-121.6	-2.9	9.6	70.4
$\Delta E^1_{\mathbf{reo}}$	-86.8	209.3	182.3	174.6	-68.5
$\Delta E^6_{\mathbf{reo}}$	-165.9	126.4	23.2	4.8	-79.1

Energies are given in kJ mol^{-1}. ΔE^6 is the energy difference between unit cells containing 6 aggregated versus distributed linker defects: $\Delta E^6 = E^6_{\text{aggregated}} - E^6_{\text{distributed}}$. $\Delta E^n_{\mathbf{reo}}$ is the energy of a **reo** unit cell relative to a cell in which n linkers only have been removed: $\Delta E^n_{\mathbf{reo}} = (\frac{6}{n} - \frac{3}{4})E_{\mathbf{fcu}} + E_{\mathbf{reo}} - \frac{6}{n}E^6$. $E_{\mathbf{fcu}}$, E^n, and $E_{\mathbf{reo}}$ are respectively the energies of the pristine unit cell and cells missing n linkers or a cluster. Table adapted from ref. [82].

Positive values for ΔE^6 imply that distributed linker defects are preferred over aggregated ones, which is the case for **e**, AA, and TFAA. It is likely that these caps required more significant structural relaxation to be accommodated for, which is only possible when they are distributed. Aggregation is favoured for Cl-H_2O as closer proximity between these species allows for stabilising electrostatic interactions and hydrogen bonding. These factors also explain the trend in $\Delta E^n_{\mathbf{reo}}$: the charge-compensating species around a missing cluster are spatially well-separated, which is favourable in the cases of **e** and TFAA. Note that the energy of the **reo** system relative to missing linkers is still low for FA and AA. Hence, the use of these acids is still expected to lead to some missing clusters, as is borne out by experiment. While this might also be expected for **e** defects, these should not play a role during synthesis given that they require framework activation under vacuum and high temperature. Finally, as the data in **Table 1** compares defective systems to each other, no conclusions can be drawn regarding their stability relative to the pristine material.

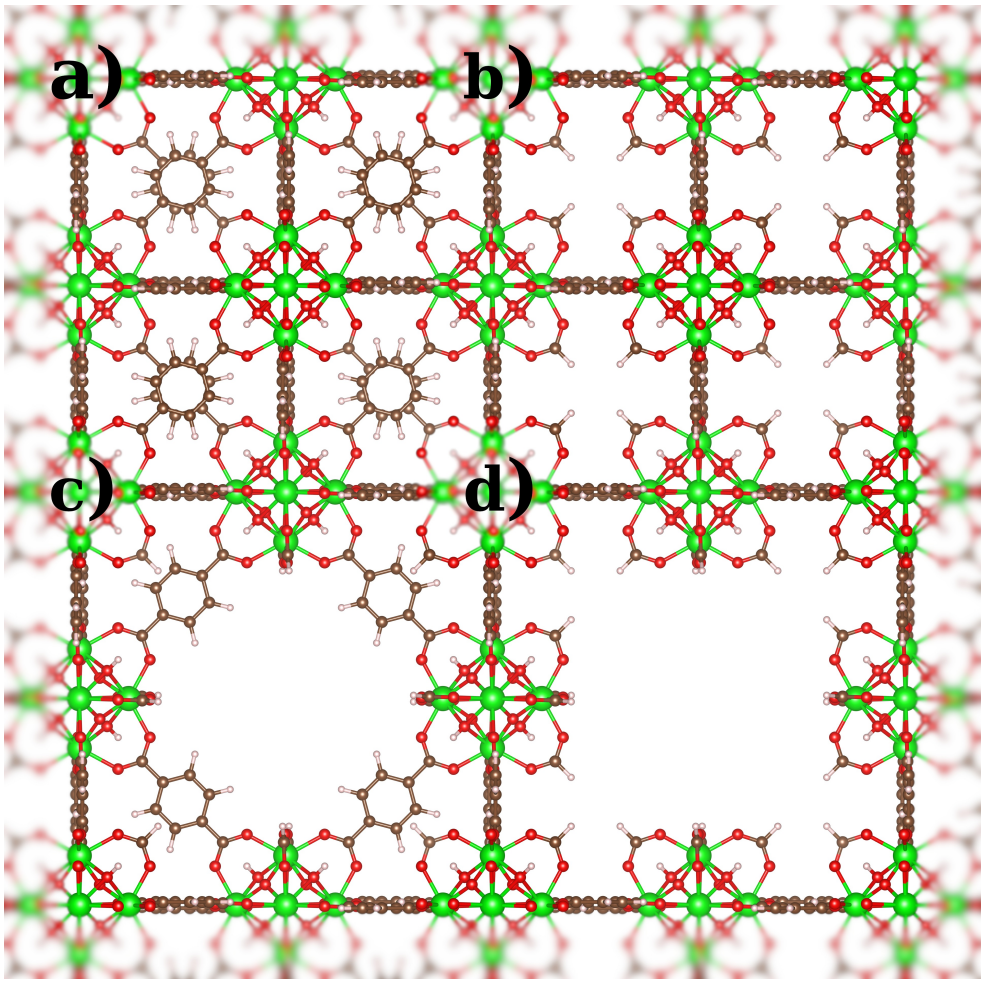


Figure 6: Different topologies of UiO-66. a) the (pristine) **fcu** topology. b) the **bcu** topology, containing 8 missing linkers. c) the **reo** topology, missing 1 cluster and 12 linkers. d) the **scu** topology, missing 1 cluster and 16 linkers.

More recently, it has become possible to see these nanodomains directly through electron imaging and diffraction techniques. Liu *et al.* made use of a very sensitive detector to capture high resolution transmission electron microscopy (HRTEM) images of samples synthesised with formic acid.[83] In this manner, they clearly identified **reo** regions and as well as 2 new defective topologies, **bcu** and **scu**, as depicted in **Figure 6** panels b and d, which also form nanoregions. Supplementary DFT calculations yielded relative (to **fcu**) energies for these unit cells with the same qualitative trend as their experimentally observed abundances. Subsequent scanning electron diffraction (SED) experiments from Johnstone *et al.*

also directly evidenced **reo** nanodomains in a geometry more blocky and lamellar than previously thought.[84] This led the authors to suggest that growth of these regions may result from variation in local BDC supersaturation perpendicular to the growth front, drawing comparisons with nucleation-free processes such as eutectic growth and spinodal decomposition.

1.4 Defect tuning

Alongside attempts to characterise these features, much effort has been directed towards controlling the type and concentration of defects in UiO-66. It is primarily the acidic modulator which has been used to achieve this; Wu *et al.*'s initial study on defects showed that a higher acetic acid to BDC ratio afforded more porous samples, which was ascribed to increased missing linker content.[63] Numerous studies have since shown the importance of the modulator concentration in creating defects.[85–87] Shearer *et al.* importantly demonstrated that the choice of modulator is equally crucial;[72] across UiO-66 samples synthesised with equal amounts of different monocarboxylates, defect content mirrored the trend in pKa: TFAA > difluoroactic acid > FA > AA. This reflects the competition for sites on Zr nodes during synthesis: more modulator leads to fewer spots where BDC can bond, and higher pKa means the monocarboxylate is more likely to be deprotonated than BDC and therefore be able to participate in bonding. Atzori *et al.* further showed in a subsequent study that using benzoic acid produced almost exclusively missing cluster defects, such that type as well as concentration can both be controlled using the modulator.[81]

There are other defect-generation strategies which do not rely on monocarboxylic acids. One approach is to accelerate the synthesis so as to promote cluster terminations via dislocations or crystal cleavages. Liang *et al.* and Katz *et al.* used HCl (without other modulators) to accelerate synthesis and achieve high defectivity.[86,88] It is argued that this

results from increased misconnections and dislocations in the material. However, it may also be the water within the HCl source that actually speeds up crystallisation, as shown by Ragon *et al.* who compared syntheses with HCl and H_2O.[89] Indeed, careful control over water content can be used to generate missing clusters during synthesis, as displayed by Chammingkwan *et al.*[90] Small amounts of water are necessary for the construction of $Zr_6O_4(OH)_4$ nodes, which can usually be drawn from the atmosphere or impurities. However by using an inert atmosphere, dry reagents and sub-stoichimetric amounts of H_2O, the intensity of the PXRD peak corresponding to correlated **reo** nanodomains can be tuned, suggesting control over missing clusters.

Finally, there are also some post-synthetic methods through which defects can be introduced. One of these is that of Bueken *et al.*, which involves synthesising the material with a mixture of BDC and 1,4-trans-cyclohexanedicarboxylate.[91] The latter is thermally unstable above 275 °C, compared to 450 °C for BDC. Hence, post-synthetic thermal treatment leads to the selective degradation of only some linkers, leading to well-distributed missing-linker defects. The same strategy was used by Feng *et al.* with BDC-NH_2 to make hierarchical pores via selective linker thermolysis.[34] The latter study notes that the distribution of the secondary linker must be tuned via synthetic conditions prior to thermolysis: well-dispersed linkers will lead to distributed defects (as seen also by Bueken *et al.*)[91] while the presence of domains will yield mesopores following thermal treatment. The same strategy can be used with a second linker that is more labile, as was done by Feng *et al.*[92] 4-sulfobenzoic acid can fulfil the same role as BDC, but is easily removed by post-synthetic treatment with sulphuric acid, again leading to distributed missing linker defects.

1.5 Metal-mixing

Another manner in which UiO-66 can be altered is by replacing Zr with another element. Using this material as a support, it is possible to integrate catalytically interesting metals which otherwise rarely form stable frameworks, such as titanium and lanthanides. In this way, electronic properties can be modified without altering the structure, maintaining porosity and some stability. This can be done by changing the synthetic procedure to include a mixture of precursors or via post-synthetic metal exchange, resulting in like-for-like replacement of Zr in the skeleton, or yielding immobilised species in the cavities.

An important family of metal-substituted UiO-66 is that which contains titanium. Ti is of interest due to its low toxicity, high abundance, and excellent photocatalytic properties, but difficult solution chemistry makes the assembly of new stable frameworks difficult.[93] This problem can be avoided by taking an existing robust material, such as UiO-66, and grafting or exchanging the metal. Work by Nguyen *et al.* exemplifies this approach, wherein Ti was introduced in 3 ways: like-for-like replacement of Zr, grafting onto the nodes, and immobilisation on a functionalised linker.[94] The authors noted improved catalytic oxidation of cyclohexene, with some control over activity via different Ti coordination environments. The degradation of dyes has also been investigated by Wang *et al.*, who noted enhanced photolytic breakdown of methylene blue following Ti substitution.[95] Such improvements are in line with findings from Zeama *et al.*, who reported much-improved activity in the reduction of CO_2 after incorporating titanium in UiO-66.[96] Supporting DFT calculations revealed a shrinking of the band gap as more Ti is added, vindicating its use in photocatalytic UiO-66.

Other transition metals have been investigated as dopants with which electronic behaviour can be tuned. Hafnium has been investigated since the synthesis of pure Hf UiO-66 by Jakobsen *et al.* for the purposes of

radioactive waste scavenging.[97] The similar ionic radii of Zr^{4+} and Hf^{4+} makes the latter completely miscible within UiO-66, but its higher oxophilicity leads to stronger Brønsted acid sites for catalysis.[98] Ta, Y, Nb, and W-doped forms of UiO-66 have also been examined in a detailed study by Syzgantseva *et al.*[99] The corresponding band gaps were evaluated with DFT and were found to cover a range of 1.7 to 2.8 eV, compared to 2.7 eV with Zr. The Nb-doped material was synthesised and exhibited a colour shift consistent with the DFT results. Lanthanides have also been the subject of interest for doping. Hendrickx *et al.* comprehensively studied UiO-66 containing Yb, Eu, and Ce; based on spectroscopic experiments, it was posited that the inclusion of these metals leads to additional decay pathways for photoexcited states, favouring charge transfer to catalytic sites.[100] Detailed DFT and time-dependent DFT calculations with Ce revealed that the empty 4f orbitals provide the alternate decay pathway, explaining the change in photoproperties upon lanthanide inclusion.

Interest in cerium has extended beyond just photocatalysis. Including this species in UiO-66 can allow for redox chemistry since it can readily move between 4+ and 3+ states, via which exotic catalytic cycles can be constructed. An attempt to synthesise the pure-Ce framework was first carried out by Lammert *et al.* and yielded a material isostructural to Zr-UiO-66.[101] Sample analysis via X-ray absorption near-edge structure (XANES) spectroscopy showed no trace of Ce^{3+}, but TGA revealed the presence of missing linker defects. In order to gauge redox-facilitated catalytic behaviour, the authors tested the aerobic oxidation of benzyl alcohol. The standard material exhibited a yield of just 8%, but this rose to up to 88% in the presence of a (2,2,6,6-tetramethylpiperidin-1-yl)oxyl (TEMPO) co-catalyst (compared to 7% with TEMPO alone). However, this promising activity came at the cost of severely compromised thermal stability; TGA experiments revealed a degradation temperature of only 300 °C, 150 degrees lower than the Zr counterpart. A subsequent

study by Nouar *et al.* showed that it was possible to combine the attractive properties of both frameworks by synthesising UiO-66 with 5% Ce content.[102] Here, XANES confirmed the presence of Ce^{3+} (making up ∼20 % of total Ce) while TGA revealed both defects and an increased thermal degradation temperature of 450 °C. The material also remained redox-active, as confirmed by temperature programmed reduction and oxidation experiments.

The improved performance of mixed Zr-Ce UiO-66 has led to much interest over the way these metals are distributed throughout the framework. Trousselet *et al.* used DFT to evaluate the cost of introducing cerium into UiO-66, exploiting symmetry to minimise computational cost.[103] The data revealed positive mixing energies for all cerium contents, in the range of 40-160 kJ mol^{-1} per conventional unit cell; the hybrid material is always less stable than separate phases. This suggests that the framework should be divided into regions of pure Ce and Zr UiO-66. However, this would be at odds with the results of Nouar *et al.*'s TGA experiments:[102] only a single thermal degradation temperature was recorded, whereas one would expect two of these for a phase-separated sample on account of the different thermal stabilities of the Zr and Ce forms of the material. Indeed, the presence of mixed SBUs was later confirmed by Lomachenko *et al.* through extended X-ray absorption fine structure (EXAFS) spectroscopy.[104] By comparing metal-metal scattering paths for pure Ce and Zr UiO-66, along with intermediate Ce/Zr ratios, they demonstrated that the majority of SBUs are of the type Zr_6, Zr_5Ce_1, and Ce_6. As Ce content increases, metal distribution follows the progression $Zr_6 + Zr_5Ce_1 \rightarrow Zr_5Ce_1 \rightarrow Zr_5Ce_1 + Ce_6 \rightarrow Ce_6$. This agrees with thermal and chemical stability trends reported by Lammert *et al.*, wherein stability of Ce-doped UiO-66 decreases until a plateau corresponding to the pure material, supposedly because Zr_5Ce_1 and Ce_6 become more prevalent.[105] Follow-up work by Atzori *et al.* came to the same conclusions using experimental and DFT-simulated IR spectroscopy.[106] The stretching frequency of μ_3-

OH groups depends on the adjacent metals, meaning a small number of distinct peaks should be present in different ratios. Upon increasing Ce content, the experimental spectra more closely follow a model with Zr_6, Ce_6, and Zr_5Ce_1 SBUs only than with all combinations allowed, though there remains evidence for small amounts of other clusters.

The factors which control the distribution of metals are therefore not precisely known. Trousselet *et al.*'s computational work points towards phase-separated Zr and Ce UiO-66 being the thermodynamic product.[103] This is inconsistent with experimental evidence of the mixed framework including heterometallic building blocks such as Zr_5Ce_1, which suggests kinetic factors must play some role in controlling the final product. However, the aforementioned simulations did not take defects into account, which should alleviate the strain arising from the discrepancy in ionic radius between Zr^{4+} and Ce^{4+}. The possible presence of Ce^{3+} has also been neglected. Thus, the existing research is not unambiguous as to the nature of the driving force behind cerium distribution in mixed UiO-66.

1.6 Impact on material properties

Much of the literature discussed in **Section 1.3** is fundamental and covers evidence for defects, correlation, and metal mixing in UiO-66. The current section aims to draw more explicit links between these features and the effects they have on material properties, so as to paint a more complete picture of how they can be used for tuning.

1.6.1 Stability

As discussed in **Section 1.2**, UiO-66 is exceptionally stable in the context of MOFs on account of strong linker-SBU bonds and a high degree of framework connectivity. However, the introduction of defects and the concurrent removal of linkers can compromise this, as was found by Thornton *et al.*[107] Periodic DFT calculations on a series of increasingly defective unit cells using a range of capping species revealed sharp

drops in Young and shear moduli along directions compromised by missing linkers; perpendicular directions saw almost no changes in these values. The authors also noted that bulky acids like AA or TFAA led to smaller decreases in elastic moduli, ascribing this to these moieties being in close contact at defect sites and there being less space for efficient compression. Interestingly, the **reo** unit cell exhibited only moderate changes to its mechanical properties, suggesting that robust defective UiO-66 can be synthesised via control over defect distribution and capping species. A similar study was carried out by Rogge *et al.* using QuickFF, an *ab initio*-parametrised forcefield.[108] Drops in bulk modulus and loss-of-crystallinity pressure were also noted with increasing defect content, and like in Thornton *et al.*'s study,[107] 'strong' and 'weak' directions were identified depending on the positions of defects in the unit cell. Experimental bulk modulus measurements by Dissegna *et al.* on defective samples appear to support these computational findings, including higher-than-expected stability for highly defective UiO-66 (where **reo** nanodomains are more prevalent).[109]

In contrast to its mechanical resilience, the chemical and thermal stabilities of UiO-66 are less drastically affected by defects. Work from Wang *et al.* on defective UiO-66 sought to evaluate the material for dye separation.[110] To verify that samples with more missing linkers remained stable, they were immersed in water for 7 days and HCl for 1 day. Subsequent PXRD showed no structural changes compared to the pristine framework. A similar study from Li *et al.* for the purpose of roxarsone capture found that the defective sample remained effective for adsorption over the tested pH range of 4 to 8.[85] The relationship between degradation temperature and defects appears more nuanced, as noted by Shearer *et al.*[56] A first set of experiments, for which the charge compensating species was ascertained to be Cl^-, revealed that thermal stability decreased with increasing content of missing clusters, as marked by partial framework degradation even at 250 °C. Follow-up work involving TGA and the use

of different modulators during synthesis yielded a similar conclusion.[72] There, missing clusters were also the predominant defect, but the range of degradation temperatures was much smaller. The authors suggested that the identity of the capping species (monocarboxylic acids in this case) was responsible for this, as these allow zirconium to maintain a chemical environment similar to that in the pristine material. Another potential reason put forward was differences in **reo** nanodomain sizes. Hence it may be both defect content and defect type (missing linker vs cluster, nature of capping species) which control degradation temperature.

As discussed in **Section 1.5**, the introduction of cerium into the framework bears similar consequences to those of defects. In addition to a decrease in thermal stability which ends in a plateau at the degradation temperature of Ce-UiO-66, mechanical properties also suffer. Rogge *et al.* carried out experiments on Hf, Ce, and Zr heterometallic UiO-66 models using QuickFF to estimate loss-of-crystallinity pressures.[53] As the ratio of Zr decreases, the material becomes less stable for Ce (mirroring the thermal trend reported by Lammert *et al.*)[101] but sees no change for Hf. Experimental measurements of these pressures revealed pure Zr to be the most resilient, followed by Hf then Ce. What can be concluded from work on both defects and cerium-substituted UiO-66 is that the material's stability depends on that of its 'weakest link', but that there are ways to mitigate this, for example by focusing on missing clusters and specific capping species.

1.6.2 Pore structure

An important property which is intrinsically linked to defects is pore structure. The removal of either linkers or clusters will result in a more open framework, as is often reported via increases in BET surface areas. However, it has been shown by Shearer *et al.* that it is actually cluster vacancies which contribute to increases in gas adsorption.[72] A set of sim-

ulated nitrogen adsorption isotherms for unit cells with increasing numbers of difluoroacetic acid-capped missing linker defects showed drops in adsorbed nitrogen compared to the pristine material. In contrast, the experimental results followed the opposite trend, agreeing instead with the isotherm for the **reo** system. Work from Bueken *et al.* has provided corroborating evidence for this.[91] By synthesising multivariate UiO-66 using a thermally unstable secondary linker and subsequently heating the material, the authors were able to generate well-dispersed linker vacancies. No increases in BET surface area were recorded after this process, confirming that missing clusters are the only contributors.

As noted in **Section 1.4**, defect-generation strategies can also be employed to synthesise hierarchically porous (HP) UiO-66. Some of these are applied during synthesis, such as the work of Cai and Jiang, which makes use of long-chain monocarboxylic acids as modulators.[111] These force the formation of large voids during crystal growth; pore sizes up to 8 nm in diameter can be reached by choosing different chain lengths. Yin *et al.* used the same method to make HP UiO-66 for seawater uranium harvesting; they reported pores as large as 100 nm in diameter.[112] Alternatively, mesopores can be formed through the aggregation of UiO-66 nanoparticles, as reported by Hao *et al.*[113] The synthesis of UiO-66 was optimised to produce different pore sizes, after which samples were examined using HRTEM. The images revealed aggregated clusters of nanoparticles within which the additional mesopores were located. Post-synthetic techniques can also be effective; Feng *et al.* used linker thermolysis in multivariate UiO-66 to generate hierarchical pores.[34] The pore structure can be controlled in this manner by tuning synthetic conditions to distribute the secondary linker in smaller or larger domains. Alternatively, pristine UiO-66 can be etched by a monocarboxylic acid after synthesis to remove linkers and clusters, as shown by Yang *et al.*[114] By controlling the etching parameters (duration, temperature, and choice of acid), the resulting void sizes can be varied, leading to pores up to 40 nm in diameter.

This relationship between defects and pore structures is important in the context of adsorption behaviour. Notably, Ghosh *et al.* have found through simulation that water loading in UiO-66 increases when more defects are present.[115] The authors attributed this to an increase in the hydrophilicity of the pore surface brought about by the hydroxyl groups used to terminate missing linkers, onto which water preferentially adsorbs. These findings have been echoed by Hossain *et al.*,[116] who carried out similar simulation work on the coadsorption of CO_2 and water in defective UiO-66. Cooperative effects were additionally found between these two molecules during adsorption which depends on the distribution of defects, such that both the number and location of missing linkers impact adsorption. Hernandez *et al.* have shown through a combination of experimental and computational work that this effect can also be observed in UiO-66 synthesised with amine-functionalised linkers.[117] Increased water loading again manifests in this material through preferential adsorption at defects such that individual H_2O molecules, oriented by nearby amine groups, join together to form bridges across missing linkers. This link between defects and both pore size and the functional groups present along the pore surface is important in the context of gas capture and water harvesting applications.[118,119]

1.6.3 Catalytic potential

The catalytic potential of UiO-66 and its relationship with defects has been the subject of numerous experiments, many of which were done by the Van Speybroeck group. The main focus of these studies has been acid-base reactions, which will be covered first, but photocatalytic applications have also received some attention. Vermoortele *et al.* reported in 2013 that modulating the synthesis of UiO-66 with TFAA and HCl strongly improved the conversion of citronellal to isopulegol compared to the pristine material.[70] This was ascribed to the presence of a larger number of unsaturated Lewis acid sites rather than a change in site strength

on account of unchanged isopulegol isomer selectivity. More detail was made available by Vandichel *et al.*'s computational study on the same reaction.[71] DFT-derived free energy differences revealed a slightly lower barrier for the reaction when TFAA is used to terminate defects, suggesting this species has some effect on Lewis acidity after all. It should be noted that in the Vermoortele study,[70] the high activation temperature may have purged TFAA from the structures prior to catalytic tests.

Other reactions have been demonstrated to be effectively catalysed by Lewis-acidic defective UiO-66 using computational and experimental methods. Hajek *et al.* have published a detailed mechanism for the condensation of aldols using DFT methods.[120] The 2 open Zr sites at an uncapped defect (cap of type **e**) are able to host aldols for self or cross condensation with moderate barriers of $\sim$100 kJ mol^{-1}, in line with experimental conversion measurements (involving benzaldehyde and propanal) of 45% after 1 hour. Caratelli *et al.* carried out similar work on the Fischer esterification of methanol and propionic acid, again starting from the defective unit cell.[121] In this mechanism again it is a pair of Zr sites which hosts the reactants, but in this case, a lower barrier is afforded when the defect is initially capped by H_2O-OH rather than uncapped, as the hydroxyl moiety enables the loss of water from the transition state. Supporting experiments for the production of ethyl levulinate from levulinic acid and ethanol evidenced improved yield in hydrated and defective UiO-66 samples. Follow-up work by Caratelli *et al.* involving metadynamics simulations on UiO-66 in water confirmed that Zr ions need not be unsaturated for them to be accessible.[122] The solvent makes the coordination environment of Zr dynamic, allowing for transient over-coordination and under-coordination via changes in linker bonding.

The capping species themselves can also be modified to act as the active sites. Nguyen *et al.* synthesised defective UiO-66 and grafted vanadium at the defects.[123] Post-synthetic treatment with vanadium acetylaceto-

nate replaced the capping species with a new metal centre available for catalysis. Experimental measurements for the gas-phase dehydrogenation of cyclohexene showed the functionalised MOF to be an effective catalyst, unlike the standard material. Similar work has been carried out by Gutov *et al.* using chromium for the self-cyclisation of acetaldehyde, and by Abdel-Mageed *et al.* with copper for the oxidation of CO.[124,125] Alternatively, the capping carboxylic acid can be replaced by another to add an additional functional group. Ji *et al.* used a procedure to replace formate caps in MOF-808 (in principle this could also be done with defective UiO-66) with H_2O-OH before replacing these with triflate anions, resulting in good catalytic activity for a range of reactions.[126]

Differences in UiO-66 catalytic activity have also been reported between UiO-66 containing missing clusters versus missing linkers. Liu *et al.* noted that equally defective samples with more missing clusters than linkers were more effective for d-glucose isomerisation to fructose.[83] Using a combination of TGA, pore-size analysis, diffuse PXRD and HRTEM, the authors were able to reliably evaluate the relative proportions of both types of defects and thus assign improved activity to missing clusters. This was also demonstrated clearly by Wang *et al.* through the comparison of two modulators.[127] UiO-66 was synthesised using TFAA and found to have significant **reo** nanodomains, as indicated by forbidden peaks in PXRD; synthesis with FA yielded similar defect content but with almost no **reo**. The TFAA sample was more active for cyclohexanone conversion to cyclohexanol, and this remained the case even after a post-synthetic exchange procedure was used to swap FA and TFAA in the two samples. Missing cluster defects can also be generated by limiting the amount of water available during synthesis, as was done by Chammingkwan *et al.*[90] Again the sample with the most missing clusters was identified via diffuse PXRD peaks and was shown to be the most effective for the Meerwein–Ponndorf–Verley reduction of 4-tert-butylcyclohexanone and alcoholysis of styrene oxide.

Defects can also be important in tuning and imparting Brønsted acidity to UiO-66. Liu *et al.* reported improvements in the ring-opening of styrene oxide by isopropanol in the presence of missing linkers.[128] Given appropriate framework activation, defects contribute additional acidic protons via H_2O-OH cappings, and the authors noted a trend between sample defectivity and styrene oxide conversion after 24 hours. The importance of these water moieties was also demonstrated by Fu *et al.* in hydrated and dehydrated defective UiO-66.[129] UiO-66 was tested for the isomerisation of α-pinene oxide, in which low selectivity is indicative of higher Brønsted acidity. Defective samples activated at low temperature in which water was not removed exhibited low selectivity, whereas the converse was observed when a high activation temperature was used. Cirujano and Llabrés I Xamena carried out a detailed study of the acidic centres in defective UiO-66 by measuring IR frequencies of CO probes and the pH of UiO-66 suspensions in methanol.[130] Their conclusions agreed with those of Liu *et al.*:[128] lower pH was measured for the hydrated defective sample than the pristine or dehydrated ones, and additional CO frequencies were observed corresponding to additional acid sites.

Finally, some authors have explored the possibility of using UiO-66 as a photocatalyst by altering its band structure through defects. Fu *et al.* found for UiO-66 made with amino-functionalised linkers that defective samples were more active in the photocatalytic reduction of CO_2.[131] It was also shown that, upon irradiation with visible light, a greater photocurrent is generated in the material when more vacancies are present, which is indicative of increased charge transfer. A similar trend was found by Xu *et al.* for the reduction of Cr(IV) to Cr(III) by amino-functionalised UiO-66.[132] For this process, the defective material outperformed the pristine one by a factor of up to 47, which the authors ascribed to vacancies reducing the energy requirement for metal to ligand charge transfer (LCMT). This is consistent with the DFT work of De Vos *et al.*,[76] who

showed that removing linkers from UiO-66 causes some unoccupied states associated with the SBUs to drop to lower energy, thereby reducing the gap which must be overcome for LCMT. It therefore appears that defects can also effectively be used to tune this material for photocatalytic applications.

References

(1) H. C. Zhou, J. R. Long and O. M. Yaghi, *Chem. Rev.*, 2012, **112**, 673–674.

(2) H. Furukawa, K. E. Cordova, M. O'Keeffe and O. M. Yaghi, *Science*, 2013, **341**, 1230444.

(3) O. M. Yaghi, D. A. Richardson, G. Li, C. E. Davis and T. L. Groy, *Mater. Res. Soc. Symp. - Proc.*, 1995, **371**, 15–19.

(4) J. L. Rowsell and O. M. Yaghi, *Micropor. Mesopor. Mat.*, 2004, **73**, 3–14.

(5) I. Huskić, I. V. Pekov, S. V. Krivovichev and T. Friščić, *Sci. Adv.*, 2016, **2**, 1–8.

(6) Clarivate - Web of Science, *Yearly MOF publications*, 2023.

(7) 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, *PNAS*, 2006, **103**, 10186–10191.

(8) T. Loiseau, C. Serre, C. Huguenard, G. Fink, F. Taulelle, M. Henry, T. Bataille and G. Férey, *Chem. Eur. J.*, 2004, **10**, 1373–1382.

(9) S. S. Chui, S. M. Lo, J. P. Charmant, A. G. Orpen and I. D. Williams, *Science*, 1999, **283**, 1148–1150.

(10) S. E. Henkelis, S. M. Vornholt, D. B. Cordes, A. M. Slawin, P. S. Wheatley and R. E. Morris, *CrystEngComm*, 2019, **21**, 1857–1861.

(11) T. F. Willems, C. H. Rycroft, M. Kazi, J. C. Meza and M. Haranczyk, *Micropor. Mesopor. Mat.*, 2012, **149**, 134–141.

(12) H. Ghasempour, K. Y. Wang, J. A. Powell, F. ZareKarizi, X. L. Lv, A. Morsali and H. C. Zhou, *Coord. Chem. Rev.*, 2021, **426**, 213542.

(13) B. Chen, Z. Yang, Y. Zhu and Y. Xia, *J. Mater. Chem. A*, 2014, **2**, 16811–16831.

(14) G. Shimizu, R. Vaidhyanathan and J. Taylor, *Chem. Soc. Rev.*, 2009, **38**, 1430–1449.

(15) Y. Bai, Y. Dou, L. H. Xie, W. Rutledge, J. R. Li and H. C. Zhou, *Chem. Soc. Rev.*, 2016, **45**, 2327–2367.

(16) X. Liu, Y. Zhou, J. Zhang, L. Tang, L. Luo and G. Zeng, *ACS Appl. Mater. Interfaces*, 2017, **9**, 20255–20275.

(17) K. Nath, A. Bin Rahaman, R. Moi, K. Maity and K. Biradha, *Chem. Commun.*, 2020, **56**, 14873–14876.

(18) N. L. Rosi, J. Kim, M. Eddaoudi, B. Chen, M. O'Keeffe and O. M. Yaghi, *J. Am. Chem. Soc.*, 2005, **127**, 1504–1518.

(19) T. Wu, N. Prasetya and K. Li, *J. Memb. Sci.*, 2020, **615**, 118493.

(20) S. Iram, M. Imran, F. Kanwal, S. Latif, Z. Iqbal and H. G. Stammler, *ChemistrySelect*, 2018, **3**, 10443–10449.

(21) M. Li, D. Li, M. O'Keeffe and O. M. Yaghi, *Chem. Rev.*, 2014, **114**, 1343–1370.

(22) H. Furukawa, Y. B. Go, N. Ko, Y. K. Park, F. J. Uribe-Romo, J. Kim, M. O'Keeffe and O. M. Yaghi, *Inorg. Chem.*, 2011, **50**, 9147–9152.

(23) H. Wu, T. Yildirim and W. Zhou, *J. Phys. Chem. Lett.*, 2013, **4**, 925–930.

(24) J. Sim, H. Yim, N. Ko, S. B. Choi, Y. Oh, H. J. Park, S. Y. Park and J. Kim, *J. Chem. Soc. Dalt. Trans.*, 2014, **43**, 18017–18024.

(25) Z. H. Rada, H. R. Abid, J. Shang, H. Sun, Y. He, P. Webley, S. Liu and S. Wang, *Ind. Eng. Chem. Res.*, 2016, **55**, 7924–7932.

(26) P. Valvekens, F. Vermoortele and D. De Vos, *Catal. Sci. Technol.*, 2013, **3**, 1435–1445.

(27) 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.

(28) Y. G. Chung, E. Haldoupis, B. J. Bucior, M. Haranczyk, S. Lee, H. Zhang, K. D. Vogiatzis, M. Milisavljevic, S. Ling, S. Camp, B. Slater, J. I. Siepmann, D. S. Sholl and R. Q. Snurr, *J. Chem. Eng. Data*, 2019, **64**, 5985–5998.

(29) 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.

(30) Y. J. Colón, D. A. Gómez-Gualdrón and R. Q. Snurr, *Cryst. Growth Des.*, 2017, **17**, 5801–5810.

(31) P. G. Boyd and T. K. Woo, *CrystEngComm*, 2016, **18**, 3777–3792.

(32) T. Islamoglu, S. Goswami, Z. Li, A. J. Howarth, O. K. Farha and J. T. Hupp, *Acc. Chem. Res.*, 2017, **50**, 805–813.

(33) T. Y. Luo, C. Liu, X. Y. Gan, P. F. Muldoon, N. A. Diemler, J. E. Millstone and N. L. Rosi, *J. Am. Chem. Soc.*, 2019, **141**, 2161–2168.

(34) L. Feng, S. Yuan, L. L. Zhang, K. Tan, J. L. Li, A. Kirchon, L. M. Liu, P. Zhang, Y. Han, Y. J. Chabal and H. C. Zhou, *J. Am. Chem. Soc.*, 2018, **140**, 2363–2372.

(35) P. Falcaro, R. Ricco, A. Yazdi, I. Imaz, S. Furukawa, D. Maspoch, R. Ameloot, J. D. Evans and C. J. Doonan, *Coord. Chem. Rev.*, 2016, **307**, 237–254.

(36) Z. Fang, B. Bueken, D. E. De Vos and R. A. Fischer, *Angew. Chemie - Int. Ed.*, 2015, **54**, 7234–7254.

(37) X. Zhao, Y. Wang, D. S. Li, X. Bu and P. Feng, *Adv. Mater.*, 2018, **30**, 1–34.

(38) H. Li, L. Li, R.-B. Lin, W. Zhou, Z. Zhang, S. Xiang and B. Chen, *EnergyChem*, 2019, **1**, 100006.

(39) W. Xu and O. M. Yaghi, *ACS Cent. Sci.*, 2020, **6**, 1348–1354.

(40) S. Ma and H. C. Zhou, *Chem. Commun.*, 2010, **46**, 44–53.

(41) A. Dhakshinamoorthy, A. Santiago-Portillo, A. M. Asiri and H. Garcia, *ChemCatChem*, 2019, **11**, 899–923.

(42) J. Lee, O. K. Farha, J. Roberts, K. A. Scheidt, S. T. Nguyen and J. T. Hupp, *Chem. Soc. Rev.*, 2009, **38**, 1450–1459.

(43) Q. Wang, Q. Gao, A. M. Al-Enizi, A. Nafady and S. Ma, *Inorg. Chem. Front.*, 2020, **7**, 300–339.

(44) J. Jacobsen, A. Ienco, R. D'Amato, F. Costantino and N. Stock, *Dalt. Trans.*, 2020, **49**, 16551–16586.

(45) J. M. Palomba, S. P. Harvey, M. Kalaj, B. R. Pimentel, J. B. Decoste, G. W. Peterson and S. M. Cohen, *ACS Appl. Mater. Interfaces*, 2020, **12**, 14672–14677.

(46) G. Mínguez Espallargas and E. Coronado, *Chem. Soc. Rev.*, 2018, **47**, 533–557.

(47) C. G. Silva, A. Corma and H. García, *J. Mater. Chem.*, 2010, **20**, 3141–3156.

(48) A. E. Baumann, D. A. Burns, B. Liu and V. S. Thoi, *Commun. Chem.*, 2019, **2**, 1–14.

(49) H. S. Wang, Y. H. Wang and Y. Ding, *Nanoscale Adv.*, 2020, **2**, 3788–3797.

(50) Y. Sun, L. Zheng, Y. Yang, X. Qian, T. Fu, X. Li, Z. Yang, H. Yan, C. Cui and W. Tan, *Nano-Micro Lett.*, 2020, **12**, 1–29.

(51) M. Cai, G. Chen, L. Qin, C. Qu, X. Dong, J. Ni and X. Yin, *Pharmaceutics*, 2020, **12**, 232.

(52) J. H. Cavka, S. Jakobsen, U. Olsbye, N. Guillou, C. Lamberti, S. Bordiga and K. P. Lillerud, *J. Am. Chem. Soc.*, 2008, **130**, 13850–13851.

(53) S. M. Rogge, P. G. Yot, J. Jacobsen, F. Muniz-Miranda, S. Vandenbrande, J. Gosch, V. Ortiz, I. E. Collings, S. Devautour-Vinot, G. Maurin, N. Stock and V. Van Speybroeck, *ACS Mater. Lett.*, 2020, **2**, 438–445.

(54) C. G. Piscopo, A. Polyzoidis, M. Schwarzer and S. Loebbecke, *Micropor. Mesopor. Mat.*, 2015, **208**, 30–35.

(55) D. Bužek, J. Demel and K. Lang, *Inorg. Chem.*, 2018, **57**, 14290–14297.

(56) G. C. Shearer, S. Chavan, J. Ethiraj, J. G. Vitillo, S. Svelle, U. Olsbye, C. Lamberti, S. Bordiga and K. P. Lillerud, *Chem. Mater.*, 2014, **26**, 4068–4071.

(57) H. Furukawa, F. Gándara, Y. B. Zhang, J. Jiang, W. L. Queen, M. R. Hudson and O. M. Yaghi, *J. Am. Chem. Soc.*, 2014, **136**, 4369–4381.

(58) J. E. Mondloch, W. Bury, D. Fairen-Jimenez, S. Kwon, E. J. Demarco, M. H. Weston, A. A. Sarjeant, S. T. Nguyen, P. C. Stair, R. Q. Snurr, O. K. Farha and J. T. Hupp, *J. Am. Chem. Soc.*, 2013, **135**, 10294–10297.

(59) N. Planas, J. E. Mondloch, S. Tussupbayev, J. Borycz, L. Gagliardi, J. T. Hupp, O. K. Farha and C. J. Cramer, *J. Phys. Chem. Lett.*, 2014, **5**, 3716–3723.

(60) J. Winarta, B. Shan, S. M. McIntyre, L. Ye, C. Wang, J. Liu and B. Mu, *Cryst. Growth Des.*, 2020, **20**, 1347–1362.

(61) D. Zou and D. Liu, *Mater. Today Chem.*, 2019, **12**, 139–165.

(62) L. Valenzano, B. Civalleri, S. Chavan, S. Bordiga, M. H. Nilsen, S. Jakobsen, K. P. Lillerud and C. Lamberti, *Chem. Mater.*, 2011, **23**, 1700–1718.

(63) H. Wu, Y. S. Chua, V. Krungleviciute, M. Tyagi, P. Chen, T. Yildirim and W. Zhou, *J. Am. Chem. Soc.*, 2013, **135**, 10525–10532.

(64) S. Øien, D. Wragg, H. Reinsch, S. Svelle, S. Bordiga, C. Lamberti and K. P. Lillerud, *Cryst. Growth Des.*, 2014, **14**, 5370–5372.

(65) C. A. Trickett, K. J. Gagnon, S. Lee, F. Gándara, H. B. Bürgi and O. M. Yaghi, *Angew. Chemie - Int. Ed.*, 2015, **54**, 11162–11167.

(66) R. C. Klet, Y. Liu, T. C. Wang, J. T. Hupp and O. K. Farha, *J. Mater. Chem. A*, 2016, **4**, 1479–1485.

(67) S. Ling and B. Slater, *Chem. Sci.*, 2016, **7**, 4706–4712.

(68) M. Vandichel, J. Hajek, A. Ghysels, A. De Vos, M. Waroquier and V. Van Speybroeck, *CrystEngComm*, 2016, **18**, 7056–7069.

(69) K. Tan, H. Pandey, H. Wang, E. Velasco, K. Y. Wang, H. C. Zhou, J. Li and T. Thonhauser, *J. Am. Chem. Soc.*, 2021, **143**, 6328–6332.

(70) F. Vermoortele, B. Bueken, G. Le Bars, B. Van De Voorde, M. Vandichel, K. Houthoofd, A. Vimont, M. Daturi, M. Waroquier, V. Van Speybroeck, C. Kirschhock and D. E. De Vos, *J. Am. Chem. Soc.*, 2013, **135**, 11465–11468.

(71) M. Vandichel, J. Hajek, F. Vermoortele, M. Waroquier, D. E. De Vos and V. Van Speybroeck, *CrystEngComm*, 2015, **17**, 395–406.

(72) G. C. Shearer, S. Chavan, S. Bordiga, S. Svelle, U. Olsbye and K. P. Lillerud, *Chem. Mater.*, 2016, **28**, 3749–3761.

(73) J. K. Bristow, K. L. Svane, D. Tiana, J. M. Skelton, J. D. Gale and A. Walsh, *J. Phys. Chem. C*, 2016, **120**, 9276–9281.

(74) R. Wei, C. A. Gaggioli, G. Li, T. Islamoglu, Z. Zhang, P. Yu, O. K. Farha, C. J. Cramer, L. Gagliardi, D. Yang and B. C. Gates, *Chem. Mater.*, 2019, **31**, 1655–1663.

(75) C. S. Cox, E. Slavich, L. K. Macreadie, L. K. McKemmish and M. Lessio, *Chem. Mater.*, 2023, **35**, 3057–3072.

(76) A. De Vos, K. Hendrickx, P. Van Der Voort, V. Van Speybroeck and K. Lejaeghere, *Chem. Mater.*, 2017, **29**, 3006–3019.

(77) F. G. Xi, Y. Yang, H. Liu, H. F. Yao and E. Q. Gao, *RSC Adv.*, 2015, **5**, 79216–79223.

(78) D. M. Driscoll, D. Troya, P. M. Usov, A. J. Maynes, A. J. Morris and J. R. Morris, *J. Phys. Chem. C*, 2018, **122**, 14582–14589.

(79) G. C. Shearer, S. Forselv, S. Chavan, S. Bordiga, K. Mathisen, M. Bjørgen, S. Svelle and K. P. Lillerud, *Top. Catal.*, 2013, **56**, 770–782.

(80) M. J. Cliffe, W. Wan, X. Zou, P. A. Chater, A. K. Kleppe, M. G. Tucker, H. Wilhelm, N. P. Funnell, F. X. Coudert and A. L. Goodwin, *Nat. Commun.*, 2014, **5**, 1–8.

(81) C. Atzori, G. C. Shearer, L. Maschio, B. Civalleri, F. Bonino, C. Lamberti, S. Svelle, K. P. Lillerud and S. Bordiga, *J. Phys. Chem. C*, 2017, **121**, 9312–9324.

(82) K. L. Svane, J. K. Bristow, J. D. Gale and A. Walsh, *J. Mater. Chem. A*, 2018, **6**, 8507–8513.

(83) L. Liu, Z. Chen, J. Wang, D. Zhang, Y. Zhu, S. Ling, K. W. Huang, Y. Belmabkhout, K. Adil, Y. Zhang, B. Slater, M. Eddaoudi and Y. Han, *Nat. Chem.*, 2019, **11**, 622–628.

(84) D. N. Johnstone, F. C. Firth, C. P. Grey, P. A. Midgley, M. J. Cliffe, S. M. Collins and D. N. Johnstone, *J. Am. Chem. Soc.*, 2020, **142**, 13081–13089.

(85) B. Li, X. Zhu, K. Hu, Y. Li, J. Feng, J. Shi and J. Gu, *J. Hazard. Mater.*, 2016, **302**, 57–64.

(86) W. Liang, C. J. Coghlan, F. Ragon, M. Rubio-Martinez, D. M. D'Alessandro and R. Babarao, *Dalt. Trans.*, 2016, **45**, 4496–4500.

(87) S. Dissegna, R. Hardian, K. Epp, G. Kieslich, M. V. Coulet, P. Llewellyn and R. A. Fischer, *CrystEngComm*, 2017, **19**, 4137–4141.

(88) M. J. Katz, Z. J. Brown, Y. J. Colón, P. W. Siu, K. A. Scheidt, R. Q. Snurr, J. T. Hupp and O. K. Farha, *Chem. Commun.*, 2013, **49**, 9449–9451.

(89) F. Ragon, P. Horcajada, H. Chevreau, Y. K. Hwang, U. H. Lee, S. R. Miller, T. Devic, J. S. Chang and C. Serre, *Inorg. Chem.*, 2014, **53**, 2491–2500.

(90) P. Chammingkwan, G. Y. Shangkum, L. T. T. Mai, P. Mohan, A. Thakur, T. Wada and T. Taniike, *RSC Adv.*, 2020, **10**, 28180–28185.

(91) B. Bueken, N. Van Velthoven, A. Krajnc, S. Smolders, F. Taulelle, C. Mellot-Draznieks, G. Mali, T. D. Bennett and D. De Vos, *Chem. Mater.*, 2017, **29**, 10478–10486.

(92) X. Feng, J. Hajek, H. S. Jena, G. Wang, S. K. Veerapandian, R. Morent, N. De Geyter, K. Leyssens, A. E. Hoffman, V. Meynen, C. Marquez, D. E. De Vos, V. Van Speybroeck, K. Leus and P. Van Der Voort, *J. Am. Chem. Soc.*, 2020, **142**, 3174–3183.

(93) S. Wang, H. Reinsch, N. Heymans, M. Wahiduzzaman, C. Martineau-Corcos, G. De Weireld, G. Maurin and C. Serre, *Matter*, 2020, **2**, 440–450.

(94) H. G. T. Nguyen, L. Mao, A. W. Peters, C. O. Audu, Z. J. Brown, O. K. Farha, J. T. Hupp and S. T. Nguyen, *Catal. Sci. Technol.*, 2015, **5**, 4444–4451.

(95) A. Wang, Y. Zhou, Z. Wang, M. Chen, L. Sun and X. Liu, *RSC Adv.*, 2016, **6**, 3671–3679.

(96) M. Zeama, M. Morsy, S. Abdel-Azeim, M. Abdelnaby, A. Alloush and Z. Yamani, *Inorganica Chim. Acta*, 2020, **501**, 119287.

(97) S. Jakobsen, D. Gianolio, D. S. Wragg, M. H. Nilsen, H. Emerich, S. Bordiga, C. Lamberti, U. Olsbye, M. Tilset and K. P. Lillerud, *Phys. Rev. B - Condens. Matter Mater. Phys.*, 2012, **86**, 125429.

(98) V. R. Bakuru, S. R. Churipard, S. P. Maradur and S. B. Kalidindi, *Dalt. Trans.*, 2019, **48**, 843–847.

(99) M. A. Syzgantseva, C. P. Ireland, F. M. Ebrahim, B. Smit and O. A. Syzgantseva, *J. Am. Chem. Soc.*, 2019, **141**, 6271–6278.

(100) K. Hendrickx, J. J. Joos, A. De Vos, D. Poelman, P. F. Smet, V. Van Speybroeck, P. Van Der Voort and K. Lejaeghere, *Inorg. Chem.*, 2018, **57**, 5463–5474.

(101) M. Lammert, M. T. Wharmby, S. Smolders, B. Bueken, A. Lieb, K. A. Lomachenko, D. De Vos and N. Stock, *Chem. Commun.*, 2015, **51**, 12578–12581.

(102) F. Nouar, M. I. Breeze, B. C. Campo, A. Vimont, G. Clet, M. Daturi, T. Devic, R. I. Walton and C. Serre, *Chem. Commun.*, 2015, **51**, 14458–14461.

(103) F. Trousselet, A. Archereau, A. Boutin and F. X. Coudert, *J. Phys. Chem. C*, 2016, **120**, 24885–24894.

(104) K. A. Lomachenko, J. Jacobsen, A. L. Bugaev, C. Atzori, F. Bonino, S. Bordiga, N. Stock and C. Lamberti, *J. Am. Chem. Soc.*, 2018, **140**, 17379–17383.

(105) M. Lammert, C. Glißmann and N. Stock, *Dalt. Trans.*, 2017, **46**, 2425–2429.

(106) C. Atzori, K. A. Lomachenko, J. Jacobsen, N. Stock, A. Damin, F. Bonino and S. Bordiga, *Dalt. Trans.*, 2020, **49**, 5794–5797.

(107) A. W. Thornton, R. Babarao, A. Jain, F. Trousselet and F. X. Coudert, *Dalt. Trans.*, 2016, **45**, 4352–4359.

(108) S. M. Rogge, J. Wieme, L. Vanduyfhuys, S. Vandenbrande, G. Maurin, T. Verstraelen, M. Waroquier and V. Van Speybroeck, *Chem. Mater.*, 2016, **28**, 5721–5732.

(109) S. Dissegna, P. Vervoorts, C. L. Hobday, T. Düren, D. Daisenberger, A. J. Smith, R. A. Fischer and G. Kieslich, *J. Am. Chem. Soc.*, 2018, **140**, 11581–11584.

(110) K. Wang, C. Li, Y. Liang, T. Han, H. Huang, Q. Yang, D. Liu and C. Zhong, *Chem. Eng. J.*, 2016, **289**, 486–493.

(111) G. Cai and H.-L. Jiang, *Angew. Chemie*, 2017, **129**, 578–582.

(112) C. Yin, Q. Liu, R. Chen, J. Liu, J. Yu, D. Song and J. Wang, *Ind. Eng. Chem. Res.*, 2019, **58**, 1159–1166.

(113) L. Hao, X. Li, M. J. Hurlock, X. Tu and Q. Zhang, *Chem. Commun.*, 2018, **54**, 11817–11820.

(114) P. Yang, F. Mao, Y. Li, Q. Zhuang and J. Gu, *Chem. Eur. J.*, 2018, **24**, 2962–2970.

(115) P. Ghosh, Y. J. Colón and R. Q. Snurr, *Chem. Commun.*, 2014, **50**, 11329–11331.

(116) M. I. Hossain, J. D. Cunningham, T. M. Becker, B. E. Grabicka, K. S. Walton, B. D. Rabideau and T. G. Glover, *Chem. Eng. Sci.*, 2019, **203**, 346–357.

(117) A. F. Hernandez, R. K. Impastato, M. I. Hossain, B. D. Rabideau and T. G. Glover, *Langmuir*, 2021, **37**, 10439–10449.

(118) M. Usman, A. Helal, M. M. Abdelnaby, A. M. Alloush, M. Zeama and Z. H. Yamani, *Chemical Record*, 2021, **21**, 1771–1791.

(119) M. J. Kalmutzki, C. S. Diercks and O. M. Yaghi, *Adv. Mater.*, 2018, **30**, 1–26.

(120) J. Hajek, M. Vandichel, B. Van De Voorde, B. Bueken, D. De Vos, M. Waroquier and V. Van Speybroeck, *J. Catal.*, 2015, **331**, 1–12.

(121) C. Caratelli, J. Hajek, F. G. Cirujano, M. Waroquier, F. X. Llabrés i Xamena and V. Van Speybroeck, *J. Catal.*, 2017, **352**, 401–414.

(122) C. Caratelli, J. Hajek, E. J. Meijer, M. Waroquier and V. Van Speybroeck, *Chem. Eur. J.*, 2019, **25**, 15315–15325.

(123) H. G. T. Nguyen, N. M. Schweitzer, C. Y. Chang, T. L. Drake, M. C. So, P. C. Stair, O. K. Farha, J. T. Hupp and S. T. Nguyen, *ACS Catal.*, 2014, **4**, 2496–2500.

(124) O. V. Gutov, M. G. Hevia, E. C. Escudero-Adán and A. Shafir, *Inorg. Chem.*, 2015, **54**, 8396–8400.

(125) A. M. Abdel-Mageed, B. Rungtaweevoranit, M. Parlinska-Wojtan, X. Pei, O. M. Yaghi and R. Jürgen Behm, *J. Am. Chem. Soc.*, 2019, **141**, 5201–5210.

(126) P. Ji, X. Feng, P. Oliveres, Z. Li, A. Murakami, C. Wang and W. Lin, *J. Am. Chem. Soc.*, 2019, **141**, 14878–14888.

(127) J. Wang, L. Liu, C. Chen, X. Dong, Q. Wang, L. Alfilfil, M. R. Alalouni, K. Yao, J. Huang, D. Zhang and Y. Han, *J. Mater. Chem. A*, 2020, **8**, 4464–4472.

(128) Y. Liu, R. C. Klet, J. T. Hupp and O. Farha, *Chem. Commun.*, 2016, **52**, 7806–7809.

(129) G. Fu, F. G. Cirujano, A. Krajnc, G. Mali, M. Henrion, S. Smolders and D. E. De Vos, *Catal. Sci. Technol.*, 2020, **10**, 4002–4009.

(130) F. G. Cirujano and F. X. Llabrés I Xamena, *J. Phys. Chem. Lett.*, 2020, **11**, 4879–4890.

(131) Y. Fu, J. Wu, R. Du, K. Guo, R. Ma, F. Zhang, W. Zhu and M. Fan, *RSC Adv.*, 2019, **9**, 37733–37738.

(132) Z. Xu, J. Cao, X. Chen, L. Shi and Z. Bian, *Trans. Tianjin Univ.*, 2021, **27**, 147–154.

2 Theoretical Background and Methods

This section of the doctoral work aims to provide the reader with sufficient knowledge of the modelling techniques used so that the works in **Chapter 3**, **Chapter 4**, and **Chapter 5** may be understood. It is not intended to be more than a high-level introductory text; more complete sources will be referred to for details and derivations. The following subjects will be covered: quantum mechanics, density functional theory, molecular dynamics, vibrational analysis, and transition state methods.

2.1 Hartree-Fock Theory

Quantum mechanics is a powerful mathematical framework for describing how a system evolves on the scale of particles and molecules. It is centrally governed by a set of postulates which act as rules that the quantum description should obey and enable the physical interpretation of mathematical results. The Schrödinger equation is often taken as one of these postulates and will be the starting point for the following introduction to quantum mechanics. Fundamentally, this equation describes the relationship between a quantum object and its energy, such that changes in the latter can be tracked as changes are made to the system (such as atoms being moved around) or when an external stimulus (such as an electric field) is applied. For simplicity, we begin with the time-independent non-relativistic formulation:

$$\hat{H}\Psi(\boldsymbol{R}) = E\Psi(\boldsymbol{R}) \tag{1}$$

where $\hat{H}$ is the Hamiltonian operator describing the sum of the kinetic and potential energy operators $\hat{T}$ and $\hat{V}$:

$$\hat{H} = \hat{T} + \hat{V} \tag{2}$$

$\Psi(\boldsymbol{R})$ is the wavefunction, a complex-valued mathematical object which describes the system and depends on a set of variables $\boldsymbol{R}$; and E is the total energy. If we consider an isolated molecule with n electrons and

N nuclei described respectively by coordinates $\boldsymbol{x}$ and $\boldsymbol{X}$, masses m_e and M, and charges e and Z, a more explicit form can be used:

$$\begin{aligned}\hat{H} = &-\sum_{i}^{N}\frac{\hbar^2}{2M_i}\boldsymbol{\nabla}^2_{X_i} - \sum_{i}^{n}\frac{\hbar^2}{2m_e}\boldsymbol{\nabla}^2_{x_i} - \sum_{i}^{n}\sum_{j}^{N}\frac{Z_i e}{4\pi\epsilon_0|\boldsymbol{X}_i - \boldsymbol{x}_j|} \\ &\sum_{i}^{N}\sum_{j>i}^{N}\frac{Z_i Z_j}{4\pi\epsilon_0|\boldsymbol{X}_i - \boldsymbol{X}_j|} + \sum_{i}^{n}\sum_{j>i}^{n}\frac{e^2}{4\pi\epsilon_0|\boldsymbol{x}_i - \boldsymbol{x}_j|}\end{aligned} \tag{3}$$

These terms correspond respectively to: the kinetic energies of nuclei and electrons, as described by the Laplacian operator $\boldsymbol{\nabla}^2$; the attractions between nuclei and electrons; and electron-electron and nuclear-nuclear repulsion. This can be rewritten more simply in operator form:

$$\hat{H} = \hat{T}_n + \hat{T}_e + \hat{V}_{ne} + \hat{V}_{nn} + \hat{V}_{ee} \tag{4}$$

One issue with **Equation 3** is that it can only be solved analytically for a limited number of simple cases. For any system containing more than two particles, a number of approximations must be made. We follow here the approach of Hartree and Fock through which a useful and tractable set of simplified equations can be reached. Firstly, given that electronic degrees of freedom vary much faster than those of nuclei on account of the lower mass of these particles, the nuclear-only terms can be treated as constants and dropped from the Hamiltonian (this is known as the Born–Oppenheimer approximation, or BO):

$$\hat{H} = \hat{T}_e + \hat{V}_{ne} + \hat{V}_{ee} \tag{5}$$

Now the Schrödinger equation depends explicitly on electronic coordinates and only parametrically on atomic coordinates in the third term of **Equation 3**. It is still not possible to derive a general solution for this simplified equation, but progress can be made if it is assumed that the wavefunction can be expressed as a combination of 1-electron wavefunctions. This entails that each electron no longer interacts explicitly with the others, instead doing so through an 'effective' potential.[1] It is also important that this total electronic wavefunction be antisymmetric to respect that these particles are indistinguishable and fermions. As each electron has both spatial and spin coordinates, it will be described

by both spatial (φ) and spin (α or β) wavefunctions. These are referred to respectively as atomic and spin orbitals, mathematical objects which describe the position and behaviour of electrons within an atom (or molecule for molecular orbitals). They can be combined together to form 1-electron atomic spin orbitals:

$$\phi_j(i) = \varphi_j(i)\alpha(i) \tag{6}$$

$$\phi_j(i) = \varphi_j(i)\beta(i) \tag{7}$$

where $\phi_j(i)$ describes electron i in atomic spin orbital j. As the Hamiltonian operator is Hermitian, in order to serve as valid solutions, atomic spin orbitals will need to be orthonormal, meaning they obey the following condition:

$$\int \phi_j \phi_k^* d\tau = \delta_{jk} \tag{8}$$

where * denotes a complex conjugate, τ runs over all coordinates, and δ_{jk} is the Kronecker delta, equal to 1 if j=k and 0 otherwise. A convenient way to reach a total electronic wavefunction which is antisymmetric with respect to swapping electron labels in either spatial or spin wavefunctions is to use a Slater determinant of such atomic spin orbitals:

$$\Psi(1, 2...n) = \frac{1}{\sqrt{N!}} \begin{vmatrix} \phi_1(1)\ \phi_2(1)\ ...\ \phi_n(1) \\ \phi_1(2)\ \phi_2(2)\ ...\ \phi_n(2) \\ \vdots \\ \phi_1(n)\ \phi_2(n)\ ...\ \phi_n(n) \end{vmatrix} \tag{9}$$

The prefactor of $\frac{1}{\sqrt{N!}}$ is required to keep Ψ normalised. A natural consequence of antisymmetric wavefunctions such as these is that they would equal 0 if any electrons were to occupy the same atomic spin orbital. This is the well-known Pauli exclusion principle. Before applying our ansatz wavefunction to **Equation 5**, it is useful to introduce the notation $|\boldsymbol{x}_i - \boldsymbol{x}_j| = x_{ij}$ and use atomic units, such that e, m_e, $\hbar$, and $4\pi\epsilon_0$ all have a numerical value of 1 (which will be used henceforth). This notably means that within this scope, charge, mass, length, and energy

are respectively measured in units of the charge on an electron, the rest electron mass, the Bohr radius a_0 ($a_0 = \frac{4\pi\epsilon_0\hbar^2}{m_e e^2}$), and the Hartree E_h ($E_h = \frac{\hbar^2}{m_e a_0^2}$). We also define the single-particle operator $\hat{h_i}$:

$$\hat{h_i} = -\frac{1}{2}\nabla^2_{\boldsymbol{x}_i} - \sum_j^N \frac{Z_j}{|\boldsymbol{X}_j - \boldsymbol{x}_i|} \tag{10}$$

The Hamiltonian can be separated into terms that depend on one and two particles:

$$\hat{H} = \sum_i^n \hat{h_i} + \sum_i^n \sum_{j>i}^n \frac{1}{x_{ij}} \tag{11}$$

We will henceforth use Dirac notation where possible. Our ansatz wavefunction can be applied in **Equation 11**. Given that our atomic spin orbitals are orthonormal, the result for the one-particle operator is straightforward:

$$\langle\Psi| \sum_i^n \hat{h_i} |\Psi\rangle = \sum_i^n \langle\phi_i(1)|\hat{h_1}|\phi_i(1)\rangle = \sum_i^n h_i \tag{12}$$

where h_i are the energies of the 1-electron atomic spin orbitals. The same can be done for the two-particle operator, which although algebraically more tricky nevertheless reduces to 2 sets of terms:[2]

$$\begin{aligned}\langle\Psi| \sum_i^n \sum_{j>i}^n \frac{1}{x_{ij}} |\Psi\rangle = \sum_i^n \sum_{j>i}^n \Big(&\langle\phi_i(1)\phi_j(2)|\frac{1}{x_{12}}|\phi_i(1)\phi_j(2)\rangle \\ &- \langle\phi_i(1)\phi_j(2)|\frac{1}{x_{12}}|\phi_i(2)\phi_j(1)\rangle \Big)\end{aligned} \tag{13}$$

At this stage, it is not clear how the Schrödinger equation can be solved or what form the atomic spin orbitals should take. However, the energy determined using an arbitrary wavefunction will only ever be greater than or equal to the ground-state energy found using the ground-state wavefunction. A solution can therefore be reached by minimising the energy through variation in some trial Ψ, in an approach known as the variational principle. However, it is necessary that the atomic spin orbitals remain orthonormal during this process. This can be done using the method of Lagrange multipliers, which will be labelled $\epsilon_{i,j}$, and the Lagrangian $\mathcal{L}$:

$$\mathcal{L} = \langle\Psi|\hat{H}|\Psi\rangle - \sum_i^n \sum_j^n \epsilon_{i,j} \langle\phi_i(1)|\phi_j(1)\rangle \tag{14}$$

We now allow variation in the atomic spin orbitals and set the Lagrangian to 0:

$$\delta\mathcal{L} = 0 = \sum_{i}^{n} \langle \delta\phi_i(1)|\hat{h_1}|\phi_i(1)\rangle + \sum_{i}^{n}\sum_{j>i}^{n} \Big(\langle \delta\phi_i(1)\phi_j(2)|\frac{1}{x_{12}}|\phi_i(1)\phi_j(2)\rangle - \langle \delta\phi_i(1)\phi_j(2)|\frac{1}{x_{12}}|\phi_i(2)\phi_j(1)\rangle \Big) - \sum_{i}^{n}\sum_{j}^{n} \epsilon_{i,j} \langle \delta\phi_i(1)|\phi_j(1)\rangle \tag{15}$$

This can be re-arranged into a simpler form by factorising $\sum_i^n \langle \delta\phi_i(1)|$:

$$0 = \sum_{i}^{n} \langle \delta\phi_i(1)| \Bigg[\hat{h_1}\,|\phi_i(1)\rangle + \sum_{j>i}^{n} \Big(|\phi_j(2)\frac{1}{x_{12}}\phi_i(1)\phi_j(2)\rangle - |\phi_j(2)\frac{1}{x_{12}}\phi_i(2)\phi_j(1)\rangle \Big) - \sum_{j}^{n} \epsilon_{i,j}\,|\phi_j(1)\rangle \Bigg] \tag{16}$$

Given that the variation $\delta\phi_i(1)$ is arbitrary, the terms inside the square braces must independently equal 0. Here we introduce the Coulomb operator $\hat{J}_j$:

$$\hat{J}_j(1)\phi_i(1) = \phi_i(1) \int |\phi_j(2)|^2 \frac{1}{x_{12}} d\boldsymbol{x_2} \tag{17}$$

where $\boldsymbol{x_2}$ are the coordinates associated with electron 2. Given the use of atomic units, normalised 1-electron atomic spin orbitals, and the general interpretation of $|\Psi|^2$ as a probability density, $|\phi_j(2)|^2$ is the electron density associated with a single electron. The Coulombic character of this operator is therefore reflected by the interaction of electron 1 with the density of electron 2 via a $\frac{1}{x_{12}}$ term. The exchange operator, which has no classical analogue and arises out of the antisymmetry of the wavefunction, is also defined:

$$\hat{K}_j(1)\phi_i(1) = \phi_j(1) \int \phi_j^*(2) \frac{1}{x_{12}} \phi_i(2) d\boldsymbol{x_2} \tag{18}$$

We re-write again **Equation 16** in terms of these operators:

$$\hat{h_1}\,|\phi_i(1)\rangle + \sum_{j>i}^{n} \Big(\hat{J}_j(1)\,|\phi_i(1)\rangle - \hat{K}_j(1)\,|\phi_i(1)\rangle \Big) = \sum_{j}^{n} \epsilon_{i,j}\,|\phi_j(1)\rangle \tag{19}$$

We can define the Fock operator as a combination of the 1-electron, Coulomb, and exchange operators:

$$\hat{f} = \hat{h_1} + \sum_{j>i}^{n} (\hat{J}_j - \hat{K}_j) \tag{20}$$

This allows us to simplify our expression for energy further:

$$\hat{f}\phi_i(1) = \sum_j^n \epsilon_{i,j}\phi_j(1) \tag{21}$$

Some further manipulation is possible which leaves $\hat{f}$ unchanged but turns ϵ into a diagonal matrix, yielding the Hartree-Fock (HF) equations:[3]

$$\hat{f}\phi_i(1) = \epsilon_i\phi_i(1) \tag{22}$$

It can be shown that ϵ_i are equal to the energies of the single-electron orbitals.[4] Finally, each atomic spin orbital can be expressed as a linear combination, described by coefficients c, of simpler functions known as basis functions (such as Gaussians or polynomials), denoted χ:

$$\phi_i(q) = \sum_l^k c_{i,l}\chi_l(q) \tag{23}$$

By doing this, **Equation 22** can be turned into a set of coupled linear equations which can be solved efficiently with computers using matrix algebra:[5]

$$\Big[\mathbf{F} - \epsilon\mathbf{S}\Big]\mathbf{c} = 0 \tag{24}$$

where $\mathbf{F}$ contains elements $\langle\chi|\hat{f}_1|\chi\rangle$; $\mathbf{S}$ is the overlap matrix of the basis functions and contains elements $\langle\chi_l|\chi_{l'}\rangle$; and $\mathbf{c}$ is a matrix of coefficients. These are known as the Roothaan-Hall equations. Given that $\mathbf{F}$ depends on the coefficients $\mathbf{c}$ via the Coulomb and exchange terms, these equations must be solved by first choosing a basis set, guessing the coefficients to define the Fock operator, then iteratively refining $\mathbf{c}$ and $\mathbf{F}$ until a self-consistent solution is reached (hence this approach is termed self-consistent field, or SCF). Different criteria may be used to signal for convergence, but these usually entail sufficiently small changes between subsequent steps in the electronic energy or atomic forces. This scheme is summarised in **Figure 7**:

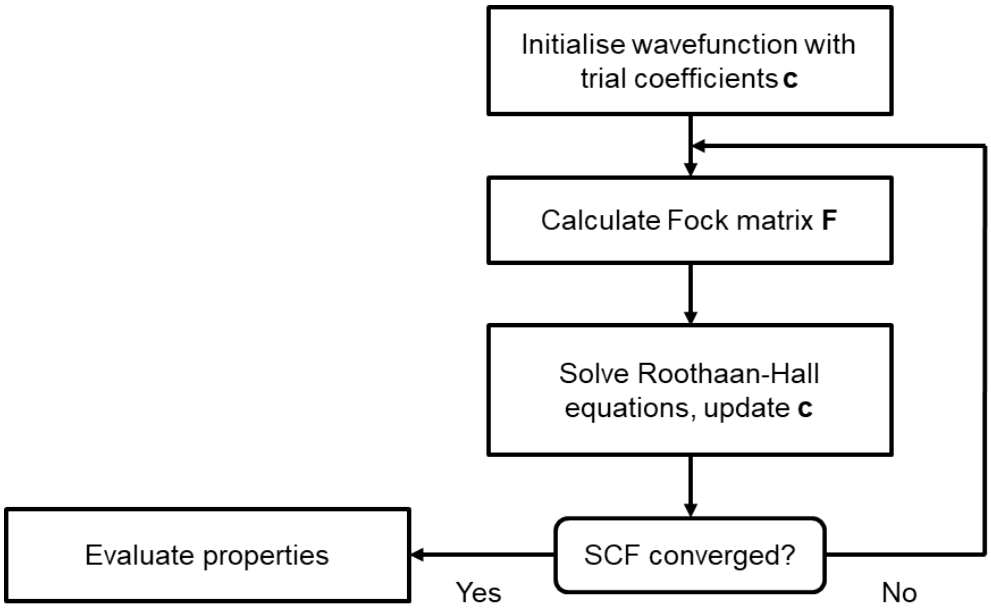


Figure 7: Schema for reaching a self-consistent electronic solution with Hartree-Fock theory. With a good guess for the trial wavefunction, fewer SCF loops are typically required before convergence is reached. A standard procedure is to use coefficients which correspond to a sum of spherically averaged densities of isolated atoms.

HF calculations typically yield good geometries and properties,[6] particularly for mononuclear species.[7] However, the single-particle treatment of the wavefunction means some electron-electron correlation is lost. While this typically represents less than 1% of a system's energy, it is sufficient to cause significant inaccuracies in a range of cases. For example, F_2 is predicted to be an unbound molecule and intermolecular interactions are underestimated in dispersion-dominated systems such as molecular crystals.[8,9] The HF method also has a tendency to overestimate the energy difference between the highest occupied and lowest unoccupied orbitals, known as the band gap.[10,11] Post-Hartree-Fock methods can reintroduce correlation in a number of ways, such as Møller–Plesset theory or coupled-cluster approaches,[12,13] but these techniques are more computationally expensive.

While the HF approach has not been used directly in this work due to the shortcomings highlighted, it represents a good introduction for more modern electronic structure techniques. The moderate size of UiO-66, which contains 456 atoms in the conventional unit cell, also precludes the use of post-HF methods due to their associated computational cost.

2.2 Density Functional Theory

An effective alternative to the HF method is density functional theory, which aims to solve the Schrödinger equation in terms of the electron density distribution rather than the wavefunction. Hohenberg and Kohn have shown that the electron density of a system uniquely defines the Hamiltonian operator, from which all its ground-state properties can be derived.[14] This is known as the first Hohenberg-Kohn theorem. For a given wavefunction representing n total electrons, the density is linked via the relation:

$$\rho(\boldsymbol{x}) = n|\Psi(\boldsymbol{x})|^2 \tag{25}$$

A key advantage of working with the density is that it is an observable quantity and remains a function of 3 variables no matter the size of the problem. In order to reach a set of computationally useful equations, we can start from **Equation 5** and reformulate it in terms of density functionals. We now also express the nuclear-electronic term as an interaction between an external potential $v(x)$, resulting from the nuclear electric fields, and the total electron density:

$$E[\rho] = T[\rho] + \int \rho(\boldsymbol{x})v(\boldsymbol{x})\mathrm{d}\boldsymbol{x} + V_{ee}[\rho] \tag{26}$$

Hohenberg and Kohn showed that this total DFT energy is variational in the density, providing a means for solving this equation.[14] This is the second Hohenberg-Kohn theorem. Unfortunately, the exact forms of $T[\rho]$ and $V_{ee}[\rho]$ are not known, but the subsequent work of Kohn and Sham provides a way forward.[15] We consider a reference system of non-interacting electrons subject to an effective potential v_{eff} such that their

density is the same as that of the real system. We can now express the non-interacting kinetic energy in terms of some 1-electron orbitals ψ:

$$T_s[\rho] = \sum_{i=1}^{N} \langle \psi_i | \frac{-\nabla^2}{2} | \psi_i \rangle \tag{27}$$

These are linked to the electron density by the relation:

$$\rho(\boldsymbol{x}) = \sum_{i=0}^{N} |\psi_i(\boldsymbol{x})|^2 \tag{28}$$

We can also separate from $V_{ee}[\rho]$ the classical Coulomb contribution $J[\rho]$:

$$J[\rho] = \frac{1}{2} \int \frac{\rho(\boldsymbol{x})\rho(\boldsymbol{x}')}{|\boldsymbol{x} - \boldsymbol{x}'|} \mathrm{d}\boldsymbol{x} \mathrm{d}\boldsymbol{x}' \tag{29}$$

The total energy of the real system can in this way be re-written as:

$$E[\rho] = T_s[\rho] + \int \rho(\boldsymbol{x}) v(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} + J[\rho] + E_{xc}[\rho] \tag{30}$$

where the final term, the exchange-correlation functional, accounts for differences between the non-interacting and real kinetic energies as well as between the Coulombic and full electronic interactions:

$$E_{xc}[\rho] = (T[\rho] - T_s[\rho]) + (V_{ee}[\rho] - J[\rho]) \tag{31}$$

Meanwhile, the energy for the reference non-interacting system is simply given by:

$$E[\rho] = T_s[\rho] + \int \rho(\boldsymbol{x}) v_{\text{eff}}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \tag{32}$$

We can now make use of the same procedure as for the Hartree-Fock equations. Namely, we minimise the energy subject to the constraint that the 1-electron orbitals remain orthonormal. Doing this for **Equation 30** yields the Kohn-Sham (KS) equations:

$$\Big(\frac{-\nabla^2}{2} + v + \int \frac{\rho(\boldsymbol{x}')}{|\boldsymbol{x} - \boldsymbol{x}'|} \mathrm{d}\boldsymbol{x} + \frac{\delta E_{\text{xc}}}{\delta \rho} \Big) \psi_i = \epsilon_i \psi_i \tag{33}$$

$\frac{\delta E_{\text{xc}}}{\delta \rho}$ is the functional derivative of the exchange-correlation functional and ϵ_i are the energies of the 1-electron states. The same process can be carried out for the reference system:

$$\Big(\frac{-\nabla^2}{2} + v_{\text{eff}} \Big) \psi_i = \epsilon_i \psi_i \tag{34}$$

It can be seen by examining both expressions that the energy of the real system can be found from the reference one through a judicious choice for the effective potential:

$$v_{\text{eff}} = v + \int \frac{\rho(\boldsymbol{x}')}{|\boldsymbol{x} - \boldsymbol{x}'|} \mathrm{d}\boldsymbol{x} + \frac{\delta E_{\text{xc}}}{\delta \rho} \tag{35}$$

The second and third terms on the right hand side are often respectively called the Hartree and exchange-correlation potentials. As done in **Section 2.1**, **Equation 34** can be recast into matrix form by expressing the single-particle orbitals as linear combinations of basis functions, such that they can computationally be solved efficiently. Note that, as with the Hartree-Fock equations, these need to be solved in a self-consistent manner given that v_{eff} depends on the density. The general scheme for this is shown in **Figure 8**.

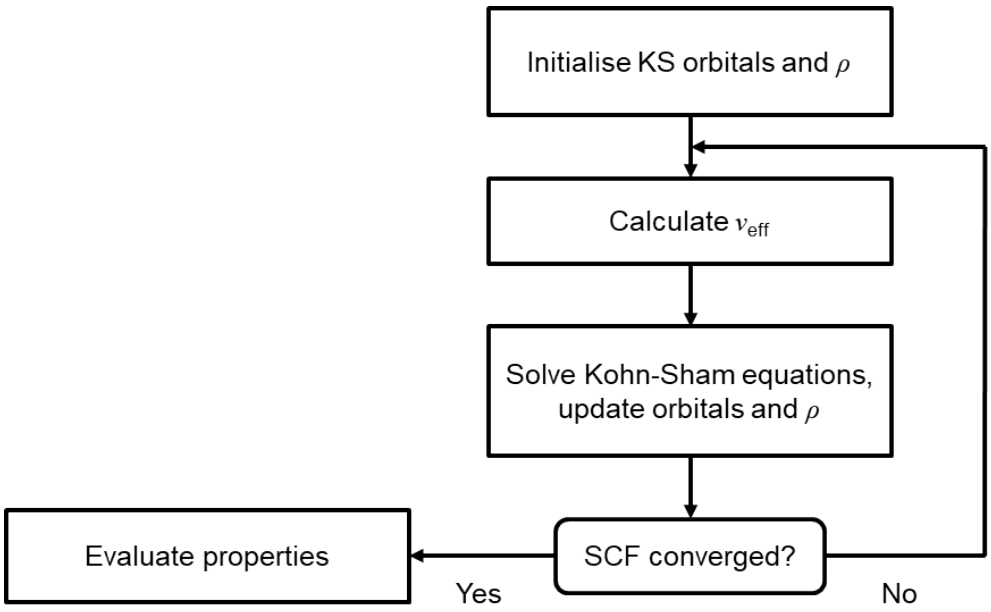


Figure 8: Schema for reaching a self-consistent electronic solution with density functional theory. As with the Hartree-Fock method, a good initial guess helps the procedure converge, with a sum of atomic densities the standard choice.

Nothing has yet been said about the functional form of E_{xc} and therefore v_{xc}, which should correct for a number of differences between the reference and real systems. Firstly, the kinetic energy is that of non-interacting electrons, which is easy to calculate but is not the same as

that of the interacting system. Next is the Hartree energy, which as formulated in **Equation 34** includes unphysical self-overlap of electrons via the total density, also known as self-interaction error. Finally there are all the electron exchange and correlation interactions not covered by the Hartree term. Unfortunately, this functional is generally not known and must be approximated, which is one of the major sources of error in DFT.

In their original formulation, Kohn and Sham used the exact exchange-correlation energy for a homogeneous electron gas,[16] which depends only on density, and argued that this is valid in the limit of slowly-varying electronic density. This is known as the local-density approximation (LDA) and works remarkably well at predicting structure for a range of systems.[17] However, this method is generally overbinding, such that atomisation energies are overestimated and lattice constants of alkali metals are underestimated.[18] LDA also leads to incorrect relative energetic ordering for hydrocarbon conformers and for phases of silica and iron.[19,20] These shortcomings have led to the development of ever more sophisticated DFT methods, which are typically described within a hierarchy referred to as Jacob's ladder.[21] The next rung on this ladder is the generalised gradient approximation (GGA), which aims to improve the chemical accuracy by making the exchange-correlation functional dependent also on the gradient of the electron density.[22–26] The most ubiquitous of these is Perdew–Burke-Ernzerhof (PBE),[27] on account of its good (but not necessarily best-in-class) performance across a large range of systems.[28] This can be further augmented by taking into account the kinetic energy density and sometimes the second gradient of the density in what is then referred to as a meta-GGA functional.[29–32] However, these sometimes suffer from slow convergence and poor numerical stability.[33] It is also possible to include a fraction of the exact Hartree-Fock exchange in the Hamiltonian, leading to yet another class referred to as hybrid functionals.[34–36] This can lead to further refinement of atomisation energies, barriers, and formation enthalpies;[37] the description of

band gaps and strongly correlated solids can also be improved.[38,39] There are further rungs on the ladder representing corrections for various shortcomings of DFT, but these will not be presented here.[40] Note also that computational cost generally increases in going up the ladder, such that for a given set of experiments, this has to be weighed against the desired chemical accuracy. Some functionals are also constructed to capture certain effects or fitted against certain classes of materials, so performance can vary depending on the system studied.[41]

All calculations in this work have been carried out at the DFT level, as justified by the good performance of this method across metal-organic frameworks.[42] More specifically, the PBE[27] functional has generally been chosen given the good compromise it offers between chemical accuracy and computational expense for modelling UiO-66.[43–52] While the use of DFT for this material is well-established and post-HF methods have been discounted on account of computational cost, it is necessary to comment on another common approach which has not been used here, that of force fields. These model bonded interactions using a range of spring or spring-like terms fitted to reproduce *ab initio* or empirical data. Non-bonded interactions such as van der Waals and electrostatic forces are described by pair potentials. While this method is computationally cheaper than DFT, its accuracy and transferability are typically limited by the quality of the fitting. The large number of different chemical environments brought about by defects in UiO-66 would require fitting a very large number of different parameters, which would be a considerable undertaking in its own right. Generic force fields such as UFF4MOF[53] and Dreiding[54] have also been deemed inadequate as these typically fail to reproduce some important chemistry, such as the behaviour around unsaturated metals.[55,56] Therefore, given these concerns around fitting and transferability, DFT has been chosen ahead of force fields.

2.2.1 Open-shell systems

The implementation of DFT presented thus far is a function only of the total electron density $\rho(\mathbf{x})$, thus neglecting the possibility for α and β spin channels having unequal occupancies or different spatial orbitals. Taking these factors into account becomes important when modelling systems containing unpaired electrons, such as magnetic materials, organic radicals, and transition metal complexes, or even for the dissociation of dimers.[57] In the context of the current work, this is relevant for modelling cerium in its +3 oxidation state, which contains an unpaired electron. The main ways in which standard KS-DFT can be modified to properly model open-shell systems are the unrestricted Kohn-Sham (UKS)[58] and restricted open-shell Kohn-Sham (ROKS)[59] approaches, which will be briefly described here. In general, the Hamiltonian, which was previously a function of the total electron density only, will now include terms that depend on the total density, as well as the densities of α and β electrons, related by:

$$\rho(\mathbf{x}) = \rho_\alpha(\mathbf{x}) + \rho_\beta(\mathbf{x}) \tag{36}$$

With the UKS method, the reference system of non-interacting electrons envisaged by Kohn and Sham is extended such that its α and β spin densities also match those of the real system. In addition, the spatial parts for the two spin channels are no longer required to be the same (hence the method is termed 'unrestricted') for a given orbital. One consequence of this is that the kinetic energy now depends on the spin densities since its corresponding operator acts on orbitals. This in turn means the exchange-correlation functional becomes dependent on the spins given that it is defined in part by the difference between the kinetic energies of the real and non-interacting systems. Some manipulation of **Equation 33** and **Equation 34** using these modified operators leads to two sets of Kohn-Sham equations,[60] coupled through the total and spin

densities, from which the α and β spatial orbitals can be determined:

$$\Big(-\frac{1}{2}\nabla^2 + v_{\text{Coulomb}}[\rho] + v + v_{\text{XC}}^{\alpha}[\rho^{\alpha}, \rho^{\beta}] \Big) \phi_i^{\alpha} = \epsilon_i^{\alpha} \phi_i^{\alpha} \tag{37}$$

$$\Big(-\frac{1}{2}\nabla^2 + v_{\text{Coulomb}}[\rho] + v + v_{\text{XC}}^{\beta}[\rho^{\alpha}, \rho^{\beta}] \Big) \phi_i^{\beta} = \epsilon_i^{\beta} \phi_i^{\beta} \tag{38}$$

The spin state, corresponding to some value for the multiplicity M_S, can be set by partially populating the corresponding orbitals. One consequence of allowing the spatial orbitals of both spins to differ is that the resulting Slater determinant solution is no longer an eigenfunction of the spin squared operator, $\langle \hat{S} \rangle^2$. This operator has an expectation value given by:

$$\langle \hat{S} \rangle^2 = M_S(M_S + 1) + N_B - \sum_{i=1}^{N_A} \sum_{j=1}^{N_B} < \phi_i^{\alpha} | \phi_j^{\beta} > \tag{39}$$

N_A and N_B are the numbers of alpha and beta electrons respectively, with the former assumed to be larger. Since the spatial parts of alpha and beta orbitals are no longer the same, they are in general not orthogonal to each other (though separately, they remain orthogonal sets). As a result, the last 2 terms in **Equation 39** only partially cancel. This effect, referred to as spin contamination, corresponds to mixing between different electronic states and can artificially lower the energy in an unphysical manner. The deviation between the expected and calculated values of $\langle \hat{S} \rangle^2$ therefore needs to be monitored; a common rule of thumb is that this difference should not exceed 10%.[61] Methods such as spin annihilation and spin projection are available to correct for this,[62,63] but their application is difficult in periodic calculations due to the common use of plane wave basis functions for orbitals.[64]

Alternatively, the spatial parts of α and β orbitals can be kept the same (restricted), as is done with the ROKS method.[59] In this case, the problem is still decomposed into two electronic subsystems corresponding to closed and open orbitals, but they are governed by a single Hamiltonian.[59,65] The contributions to this operator are instead varied according

to the occupancy of the shell, which will be set by the overall multiplicity. This approach eliminates spin contamination, as the last 2 terms in **Equation 39** cancel out, but this does not yield the correct spin density as this is not a criterion for the non-interacting reference system.

2.2.2 Dispersion Interactions

One important shortcoming of standard DFT is its failure to properly model dispersion interactions.[66] LDA, GGA, and hybrid functionals typically fail to reproduce these effects which arise from long-range electronic correlation. There are two main approaches through which this can be taken into account: semi-empirical corrections and dedicated density-dependent functionals.

The idea behind semi-empirical dispersion corrections is simple. London's formulae for the interaction energies between temporary induced multipoles are used to calculate interaction energies for all atom pairs,[67] which are straightforwardly added to the DFT energy. The functional form below is generally followed:

$$E_{\text{disp.}} \propto \sum_{i} \sum_{k=0} \frac{-C_{i,k}}{r_i^{6+2k}} f(r_i) \tag{40}$$

The index i runs over pairs of atoms, separated by a distance r_i and which have an atom type-dependent coefficient $C_{i,k}$. These also depend on the order of multipole considered, which runs over index k. Finally, a damping function f is required to avoid the singularity that would otherwise arise near $r = 0$. Usually, only the dipole terms are used, but given the low computational cost, other multipoles can be included as long as the coefficients are available. The most common schemes which make use of this straightforward approach are those of the group of Grimme, known as DFT-D.[68–71] Briefly, these differ in the choice of damping function, the manner in which C coefficients are tabulated, and the functional-dependent fitting of scaling parameters. The most widely used of these is D3, in which different coefficients are used for a given atom depending on its coordination environment, which greatly improves the DFT

description of non-covalent interactions.[72] Other similar schemes include those of Tkatchenko and Scheffler or of Becke and Johnson.[73,74]

Another common approach is to use a van der Waals density functional (vdW-DF). The exchange-correlation energy is decomposed into 3 terms:

$$E_{\mathrm{xc}} = E_{\mathrm{x}}^{GGA} + E_{\mathrm{c}}^{LDA} + E_{\mathrm{c}}^{nl} \tag{41}$$

which respectively correspond to the exchange of a chosen GGA functional, the correlation within the LDA, and non-local correlation. This last term is a double space integral of the density and some kernel function ϕ, which is intended to capture dispersion interactions in an *ab initio* manner:

$$E_{\mathrm{c}}^{nl} = \int\int \rho(\boldsymbol{x})\phi(\boldsymbol{x},\boldsymbol{x}')\rho(\boldsymbol{x}')\mathrm{d}\boldsymbol{x}\mathrm{d}\boldsymbol{x}' \tag{42}$$

Two main sets of kernels are used, those of Dion *et al.* and those of Vydrov and Van Voorhis, both of which make use of the polarisation operator.[75,76] This method leads to a good description of materials, even when dispersion forces are not expected to be dominant, but at higher computational cost than standard GGA functionals.[77,78] Note that these do not necessarily improve over the semi-empirical pairwise correction schemes.[79] More sophisticated approaches exist, though these overlap with the upper rungs of Jacob's ladder of DFT.

In the current study, the semi-empirical D3 correction has been used for all calculations.[70] This choice, particularly combined with the PBE functional, has been found to yield good results for dispersion-dominated phenomena in MOFs.[80,81]

2.2.3 Periodic calculations

Thus far, nothing has been said in **Section 2.1** and **Section 2.2** about the nature of the systems being modelled. The equations derived are straightforward for isolated molecules as they represent a finite number of electrons. In an object of macroscopic size, the number of relevant particles is virtually infinite and the HF or KS equations are intractable. It is possible to sidestep this problem by relying on translational and other symmetry to reduce a given periodic system to a much smaller building block known as the unit cell. This is a parallelepiped, bounded by 3 edge vectors **a**, **b**, and **c**, which can be tesselated in 3 dimensions to recover the macroscopic material. The resulting grid is known as the lattice. A complementary description in reciprocal space can be defined from the real space unit cell edge (or lattice) vectors, giving the reciprocal lattice vectors:

$$\boldsymbol{a}^* = \frac{2\pi \boldsymbol{b} \times \boldsymbol{c}}{\boldsymbol{a} \cdot (\boldsymbol{b} \times \boldsymbol{c})} \quad \boldsymbol{b}^* = \frac{2\pi \boldsymbol{a} \times \boldsymbol{c}}{\boldsymbol{b} \cdot (\boldsymbol{a} \times \boldsymbol{c})} \quad \boldsymbol{c}^* = \frac{2\pi \boldsymbol{a} \times \boldsymbol{b}}{\boldsymbol{c} \cdot (\boldsymbol{a} \times \boldsymbol{b})} \tag{43}$$

These describe the reciprocal unit cell and the reciprocal lattice. The reader is referred to ref. [82] for a more complete discussion. Representing a periodic system in this way leaves the Schrödinger equation unchanged except that the potential has the same periodicity as the unit cell:

$$V(\boldsymbol{x}) = V(\boldsymbol{x} + \boldsymbol{X}) \tag{44}$$

where $\boldsymbol{X}$ is an integer combination of lattice vectors. Given this, Bloch's Theorem states that acceptable one-electron wavefunctions take the form below:[82]

$$\psi_{n,\boldsymbol{k}}(\boldsymbol{x}) = u_{n,\boldsymbol{k}}(\boldsymbol{x}) e^{i\boldsymbol{k}\bullet\boldsymbol{x}}, \quad u_{n,\boldsymbol{k}}(\boldsymbol{x}) = u_{n,\boldsymbol{k}}(\boldsymbol{x} + \boldsymbol{X}) \tag{45}$$

n is a quantum number describing the energy level of the given eigenstate; $\boldsymbol{k}$ is a vector in reciprocal space known as the wavevector; $u_{\boldsymbol{k}}$ are parametric functions in $\boldsymbol{k}$ known as Bloch functions; and the exponential term is a complex phase factor. As this description of the system is translationally symmetric, $\boldsymbol{k}$ can always be expressed as $\boldsymbol{k}' = \boldsymbol{k} - \boldsymbol{K}$, where

$\boldsymbol{K}$ is an integer combination of reciprocal lattice vectors, such that $\boldsymbol{k}'$ lies within the Wigner-Seitz primitive cell of the reciprocal lattice. This domain, known as the first Brillouin zone (1BZ), contains all values of the wavevector necessary to define the system's electronic states. Most of the material's properties and behaviour can be described using the information within the 1BZ. Formally, $\boldsymbol{k}$ is also quantised, with increments inversely proportional to the real-space size of the material as imposed by Born–von Karman boundary conditions. In practice, if the crystal of interest is macroscopic in size, these increments are sufficiently fine that the wavevector may be treated as a continuous variable.

One implication of Bloch's theorem is that the wavefunction now depends on $\boldsymbol{k}$ as an additional quantum number. It can be shown that the exponential term in **Equation 45** introduces a wavevector dependence for the kinetic energy, such that the Schrödinger equation for a Bloch function becomes:[83]

$$\hat{\mathcal{H}}_{\mathbf{k}} u_{\mathbf{k},n}(\mathbf{x}) = \epsilon_{\mathbf{k},n} u_{\mathbf{k},n}(\mathbf{x}) \tag{46}$$

For a fixed n, the energy varies continuously with the wavevector and leads to a family of wavefunction solutions, referred to as a band (hence n is known as the band index). As a result, total energies (and other quantities which depend on $\boldsymbol{k}$) must be evaluated as an integral over the first Brillouin Zone:[84]

$$\epsilon_n = A \int_{\mathrm{BZ}} \epsilon_{\boldsymbol{k},n} \mathrm{d}\boldsymbol{k} \tag{47}$$

where A is a constant related to the volume of the unit cell. Unfortunately, the dependence of the energy on k is not known analytically, such that it has to be estimated. It therefore appears that the general problem has shifted from one with an infinite number of electrons to one which requires evaluation over an infinite number of k-points. However, the

pseudo-continuous nature of k means **Equation 47** can be approximated as a finite sum rather than an integral. Furthermore, the wavefunction varies quite slowly with $\boldsymbol{k}$ for insulators and semiconductors, such that only a small number of summation terms, chosen using schemes such as Monkhorst-Pack,[85] are needed to accurately determine the energy. Even fewer points are needed if the symmetry of the 1BZ is exploited: the possible values of $\boldsymbol{k}$ can be covered by a set of high-symmetry locations, usually given special notation, and the paths which interconnect them. Given also the reciprocal relationship between the sizes of the real-space and the reciprocal unit cells, often a single k-point will suffice (where $\boldsymbol{k} = 0$, denoted as Γ) to converge properties. This is particularly appropriate for insulators with large cell parameters ($>\sim$10 Å), such as MOFs, where this choice leads to a sampling density of around 0.5 $Å^{-1}$ or better.[86] Such a choice should therefore be suitable for the current study's system of interest, UiO-66, given its experimental cell parameter of 20.7004 Å.[87] For conductors, where the energy varies rapidly as a function of $\boldsymbol{k}$, a much denser net of k-points is usually required, along with methods which smooth out discontinuities.[88,89]

2.2.4 Basis sets

The approaches discussed in **Section 2.1** and **Section 2.2** both ultimately represent orbitals using linear combinations of basis functions. These basis sets will be discussed here, first for isolated systems, and subsequently for periodic materials where Bloch's theorem must additionally be respected.

Given the non-interacting treatment of electrons involved in the HF method and DFT, one might intuitively think to use hydrogenic orbitals (solutions to the Schrödinger equation for the hydrogen atom) for the basis functions. However, these do not ordinarily represent a complete set, meaning it is not possible to decompose an arbitrary wavefunction into a linear combination of these functions without loss of information.[90]

An improvement is to simplify the polynomial part of hydrogen's radial wavefunction as is done for Slater-Type Orbitals:

$$\phi(\boldsymbol{r}) = Y_l^m(\boldsymbol{r})R(r) = Y_l^m(\boldsymbol{r}) \times Ar^{n-1}\exp^{-\zeta r} \tag{48}$$

where $Y_l^m(\boldsymbol{r})$ is a spherical harmonic and describes angular behaviour, while A and ζ are constants (the latter of which is positive). These share key features with the hydrogenic orbitals (exponential decay with r and a Kato cusp at $r = 0$) but have no radial nodes, making them easier to form a complete basis with. However, multi-centre integrals involving such functions are difficult to solve, such that it is more practical to use Gaussians, for which these can be resolved analytically. This comes at the cost of faster-than-exponential decay and the loss of the cusp, hence a single basis function is typically represented as a linear combination of Gaussians:

$$\phi_\alpha = \sum_i^k d_{i,\alpha} g_i \tag{49}$$

This is known as a contracted Gaussian and is described by coefficients $d_{i\alpha}$ and primitive Gaussians g_i which remain fixed during calculations. Such a basis can be improved by adding more primitives to the contractions, or by using more than one contraction per atomic orbital (AO). A basis with 1 contraction per AO is referred to as single-ζ (SZ), 2 contractions as double-ζ (DZ) etc. This representation is convenient as many of the resulting integrals can be solved analytically, and the atom-like nature of these basis functions means the wavefunctions can be interpreted similarly to orbitals. However, they form a non-orthogonal set, which means it is possible for there to be linear dependencies between basis functions. Molecular fragments can also 'borrow' basis functions from each other upon approach, artificially lowering their energies relative to their isolated states in the process. This is known as basis set superposition errors, or BSSE. There are other possibilities for basis functions

in molecular systems, such as piecewise polynomials or numerically tabulated orbitals,[91,92] but Gaussians remain the most widespread.

When dealing with periodic systems, it is also necessary for the wavefunction to follow Bloch's theorem, meaning the basis functions themselves must also be periodic. The most natural form for this is to use plane waves for the Bloch functions:

$$u_{\boldsymbol{k}}(\boldsymbol{x}) = \sum_{\boldsymbol{G}} c_{\boldsymbol{k},\boldsymbol{G}} e^{i\boldsymbol{G}\bullet\boldsymbol{x}} \tag{50}$$

where the sum $\boldsymbol{G}$ runs over reciprocal lattice vectors. In principle, this is an infinite sum, with larger $\boldsymbol{G}$ corresponding a shorter distance between adjacent maxima of the same set of plane waves. In practice, this set is truncated at an upper limit, the plane-wave cutoff, which controls how detailed the basis is. This expansion for the Bloch functions is mathematically convenient: plane waves naturally fulfil the requirement of periodicity; the basis set can be improved systematically by increasing the cutoff; orthogonality and completeness are ensured; and no linear dependencies or superposition errors are possible. However, a notable downside is that rapidly varying atomic core regions require high cutoffs. This is usually mitigated through the frozen-core approximation: core electrons are modelled as static and replaced by a pseudopotential that mimics their corresponding density distribution.

It is also possible to use Gaussians for the Bloch functions in the form of 3D combs:

$$u_{\boldsymbol{k}}(\boldsymbol{x}) = \sum_{i} \phi_{\alpha}(\boldsymbol{x} - \boldsymbol{X}_i) \tag{51}$$

where the summation runs over all integer combinations of the lattice vectors, represented as $\boldsymbol{X}_i$. As there is an infinite number of such combinations, the sum has to be truncated, resulting in approximate Bloch functions.

2.2.5 Forces at the quantum mechanical level

The HF and DFT methods provide recipes for determining the electronic energy of a system for which atoms have been fixed as per the BO approximation. This requires an initial trial geometry for which the atomic positions are generally not optimal, meaning the nuclear energy is not at a minimum. If we are interested in finding the equilibrium geometry of a system, it is necessary to minimise the total energy, which is a function of both nuclear and electronic coordinates. In order to minimise the nuclear component, atoms need to be moved in directions which lower the total energy, which can be done if the forces acting upon them are known. Taking the derivative of the total energy in **Equation 3** with respect to the position of nucleus j yields, for component k:

$$F_{j,k} = -\frac{\mathrm{d}E}{\mathrm{d}X_{j,k}} = -\frac{\mathrm{d}E_e}{\mathrm{d}X_{j,k}} + \sum_{l}^{N}\sum_{i>l}^{N} \frac{Z_l Z_i}{|\boldsymbol{X}_l - \boldsymbol{X}_i|^2}(X_{l,k} - X_{i,k}) \qquad (52)$$

Note that there is no term involving the nuclear kinetic energy given the use of the BO approximation. The electronic energy term can be calculated analytically using a modified form of the Hellmann-Feynman theorem.[93] However, this requires the use of a complete or non-localised set of basis functions. The dependence otherwise of the wavefunction on atomic coordinates will lead to a discrepancy between the Hellmann-Feynman and true forces, which need to be corrected for (these are known as Pulay forces).[94,95] The second term in **Equation 52** is trivial given that atomic charges and nuclear positions are known from the trial geometry.

Once the forces are known, following them allows atoms to be moved towards their equilibrium positions. Given the complexity of the potential energy surface for all but the simplest systems, a number of such steps will be required before the structure reaches a minimum. Convergence is signalled once some criteria have been met; typically this means that changes in energy and positions compared to the previous step, as

well as forces on all atoms, are all below some chosen thresholds. This process is summarised below:

1. Initialise trial geometry and wavefunction
2. Calculate electronic energy and forces
3. Use some gradient-following scheme to move atoms in the direction of net forces
4. Repeat steps 2-3 until chosen convergence criteria have been met

For periodic systems, the stress tensor can also be calculated analytically to optimise the dimensions and shape of the unit cell.[96] In this way, the system size and shape can be allowed to change, which may be necessary in a number of cases, such as: the trial geometry is poor; the simulation cell is subjected to an external stimulus; or modifications have been applied to the structure, like the introduction of interstitial atoms or vacancies.

2.3 Molecular Dynamics

Often, a phenomenon of interest may be dynamic in nature and require atoms to move in a natural manner. This process of allowing atomic positions to evolve with time is referred to as molecular dynamics (MD). At its core, this method makes use of Newton's laws of motion and (numerical) propagation algorithms to update coordinates and velocities given their current values. The key equations are given as 53, 54, and 55:

$$\mathbf{x}(t+\delta t) = \mathbf{x}(t) + \mathbf{v}(t)\delta t + \frac{1}{2}\mathbf{a}(t)\delta t^2 \tag{53}$$

$$\mathbf{v}(t+\delta t) = \mathbf{v}(t) + \frac{1}{2}\Big[\mathbf{a}(t) + \mathbf{a}(t+\delta t)\Big]\delta t \tag{54}$$

$$\mathbf{a}(t) = \frac{1}{m}\mathbf{F}(\mathbf{x}(t)) \tag{55}$$

$\mathbf{x}$, $\mathbf{v}$, $\mathbf{a}$, and $\mathbf{F}$ denote position, velocity, acceleration, and force vectors

for a given particle of mass m using a short time step of duration δt. Varied forms can be used for equations 53 and 54; the ones shown here are the position and velocity Verlet algorithms and are commonly used in MD approaches. Updating the system over the course of a single step can be summarised by:

1. Calculate forces for the current positions; determine accelerations
2. Update positions using current velocities and accelerations
3. Calculate forces for updated positions; determine new accelerations (which can be used for step 1 of following cycle)
4. Calculate velocities for new positions

The forces invoked in **Equation 55** are usually calculated using one of two approaches. The first is to use a force-field, and the second is to use a quantum mechanical method such as HF or DFT to calculate the forces, in which case this is referred to as *ab initio* molecular dynamics (AIMD). The BO approximation is applied such that at each step, the electron density is assumed to instantaneously adjust to the new configuration of atoms and is determined with the SCF procedure, following which forces are calculated at the quantum mechanical level.

Given the different recipes that can be followed to calculate forces, the chemical accuracy of an MD run depends on both the overall method as well as the quality of the force field, functional and/or basis set (whichever are applicable). Care must also be taken with respect to the choice of time step δt; events which take place on a shorter time scale than this will not be observed. For phenomena which take place over long times and distances, such as diffusion, picosecond intervals may be used. For molecular vibrations, which represent terahertz frequencies, this corresponds to a step size of a femtosecond or less. In all cases, the step size should be at least half as small as the smallest frequency

exhibited by the system (as prescribed by Nyquist's theorem) in order for these to be observed. Note also that equations 53 and 54 are not analytical and neglect terms in $(\delta t)^3$ and smaller, so larger timesteps may also contribute to numerical error.

2.3.1 Ensembles and Thermostats

Atoms may be straightforwardly made to move using propagation algorithms, but the objective is ultimately to make coordinates evolve in a manner that is representative of a real system of interest. This may be done by using different statistical mechanical ensembles to determine what the distribution of particle velocities and other system properties should follow in order to mimic a given set of experimental conditions. Some commonly used ensembles are listed in **Table 2**:

Table 2: Thermodynamic ensembles used in molecular dynamics

Ensemble	Shorthand	Fixed system quantities
Microcanonical	NVE	Number of particles, volume, total energy
Canonical	NVT	Number of particles, volume, temperature
Isothermal-isobaric	NPT	Number of particles, pressure, temperature
Grand canonical	μVT	Chemical potential, volume, temperature

Coupling with an external reservoir is necessary in order to maintain certain variables fixed; this is achieved through the use of thermostats and barostats. These rescale particles velocities and the cell volume to maintain the chosen conditions while allowing small fluctuations.

The microcanonical ensemble is often too limited to reflect experimental conditions (isolated system) and will not be discussed here, though more information can be found in ref. [97]. For the remaining ensembles, the system is allowed to exchange some combination of energy and

molecules with some imaginary external reservoir (held at fixed T and μ) to which it is coupled. Particle velocities are initialised according to a Maxwell-Boltzmann distribution at the chosen temperature. As a simulation progresses, the instantaneous temperature (defined by average kinetic energy via the equipartition theorem) will diverge from the chosen value due to exchanges between kinetic and potential forms. This means velocity rescaling is required in order to maintain the average temperature constant. A thermostat needs to be used in order for this to be done in a time-reversible, stochastic, and physically sensible manner. These typically involve fictitious degrees of freedom and some parameter that describes the strength of coupling between the system and the heat reservoir and make their way into the equations of motion. Details on specific thermostats can be found in ref. [98].

Similarly, if a flexible simulation cell is needed, a barostat must be defined to allow the volume to change while regulating pressure. Broadly, two approaches can be used. The first involves re-scaling volume and interparticle distances at a regular interval based on the difference between instantaneous and target pressures.[99] This leads to a barostat which is independent of the thermostat. The second introduces cell size as a variable in the equations of motion by using volume-scaled coordinates.[100,101] In this case, the barostat and thermostat are interdependent. Coupling strength between the system and reservoir is again controlled by a parameter.

Where AIMD has been used in this study, it has been within the canonical ensemble. This is a sensible choice if the pressure dependence of the system is not of interest or if fluctuations in unit cell shape and size are not deemed to be important. This is the case here, where MD has been used only to generate a set of trial coordinates for subsequent calculations in unit cells that share the same fixed dimensions.

2.4 Vibrational Analysis

Another useful description of how atoms move in molecular and periodic systems is in terms of vibrations. At finite temperature, each atom has some kinetic energy, such that it oscillates about its equilibrium position. The direction and amplitude of this motion are influenced by neighbouring atoms either directly through bonds or indirectly via dispersion or electrostatic interactions. Movement is therefore often collective, involving multiple particles oscillating together at a given frequency in what is known as a vibrational normal mode. These can be determined from the equilibrium geometry of a system using Lagrangian mechanics if the nuclei are treated as classical particles. The kinetic energy can be expressed in matrix form using a diagonal mass matrix $\mathbf{M}$:

$$T = \sum_i \frac{1}{2} m_i v_i^2 = \sum_i \frac{1}{2} m_i \dot{x}_i^2 = T = \frac{1}{2} \dot{\mathbf{x}} \cdot \mathbf{M} \cdot \dot{\mathbf{x}} \tag{56}$$

where x_i and m_i are displacement from equilibrium of some atom along some coordinate and its associated mass. The potential energy due to the small displacement of a single atom about its point of equilibrium may be expanded as a Taylor series:

$$U = U_0 + x_i \frac{\partial U}{\partial x_i}\Big|_{x_i=0} + \frac{x_i^2}{2} \frac{\partial^2 U}{\partial x_i^2}\Big|_{x_i=0} + ... \tag{57}$$

The second term vanishes by construction since the force is zero at equilibrium, and the cubic and higher-order terms can be discarded if the displacement is small (in doing so, the harmonic approximation is being applied). If the displacements of all atoms are considered then:

$$U = U_0 + \sum_i \frac{x_i^2}{2} \frac{\partial^2 U}{\partial x_i^2}\Big|_{x_i=0} = U_0 + \frac{1}{2} \mathbf{x} \cdot \mathbf{H} \cdot \mathbf{x} \tag{58}$$

where $\mathbf{H}$ is the Hessian matrix of the potential energy. T and U now define the Lagrangian and the Euler-Lagrange equation can be used to

determine the equations of motion:

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\frac{\partial \mathcal{L}}{\partial \dot{\mathbf{x}}}\right) = \frac{\partial \mathcal{L}}{\partial \mathbf{x}} \tag{59}$$

With $\mathcal{L} = T - U$ this yields:

$$\mathbf{M}\cdot \ddot{\mathbf{x}} = -\mathbf{H}\cdot \mathbf{x} \tag{60}$$

Using a trial solution $\mathbf{x} = \mathbf{X}e^{i\omega t}$ and re-arranging terms leads to a matrix equation of the form:

$$\Big(\mathbf{H} - \omega^2\mathbf{M}\Big)\mathbf{X} = 0 \tag{61}$$

Equation 61 can be solved to find a set of eigenvalues (ω) and eigenvectors corresponding to the vibrational normal modes and their frequencies. These represent a natural coordinate for the system which can be used to describe any motion which respects the harmonic approximation (sufficiently small displacements about equilibrium). Within this regime, the eigenvectors and frequencies also correspond to the absorption bands which are observed in IR spectroscopy. This analysis is therefore also convenient as a means by which a given approach for calculating forces can be compared against and validated by experiment.

No assumptions have been made so far about the nature of the system, other than that nuclei are treated as classical particles. This means the same formalism is valid with both force fields and *ab initio* approaches (making use of the BO approximation for the latter). The mass matrix is defined by the atom types in the system, so only the Hessian needs to be evaluated, usually through numerical differentiation. As with other dynamical methods discussed in this section, chemical accuracy is determined by the quality of the underlying technique used to calculate energy. Care also needs to be taken with regards to the harmonic approximation

made in **Equation 58**, which can break down near phase transitions or in inherently anharmonic materials such as perovskites.[102,103] Generally, frequencies evaluated via DFT will also be slightly larger than those found via experiment, necessitating that they be scaled down by a functional and basis set dependent constant factor.[104,105] This can be determined by comparing clearly identifiable characteristic frequencies between the simulation and experiment, but care needs to be taken that anharmonicity does not vary too much from mode to mode, otherwise uniform scaling will not be appropriate.[102,103] It should also be noted that solving **Equation 61** in itself does not yield all the information necessary to construct an infrared spectrum, for which proper consideration of symmetry and electrostatics are also required. Finally, this approach is not the only one through which a system's normal modes may be determined. MD can be used to directly sample the energy landscape for this purpose, naturally taking into account any anharmonicity in the process.[106]

In this work, normal modes have been identified through vibrational analysis. This approach was chosen over MD given the latter's high computational cost and previous reports of good agreement between experimental spectra and those generated within the harmonic approximation for UiO-66.[107–109] In addition, it is useful to have access to the Hessian matrix for post-processing with software packages such as TAMkin.[110]

2.5 Transition States

As described in **Section 2.2.5**, minima in the potential energy landscape can readily be found if chemically accurate forces and a suitable initial guess for the system geometry are available. In the context of atomic re-arrangements or chemical reactions, these correspond to systems in the reactant or product states. Often, the minimum energy pathway (MEP) through the potential energy surface (PES) that leads from one stable configuration to another is of equal interest. With knowledge of

the MEP and the transition state which crests it, it is possible to evaluate the energy barriers of the given chemical process. This can also offer information about the associated reaction kinetics via transition state theory.[111] An illustration of a model PES and the points of interest along the MEP is given in **Figure 9**.

In order to identify the transition state and other points along an MEP, it is necessary to move beyond the gradient descent techniques used to identify minima in the PES. The transition state is a saddle point on the PES, meaning it is a maximum along the reaction coordinate, and a minimum along all transverse coordinates. Therefore, vibrational analysis of such a system would yield exactly one negative eigenvalue (the rest being positive), for which the associated eigenvector corresponds to the reaction coordinate. This property can be used to confirm when a transition state has been identified. Two approaches will be described here for identifying transition states along MEPs, the elastic band and dimer methods. Both of these have been used in **Chapter 4** to identify the MEP in the isomerisation of citronellal to isopulegol, an often-used experimental yardstick for the Lewis catalytic activity of UiO-66.

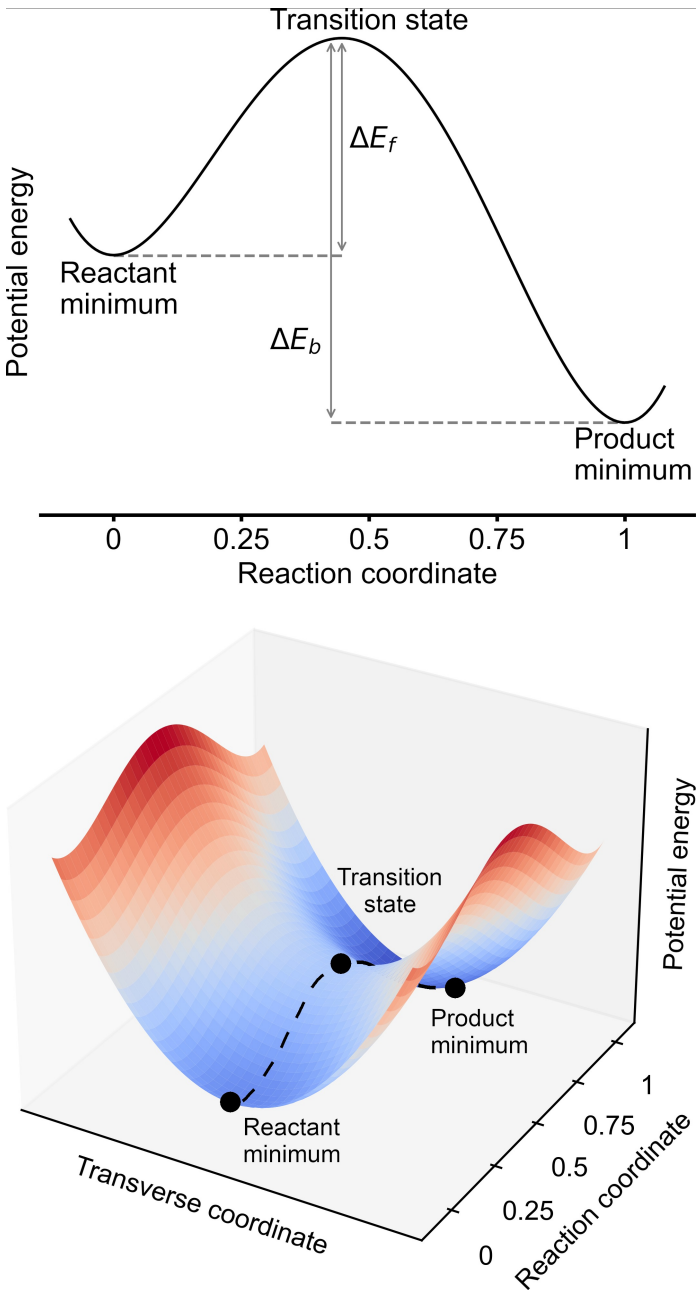


Figure 9: Top panel: an MEP, projected along the reaction coordinate, between two minima. The reaction coordinate is a collective variable describing the change in the atomic configuration when moving from one state to the other. If the extrema along the path are known, the forward and backward energy barriers ΔE_f and ΔE_b can be determined. Bottom panel: the same MEP, marked by a dotted line, shown with an additional coordinate transverse to the path. Along the MEP, the energy is minimised with respect to all variables which are not part of the reaction coordinate.

2.5.1 Elastic Band Method

To generally describe elastic band methods, we first follow the work of Elber and Karplus.[112] The problem of finding a minimum energy pathway between two minima is similar to that of minimising the line integral of a function $\boldsymbol{G}$ between two points:

$$S(\boldsymbol{r}_A, \boldsymbol{r}_B) = \frac{1}{L} \int_{\boldsymbol{r}_A}^{\boldsymbol{r}_B} \boldsymbol{G}(\boldsymbol{r}) \cdot \mathrm{d}\boldsymbol{l}(\boldsymbol{r}) \tag{62}$$

A and B are the fixed endpoints on a path of length L, which exists within a coordinate space $\boldsymbol{r}$ and is defined by line elements $\mathrm{d}\boldsymbol{l}(\boldsymbol{r})$. This line integral can be discretised using a fixed number of line segments:

$$S(\boldsymbol{r}_A, \boldsymbol{r}_B) \approx \frac{1}{L} \sum_{i=1}^{M} \boldsymbol{G}(\boldsymbol{r}_i) \cdot \Delta\boldsymbol{l}_i \tag{63}$$

M is the number of line segments used, $\boldsymbol{r}_i$ are the coordinates at the end of segment i, which are themselves described by:

$$\Delta\boldsymbol{l}_i = ((\boldsymbol{r}_i - \boldsymbol{r}_{i-1})^2)^{\frac{1}{2}} \boldsymbol{u}_{i,i-1} \tag{64}$$

where $u_{i,i-1}$ are unit vectors along the direction of the line segments. In order for this discretisation to remain valid, the segment lengths should remain similar. Therefore, an additional term is introduced to **Equation 63** to enforce this:

$$S'(\boldsymbol{r}_A, \boldsymbol{r}_B) = S(\boldsymbol{r}_A, \boldsymbol{r}_B) + \sum_{i}^{M} \lambda(\Delta l_i - \Delta l)^2 \tag{65}$$

Δl is the root mean square segment length along the path, and λ is a parameter which controls how much Δl_i are allowed to deviate. In order to now apply this formalism to the case of an MEP along a potential energy surface, we simply choose for the target function of our line integral to be the potential energy:

$$\boldsymbol{G}(\boldsymbol{r}) = V(\boldsymbol{r})\boldsymbol{u}_{i,i-1} \tag{66}$$

The segments meanwhile can be generated by interpolating the structure between points A and B, or via some prior calculation, such that there is a string of 'images' along the path. These correspond to coordinates $\boldsymbol{r}_i$

and define the segments. Each image is held close to its neighbours by the final term in **Equation 65**, which has the form of a Hookian spring with constant λ. During the minimisation of $S'(\boldsymbol{r}_A, \boldsymbol{r}_B)$ these introduce additional forces on the images which prevent them from free-falling directly towards minima A and B. Hence, this framework is referred to as the plain elastic band method. Note that additional terms may be required in **Equation 65** if Cartesian coordinates are used in order to penalise arbitrary translations and rotations of the whole system.

Unfortunately, this plain elastic band (PEB) method has two major shortcomings that mean a given path may not converge to an MEP.[113] Firstly, off-path components of the spring forces which do not cancel out may push images uphill, leading to 'corner-cutting'. Secondly, components of the true force on an image which lie along the path will pull images downwards, increasing spacing between them near the transition state and thereby reducing resolution. Mills and Jónsson have addressed these issues by projecting out the parts of the spring and true forces which are respectively perpendicular and parallel to the path.[114–116] The force on a given image in a PEB is given by the gradient of the potential energy surface at that point and the sum of forces from the attached springs:

$$\boldsymbol{F}_i = -\nabla V(\boldsymbol{r}_i) + \lambda(\boldsymbol{r}_{i+1} - \boldsymbol{r}_i) - \lambda(\boldsymbol{r}_i - \boldsymbol{r}_{i-1}) \tag{67}$$

$$\boldsymbol{F}_i = -\nabla V(\boldsymbol{r}_i) + \boldsymbol{F}_i^s \tag{68}$$

We can now project out the problematic force components using a unit vector parallel to the path, denoted by $\boldsymbol{\tau}_i^{\parallel}$:

$$\boldsymbol{F}_i^{\mathrm{NEB}} = -[\nabla V(\boldsymbol{r}_i) - \nabla(V(\boldsymbol{r}_i) \cdot \boldsymbol{\tau}_i^{\parallel})\boldsymbol{\tau}_i^{\parallel}] + (\boldsymbol{F}_i^s \cdot \boldsymbol{\tau}_i^{\parallel})\boldsymbol{\tau}_i^{\parallel} \tag{69}$$

The elastic band method and the terms in **Equation 69** are shown schematically in **Figure 10**.

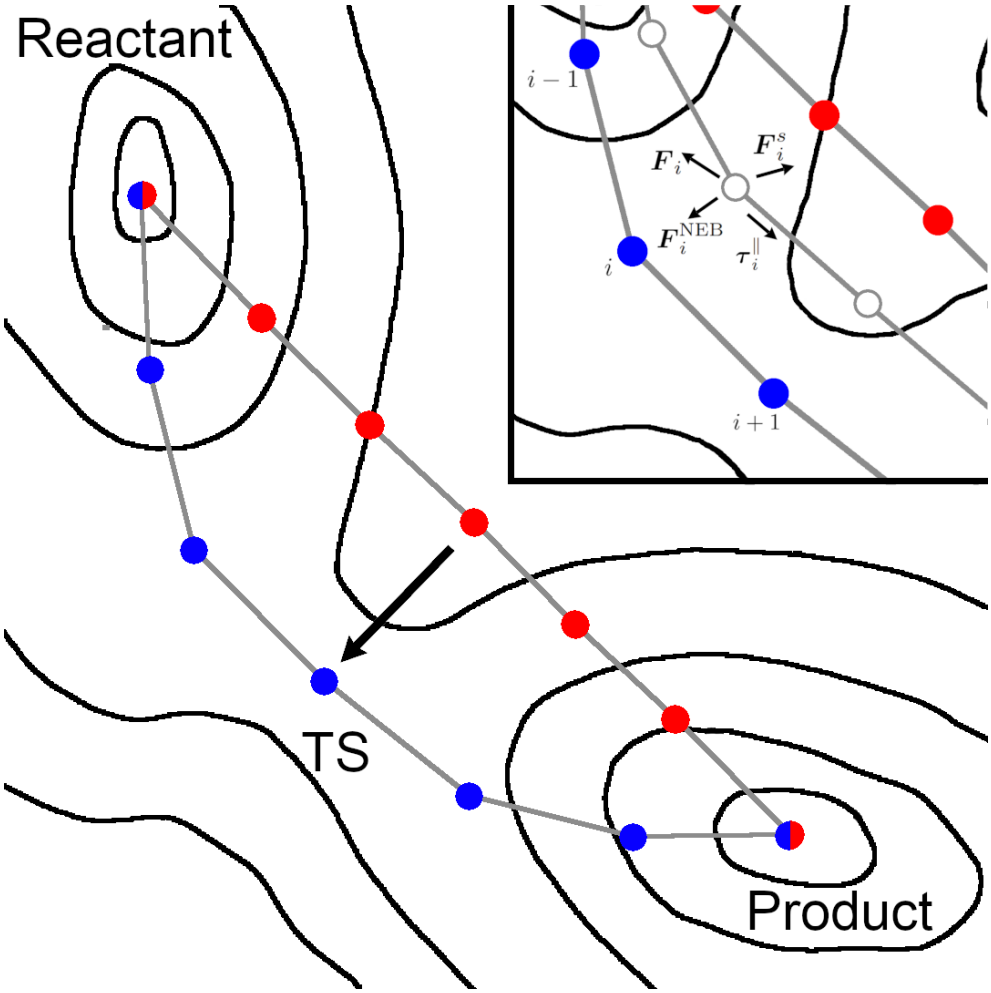


Figure 10: Schematic illustration of the elastic band method. The red points correspond to images at the start of the simulation, which might have been found for example through interpolation. The blue points are the images once the MEP has been found. The inset additionally shows an intermediate set of images and illustrates the vectors and forces involved in the calculation.

This process of projecting out certain forces is referred to as 'nudging', hence this improved method is referred to as the nudged elastic band (NEB). However, certain types of energy landscapes can still cause the band to buckle as a result of forces parallel and perpendicular to the path respectively being large and small. In such cases, the definition of a unit tangent vector becomes ambiguous, so some fraction of the perpendicular spring force should be retained by using a switching function that depends on the angle between two adjacent band segments.

Two further modifications have been made to the NEB technique which are commonly employed. The first is an improved definition of the tangent (improved-tangent, or IT-NEB), which is normally defined as the normalised vector between an image's neighbours.[117] Using the normalised bisector of vectors between an image and both its neighbours gives a more robust tangent when the energy landscape is highly curved. The second is that once the highest-energy image has been identified, the true force parallel to the path and the spring forces acting on it are respectively inverted and switched off. This ensures that the maximal image will climb in energy along the path (climbing-image, or CI-NEB), and slide down perpendicular to it, thus rigorously converging to a transition state.[118] In practice this requires first running IT-NEB for some iterations such that this maximum is unambiguous.

2.5.2 The Dimer Method

An alternative to elastic band techniques is the dimer method of Henkelman and Jónsson.[119] Within this framework, one begins with a single configuration on the energy landscape, usually near a minimum, unlike the start and end states required for NEB. From this initial set of coordinates, denoted by $\boldsymbol{r}$, images A and B are generated at configurations $\boldsymbol{r}_A$ and $\boldsymbol{r}_B$ by moving a distance $\pm\Delta r$ along a random unit vector $\hat{\boldsymbol{n}}$. This construction, shown schematically in **Figure 11**, is referred to as the dimer. In order to find a transition state, this object is driven uphill through a series of rotation and translation steps.

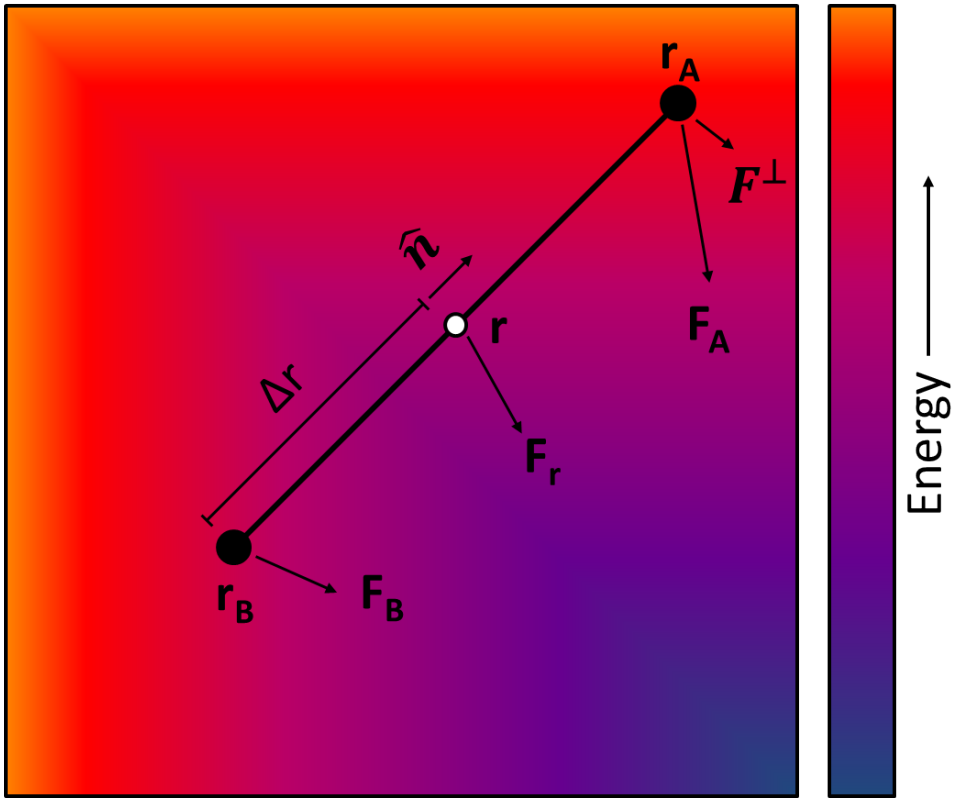


Figure 11: Schematic of the dimer method construction overlaid on top of a dummy energy landscape. The turning force will generally align the dimer parallel to the direction of least curvature. Subsequent translation steps will move the midpoint closer towards a saddle point. The images are depicted by filled black circles, while the midpoint is a white circle.

The first type of step, rotation, is used to find the direction of least curvature, so as to subsequently make translation steps more efficient. We begin by evaluating the forces on both images, $\boldsymbol{F}_A$ and $\boldsymbol{F}_B$. From these, we can then project out components which are parallel to the dimer so as to find the turning force $\boldsymbol{F}^{\perp}$:

$$\boldsymbol{F}_i^{\perp} = \boldsymbol{F}_i - (\boldsymbol{F}_i \cdot \hat{\boldsymbol{n}})\hat{\boldsymbol{n}} \tag{70}$$

$$\boldsymbol{F}^{\perp} = \boldsymbol{F}_A^{\perp} - \boldsymbol{F}_B^{\perp} \tag{71}$$

If we now define a unit vector $\hat{\boldsymbol{\Theta}}$ perpendicular to $\hat{\boldsymbol{n}}$, then now we have an orthonormal basis which describes the plane of rotation. We can use this to evaluate the new position of image A if the dimer were to be

rotated by some angle $\mathrm{d}\theta$:

$$\boldsymbol{r}_A^* = \boldsymbol{r}_A + [\hat{\boldsymbol{n}}\cos(\mathrm{d}\theta) + \hat{\boldsymbol{\Theta}}\sin(\mathrm{d}\theta)] \tag{72}$$

In order to complete the rotation, we update the orientation vector and the position of image B:

$$\hat{\boldsymbol{n}}^* = \frac{\boldsymbol{r}_A^* - \boldsymbol{r}}{|\boldsymbol{r}_A^* - \boldsymbol{r}|} \tag{73}$$

$$\boldsymbol{r}_B = \boldsymbol{r} - \hat{\boldsymbol{n}}^* \Delta r \tag{74}$$

It is now necessary to rotate the dimer until the turning force vanishes. Practically, this is done by first defining a scalar turning force:

$$F = \frac{\boldsymbol{F}^{\perp} \cdot \hat{\boldsymbol{\Theta}}}{\Delta r} \tag{75}$$

Division by the dimer radius ensures that this scalar force is independent of the chosen separation between images. Any suitable minimisation algorithm can then be used to bring F to zero. Once the optimal orientation has been found, then the dimer can be translated. Defining the force on the dimer as the average of the dimer forces:

$$\boldsymbol{F}_r = \frac{\boldsymbol{F}_A + \boldsymbol{F}_B}{2} \tag{76}$$

The component of this force parallel to the dimer can then be inverted:

$$\boldsymbol{F}^{\dagger} = \boldsymbol{F}_r - 2(\boldsymbol{F}_r \cdot \hat{\boldsymbol{n}})\hat{\boldsymbol{n}} \tag{77}$$

If the dimer now follows this modified force, it will move uphill along a potential reaction coordinate, and downhill along others. This remains true as long as the orientation is updated at each step so as to be along the direction of least curvature. The procedure for reaching a transition state can be summarised:

1. Choose initial configuration r; initialise the dimer
2. Iteratively rotate the dimer until the turning force vanishes
3. Single translation step in direction of $\boldsymbol{F}^{\dagger}$

4. Repeat steps 2-3 until convergence as prescribed by chosen criteria on atomic forces and displacements

It is important to note that, unlike with an elastic band, the end state of this method is not known *a priori*. Given that there are in general multiple MEPs leading out of a minimum, the transition state which the dimer will move towards will depend on the initial configuration. However, if some information is already known about a given MEP of interest (for example through prior calculations), then a simulation can be steered towards a chosen transition state by setting the initial coordinates to be close by. This means that the dimer method can notably be used to refine a nearly-converged geometry at low cost, for example from a NEB calculation. In this work, vibrational analysis revealed for a number of maxima along NEB-derived MEPs more than one negative eigenvalue. This corresponds to atomic configurations close to but not exactly at transition states. In such cases, these geometries were used as initial guesses for the dimer method with the aim of refining the structures towards true transition states.

2.6 Software details

The primary tool used in this thesis has been the CP2K package.[120] This software uses two basis sets to efficiently evaluate different components of the DFT energy. A primary Gaussian basis set is used to represent the Kohn-Sham orbitals, via which the kinetic energy and external potential are calculated. A secondary plane-wave basis set is used to represent the electron density, which makes the numerical integration involved in the exchange-correlation energy computationally cheaper. This representation of the density can also easily be Fourier-transformed to compute the long-range part of the Hartree energy in reciprocal space. The use of such a dual basis set requires that the plane-wave coefficients be calculated from the Gaussian ones at each SCF step, a process which is efficiently implemented in CP2K.[121] These features allow this software to scale favourably with system size, which makes it ideal for studying a

material like UiO-66, particularly when defects and supercells are being examined such that a large number of atoms need to be modelled.

Except where mentioned, all CP2K calculations in this work have been carried out with general parameters which will be described here. For most work, the PBE functional has been used together with its semi-empirical D3 corrections.[26,70] The Kohn-Sham wavefunction was represented using the double–ζ quality MOLOPT Gaussian basis sets (except for Zr, for which the short-range MOLOPT-SR variant was used)[122] while plane waves up to a cutoff of 1200 Ry were used for the electron density. Goedecker–Teter–Hutter (GTH) pseudopotentials were used to model core electrons for all elements.[123] K-point sampling was restricted to the Γ-point as this has previously been reported to be sufficient for the system of interest.[48] The SCF procedure was controlled by an energy tolerance of 10^{-8} Ha between consecutive steps. During structural optimisations, atomic positions as well as cell parameters were allowed to change independently from each other. Convergence criteria of 4.5×10^{-4} Bohr, 4.5×10^{-4} Ha Bohr^{-1}, 4.5×10^{-4} Bohr, and 1.0×10^{-4} Ha Bohr^{-1} were applied for root-mean-square (RMS) displacement, RMS force, maximum displacement, and maximum force during such calculations. When modelling neutral molecules, this was done inside empty cells of fixed size 20×20×20 Å, corresponding to a minimum vacuum gap of ∼15 Å, thereby minimising interactions between periodic images. During vibrational analysis, elements of the Hessian matrix were evaluated via numerical differentiation using displacements of 0.001 bohr and a stricter SCF tolerance of 10^{-10} Ha to reduce noise.

The above set of parameters for the Hamiltonian, basis set, and convergence parameters were found to yield cell parameters for the pristine conventional unit cell of UiO-66 within 1% of the experimental ones reported by Cavka *et al.*[87] The plane-wave cutoff is high for the unit cell of UiO-66, which has dimensions ∼21 Å, but this choice means no change of

basis set is required when scaling up to supercells up to 2×2×2. Results from single-cell calculations can therefore be directly compared to those involving supercells.

A smaller number of calculations has also been carried out using ORCA.[124] This software package offers HF, post-HF, DFT, and semi-empirical methods for use on aperiodic systems. It is therefore suitable for modelling charged molecules, for which the long-range nature of electrostatic forces would otherwise cause problems in periodic programs such as CP2K due to interactions between periodic images. For such calculations, the PBE functional has been used together with D3 semi-empirical corrections and Becke-Johnson damping functions. Primary def2 triple–ζ and auxiliary def2/J basis sets (both Gaussian in nature) from Weigend and Ahlrichs were used.[125,126] A convergence criteria of 10^{-8} Ha was applied for the SCF procedure and 4.5×10^{-4} Bohr, 4.5×10^{-4} Ha Bohr^{-1}, 4.5×10^{-4} Bohr, and 1.0×10^{-4} Ha Bohr^{-1} for root-mean-square (RMS) displacement, RMS force, maximum displacement, and maximum force during geometry optimisations.

References

(1) S. Blinder, in *Mathematical Physics in Theoretical Chemistry*, ed. S. Blinder and J. House, Elsevier, Amsterdam, 2019, ch. 1, p. 21.

(2) A. Leach, in *Molecular Modelling: Principles and Applications*, Pearson Education Limited, Harlow, England, 2nd, 2001, ch. 2.4, pp. 41–51.

(3) A. Szabo and N. Ostlund, in *Modern quantum chemistry : introduction to advanced electronic structure theory*, Dover Publications, Mineola, New York, Dover Edit, 1996, ch. 3.1, pp. 212–216.

(4) S. Blinder, in *Mathematical Physics in Theoretical Chemistry*, ed. S. Blinder and J. House, Elsevier, Amsterdam, 2019, ch. 1, pp. 11–12.

(5) S. Blinder, in *Mathematical Physics in Theoretical Chemistry*, ed. S. Blinder and J. House, Elsevier, Amsterdam, 2019, ch. 1, pp. 16–21.

(6) C. F. Matta, *J. Comput. Chem.*, 2010, **31**, 1297–1311.

(7) Y. B. Malykhanov and M. V. Gorshunov, *J. Appl. Spectrosc.*, 2013, **80**, 631–636.

(8) J. F. Harrison and D. B. Lawson, *J. Chem. Educ.*, 2005, **82**, 1205–1209.

(9) J. G. Brandenburg and S. Grimme, *Top. Curr. Chem.*, 2014, **345**, 1–24.

(10) M. Jain, J. R. Chelikowsky and S. G. Louie, *Phys. Rev. Lett.*, 2011, **107**, 1–5.

(11) K. Sohlberg and M. E. Foster, *RSC Adv.*, 2020, **10**, 36887–36896.

(12) C. Møller and M. S. Plesset, *Phys. Rev.*, 1934, **46**, 618–622.

(13) P. G. Szalay, T. Müller, G. Gidofalvi, H. Lischka and R. Shepard, *Chem. Rev.*, 2012, **112**, 108–181.

(14) P. Hohenberg and W. Kohn, *Phys. Rev.*, 1964, **136**, B864–B871.

(15) W. Kohn and L. J. Sham, *Phys. Rev.*, 1965, **140**, 1133.

(16) E. H. Lieb and B. Simon, *Adv. Math. (N. Y).*, 1977, **23**, 22–116.

(17) R. O. Jones and O. Gunnarsson, *Rev. Mod. Phys.*, 1989, **61**, 689–746.

(18) P. Ziesche, S. Kurth and J. P. Perdew, *Comput. Mater. Sci.*, 1998, **11**, 122–127.

(19) J. C. Grossman, L. Mitas and K. Raghavachari, *Phys. Rev. Lett.*, 1995, **75**, 3870–3873.

(20) A. Zupan, P. Blaha and K. Schwarz, *Phys. Rev. B - Condens. Matter Mater. Phys.*, 1998, **58**, 11266–11272.

(21) J. P. Perdew and K. Schmidt, *AIP Conf. Proc.*, 2001, **577**, 1–20.

(22) A. D. Becke, *J. Chem. Phys.*, 1986, **84**, 4524–4529.

(23) J. P. Perdew and W. Yue, *Phys. Rev. B*, 1986, **33**, 8800–8802.

(24) A. D. Becke, *Phys. Rev. A*, 1988, **38**, 3098–3100.

(25) C. Lee, W. Yang and R. G. Parr, *Phys. Rev. B*, 1988, **37**, 785–789.

(26) J. P. Perdew, K. Burke and M. Ernzerhof, *Phys. Rev. Lett.*, 1996, **77**, 3865–3868.

(27) L. Goerigk and N. Mehta, *Aust. J. Chem.*, 2019, **72**, 563–573.

(28) N. Mardirossian and M. Head-Gordon, *Mol. Phys.*, 2017, **115**, 2315–2372.

(29) J. P. Perdew, S. Kurth, A. Zupan and P. Blaha, *Phys. Rev. Lett.*, 1999, **82**, 2544–2547.

(30) J. Tao, J. P. Perdew, V. N. Staroverov and G. E. Scuseria, *Phys. Rev. Lett.*, 2003, **91**, 3–6.

(31) Y. Zhao and D. G. Truhlar, *J. Chem. Phys.*, 2006, **125**, 194101.

(32) J. Sun, A. Ruzsinszky and J. Perdew, *Phys. Rev. Lett.*, 2015, **115**, 1–6.

(33) S. Lehtola and M. A. Marques, *J. Chem. Phys.*, 2022, **157**, 174114.

(34) A. D. Becke, *J. Chem. Phys.*, 1993, **98**, 5648–5652.

(35) J. P. Perdew, M. Ernzerhof and K. Burke, *J. Chem. Phys.*, 1996, **105**, 9982–9985.

(36) J. Heyd, G. E. Scuseria and M. Ernzerhof, *J. Chem. Phys.*, 2003, **118**, 8207–8215.

(37) G. I. Csonka, J. P. Perdew and A. Ruzsinszky, *J. Chem. Theory Comput.*, 2010, **6**, 3688–3703.

(38) E. N. Brothers, A. F. Izmaylov, J. O. Normand, V. Barone and G. E. Scuseria, *J. Chem. Phys.*, 2008, **129**, 011102.

(39) K. N. Kudin, G. E. Scuseria and R. L. Martin, *Phys. Rev. Lett.*, 2002, **89**, 21–24.

(40) A. D. Becke, *J. Chem. Phys.*, 2014, **140**, 18A301.

(41) D. Rappoport, N. R. M. Crawford, F. Furche and K. Burke, in *Encyclopedia of Inorganic Chemistry*, John Wiley & Sons, 2009.

(42) D. Nazarian, P. Ganesh and D. S. Sholl, *J. Mater. Chem. A*, 2015, **3**, 22432–22440.

(43) L. Liu, Z. Chen, J. Wang, D. Zhang, Y. Zhu, S. Ling, K. W. Huang, Y. Belmabkhout, K. Adil, Y. Zhang, B. Slater, M. Eddaoudi and Y. Han, *Nat. Chem.*, 2019, **11**, 622–628.

(44) M. Vandichel, J. Hajek, F. Vermoortele, M. Waroquier, D. E. De Vos and V. Van Speybroeck, *CrystEngComm*, 2015, **17**, 395–406.

(45) X. Feng, J. Hajek, H. S. Jena, G. Wang, S. K. Veerapandian, R. Morent, N. De Geyter, K. Leyssens, A. E. Hoffman, V. Meynen, C. Marquez, D. E. De Vos, V. Van Speybroeck, K. Leus and P. Van Der Voort, *J. Am. Chem. Soc.*, 2020, **142**, 3174–3183.

(46) A. De Vos, K. Hendrickx, P. Van Der Voort, V. Van Speybroeck and K. Lejaeghere, *Chem. Mater.*, 2017, **29**, 3006–3019.

(47) C. Caratelli, J. Hajek, E. J. Meijer, M. Waroquier and V. Van Speybroeck, *Chem. Eur. J.*, 2019, **25**, 15315–15325.

(48) J. Hajek, M. Vandichel, B. Van De Voorde, B. Bueken, D. De Vos, M. Waroquier and V. Van Speybroeck, *J. Catal.*, 2015, **331**, 1–12.

(49) J. Marreiros, C. Caratelli, J. Hajek, A. Krajnc, G. Fleury, B. Bueken, D. E. De Vos, G. Mali, M. B. Roeffaers, V. Van Speybroeck and R. Ameloot, *Chem. Mater.*, 2019, **31**, 1359–1369.

(50) M. Vandichel, J. Hajek, A. Ghysels, A. De Vos, M. Waroquier and V. Van Speybroeck, *CrystEngComm*, 2016, **18**, 7056–7069.

(51) C. Caratelli, J. Hajek, F. G. Cirujano, M. Waroquier, F. X. Llabrés i Xamena and V. Van Speybroeck, *J. Catal.*, 2017, **352**, 401–414.

(52) F. Vermoortele, B. Bueken, G. Le Bars, B. Van De Voorde, M. Vandichel, K. Houthoofd, A. Vimont, M. Daturi, M. Waroquier, V. Van Speybroeck, C. Kirschhock and D. E. De Vos, *J. Am. Chem. Soc.*, 2013, **135**, 11465–11468.

(53) S. L. Mayo, B. D. Olafson and W. A. Goddard, *J. Phys. Chem.*, 1990, **94**, 8897–8909.

(54) M. A. Addicoat, N. Vankova, I. F. Akter and T. Heine, *J. Chem. Theory Comput.*, 2014, **10**, 880–891.

(55) R. B. Getman, Y. S. Bae, C. E. Wilmer and R. Q. Snurr, *Chem. Rev.*, 2012, **112**, 703–723.

(56) H. Daglar and S. Keskin, *Coord. Chem. Rev.*, 2020, **422**, 213470.

(57) A. Leach, in *Molecular Modelling: Principles and Applications*, Pearson Education Limited, Harlow, England, 2nd, 2001, ch. 3.2, pp. 108–110.

(58) U. Von Barth and L. Hedin, *Journal of Physics C: Solid State Physics*, 1972, **5**, 1629–1642.

(59) M. Filatov and S. Shaik, *Chem. Phys. Lett.*, 1998, **288**, 689–697.

(60) J. P. Perdew and A. Zunger, *Phys. Rev. B*, 1981, **23**, 5048–5079.

(61) A. Rettig, D. Hait, L. W. Bertels and M. Head-Gordon, *J. Chem. Theory Comput.*, 2020, **16**, 7473–7489.

(62) H. B. Schlegel, *J. Phys. Chem.*, 1988, **92**, 3075–3078.

(63) C. J. Cramer, F. J. Dulles, D. J. Giesen and J. Almlöf, *Chem. Phys. Lett.*, 1995, **245**, 165–170.

(64) K. Tada, S. Yamanaka, T. Kawakami, Y. Kitagawa, M. Okumura, K. Yamaguchi and S. Tanaka, *Chem. Phys. Lett.*, 2021, **765**, 138291.

(65) M. Filatov and S. Shaik, *J. Chem. Phys.*, 1999, **110**, 116–125.

(66) J. Klimeš and A. Michaelides, *J. Chem. Phys.*, 2012, **137**, 120901.

(67) F. London, *Trans. Faraday Soc.*, 1937, **33**, 8–26.

(68) S. Grimme, *J. Comput. Chem.*, 2004, **25**, 1463–1473.

(69) S. Grimme, *J. Comput. Chem.*, 2006, **27**, 1787–1799.

(70) S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.

(71) E. Caldeweyher, C. Bannwarth and S. Grimme, *J. Chem. Phys.*, 2017, **147**, 034112.

(72) L. Goerigk, H. Kruse and S. Grimme, *ChemPhysChem*, 2011, **12**, 3421–3433.

(73) A. Tkatchenko and M. Scheffler, *Phys. Rev. Lett.*, 2009, **102**, 6–9.

(74) A. D. Becke and E. R. Johnson, *J. Chem. Phys.*, 2005, **122**, 154104.

(75) M. Dion, H. Rydberg, E. Schröder, D. C. Langreth and B. I. Lundqvist, *Phys. Rev. Lett.*, 2004, **92**, 22–25.

(76) O. A. Vydrov and T. Van Voorhis, *Phys. Rev. Lett.*, 2009, **103**, 7–10.

(77) J. Klimeš, D. R. Bowler and A. Michaelides, *J. Phys. Condens. Matter*, 2010, **22**, 022201.

(78) J. Klime, D. R. Bowler and A. Michaelides, *Phys. Rev. B - Condens. Matter Mater. Phys.*, 2011, **83**, 195131.

(79) T. Risthaus and S. Grimme, *J. Chem. Theory Comput.*, 2013, **9**, 1580–1591.

(80) B. Vlaisavljevich, J. Huck, Z. Hulvey, K. Lee, J. A. Mason, J. B. Neaton, J. R. Long, C. M. Brown, D. Alfè, A. Michaelides and B. Smit, *J. Phys. Chem. A*, 2017, **121**, 4139–4151.

(81) J. Wieme, K. Lejaeghere, G. Kresse and V. Van Speybroeck, *Nat. Commun.*, 2018, **9**, 4899.

(82) N. W. Ashcroft and N. D. Mermin, in *Solid State Physics*, ed. N. W. Ashcroft and N. D. Mermin, Holt, Rinehart and Winston, NewYork; London, Print, 1977, ch. 4-7.

(83) N. W. Ashcroft and N. D. Mermin, in *Solid State Physics*, ed. N. W. Ashcroft and N. D. Mermin, Holt, Rinehart and Winston, New York; London, Print, 1977, ch. 8.

(84) P. Kratzer and J. Neugebauer, *Front. Chem.*, 2019, **7**, 1–18.

(85) H. J. Monkhorst and J. D. Pack, *Phys. Rev. B*, 1976, **13**, 5188–5192.

(86) F. Zanca, L. T. Glasby, S. Chong, S. Chen, J. Kim, D. Fairen-Jimenez, B. Monserrat and P. Z. Moghadam, *J. Mater. Chem. C*, 2021, **9**, 13584–13599.

(87) J. H. Cavka, S. Jakobsen, U. Olsbye, N. Guillou, C. Lamberti, S. Bordiga and K. P. Lillerud, *J. Am. Chem. Soc.*, 2008, **130**, 13850–13851.

(88) D. S. Sholl and J. A. Steckel, in *Density Functional Theory: A Practical Introduction*, ed. D. S. Sholl and J. A. Steckel, Wiley, Hoboken, New Jersey, Print, 2009, ch. 3.

(89) M. Springborg, in *Methods of Electronic Structure Calculations: From Molecules to Solids*, Wiley & Sons, Chichester, 2000, ch. 19.

(90) H. Shull and P. O. Löwdin, *J. Chem. Phys.*, 1955, **23**, 1362.

(91) H. J. Silverstone, D. P. Carroll and D. M. Silver, *J. Chem. Phys.*, 1978, **68**, 616–618.

(92) V. Blum, R. Gehrke, F. Hanke, P. Havu, V. Havu, X. Ren, K. Reuter and M. Scheffler, *Comput. Phys. Commun.*, 2009, **180**, 2175–2196.

(93) M. Springborg, in *Methods of Electronic Structure Calculations: From Molecules to Solids*, Wiley & Sons, Chichester, 2000, ch. 20.

(94) P. Pulay, *Mol. Phys.*, 1969, **17**, 197–204.

(95) Á. Ruiz-Serrano, N. D. Hine and C. K. Skylaris, *J. Chem. Phys.*, 2012, **136**, 234101.

(96) O. H. Nielsen and R. M. Martin, *Phys. Rev. Lett.*, 1983, **50**, 697–700.

(97) D. Frenkel and B. Smit, in *Understanding molecular simulation: from algorithms to applications*, Academic Press, San Diego, 2nd, 2002, ch. 5.

(98) M. Sri Harish and P. K. Patra, *Mol. Simul.*, 2021, **47**, 701–729.

(99) H. J. Berendsen, J. P. Postma, W. F. Van Gunsteren, A. Dinola and J. R. Haak, *J. Chem. Phys.*, 1984, **81**, 3684–3690.

(100) G. J. Martyna, M. E. Tuckerman, D. J. Tobias and M. L. Klein, *Mol. Phys.*, 1996, **87**, 1117–1157.

(101) M. Parrinello and A. Rahman, *J. Appl. Phys.*, 1981, **52**, 7182–7190.

(102) K. K. Irikura, *Chem. Phys. Lett.*, 2005, **403**, 275–279.

(103) O. T. Beynon, A. Owens, C. Carbogno and A. J. Logsdail, *J. Phys. Chem. C*, 2023, **127**, 16030–16040.

(104) I. M. Alecu, J. Zheng, Y. Zhao and D. G. Truhlar, *J. Chem. Theory Comput.*, 2010, **6**, 2872–2887.

(105) M. L. Laury, M. J. Carlson and A. K. Wilson, *J. Comput. Chem.*, 2012, **33**, 2380–2387.

(106) E. Ditler and S. Luber, *WIREs Comput. Mol. Sci.*, 2022, **12**, 1–29.

(107) S. Taherivardanjani, R. Elfgen, W. Reckien, E. Suarez, E. Perlt and B. Kirchner, *Adv. Theory Simulations*, 2022, **5**, DOI: 10.1002/adts.202100293.

(108) L. Valenzano, B. Civalleri, S. Chavan, S. Bordiga, M. H. Nilsen, S. Jakobsen, K. P. Lillerud and C. Lamberti, *Chem. Mater.*, 2011, **23**, 1700–1718.

(109) C. Atzori, G. C. Shearer, L. Maschio, B. Civalleri, F. Bonino, C. Lamberti, S. Svelle, K. P. Lillerud and S. Bordiga, *J. Phys. Chem. C*, 2017, **121**, 9312–9324.

(110) A. Ghysels, T. Verstraelen, K. Hemelsoet, M. Waroquier and V. Van Speybroeck, *J. Chem. Inf. Model.*, 2010, **50**, 1736–1750.

(111) J. L. Bao and D. G. Truhlar, *Chem. Soc. Rev.*, 2017, **46**, 7548–7596.

(112) R. Elber and M. Karplus, *Chem. Phys. Lett.*, 1987, **139**, 375–380.

(113) R. E. Gillilan and K. R. Wilson, *J. Chem. Phys.*, 1992, **97**, 1757–1772.

(114) G. Mills and H. Jónsson, *Phys. Rev. Lett.*, 1994, **72**, 1124–1127.

(115) G. Mills, H. Jónsson and G. K. Schenter, *Surf. Sci.*, 1995, **324**, 305–337.

(116) H. Jónsson, G. Mills and K. W. Jacobsen, in *Classical and Quantum Dynamics in Condensed Phase Simulations*, ed. B. J. Berne, G. Ciccotti and D. Coker, World Scientific, Singapore, 1998, pp. 385–404.

(117) G. Henkelman and H. Jónsson, *J. Chem. Phys.*, 2000, **113**, 9978–9985.

(118) G. Henkelman, B. P. Uberuaga and H. Jónsson, *J. Chem. Phys.*, 2000, **113**, 9901–9904.

(119) G. Henkelman and H. Jónsson, *J. Chem. Phys.*, 1999, **111**, 7010–7022.

(120) T. D. Kühne, M. Iannuzzi, M. Del Ben, V. V. Rybkin, P. Seewald, F. Stein, T. Laino, R. Z. Khaliullin, O. Schütt, F. Schiffmann, D. Golze, J. Wilhelm, S. Chulkov, M. H. Bani-Hashemian, V. Weber, U. Borštnik, M. Taillefumier, A. S. Jakobovits, A. Lazzaro, H. Pabst, T. Müller, R. Schade, M. Guidon, S. Andermatt, N. Holmberg, G. K. Schenter, A. Hehn, A. Bussy, F. Belleflamme, G. Tabacchi, A. Glöß, M. Lass, I. Bethune, C. J. Mundy, C. Plessl,

M. Watkins, J. VandeVondele, M. Krack and J. Hutter, *J. Chem. Phys.*, 2020, **152**, 1–47.

(121) J. Vandevondele, M. Krack, F. Mohamed, M. Parrinello, T. Chassaing and J. Hutter, *Comput. Phys. Commun.*, 2005, **167**, 103–128.

(122) J. VandeVondele and J. Hutter, *J. Chem. Phys.*, 2007, **127**, 114105.

(123) S. Goedecker and M. Teter, *Phys. Rev. B - Condens. Matter Mater. Phys.*, 1996, **54**, 1703–1710.

(124) F. Neese, *WIREs Comput. Mol. Sci.*, 2012, **2**, 73–78.

(125) F. Weigend and R. Ahlrichs, *Phys. Chem. Chem. Phys.*, 2005, **7**, 3297–3305.

(126) F. Weigend, *Phys. Chem. Chem. Phys.*, 2006, **8**, 1057–1065.

3 Defects and Correlated Nanodomains

3.1 Motivation

Defects in UiO-66 can be found as nanodomains involving missing linkers or clusters.[1–4] As noted in **Section 1.3.2**, such features are known to increase porosity and improve catalytic properties without compromising the material's exceptional thermal and chemical stability. Although some research has been dedicated to controlling the presence of these defective regions,[2,5] along with some attempts at theoretical rationalisation,[6,7] little is known about how or why nanodomains form. Such an understanding could afford greater control over where and when areas of **reo**, **bcu**, or **scu** defects form during synthesis. Given the properties of for example **reo**, which in many ways outperforms the pristine framework,[8–10] this would offer an additional tool through which UiO-66 could be improved for catalytic or water harvesting applications.

The missing cluster defective topology has so far been the most widely studied. These features were evidenced by Cliffe *et al.* through symmetry-forbidden PXRD peaks,[1] an approach which has also been used in subsequent studies.[11–13] With this method, the integrated area of the forbidden feature can be used for quantitative analysis and is found to correlate with missing cluster content. This contrasts with the routine technique used to determine the presence of missing linkers, TGA. In this case, the presence of vacancies is inferred via sub-stoichiometric weight loss relative to the amount of pure ZrO_2 left at the end of the experiment. As there is no way to determine (from TGA alone) what the stoichiometric amount of residue should be, this method cannot distinguish between missing linkers and missing clusters, as noted by Shearer *et al.*[2] Different topologies can also barely be distinguished using HRTEM, as done by Liu *et al.*[3] Small domains of **reo** can be directly identified within the material, of which images are shown in **Figure 12**, along with **bcu** and **scu**. Schematics for different linker defects are also shown in **Figure 13**.

This technique provides useful information on vacancies, but requires advanced instrumentation and significant amounts of post-processing. In contrast, PXRD is straightforward but provides limited detail; it is also possible that the subtle symmetry-forbidden features may be dismissed as noise. One therefore suspects that these defective topologies are underreported.

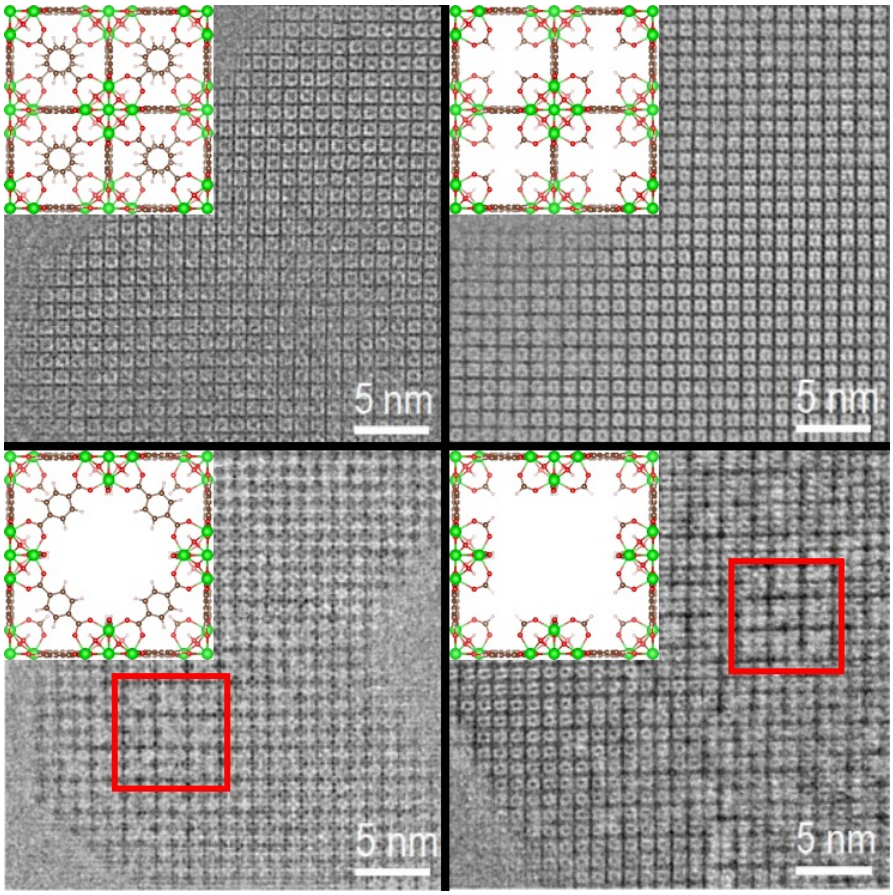


Figure 12: HRTEM contrast transfer function-corrected images of defective topologies in UiO-66. Regions of **fcu**, **bcu**, **reo**, and **scu** are shown respectively top left, top right, bottom left, and bottom right. Corresponding structural schematics have been embedded for easier interpretation. The red squares for the latter two topologies highlight extended defective regions covering several unit cells. Note that these are 2D projections and extent in the perpendicular direction may differ between images. Adapted from ref. [3].

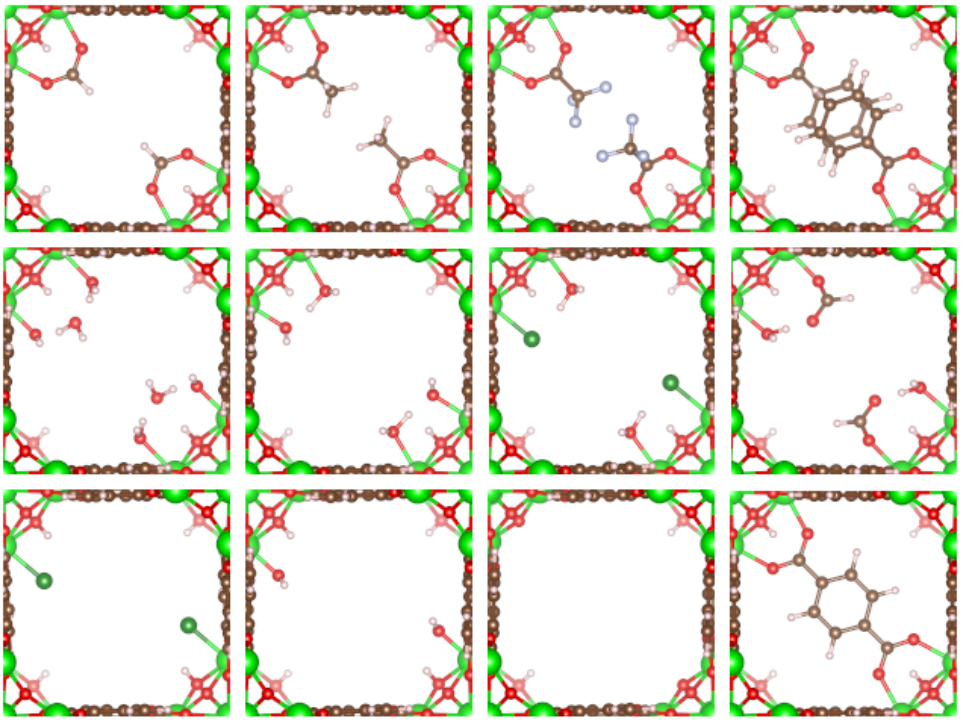


Figure 13: Schematics of the DFT-optimised geometries of linker vacancies in UiO-66. The removal of a BDC linker leaves the two SBUs it was connected to uncapped. Hence, defect termination requires the insertion of two sets of charge-compensating species, one at each zirconium cluster. The top row from left to right shows FA, AA, TFAA, and BA. This is followed by 2 H_2O–OH, H_2O–OH, Cl–H_2O, and FA–H_2O. The final row depicts Cl, OH, **e**, and the reference structure with a BDC linker. Note that in each case, additional linkers exist above and below the regions shown, but these have been obscured to enhance visibility.

Since the discovery of missing cluster nanodomains and the establishment of a method through which they can be analysed, some studies have focused on tuning these defects. One way to generate the **reo** topology is to use a large amount of monocarboxylic acid modulator. Larger ratios of this additive to BDC lead to more prominent tell-tale PXRD peaks, but different modulators require different ratios for the same effect: the forbidden peak area follows the trend benzoic acid (BA) > TFAA > FA > AA.[2,5] Arguments have been made that the preferential deprotonation of these acids over the linker or that steric constraints are what

lead to cluster vacancies, but this does not explain why these should form nanoregions. Alternatively, the modulator may be eschewed altogether and missing cluster domains may form by strictly controlling the moisture content in the reaction mixture.[14] Water is required in order to form Zr_6O_8 clusters and typically this is found in the precursor (e.g. $ZrOCl_2 \cdot 8\,H_3O$), in the modulator (aqueous HCl or impurities in the carboxylic acid), or as a deliberately added reagent. If none is present, only trace UiO-66 forms following synthesis. With sub-stoichiometric H_2O, insufficient clusters form, leading to vacancies and symmetry-forbidden PXRD peaks indicative of the **reo** structure. With enough water, these features disappear, indicating only linker defects form. Again this explanation does not prescribe that missing SBUs should be found as nanodomains. The role of the solvent in these processes is also unknown; a survey by Venturi *et al.* suggests it has some influence on the formation of defects, but missing clusters were not considered by the authors.[15] With both approaches, the manner in which the modulator or water control cluster vacancies appears straightforward, but this does not explain why these form in a spatially aggregated way.

In contrast to **reo**, the other topologies have been encountered more sparingly. While several Zr_6O_8 MOFs are known to adopt **bcu** connectivity, the structure arises from the use of different linkers rather than defectivity.[16–19] Apart from Liu *et al.*'s study,[3] there has only been one more recent report of this structure in relation to vacancies, a combined DFT and inelastic neutron scattering (INS) experiment by Cavalcante *et al.*[4] The authors confirmed this specific defect motif (and excluded **reo** and **scu**) by comparing spectra between the real and DFT-simulated materials, shown in **Figure 14**. The 'fingerprint' region from 0-100 cm^{-1} contained features consistent only with a combination of acetic and formic acid in the **bcu** pattern. **scu** is rarer still, having been reported just a single time.[3] Further complicating the analysis for both of these is that **bcu** shares the same PXRD pattern as **fcu**, and **scu** as **reo**. Without

access to specialised techniques such as HRTEM or INS, it is therefore difficult to study these structures. Little is known about why they arise, other than that the use of formic or acetic acid appears necessary.

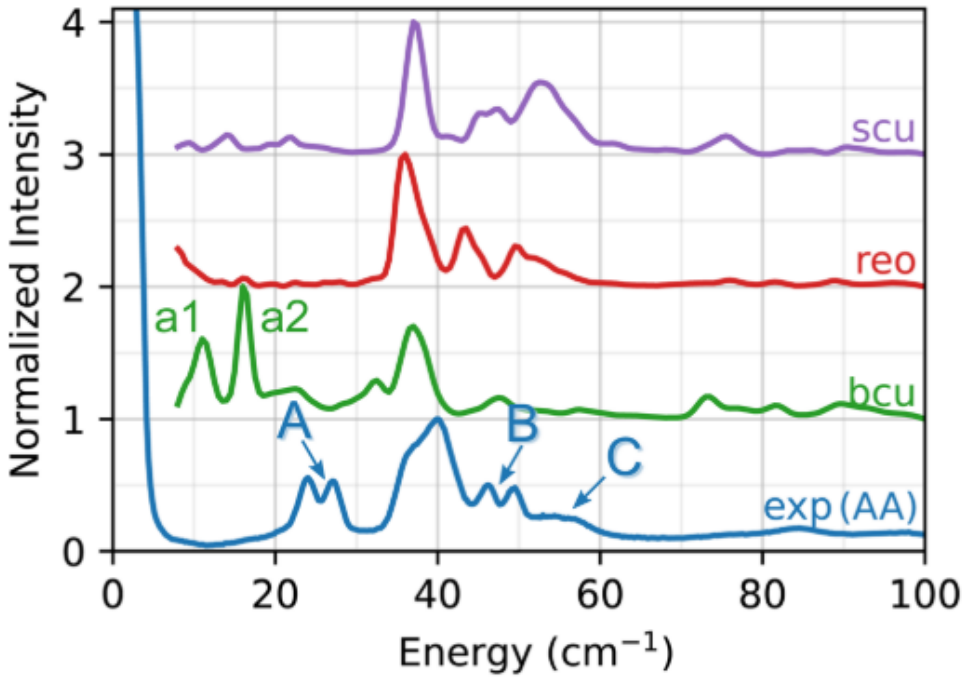


Figure 14: Inelastic neutron scattering fingerprints of AA-modulated UiO-66 and simulated topologies. Peaks a1 and a2, which match A, are associated with the correlated rotation of methyl groups on a **bcu** node, which confirms the presence of this structural motif. The central peak corresponds to linker rotation, while B and C are assigned to mixed FA and AA-capped SBUs. Adapted from ref. [4].

There have also been some purely computational studies targeted at missing cluster defects in UiO-66. Thornton *et al.* exploited the high symmetry of the **fcu** and **reo** unit cells to identify preferences between various vacancy distributions at the DFT level.[8] In this way they found large differences between configurations in the Young and shear moduli, implying the existence of planes from which linkers could be removed without compromising the structure. **reo** was also shown to outperform missing linkers in terms of mechanical stability. This was also investigated by Rogge *et al.* using a DFT-parametrised force field.[20] Having evaluated

all symmetry-inequivalent singly and doubly defective unit cells, the authors came to similar conclusions as Thornton *et al.* in that the material has 'weak' and 'strong' directions. They explained in this way differences in loss-of-crystallinity pressures, which decreased from 1.83 GPa for **fcu** down to 1.17 GPa for the worst-performer. These preferences point towards some defect distributions being more mechanically stable than others, which may play some role in the formation of nanodomains.

Some further work has been carried out on the energetics of linker and cluster vacancies. Bristow *et al.* parametrised a force field against DFT data to compare different capping species across several defect geometries.[6] Of the caps considered, it was reported that AA and Cl–H_2O were spatially correlated such that removing linkers from the same tetrahedral cage was preferable by 20-40 kJ mol^{-1}. Additionally, by carefully choosing which linkers to remove, it was possible to construct a minimum energy path along which the cost of adding defects decreases past the first and which ends with the SBUs seen in **bcu** and **reo**. In this way, the presence of a single vacancy would then promote the presence of others in a correlated manner resulting in nanodomains. Svane *et al.* subsequently used DFT to examine the energetics of linker and cluster defects using a range of capping species.[7] They confirmed in this way the correlation between Cl–H_2O caps as being due to attractive hydrogen bonding, while it was also found that TFAA should be dispersed on account of steric repulsion. The authors additionally evaluated the energy required to form a missing cluster from cells with 1 or 6 linkers removed: this was negative for TFAA, pointing to a preference for **reo** with this cap, in agreement with the experimental findings of Shearer *et al.*[2] There is therefore good reason to expect that the formation of defects and nanodomains might be rooted in thermodynamics.

The current chapter aims to build upon the existing experimental and computational work to progress our understanding of defects and nanodomains in UiO-66. This is motivated by a number of open questions on the subject: what links can be drawn between the stability of defects and the identity of the capping species? Can such trends help to rationalise why different modulators lead to different types and concentrations of vacancies? Concurrently, is it possible to identify the driving force behind nanodomains, such that the formation of these features might ultimately be controlled? To these ends, all known capping species are initially considered and evaluated using DFT across a small set of defect geometries (single missing linker, **bcu**, and **reo**) within the periodic material. These are then supplemented by partial vibrational calculations to also estimate the free energy costs of different linker vacancies. The boundaries between defective and pristine regions are also simulated to evaluate whether an interfacial energy might be responsible for defect aggregation and correlation. Finally, some consideration is given to aperiodic clusters to determine whether the thermodynamic trends seen for the periodic material change for isolated building blocks. Throughout this work, the **scu** topology is excluded on account of its rarity and the lack of experimental information about it.

3.2 Approach

3.2.1 Defining relative energies of defects

In searching for a driving force behind the formation of nanodomains in UiO-66, it will be necessary to examine the thermodynamics of defects. One way to define the energy cost of a missing linker is to consider the pristine and defective structures as being linked by an exchange of molecules. This is illustrated for a single missing linker terminated with formic acid in **Figure 15** and **Equation 78**. Note that as BDC is connected to two separate SBUs, the resulting defect needs to be terminated in two places, which therefore necessitates two equivalents of the capping species.

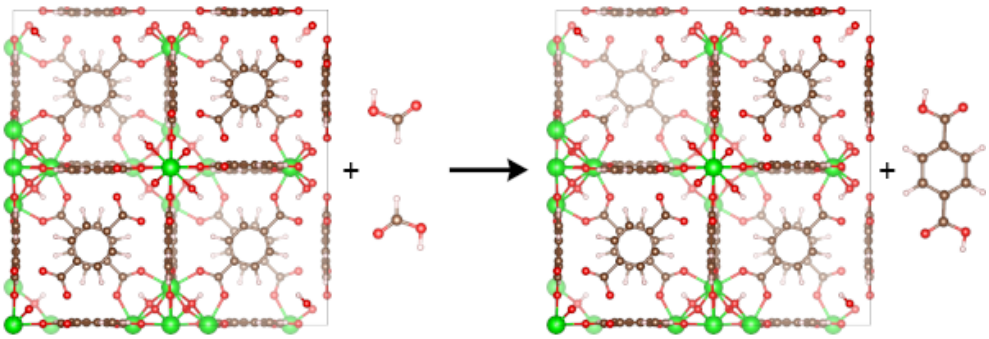


Figure 15: Schematic representation of the exchange of molecules which takes place to construct a formate-terminated missing linker defect. Formic acid and H_2BDC are treated as neutral and isolated gas-phase molecules. Translationally equivalent atoms have been obscured for the periodic systems, such that only the unit cell contents are shown. This schematic is therefore equivalent to **Equation 78**.

$$\begin{gathered}\textbf{fcu} + 2 \text{ formic acid} \rightarrow \text{missing linker defect} + H_2BDC \\ \begin{matrix}(Zr_6O_4(OH)_4)_4(C_8O_4H_4)_{24} \\ + 2\,HCOOH\end{matrix} \rightarrow \begin{matrix}(Zr_6O_4(OH)_4)_4(C_8O_4H_4)_{23}(HCOO)_2 \\ + H_2C_8O_4H_4\end{matrix}\end{gathered} \tag{78}$$

A full set of such equations for all the terminations considered can be found in **Appendix A**. The defect energy relative to the pristine structure can be evaluated for formic acid using **Equation 79**:

$$\Delta E_1 = E_1 + E_{H_2BDC} - E_{\textbf{fcu}} - 2E_{FA} \tag{79}$$

E_1 is the energy of a unit cell with a single formate-terminated linker defect, E_{H_2BDC} is the energy of an isolated gas-phase molecule of H_2BDC, $E_{\textbf{fcu}}$ is the energy of the pristine unit cell, and E_{FA} is that of an isolated gas-phase molecule of formic acid. This can be generalised to n missing linkers using an arbitrary capping species:

$$\Delta E_n = E_n + nE_{H_2BDC} - E_{\textbf{fcu}} - 2nE_{cap} \tag{80}$$

E_n is now the energy of a unit cell containing n linker defects terminated with the chosen cap, and E_{cap} is the energy of the neutral isolated capping species. In the cases that this is made up of more than one molecule,

for example with Cl-H_2O, this energy should be evaluated as the sum of energies of its isolated neutral gas-phase components:

$$E_{\mathrm{cap,Cl–H_2O}} = E_{\mathrm{HCl}} + E_{\mathrm{H_2O}} \tag{81}$$

Finally, special mention needs to be made of the **e** defect termination, since this does not involve capping species. Charge neutrality is maintained by removing framework protons from the two μ_3 OH groups closest to the removed linker, the corresponding equations for which can be found in **Appendix A.11**.

Equation 80 works for linker vacancies, of which the **bcu** topology is a special case (with n=8). However, some modifications are needed in order to take into account the missing cluster in the **reo** topology as it contains only 3 SBUs. A relative energy can still be defined by scaling down the **fcu** system, as detailed generally in **Figure 16** and specifically for the example of formic acid in **Equation 82**.

¾ **fcu** unit cell		+	12 caps		⟶	**reo** unit cell		+	6 linkers	
linkers	18		linkers	0		linkers	12		linkers	6
caps	0		caps	12		caps	12		caps	0
SBUs	3		SBUs	0		SBUs	3		SBUs	0

Figure 16: A chemically balanced scheme for defining the relative energy of the missing cluster defect. The difference in zirconium clusters between the **fcu** and **reo** unit cells means it is necessary to scale the former by a factor of $\frac{3}{4}$. The remaining differences between both structures are accounted for through an exchange of capping and linker molecules.

$$\begin{gathered}\frac{3}{4}\mathbf{fcu} + 12 \text{ formic acid} \rightarrow \mathbf{reo} + 6\mathrm{H_2BDC} \\ \begin{array}{l}(\mathrm{Zr_6O_4(OH)_4})_3(\mathrm{C_8O_4H_4})_{18} \\ +\,12\,\mathrm{HCOOH}\end{array} \rightarrow \begin{array}{l}(\mathrm{Zr_6O_4(OH)_4})_3(\mathrm{C_8O_4H_4})_{12}(\mathrm{HCOO})_{12} \\ +\,6\,\mathrm{H_2C_8O_4H_4}\end{array}\end{gathered} \tag{82}$$

An equation similar to **Equation 80** can therefore be defined:

$$\Delta E_{\mathbf{reo}} = E_{\mathbf{reo}} + 6E_{\mathrm{H_2BDC}} - \frac{3}{4}E_{\mathbf{fcu}} - 12E_{\mathrm{cap}} \tag{83}$$

$E_{\mathbf{reo}}$ is now the energy of the **reo** unit cell. For both linker and cluster vacancies, it is convenient to normalise the defect energy by the number of linkers missing in the defective system relative to the reference pristine unit cell. This corresponds to n for linker vacancies and 6 for the **reo** topology, given that in this latter case the reference state is only three quarters of the **fcu** unit cell. Dividing **Equation 80** and **Equation 83** by n and 6 respectively yields normalised relative energies through **Equation 84** and **Equation 85**.

$$\Delta\varepsilon_n = \frac{E_n}{n} + E_{\mathrm{H_2BDC}} - \frac{E_{\mathbf{fcu}}}{n} - 2E_{\mathrm{cap}} \tag{84}$$

$$\Delta\varepsilon_{\mathbf{reo}} = \frac{E_{\mathbf{reo}}}{6} + E_{\mathrm{H_2BDC}} - \frac{E_{\mathbf{fcu}}}{8} - 2E_{\mathrm{cap}} \tag{85}$$

The approach described here, which involves an exchange of neutral isolated gas-phase molecules, is simple and captures differences between the defective and pristine periodic structures. This modelling strategy is therefore a good starting point for these defective systems and has previously been used to good effect by Vandichel *et al.* and Liu *et al.*[3,21] However, the synthesis of UiO-66 typically takes place in solvent and involves deprotonated forms of the linker and capping molecules. This means differences in the stabilities of these anions and their solvation are not being taken into account when gas-phase reference states are used. It is important to remain mindful of these shortcomings when analysing results derived from the equations presented here.

3.2.2 Free energies

It is possible to build upon electronic defect energies by also considering differences in other degrees of freedom between the defective (defective unit cell + free linker) and pristine (**fcu** unit cell + free capping species) states. Solvation has already been mentioned, but this would significantly increases modelling complexity. The other important degrees of freedom are vibration, rotation, and translation. Taking these into account, it is possible to define defect free energies ΔG:

$$\Delta G_n = \frac{G_n}{n} + G_{\mathrm{H_2BDC}} - \frac{G_{\mathbf{fcu}}}{n} - 2G_{\mathrm{cap}} \tag{86}$$

$$\Delta G_{\mathbf{reo}} = \frac{G_{\mathbf{reo}}}{6} + G_{\mathrm{H_2BDC}} - \frac{G_{\mathbf{fcu}}}{8} - 2G_{\mathrm{cap}} \tag{87}$$

The individual components in **Equation 86** and **Equation 87** can be evaluated directly by passing the output of a CP2K vibrational calculation to the TAMkin Python package.[22] One choice that needs to be made at this stage is whether the reference state for molecules should be that of a truly isolated neutral gas-phase species, or whether it should be part of a pool of neutral ideal gas molecules. We opt here for the latter, as this is closer to the true reference state (charged molecules in solution).

One complication with this approach is that the periodic system contains between 300 to 500 atoms, which makes vibrational calculations expensive. For the missing linker this can be alleviated by using a smaller unit cell, such as the half cell (2 SBUs, 12 linkers) or the primitive cell (1 SBU, 6 linkers). However a missing linker in these systems corresponds respectively to 2 and 4 defects in the conventional unit cell. Here we choose the half cell as a compromise between reducing the system size and increasing the minimum number of defects. Additionally, as most of rest of the structure remains unchanged when a linker or cluster vacancy is introduced, it is possible to keep the corresponding atoms fixed so as to reduce computational cost without compromising on chemical accuracy. This is known as partial Hessian vibrational analysis (PHVA). This approach has previously been used to good effect on UiO-66 by Vandichel *et al.*[21] Some benchmarking has been carried out on missing linker systems terminated by FA and OH to determine the minimum number of atoms which should be allowed to move. These capping species represent bonding that is respectively very similar and very different from that of BDC to the SBU. 3 sets of mobile atoms have been considered: all atoms mobile (the full set); all linker or capping species atoms mobile, along with the 2 closest Zr, the closest bridging oxygen, and closest bridging hydroxyl group on both adjacent SBUs (the extended set); and all linker or capping species atoms mobile (the minimal set). The extended and minimal sets are shown schematically in **Figure 17**.

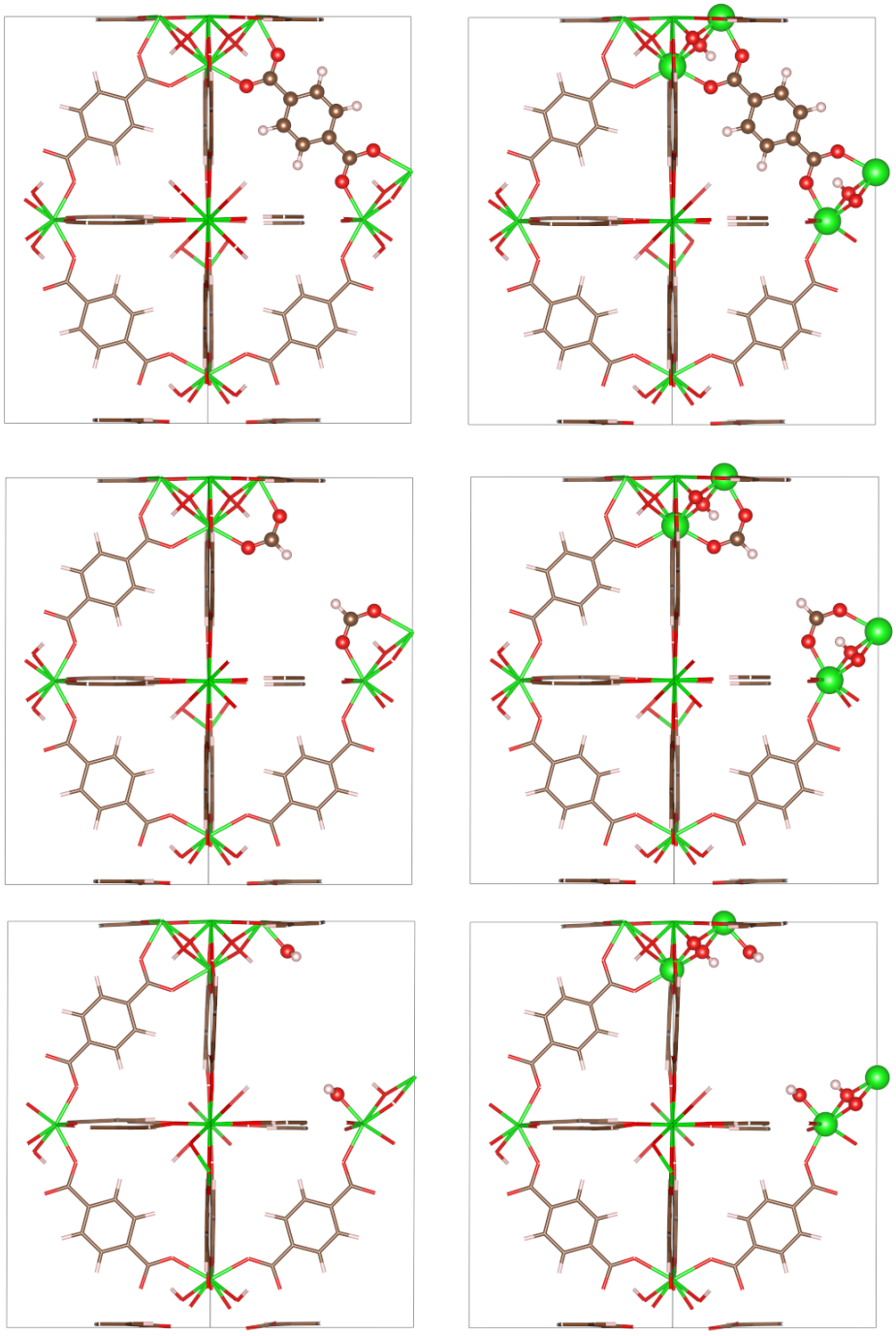


Figure 17: Schematics of the minimal (left column) and extended (right column) sets of mobile atoms. The structures correspond to half cells of the **fcu** and missing linker FA and OH systems respectively from top to bottom. For the fixed atoms, only the skeleton has been drawn.

A series of vibrational calculations using these schemes for fixed atoms has been carried out on the optimised **fcu** and missing linker FA and OH systems. PHVA was carried out with TAMkin to evaluate defect free energies via **Equation 86**. The results are shown in **Table 3**.

Table 3: PHVA defect free energy benchmark

Set of mobile atoms	Defect free energy / kJ mol^{-1}	
	FA	OH
Full	83.6	257.0
Extended	84.6	259.8
Minimal	84.9	266.5

Defect free energies evaluated via PHVA using Hessians generated with different numbers of mobile atoms. The sets of mobile atoms correspond to those shown in **Figure 17**.

With FA as the capping species, the defect free energy is relatively insensitive to number of atoms kept fixed. Given that bonding between the carboxylate and the SBU in this case is similar to that for BDC, this is not surprising. The minimal set of mobile atoms should therefore be suitable for defects terminated by FA, but also for AA, TFAA, and BA. In contrast, for OH, where bonding to the SBU and local structure significantly deviate from that of the pristine material, more mobile atoms are needed. This concern is likely to be valid for other non-carboxylate capping species, for which the extended set of mobile atoms will be used. The resulting error of $\sim$3 kJ mol^{-1} is acceptable given the reduction in computational cost. Note that these sets of mobile atoms must be extended to cover all capping species in the unit cell when modelling the **reo** system.

Finally, for analysis purposes, it will be useful to break down the overall defect free energy into contributions from different degrees of freedom. This is described by **Equation 88**:

$$\Delta G = \Delta E_{\text{electronic}} + \Delta G_{\text{vibration}} + \Delta G_{\text{rotation}} + \Delta G_{\text{translation}} \qquad (88)$$

Note that for the electronic degree of freedom, the energy only is used as the systems considered here are all insulators. This means that electron density cannot be driven (to any meaningful extent) by thermal energy to spill from the highest occupied into the lowest occupied states as the gap between them is too large. This means the associated entropy term is negligible, such that the electronic free energy is simply the electronic energy.

A final caveat with this method of evaluating the vibrational, rotational, and translational degrees of freedom regards the behaviour of capping species molecules in their reference states. These have been chosen here to be in the gas-phase, whereas under true synthetic conditions they should be solvated. The presence of solvent molecules would be expected to have a damping effect on these degrees of freedom which is being neglected with the current approach.

3.2.3 Evaluating linker strain

Some structural modifications of UiO-66 can introduce strain through distortions of the linkers. It is therefore useful to be able to quantify how unfavourable a given linker conformation is relative to that which it would have in the pristine unmodified framework. One way to do this is to extract the atomic coordinates corresponding to a chosen terephthalate from an optimised unit cell of UiO-66, as shown schematically in **Figure 18**, and evaluate its energy as an isolated species in this fixed configuration.

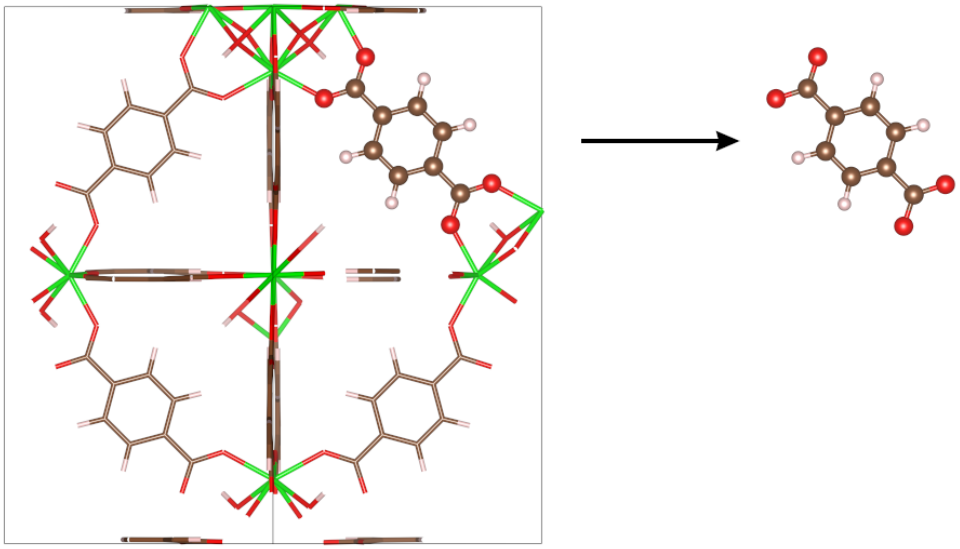


Figure 18: The atomic coordinates corresponding to a given linker can easily be extracted from an optimised periodic system. Subsequently evaluating the energy of this isolated species while keeping atoms fixed means the effect of the conformation is reflected in the total energy.

This can be repeated for another linker from a different structure such that energies can be compared to order conformations by stability. For this purpose, it is easier to model the -2 anion rather than the protonated molecule as this avoids the constrained geometry optimisation steps which would be required to optimally position the protons without relaxing the strain in the benzene backbone. Therefore, the aperiodic ORCA code was used to avoid difficulties associated with the long range electrostatic interactions between a charged system and its periodic images.

The relative energy of one terephthalate conformation with respect to another can then be evaluated as:

$$\Delta E_{\text{strain}} = E^{1}_{\text{BDC}^{2-}} - E^{2}_{\text{BDC}^{2-}} \tag{89}$$

where $E^{i}_{\text{BDC}^{2-}}$ is the energy of the unoptimised linker anion i. It is also possible to optimise the geometry of the BDC^{-2} anion in vacuum to compare strained linkers with the optimal conformation.

3.2.4 Constructing defects

The experimental conventional unit cell of Cavka *et al.* was used as the guess of coordinates for the optimisation of the pristine system.[23] The original space group of $Fm\bar{3}m$ corresponds to a random distribution of μ_3 hydroxyl groups on SBUs; modelling this with DFT would require a large supercell. Therefore, the symmetry was reduced to $F\bar{4}3m$, such that all SBUs have the same alternating pattern of bridging hydroxyls in the same orientation. This space group was only imposed upon the initial coordinates and was not enforced. Subsequently, defects were generated in a semi-automated manner using Python scripts written by the author to remove selected linkers and insert, at the SBUs, the chosen capping species in an orientation that most closely matches the original carboxylic acid. A sample of the code used can be found in **Appendix B**. When bulky molecules are used as caps, it is also necessary to rotate one of the SBUs to which the linker was originally connected by 90° along one of the crystallographic axes. This allows one of the charge-compensating species to point up, and the other down, as shown in **Figure 19**, thereby reducing steric hindrance without being pushed towards the adjacent hydroxyl.

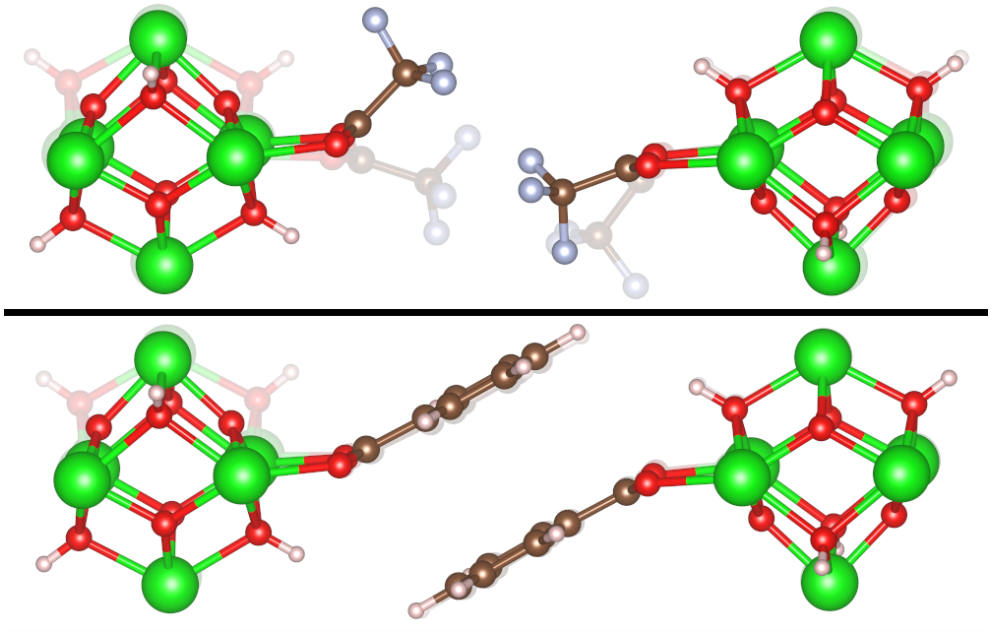


Figure 19: Schematic representation of cap orientations across a missing linker for TFAA (top) and BA (bottom). Structures using uncorrelated and correlated SBUs are drawn opaque and transparent respectively. The capping species generally bends away from its nearby μ_3 OH, and this can be simultaneously satisfied for both carboxylates only with uncorrelated SBUs. Although the geometries look similar when BA is used, bending towards the adjacent hydroxyl remains unfavourable.

Some preliminary calculations were carried out the determine for which species this procedure was necessary. For an AA-terminated missing linker, one of the SBUs adjacent to the defect was rotated by 90° about the c axis, and the system was re-optimised (both atomic positions and cell parameters). This was repeated for single linker defects capped with TFAA and BA, as well as for the pristine material. FA and other small capping species were not evaluated as they are well-separated across the gap of a linker defect. The energy cost of this distortion was evaluated as:

$$\Delta E = E_{\text{distorted}} - E_{\text{undistorted}} \tag{90}$$

The results are shown in **Table 4**:

Table 4: Energy cost of rotating SBUs in UiO-66

System	Energy cost of distortion / kJ mol^{-1}
AA	5.8
TFAA	-3.0
BA	-0.7
fcu	7.2

Energy costs were evaluated using **Equation 90** for a single missing linker terminated by the bulkiest caps. Positive values imply that this configuration is unfavourable compared to the standard SBU orientation.

Although these results suggest that this distortion does introduce additional strain in the structure, the overall energy scale is small, such that SBUs in the framework should be randomly oriented. This is consistent with the neutron diffraction findings of Wu *et al.*[24] Nevertheless, as more TFAA and BA-terminated defects are added to a unit cell, this distortion should become more favourable, whereas it will remain unfavourable for all other species. Therefore, such structures will be used in the cases of TFAA and BA, while the standard orientation of SBUs will be used for the remaining capping species.

Symmetry has also been exploited at times to reduce the configurational space represented by linkers in the conventional cell. Given that 24 of these are nominally present, the introduction of defects rapidly leads to a combinatorial explosion. If only a limited number of these are removed, a sufficient number of the $F\bar{4}3m$ space group's 96 symmetry elements remain, such that only a few distributions are inequivalent. When needed, the bsym Python package has been used to generate this subset of structures.[25]

3.2.5 Interfacial energies

It is possible to estimate the energy penalty associated with the interfaces between different topologies by comparing supercells containing a mix of **fcu** and defective regions to the respective phase-pure materials. To illustrate for the **reo** topology, this involves constructing supercells in which a single missing cluster is embedded in the pristine structure, as shown diagrammatically in **Figure 20**. Schematics can also be found in **Appendix C.1**.

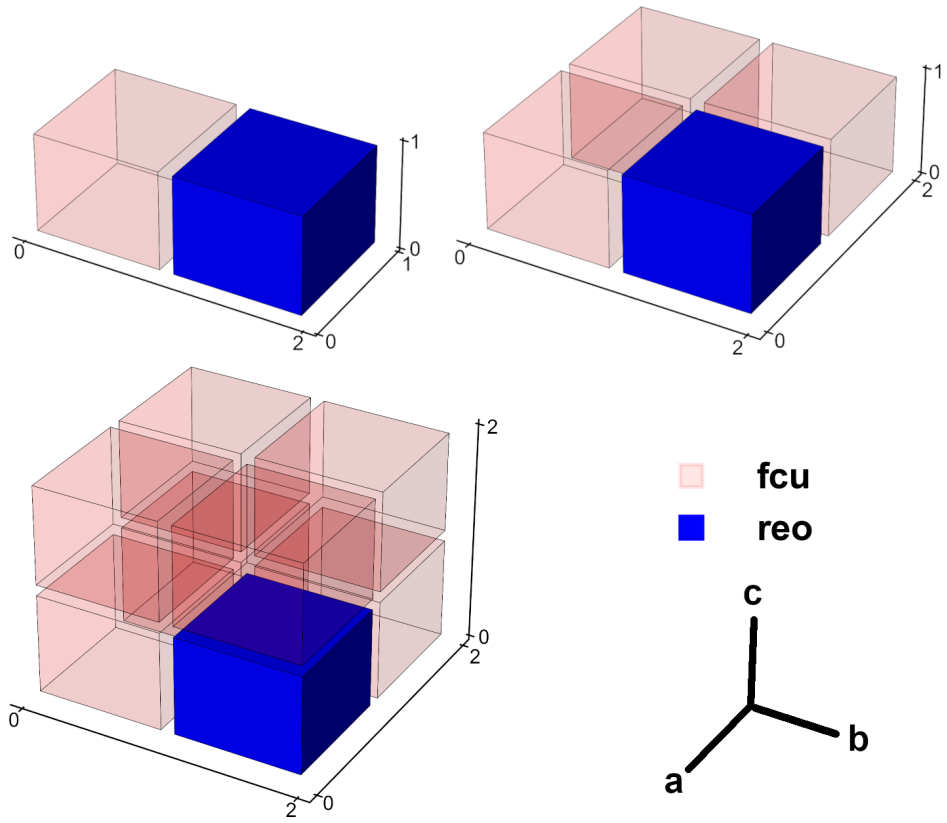


Figure 20: Diagrams representing defective supercells. Each one contains a single missing cluster embedded in the pristine structure. By virtue of the periodic boundary conditions used, 2, 4, and 6 faces of the **reo** subcell are in contact with the **fcu** structure for the 2×1×1, 2×2×1, and 2×2×2 supercell respectively. Each such face will be referred to as an interface.

We can then define the associated energy penalty in the following way:

$$\Delta E_{\text{interface}} = E^{\text{2XY}}_{\text{defective}} - E^{111}_{\boldsymbol{reo}} - (\text{2XY} - 1)E^{111}_{\boldsymbol{fcu}} \tag{91}$$

2XY describes the size of the supercell (2×1×1, 2×2×1, or 2×2×2), $E^{\text{2XY}}_{\text{defective}}$ is the energy of the defective supercell, $E^{111}_{\boldsymbol{reo}}$ is the energy of the standard **reo** unit cell, and $E^{111}_{\boldsymbol{fcu}}$ is the energy of the pristine unit cell. Unlike in **Section 3.2.1**, $\Delta E_{\text{interface}}$ here describes an interfacial energy rather than a defect energy. **Equation 91** can be adapted for interfaces between **fcu** and **bcu** regions:

$$\Delta E_{\text{interface}} = E^{\text{2XY}}_{\text{defective}} - E^{111}_{\boldsymbol{bcu}} - (\text{2XY} - 1)E^{111}_{\boldsymbol{fcu}} \tag{92}$$

where the supercell should now contain an embedded **bcu** subcell instead. Note that as this missing linker topology has an orientation, multiple supercells can be defined for 2×1×1 and 2×2×1. It is also possible to evaluate the interface between different **bcu** orientations using a 2×1×1 supercell containing 2 misaligned subcells (and no **fcu**). In such cases, the energy penalty is defined as:

$$\Delta E_{\text{interface}} = E^{211}_{\boldsymbol{bcu-bcu}} - 2E^{111}_{\boldsymbol{bcu}} \tag{93}$$

Schematics for these **fcu-bcu** and **bcu-bcu** supercells can be found in **Appendix C.2**.

3.2.6 Specific computational details

The majority of calculations in this chapter were carried out using the parameters described in **Section 2.6**. An exception is that for the modelling of aperiodic clusters with CP2K, the dimensions of the empty simulation box were increased to 25×25×25 Å given the larger size of the systems. During vibrational analysis, the majority of framework atoms were also held fixed as described in **Section 3.2.2**. For post-processing with TAMkin, free energies were evaluated at a temperature of 400 K and a pressure of 1 bar was used for the pools of gas-phase molecules. Linker strain was evaluated using ORCA with the settings given in **Section 2.6** according to the method described in **Section 3.2.3**.

3.3 Results and discussion

3.3.1 Periodic defect structures

In the search for the driving force behind nanodomain formation in UiO-66, we first examine simple defect arrangements. These are split into 3 categories: a single linker vacancy; the **bcu** defective motif, which is representative of multiple correlated missing linkers; and **reo**, the missing cluster. In each case, charge compensation can be achieved in a number of different ways involving monocarboxylic acids, water, chloride, and hydroxide. These have been enumerated in **Figure 13**. Differences in stability can be expected based on at least two criteria: differences in bonding between the cap and linker to zirconium; and differences in secondary electrostatic or dispersion interactions between the cap and the surrounding linkers and SBU. The possible defect–charge-compensation combinations have been generated and their energy costs evaluated using **Equation 84** and **Equation 85**, the results of which are shown in **Table 5**.

Table 5: UiO-66 DFT defect energies

Capping	Energy cost per missing linker / kJ mol^{-1}		
	Missing linker	**bcu**	**reo**
FA	17.2	17.5	20.1
AA	20.3	20.1	20.0
TFAA	-14.7*	-11.4**	-11.5
BA	-46.2*	-49.3**	-6.2
$2H_2O$–OH	-157.4	-156.3	-152.8
H_2O–OH	-31.2	-34.9	-33.7
Cl–H_2O	-108.9	-101.7	-99.2
FA–H_2O	-107.4	-104.0	-100.6
Cl	118.5	108.7	110.1
OH	222.7	201.0	201.7
e	408.1	373.2	374.8

Defect energies of different capping species, as evaluated using **Equation 84** and **Equation 85**. Positive and negative values respectively correspond to structures which are less and more stable than the pristine material. The latter implies that UiO-66 should be almost fully defective, which is at odds with experimental evidence. *One or **two SBUs have been rotated 90° about one of the crystallographic axes

We can begin to break down the results in **Table 5** by first examining the monocarboxylates. Removing a linker and moving it to vacuum leads to a loss of dispersion interactions between it and its neighbouring terephthalates. FA, which holds low electron density, compensates poorly in this regard, while AA, TFAA, and BA do so increasingly well. The latter 3 are also large enough for there to be stabilising attractions across the gap of a missing linker. This can be confirmed by breaking down the

defect energies into D3 dispersion and pure DFT components:

$$\Delta E = \Delta E_{\mathrm{D3}} + \Delta E_{\mathrm{DFT}} \tag{94}$$

These have been tabulated in **Table 6** and show clearly that the bulkier acids are better able to replace the missing van der Waals interactions. With BA, the pairs of benzoate groups are also oriented such that the structure benefits from a stabilising π-π overlap, as was shown in **Figure 19**. When moving to the **reo** structure, these attractive effects across the linker vacancy disappear, and dispersion components become less favourable for all species. For BA it remains negative as this species is large enough that by bending towards nearby terephthalates, some stabilisation is recovered. The remaining contribution to the defect energy is the pure DFT one. For both missing linkers and clusters, replacing a BDC with FA or TFAA is electronically favourable, whereas the converse is true for AA and BA. This is in line with the species' pKAs: (3.54,4.46), 3.75, 0.50, 4.76, and 4.20 for terephthalic, formic, trifluoroacetic, acetic, and benzoic acid respectively.[26] This means it is preferable to protonate BDC instead of FA or TFAA, but not of AA and BA. For the missing cluster this component of the defect energy becomes smaller as, with more space in the structure, the caps are able to adopt their optimal conformations and bond more strongly with the framework. The overall balance of all these effects is that there is generally little difference between the energies of linker and cluster vacancies except in the case of BA.

Next, more broadly considering all the caps, the energy cost sees little change in going from a single missing linker to **bcu**. This implies a lack of correlation between vacancies within this defect distribution, in contrast to the trend reported for AA and Cl–H_2O by Bristow *et al.*[6] In this configuration, vacancies are well-separated rather than being aggregated in tetrahedral cavities, such that secondary interactions are not possible. Nevertheless, this implies that there is no thermodynamically preferable pathway through which this motif might be constructed. To some extent, this is also true for **reo**, though we know from the prior

Table 6: Dispersion contributions to defect energies

Defect type	Carboxylic acid	Defect energy / kJ mol^{-1}	
		ΔE_{D3}	ΔE_{DFT}
Missing linker	FA	18.6	-1.3
	AA	3.9	16.4
	TFAA	-3.33	-11.4
	BA	-59.0	12.7
reo	FA	20.7	-0.5
	AA	13.8	6.3
	TFAA	12.1	-23.6
	BA	-12.9	6.7

Breakdown of the defect energies presented in **Table 5** into dispersion and electronic components. The former of these was evaluated using the D3 scheme, which depends only implicitly on the DFT results via the system geometry.[27] Replacing BDC with FA or AA generally leads to a net loss in van der Waals interactions, while with AA or BA, the electronic structure becomes less favourable.

analysis on carboxylic acids that the cancellation of differences in dispersion and electronic effects may simply be fortuitous. Nevertheless, some of these results are encouraging, as the missing linker defect energies agree in part with the experimental findings of Shearer *et al.*,[2] wherein defect concentration follows the trend BA > TFAA > FA > AA for equal concentrations of modulator. Caps involving open Zr sites (OH, Cl, and **e**) all incur high costs, which is in line with the short-lived nature of these states and reflects a net decrease in interactions with the framework compared to the linker, both in terms of bonding and in terms of dispersion.

The final important remark about **Table 5** is that a number of caps exhibit negative defect energies. This implies that for these terminations, the defective material is more stable than the pristine one. However, this is inconsistent with typical experimental missing linker concentrations, which do not exceed 30-35% even when synthesis conditions have been optimised towards defects.[2,28] There are a number of possible explanations for this. One is that terephthalates inside the material are strained relative to their gas-phase conformations. The optimised BDC^{2-} anion is indeed ∼13 kJ mol^{-1} more stable than the equivalent species from the framework. This represents a significant component of the defect energy for monocarboxylates and H_2O–OH, but is less important for other capping species. A further factor in this is that the current approach neglects the loss of translational and rotational degrees of freedom when capping species are immobilised in the framework. The exchange of one linker for two caps (which each involve one or more molecules) leads to a net loss of degrees of freedom, for which the defects are not penalised. In addition, there will be differences in vibrational behaviour between the defective (defective unit cell + free linker) and pristine (**fcu** unit cell + free capping species) states. All defect energies should be positive, given that pristine UiO-66 is known to be more stable than its defective form, so it appears necessary to evaluate defect stability using free energies with **Equation 88** and the method described in **Section 3.2.2**. It is also important to remember that the neutral isolated gas-phase molecules have been chosen as the reference state for the linkers and caps. During synthesis, these would be deprotonated and solvated, and this discrepancy likely contributes to the negative defect energies seen here.

3.3.2 Defect free energies

A more complete thermodynamic picture arises when differences in translational, rotational, and vibrational contributions are taken into account to yield free energies. PHVA was carried out on optimised missing linker

and cluster defects for all the capping species discussed in **Section 3.3.1**. The **bcu** structure was not examined in this way in order to limit computational cost. Thermodynamic corrections were then evaluated at a temperature of 400 K, which is representative of typical synthesis conditions. The resulting free energies are presented in **Table 7**. Some data from a similar study by Vandichel *et al.* is also included for comparison.[21]

Table 7: Free energies of linker and cluster defects

Capping	Free energy cost per missing linker / kJ mol^{-1}		
	Missing linker		
	This study	From ref. [21]	**reo**
FA	84.9		84.7
AA	91.4		83.4
TFAA	81.8*	85.9	88.9
BA	27.8*		62.0
$2H_2O$–OH	144.2		144.8
H_2O–OH	227.0		141.0
H_2O–Cl	73.5		76.9
H_2O–FA	84.7		84.8
Cl	165.4	140.0	149.0
OH	259.8	240.2	226.4
e	312.8	300.5	277.2

Defect free energies for missing clusters and linkers. All results are now positive, which correctly reflects the unfavourable nature of such vacancies. The small values for BA agree with the findings of Atzori *et al.*: this is the best modulator for generating missing clusters.[5] *One SBU rotated 90° about one of the crystallographic axes.

Using free energies, it is now possible to compare different capping species on an equal thermodynamic footing, with all defect terminations now exhibiting positive energy costs. These can be broken down into their different components for easier interpretation, as described by **Equation 88** and shown in **Figure 21** for missing linkers. For monocarboxylates, it is generally rotation and translation which dominate, with DFT energies and the vibrational contribution relatively small. BA has a larger DFT component due to strong dispersion and an additional π-π interaction, as previously discussed. This acid is now the most favourable cap across all species for forming linker and cluster vacancies. However, linker vacancies are heavily favoured, which is surprising given the findings of Atzori *et al.* that this modulator leads to the strongest **reo** PXRD peaks.[5] It is possible that there is a large energy barrier, on account of steric hindrance, towards adopting the stable configuration seen for the missing linker, leading instead to cluster vacancies. A similar issue might arise for TFAA, which has also been reported to result in more **reo** formation than FA and AA, wherein the specific conformation of CF_3 groups in the linker vacancy might be inaccessible due to a large barrier. In contrast, FA and AA would be expected to lead to a balanced mix of both types of defects.

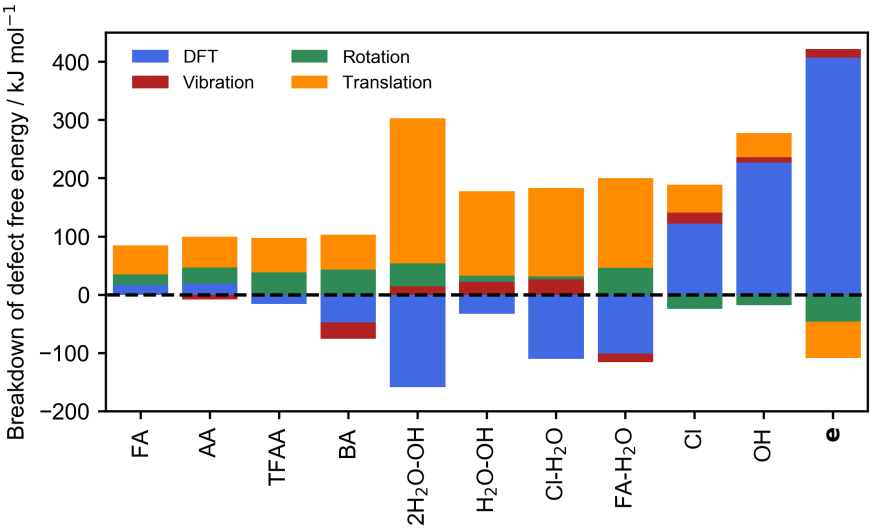


Figure 21: Breakdown of defect free energies for missing linkers. Positive and negative contributions have respectively been plotted above and below the x-axis, marked by a dotted line. Different types of capping species are dominated by different components, which reflects differences in how they bond to the framework and the resulting atomic structure. The relative proportions shown here are unchanged for the **reo** topology.

The largest difference when using free energies is seen for caps containing water, for which defect formation is now always endothermic. A linker vacancy removes 4 or 6 molecules from the gas-phase while returning only a single BDC, leading to a larger net loss of rotational and translational degrees of freedom than for monocarboxylates. However, water has a small moment of inertia, such that translation is a more significant contribution than rotation. This exchange more than compensates for the negative DFT component, but Cl–H_2O and FA–H_2O retain two of the lowest formation energies. The latter result is in good agreement with the findings of Tan *et al.*,[29] who found the water-formate configuration to be more stable than FA alone. However, these caps comprise two molecules, so it is hypothesised that they would form either in a piecewise manner or via an improbable concerted collision, both of which are more complicated than the replacement of BDC with monocarboxylates. The initial attachment of FA before the insertion of H_2O is plausible, given the stability of the former on zirconium, but for Cl–H_2O this would leave

an unfavourable undercoordinated metal, entailing a large energy barrier. Water is also required for the construction of SBUs, so there would be some competition for a reactant that is typically in short supply for this reaction. Therefore, although these charge-compensation schemes exhibit comparable free energies to the monocarboxylates, depending on the synthesis and activation conditions, they may not be observed at all in the framework.

Finally, all the geometries involving undercoordinated Zr (Cl, OH, and **e**) are highly endothermic, and the very large DFT components previously found in **Section 3.3.1** still dominate. This again confirms the inherently unstable nature of open Zr sites, which has implications for catalysis. While these features may be straightforwardly generated through heating, subsequent exposure to moisture or solvent molecules should quickly repopulate these sites given that fully saturated caps are thermodynamically preferable. It is also encouraging to see that the values from this study broadly agree with those of Vandichel *et al.*, who made use of a similar method.[21] However, here the work has been extended to cover all likely capping species across both missing linkers and clusters, allowing for conclusions to be drawn also on the relative stability of the **reo** topology. The small quantitative differences seen between the two sets of data are likely due in part to the use of a different DFT code but also to different atoms being fixed during vibrational calculations.

Ultimately however, the data in **Table 7** does not reveal a systematic preference for missing clusters for certain caps. The results are also generally not consistent with the trend reported by Shearer *et al.* for symmetry-forbidden PXRD peak intensities,[2] which implies that defect energies for the monocarboxylates should be ordered as AA > FA > TFAA > BA. While BA has indeed been found here to have the lowest formation free energy, linker vacancies are favoured over missing clusters, whereas this should be reversed. It therefore appears necessary to exam-

ine other aspects of the material or to improve the method so as to take into account anionic molecular reference states and solvation effects.

3.3.3 Nanodomain interfaces

Another aspect of defective UiO-66 which can be investigated is the interface between different nanoregions. A sufficiently high energy penalty for such features would favour the aggregation of **reo** and **bcu** defective structures into nanodomains so as to minimise interfacial areas. This is therefore a possible thermodynamic driving force behind the formation of extended defective regions in the material. We first consider the **reo** topology terminated by the monocarboxylates FA, AA, TFAA, or BA, all of which have been confirmed by Shearer *et al.* to lead to missing cluster domains when used as modulators during the synthesis.[2] Interfacial energy penalties have been evaluated for these systems using **Equation 91** and the approach described in **Section 3.2.5**; the results are shown in **Table 8**.

Table 8: fcu-reo interfacial energies

System	Interfaces	Total interfacial energy penalty / kJ mol^{-1}				Error
		FA	AA	TFAA	BA	
$2 \times 1 \times 1$	2	-0.14	0.53	-6.96	-0.05	0.06
$2 \times 2 \times 1$	4	-0.12	0.74	-13.69	-0.21	0.16
$2 \times 2 \times 2$	6	-0.44	1.03	-20.31	-0.47	0.54

Energy penalties associated with the interfaces between **fcu** and **reo** nanoregions. Negative values imply that these surfaces are favourable, such that missing clusters should therefore be dispersed so as to maximise interfacial area. Errors have been evaluated by comparing the energies of the pure **fcu** 1×1×1, 2×1×1, 2×2×1, and 2×2×2 systems.

It appears from **Table 8** that the energy penalty for mixing the defective and pristine materials is minimal at this length scale. The figures for FA, AA, and BA are sufficiently small that they cannot be untangled from the errors in the modelling method used. This apparent miscibility between the **fcu** and **reo** unit cells reflects the structural similarity between these systems: the change in cell parameters upon introducing a cluster vacancy is of the order of 0.1%, while the linkers are replaced like-for-like by other carboxylic acids. With TFAA, the total interfacial energy is negative and larger in magnitude, indicating that **fcu-reo** surfaces in this case have a stabilising influence. This also appears to scale with the number of interfaces in the supercell, such that one would expect missing clusters terminated by TFAA to be well-dispersed rather than aggregated in nanodomains. The origin of this stabilisation is not clear, but it may be that the SBUs in this structure which are capped by 1 or 2 TFAAs are more stable than those terminated by 4 in the pure **reo** topology. Nevertheless, this result and those of the other monocarboxylates do not suggest that missing clusters should be aggregated, which disagrees with experimental findings such as those of Shearer *et al.*[2]

The **bcu** topology can also be considered in the case of FA. This structure is anisotropic, as the removal of linkers from a pair of parallel planes leads to a small tetragonal distortion of the unit cell. The mismatch between the **fcu** and **bcu** unit cells is more significant such that interfaces would lead to more strain and therefore larger energy penalties than for supercells containing missing clusters. A new set of supercells has been constructed, for which schematics can be found in **Appendix C.2**, and optimised, such that interfacial energies have been evaluated using **Equation 93**. The results are shown in **Figure 22**.

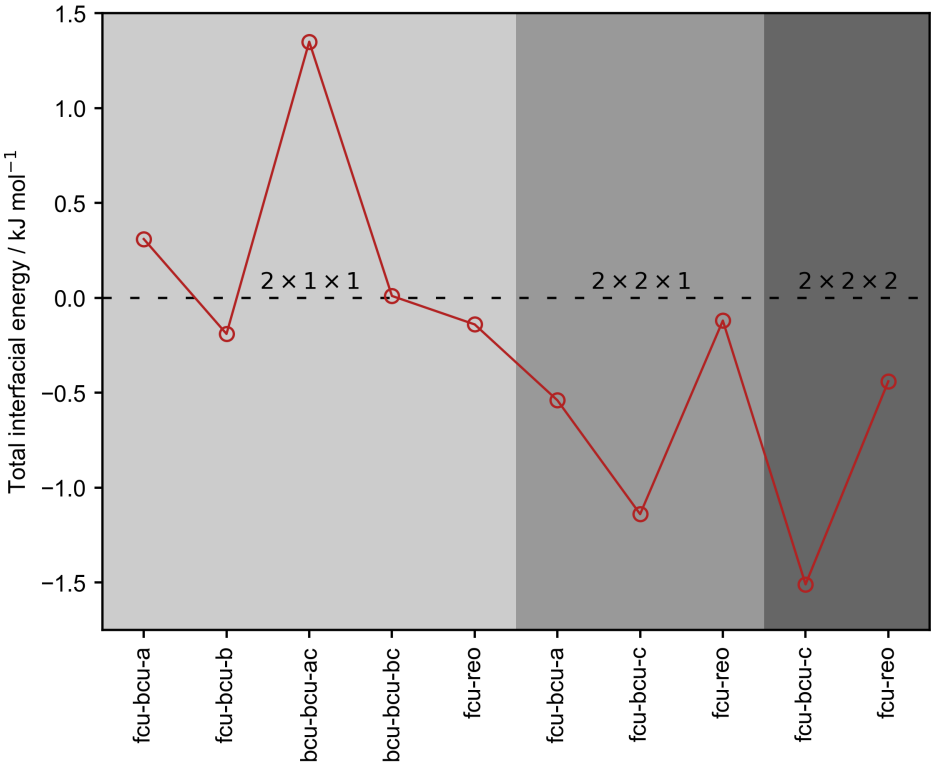


Figure 22: Relative energies of boundaries between **fcu** and formate-terminated **bcu** regions. The **fcu-reo** results have been included for comparison. **bcu** orientations are labelled via the crystallographic axis perpendicular to the planes from which linkers have been removed.

The results of this analysis are similar to those found for **fcu-reo** supercells. The total interfacial energy is small and generally sits within a range of -0.5–0.5 kJ mol^{-1}. Slightly larger penalties are associated with the 2×1×1 **bcu-bcu**-ac, 2×2×1 **fcu-bcu**-c, and 2×2×2 **fcu-bcu**-c interfaces, but these remain smaller than thermal energy. Therefore, based on these thermodynamic grounds, one would expect a range of distributions for defective supercells, rather than primarily as aggregated nanodomains. However, it should be noted that the model systems used for **fcu-reo**/**bcu** boundaries are not necessarily representative of the real material. Cliffe *et al.*'s work on missing clusters estimated the domain size to be ∼7 nm (3 to 4 unit cells),[1] while later studies by Liu *et al.* and Johnstone *et al.* have reported extended blocky and lamellar nanoregions.[3,30] Additionally, differences in vibrational behaviour and solvation arising from interfaces have not been taken into account. These

factors are likely to be more important for **fcu-bcu** and **bcu-bcu** interfaces given that the anisotropy of this missing linker topology leads to greater variation in local structure at domain boundaries.

3.3.4 Enumeration of defects around building blocks

This work has thus far focused on single missing defects and the highly symmetric **bcu** and **reo** unit cells, from which 8 and 12 linkers are respectively missing. There are of course an immense number of asymmetric structures containing an intermediate number of linker vacancies. These remain of interest given Bristow *et al.*'s computational findings that it is possible to construct a sequence of configurations such that the removal of additional linkers eventually leads to a decrease in defect energy.[31] This was demonstrated for the successive replacement of BDCs with AA, Cl-H_2O, or H_2O-OH around a single SBU in a manner that maintained the highest symmetry possible. We consider here a similar approach using FA, but instead of sampling only the highest-symmetry sequence, the full symmetry-inequivalent configurational space is enumerated using the method described in **Section 3.2.4**. The electronic defect energies from these systems have been evaluated using **Equation 84** and are shown in **Figure 23**.

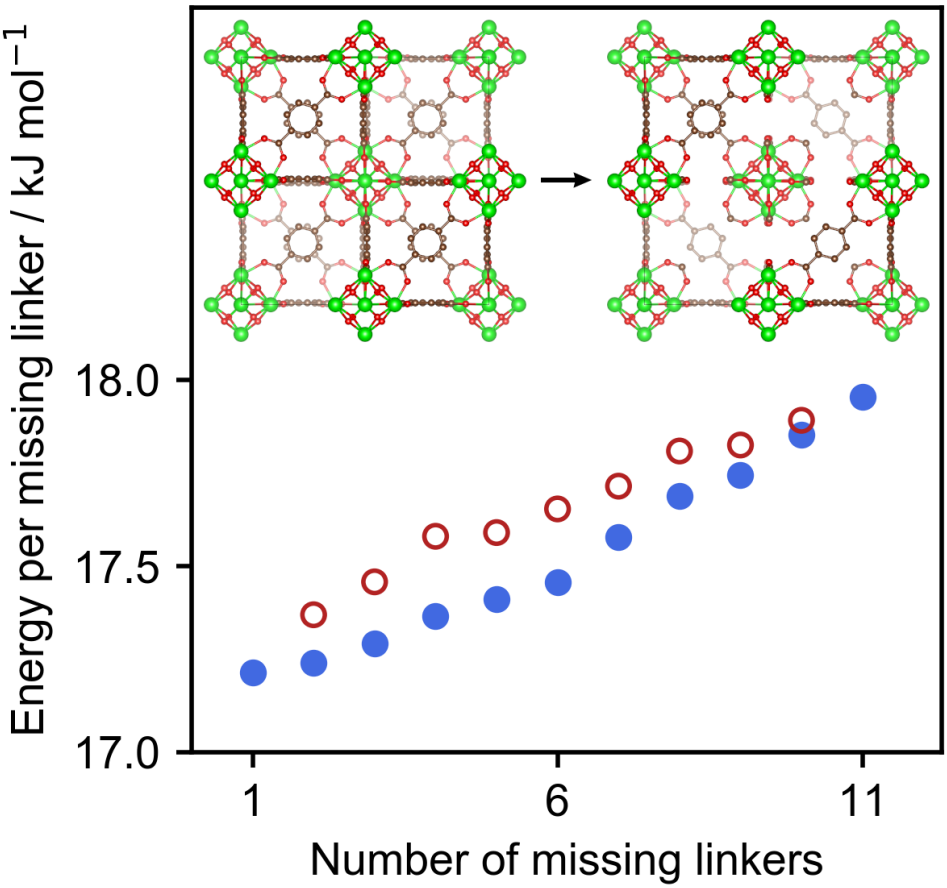


Figure 23: Enumeration of the replacement of linkers around a single SBU in the conventional unit cell with FA. The most and least favourable structures for a given number of vacancies are marked by filled blue and open red circles respectively. The inset schematics illustrate the progression from a fully coordinated to almost fully defective inorganic building block. Protons have been hidden for clarity.

In contrast with what Bristow *et al.* found for other capping species,[31] the cost per defect does not decrease as more linkers are replaced with formic acid. The variation seen here is minute, and the energy difference between the most and least stable configurations for a given number of vacancies would be easily overcome at room temperature. This data does not suggest that there is a thermodynamically preferable sequence of defective structures which would lead to the formation of missing clusters. The discrepancy between the results here and those of Bristow *et al.* may

be due to the choice of capping species or method.[31] These authors noted that close proximity between Cl-H_2O or AA caps is favourable as this leads to stabilising short-range interactions between these species; this is not the case with FA due to its small size. The authors also made use of a force-field to evaluate defect free energies, taking into account in this way differences in both structure and vibrational behaviour. Given that UiO-66 is known to have 'weak' and 'strong' crystallographic directions, as found by Thornton *et al.* and Rogge *et al.*,[8,20] it is possible that some defective configurations preferentially lead to lower energy normal modes. Vibrational analysis may therefore be crucial in identifying a favourable pathway for forming the **reo** topology. However, criticism can also be drawn to Bristow *et al.*'s approach: their force-field was parametrised to reproduce the properties of pristine UiO-66 and it may therefore not be transferable to vacancies.

This analysis can be taken further to consider individual building blocks of UiO-66 prior to framework assembly. It is known that in a solution of DMF and water, Zr readily assembles into octahedral clusters, which can further combine into oligomers upon the addition of a sufficient amount of H_2BDC.[32,33] The initial distribution of species in these systems, governed by different factors than in the periodic material, would then become 'locked in' by strong linker-SBU bonds during crystallisation. Some insight into the distribution of defects in the periodic material may therefore be gleaned by considering just the Zr_6 building blocks. In that regard, it is important to note that both **bcu** and **reo** contain SBUs around which 4 specific linkers have been removed, as shown in the left panel of **Figure 24**. The fully defective unit is also shown in the right panel. A preference for clusters of this type would then help to explain the formation of nanodomains.

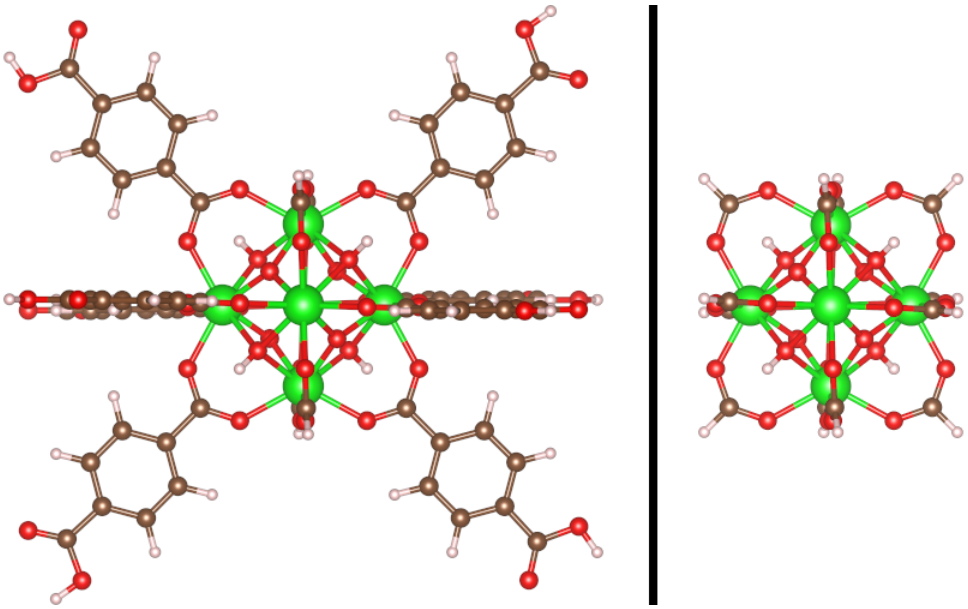


Figure 24: Schematics of unoptimised defective aperiodic building blocks terminated by FA. The left panel shows the type of SBU present in **bcu** and **reo** nanodomains, where a difference in relative orientation leads to the difference in topologies. The right panel shows a fully defective building block capped by FA.

Even for this relatively small model, the possible combinations of caps make exhaustive sampling unfeasible. This problem can be mitigated by again exploiting symmetry and limiting the possible means of termination to the linker and a given modulator acid. Starting from a cluster surrounded by 12 dangling terephthalates, 'defective' structures were generated by replacing some number of the dicarboxylic acid with FA, AA, TFAA, or BA. These were then optimised without pinning the ends of the linkers. Subsequently, relative electronic energies, shown in **Figure 25**, were evaluated using a modified version of **Equation 84**.

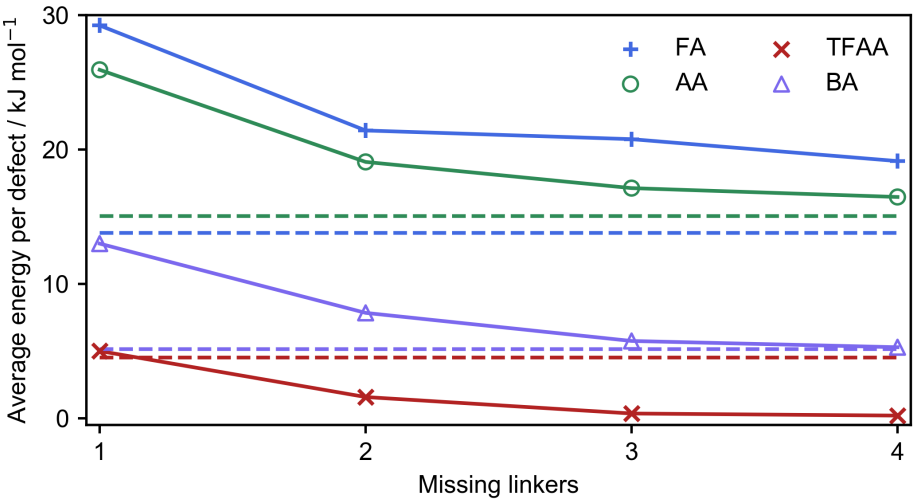


Figure 25: Electronic energy costs of linker vacancies around aperiodic clusters. For a given number of defects, all symmetry-inequivalent structures were taken into account in the average. Results for the fully defective units are marked by dotted lines. Unlike with the periodic material, it generally becomes progressively easier to replace linkers with monocarboxylates.

The behaviour and trends exhibited by the Zr_6 building block are markedly different from those of the periodic material. The energy scale involved with defects here is larger than before, which can be attributed to the additional flexibility afforded to dangling terephthalates. Over the course of optimisation, triplets of linkers around bridging oxygens approach each other so as to adopt a geometry, shown in **Figure 26**, which is similar to the slipped triangle seen in benzene trimers.[34,35] For these molecules in the gas phase, this arrangement benefits from a stabilisation of ∼20 kJ mol^{-1} relative to the isolated constituents. The subsequent introduction of a defect therefore generally incurs a larger penalty since this favourable interaction is disrupted. Isolating the dispersion component of the defect energy for the singly defective cluster, shown in **Table 9**, confirms this. FA poorly compensates for the lost linker, whereas the increasingly bulky AA, TFAA, and BA make for better replacements.

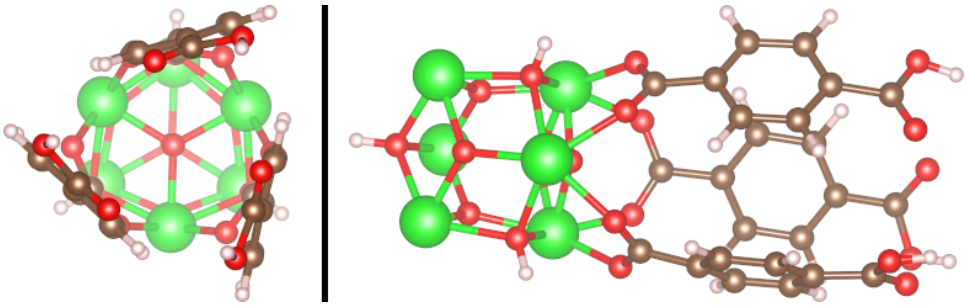


Figure 26: Slipped triangle geometry adopted by linkers in the aperiodic cluster. This structure, which can only be formed around a bridging oxygen, is affected by competing effects: stabilising dispersion interactions are partly offset by the strain introduced into the linker molecules. The slipped triangle is most easily seen in the left panel, which shows a head-on view, while distortions in the bonds can be identified in the right panel, which depicts a side-on view.

Another difference is that the energy cost decreases with each additional defect, unlike in the periodic material where this trend was reversed. Examining the electronic components in **Table 9** shows that replacement is now electronically favourable for all acids, whereas in **Section 3.3.1** this was previously the case only for FA and TFAA. This likely reflects that moving a linker from a triangle to vacuum allows it to return to its optimal planar geometry, helping to compensate for the loss of dispersion interactions. Comparing the distorted conformation with the gas-phase planar one for $HBDC^-$ shows that the latter is indeed more stable by ~10 kJ mol^{-1}. In addition to this, while the first defect disrupts a slipped triangular arrangement, subsequent vacancies do not necessarily do so as they may be taken from the same cage. Both these effects mean that, on average, the energy cost decreases with additional substitutions. For the caps tested here, the replacement of linkers becomes progressively less unfavourable until a minimum energy cost is eventually reached for the fully defective building block (dotted lines in **Figure 25**). This is not the case for TFAA, which reaches an optimum at some intermediate defect content before increasing towards the fully substituted value. It is also notable that the energy cost for TFAA and BA rapidly becomes

comparable to thermal energy, such that a range of defective clusters would be expected in solution.

Table 9: Dispersion components of isolated cluster defect energies

Carboxylic acid	Defect energy / kJ mol^{-1}	
	ΔE_{D3}	$\Delta E_{\mathrm{Electronic}}$
FA	43.0	-13.7
AA	34.7	-8.8
TFAA	28.0	-23.0
BA	20.6	-7.6

Breakdown of the defect energy of a single vacancy into dispersion and electronic components. Unlike for the periodic material, there is now always a net loss of stabilising van der Waals interactions when a linker is replaced. This is offset by the structure of the linker, which is highly distorted in the cluster, but can adopt its optimal planar conformation in vacuum, lowering the electronic energy.

Given the trends presented here and the competition the linker would face with the modulator acids in becoming deprotonated, it is expected that all building blocks in the solution will be at least partly defective. Vacancies at this stage are eventually locked in when clusters become incorporated in the growing crystal, such that the presence of missing linkers in the finished material is not surprising. However, examining the lowest-energy configurations containing 4 defects does not reveal any preferences for the **bcu**/**reo**-type SBU. Instead, with FA and AA there is a clear drive for the removal of linkers from the same tetrahedral cage, as previously reported by Bristow *et al.*, while this is not necessarily the case for TFAA and BA.[6] Therefore, while the mechanism for how vacancies arise appears to stem in part from disorder in the building blocks, this does not explain correlated defective nanodomains. It should be noted

that the results from this analysis imply that secondary electrostatic and dispersion interactions play an important role in the distribution of capping species. The semi-empirical D3 scheme used here is not strongly system-dependent, and may not fully capture the chemistry specific to benzene trimers. Additionally, it is not known whether the slipped triangular arrangements seen here would still form in the presence of solvent, which has been neglected. These might inhibit distortions through either steric hindrance or by shielding the benzoate groups from each other. The method used here may therefore paint too simple a picture of how caps and dangling linkers behave in the aperiodic building block.

3.4 Conclusions

The objective of this work was to answer questions about the origins and driving force behind defective nanodomains in UiO-66. Following modulated synthesis, the material may exhibit defective nanoregions characterised by either missing linkers or missing clusters, regularly distributed so as to lead to either the **bcu** or **reo** topology. The size and type of domain are dependent on the exact conditions of modulation. In an attempt to unravel the driving forces behind this behaviour, DFT experiments were carried out across different defect terminations, which was followed by an examination of the interfaces between topologies before the removal of linkers around individual building blocks was enumerated.

Initially, a large array of possible charge compensation schemes, derived from both experimental reports and theoretical studies, was considered. DFT-D3 defect energies revealed some trends which broadly agree with experiments: TFAA and BA outperform FA and AA in the creation of vacancies, while geometries involving undercoordinated zirconium are very unstable. On the other hand, half the caps exhibited negative energy costs, which reflects that the replacement of a linker by 2 or more molecules neglects a loss in degrees of freedom unless translation, rotation, and vibration are accounted for. This should be considered carefully,

especially when caps of different types are being compared, though the evaluation of free energies is not always computationally feasible. Additionally, the exclusion of dispersion interactions, evaluated in this case by the semi-empirical D3 correction, can reverse the observed trends. This warrants further investigation through means of more robust schemes such as vdW-DFT, despite the extensive benchmarking that has been carried out for dispersion-dominated MOFs.[36]

Partial vibrational calculations were carried out to extend this initial set of experiments with the corrections necessary for free energies. The results of this approach are broadly in line with those of Vandichel *et al.*,[21] who previously used it for the same system. With the proper treatment of all degrees of freedom, the H_2O–Cl and H_2O–FA can be confirmed to be competitive with carboxylic acids for both missing linkers and clusters. However, defective nanodomains have not been reported for these species, suggesting possible shortcomings in experimental characterisation or that water, typically in short supply during the synthesis, may be insufficiently available given it is also required to form SBUs. The monocarboxylates broadly follow the same trends reported by Shearer *et al.* for missing linkers, but this falls apart for the **reo** structure,[2,5] though BA remains the most favourable cap. This implies either that the thermodynamics of the periodic material is not crucial to the formation of defects and nanodomains, or that some chemistry is missing from the overall picture. The same considerations also apply to the H_2O–Cl and H_2O–FA charge compensations. It remains useful to point out that undercoordinated Zr is highly unfavourable and can be stabilised through the adsorption of moisture. This has implications in catalysis, where such active sites are often deliberately generated.

The system was also extended in size to examine the interfaces between **fcu** and **reo** regions. Supercells up to 2×2×2, containing around 3500 atoms, were constructed using the monocarboxylate capping species for

which defective nanodomains have been reported. The resulting interfacial energies represent penalties which are comparable to thermal energy, such that these surfaces should have no influence on the distribution of missing clusters. This analysis was repeated for boundaries between **fcu** and **bcu** regions using FA as the capping species. It was again found that the corresponding energy penalties are too small to drive any correlation or aggregation of defects in the material. It is possible that differences in vibrational behaviour due to structural changes at the interfaces may provide additional impetus for the formation of nanodomains, but such analysis is computationally intractable with the current methodology.

Finally, the replacement of linkers with FA around a single SBU within the periodic system was investigated. The full set of possible substitutions was considered through the use of symmetry, which revealed that the defect energy per missing linker does not change as more vacancies are introduced. This contrasts with the findings of Bristow *et al.*,[31] who found that this decreases when AA, Cl–H_2O, or H_2O–OH are used as capping species. This may be because of a difference in methods, in particular the inclusion of vibrational free energy, which may be critical in such analysis. This enumeration approach was repeated for the aperiodic building blocks terminated by FA, AA, TFAA, and BA for up to 4 missing linkers. It was found in this way that, in the absence of solvent, the dangling ends of linkers move towards each other so as to fall into the slipped triangle arrangement seen in benzene trimers. The replacement of a terephthalate with a monocarboxylate, which disrupts this geometry, incurs a larger energy penalty than seen in the periodic material. This cost decreases with subsequent substitutions as a result of the balance between dispersion and electronic effects. While this explains how highly disordered building blocks might arise and thus how defects might become locked in during the synthesis, no preference was found for the specific SBU present in **bcu** and **reo** domains. However, the dominance of dispersion forces in these energies highlights again the

possible need to use a more robust dispersion scheme. Additionally, the lack of any solvent may have contributed to the results seen here, as the presence of explicit solvent molecules would likely inhibit the relaxation of dangling linkers into slipped triangular arrangements.

The overall results of this study have failed to answer some of the questions regarding defective nanodomains. It remains unclear why or how regions of **bcu** and **reo** form, and why some modulators are more conducive to this. Nevertheless, the work carried out here has ruled out some possibilities and laid the groundwork for future studies. DFT analysis has shown that correlated defect motifs are not inherently more stable, but the role of dispersion interactions in UiO-66 may be more important than previously thought. Extending this analysis to consider defect free energies did not significantly alter the overall understanding of nanodomains, but it allows for a useful comparison of the thermodynamics of different caps. This has implications for how the structure of vacancies will be affected by synthesis and activation conditions. Other calculations also suggest that interfaces between different regions are unlikely to drive the formation of **bcu** or **reo** topologies, with their corresponding penalties comparable to thermal energy. Finally, examining the individual building blocks did not reveal a thermodynamically favourable pathway towards the formation of defective topologies, but this has helped to progress our understanding of how defects are incorporated and become locked into the material. Ultimately, it may be necessary to be more detailed in studying building blocks in solution; solvent molecules may be directly involved, or Zr_6 oligomers may behave differently from the monomers. It is also possible that the answer may lie in the nucleation and growth mechanism, though such phenomena are difficult to probe and would require detailed experimental support. Indeed, local supersaturation of the solution with the modulator would explain why defects become spatially aggregated, but this would not clarify why the specific **bcu** or **reo** structures are adopted.

References

(1) M. J. Cliffe, W. Wan, X. Zou, P. A. Chater, A. K. Kleppe, M. G. Tucker, H. Wilhelm, N. P. Funnell, F. X. Coudert and A. L. Goodwin, *Nat. Commun.*, 2014, **5**, 1–8.

(2) G. C. Shearer, S. Chavan, S. Bordiga, S. Svelle, U. Olsbye and K. P. Lillerud, *Chem. Mater.*, 2016, **28**, 3749–3761.

(3) L. Liu, Z. Chen, J. Wang, D. Zhang, Y. Zhu, S. Ling, K. W. Huang, Y. Belmabkhout, K. Adil, Y. Zhang, B. Slater, M. Eddaoudi and Y. Han, *Nat. Chem.*, 2019, **11**, 622–628.

(4) L. S. R. Cavalcante, M. A. Dettmann, T. Sours, D. Yang, L. L. Daemen, B. C. Gates, A. R. Kulkarni and A. J. Moulé, *Mater. Horizons*, 2022, 187–196.

(5) C. Atzori, G. C. Shearer, L. Maschio, B. Civalleri, F. Bonino, C. Lamberti, S. Svelle, K. P. Lillerud and S. Bordiga, *J. Phys. Chem. C*, 2017, **121**, 9312–9324.

(6) J. K. Bristow, K. L. Svane, D. Tiana, J. M. Skelton, J. D. Gale and A. Walsh, *J. Phys. Chem. C*, 2016, **120**, 9276–9281.

(7) K. L. Svane, J. K. Bristow, J. D. Gale and A. Walsh, *J. Mater. Chem. A*, 2018, **6**, 8507–8513.

(8) A. W. Thornton, R. Babarao, A. Jain, F. Trousselet and F. X. Coudert, *Dalt. Trans.*, 2016, **45**, 4352–4359.

(9) M. Peng, D. You, H. Shi, P. Shao, W. Ren, L. Yang, X. Sheng, J. Shao, X. Ding, L. Ding, H. Wang, K. Yu and X. Luo, *Chem. Eng. J.*, 2022, **448**, 137612.

(10) Y. Jiao, Y. Liu, G. Zhu, J. T. Hungerford, S. Bhattacharyya, R. P. Lively, D. S. Sholl and K. S. Walton, *J. Phys. Chem. C*, 2017, **121**, 23471–23479.

(11) G. C. Shearer, J. G. Vitillo, S. Bordiga, S. Svelle, U. Olsbye and K. P. Lillerud, *Chem. Mater.*, 2016, **28**, 7190–7193.

(12) J. Li, Y. Liu, X. Wang, G. Zhao, Y. Ai, B. Han, T. Wen, T. Hayat, A. Alsaedi and X. Wang, *Chem. Eng. J.*, 2017, **330**, 1012–1021.

(13) M. Taddei, R. J. Wakeham, A. Koutsianos, E. Andreoli and A. R. Barron, *Angew. Chemie - Int. Ed.*, 2018, **57**, 11706–11710.

(14) P. Chammingkwan, G. Y. Shangkum, L. T. T. Mai, P. Mohan, A. Thakur, T. Wada and T. Taniike, *RSC Adv.*, 2020, **10**, 28180–28185.

(15) D. M. Venturi, F. Campana, F. Marmottini, F. Costantino and L. Vaccaro, *ACS Sustain. Chem. Eng.*, 2020, **8**, 17154–17164.

(16) V. Bon, I. Senkovska, M. S. Weiss and S. Kaskel, *CrystEngComm*, 2013, **15**, 9572–9577.

(17) H. Reinsch, B. Bueken, F. Vermoortele, I. Stassen, A. Lieb, K. P. Lillerud and D. De Vos, *CrystEngComm*, 2015, **17**, 4070–4074.

(18) J. Jacobsen, H. Reinsch and N. Stock, *Inorg. Chem.*, 2018, **57**, 12820–12826.

(19) Y. Wang, S. Yuan, Z. Hu, T. Kundu, J. Zhang, S. B. Peh, Y. Cheng, J. Dong, D. Yuan, H. C. Zhou and D. Zhao, *ACS Sustain. Chem. Eng.*, 2019, **7**, 7118–7126.

(20) S. M. Rogge, J. Wieme, L. Vanduyfhuys, S. Vandenbrande, G. Maurin, T. Verstraelen, M. Waroquier and V. Van Speybroeck, *Chem. Mater.*, 2016, **28**, 5721–5732.

(21) M. Vandichel, J. Hajek, F. Vermoortele, M. Waroquier, D. E. De Vos and V. Van Speybroeck, *CrystEngComm*, 2015, **17**, 395–406.

(22) A. Ghysels, T. Verstraelen, K. Hemelsoet, M. Waroquier and V. Van Speybroeck, *J. Chem. Inf. Model.*, 2010, **50**, 1736–1750.

(23) J. H. Cavka, S. Jakobsen, U. Olsbye, N. Guillou, C. Lamberti, S. Bordiga and K. P. Lillerud, *J. Am. Chem. Soc.*, 2008, **130**, 13850–13851.

(24) H. Wu, Y. S. Chua, V. Krungleviciute, M. Tyagi, P. Chen, T. Yildirim and W. Zhou, *J. Am. Chem. Soc.*, 2013, **135**, 10525–10532.

(25) B. J. Morgan, *J. Open Source Softw.*, 2017, **2**, 370.

(26) J. Speight, N. A. Lange and J. A. Dean, in *Lange's Handbook of Chemistry*, McGraw Hill, New York, United States, 16th edn., 2005, ch. 2.15, pp. 2.620–2.669.

(27) S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.

(28) G. Ye, D. Zhang, X. Li, K. Leng, W. Zhang, J. Ma, Y. Sun, W. Xu and S. Ma, *ACS Appl. Mater. Interfaces*, 2017, **9**, 34937–34943.

(29) K. Tan, H. Pandey, H. Wang, E. Velasco, K. Y. Wang, H. C. Zhou, J. Li and T. Thonhauser, *J. Am. Chem. Soc.*, 2021, **143**, 6328–6332.

(30) D. N. Johnstone, F. C. Firth, C. P. Grey, P. A. Midgley, M. J. Cliffe, S. M. Collins and D. N. Johnstone, *J. Am. Chem. Soc.*, 2020, **142**, 13081–13089.

(31) J. K. Bristow, K. L. Svane, D. Tiana, J. M. Skelton, J. D. Gale and A. Walsh, *J. Phys. Chem. C*, 2016, **120**, 9276–9281.

(32) H. Xu, S. Sommer, N. L. N. Broge, J. Gao and B. B. Iversen, *Chem. Eur. J.*, 2019, **25**, 2051–2058.

(33) C. Hennig, S. Weiss, W. Kraus, J. Kretzschmar and A. C. Scheinost, *Inorg. Chem.*, 2017, **56**, 2473–2480.

(34) J. B. Hopkins, D. E. Powers and R. E. Smalley, *J. Phys. Chem.*, 1981, **85**, 3739–3742.

(35) O. Engkvist, P. Hobza, H. L. Selzle and E. W. Schlag, *J. Chem. Phys.*, 1999, **110**, 5758–5762.

(36) J. Wieme, K. Lejaeghere, G. Kresse and V. Van Speybroeck, *Nat. Commun.*, 2018, **9**, 4899.

4 The Influence of Defects on Acidic Properties

4.1 Motivation

One facet of UiO-66 that has attracted much attention is its potential as a solid-state acidic catalyst. The material nominally contains two types of acid sites: the Brønsted-acidic bridging hydroxyl groups, and the Lewis acidic Zr^{4+} ions, both in the SBUs. These well-defined atomic sites, combined with the relatively open structure of the framework, make this material (and others with the same architecture) an ideal catalyst platform.[1] While these acidic properties can be tuned by modifying the linkers,[2–4] it is also possible to alter catalytic performance by introducing linker or cluster vacancies, the latter of which can be more effective.[5,6] The relationship between acidity and defectivity is therefore an important factor to consider when using UiO-66 as a catalyst.

The native Brønsted acidity of UiO-66 has been the subject of some studies aimed at quantifying acid strength. Potentiometric titrations by Klet *et al.* on nearly defect-free UiO-67 revealed only a single equivalence point at a pKA of 3.44, corresponding to μ_3 hydroxyl groups.[7] In contrast, defective UiO-66 yielded a further 2 peaks at 6.79 and 8.30, which were assigned to water and hydroxyl moieties at defect sites. This has been corroborated by Cirujano and Llabrés I Xamena,[8] who reported distinct frequencies of 2150, 2142, and 2135 cm^{-1} for CO adsorbed onto water-capped defects, as shown diagrammatically in **Figure 27**. A recent study by Chen,[9] which calculated pKAs of these sites using computational free energies of cluster models, gave results which disagree with previous assignments. This highlights that evaluating and interpreting the relative Brønsted acidity of sites in UiO-66 is not an unambiguous exercise.

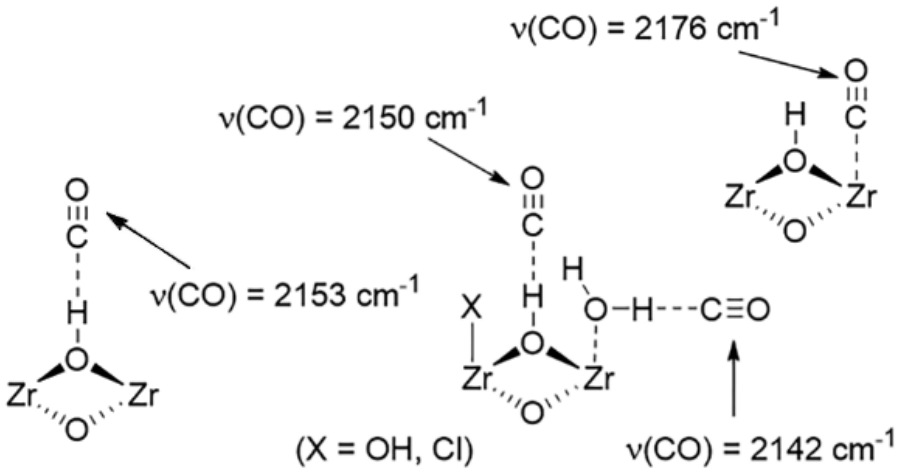


Figure 27: The Brønsted and Lewis acidic sites in defective UiO-66, as proposed by Cirujano and Llabrés I Xamena. An additional frequency of 2135 cm^{-1} is additionally reported and attributed to physisorbed CO; this may also correspond to carbon monoxide on X = OH. Diagram adapted from ref. [8].

One question which remains regarding the μ_3 hydroxyl and defects is whether the former's potential as an acid site can be tuned by adjusting the capping species at an adjacent linker vacancy. Bakuru *et al.* have demonstrated that when zirconium is replaced with Ce or Hf, the equivalence point of the bridging OH shifts to lower pKA.[10] This is in line with the oxophilicity of each metal, reflected in the strength of bonds with oxygen (760, 791, and 795 kJ mol^{-1} for Zr, Hf, and Ce respectively),[11] which weakens binding to the proton. It has also been shown that the electronic structure of Zr can be altered by adding electronegative or electropositive groups on the linker.[12] The manner in which defects are terminated, which covers a range of possible capping species, may therefore influence the Brønsted acidity of μ_3 OH groups beyond simply adding more sites.

The relationship between defects and Lewis acidity in UiO-66 has been studied more directly through catalytic tests. Vermoortele *et al.* have reported that yield in the isomerisation of citronellal to isopulegol increased with the concentration of defects resulting from modulation with

TFAA.[13] The selectivity for this reaction, an indicator of Lewis acid strength, did not vary between samples, which suggests that it was the density of active sites rather than their activity which was affected. This trend in activity has been shown to hold for defects resulting from modulation with other carboxylic acids and even from post-synthetic methods.[14–16] More effective still are cluster vacancies, which outperform linker defects in the Lewis acid catalysed isomerisation of glucose to fructose and the MPV reduction of *tert*-butylcyclohexanone.[5,17] This was also reported by Wang *et al.* for equally defective UiO-66 samples which had been modulated by either FA or TFAA.[18] The latter, known to lead to more missing clusters and confirmed in this case via PXRD,[19] was more effective in the conversion of cyclohexanone to cyclohexanol.

The aforementioned cases of Lewis acid catalysis by UiO-66 leave out an important detail: Zr is nominally coordinatively saturated, even when defects are present. It is possible to generate open sites via sample activation and dehydroxylation, but in both cases, this can be reversed by exposure to moisture (for example in air or as an impurity).[20–22] Vandichel *et al.* noted in a computational study on the conversion of citronellal to isopulegol that sensible transition states could only be found when the cap on a defect was either missing or significantly skewed so as to undercoordinate Zr.[23] Subsequent work by Caratelli *et al.* on linker dynamics in aqueous media suggests the latter is not unreasonable.[24] The solvent can allow for transient changes in Zr coordination, the exchange of capping molecules, and changes in linker bonding modes. Some examples are shown in **Figure 28**. Similar processes should also be possible for monocarboxylate caps. This therefore represents a plausible means through which coordinatively saturated active sites in UiO-66 can participate in catalysis.

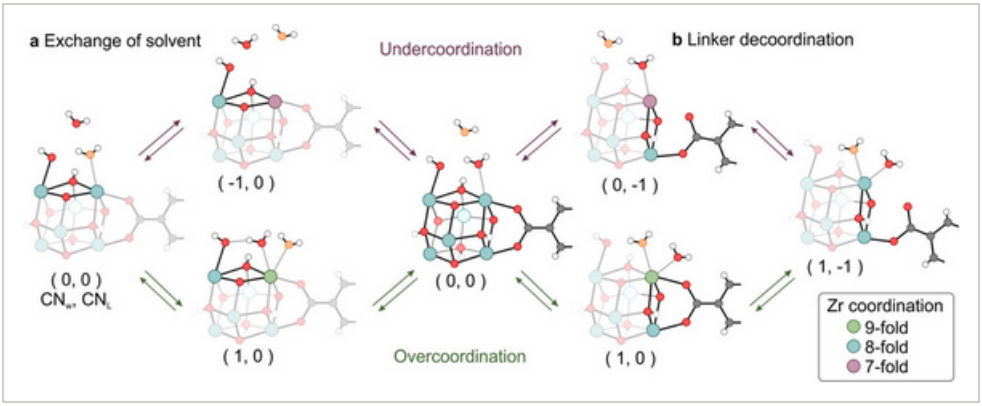


Figure 28: Over and under-coordination of Zr in water solvent found via metadynamics. A single H_2O–OH defect is initially shown, along with its possible evolutions over the course of the simulation. The dynamic nature of this capping has also been reported by Ling and Slater.[25] H, C, and O atoms are coloured white, grey, and (red, orange); Zr is coloured according to its neighbours and changes to the coordination numbers of highlighted metals are labelled via parentheses. Diagram taken from ref. [24].

The experiments conducted in this chapter were carried out to resolve questions prompted by the studies discussed. The full impact of the capping species on the Brønsted acidity of UiO-66's native μ_3-OH sites is not known: can this be tuned by changing the charge compensation schemes, and what acidic species are really present in the defective material? Enquiry is also possible with regards to Lewis acidity. Is the catalytic potential of an active site dependent on the capping species, or are experimental differences between modulators only due to the presence of missing SBUs? Finally, there remains the question of the accessibility of coordinatively saturated Zr; how can these be accessed, and does the manner in which this happens change depending on the cap? These objectives will be addressed through comparative DFT studies, wherein different defective models of UiO-66 will be evaluated across the same experiments.

4.2 Approach

4.2.1 Quantifying acidity

A well-established method for computationally estimating Brønsted acidity in porous materials is to insert an appropriate probe molecule, typically ammonia or pyridine, atop a site of interest and evaluate the resulting binding energy. Although this has been done mainly for acidic zeolites,[26,27] a MOF such as UiO-66 contains only a limited number of Brønsted acid sites, all of which are well-defined, such that this approach should remain effective. Initially, the atomic positions and unit cell of the target system are optimised, following which the probe molecule is inserted near the site of interest. A subsequent optimisation then allows the binding energy to be straightforwardly calculated via **Equation 95**:

$$E_{\text{binding}} = E_{\text{probe+framework}} - E_{\text{framework}} - E_{\text{probe}} \tag{95}$$

$E_{\text{probe+framework}}$ is the energy of the optimised framework-probe molecule system, while $E_{\text{framework}}$ is that of the optimised empty framework and E_{probe} is that of the optimised isolated gas-phase probe. A more negative binding energy is indicative of a stronger bond being formed with the framework proton, corresponding to higher Brønsted acidity, while a positive value reflects that adsorption is unfavourable. Different defect types can be tabulated by simply starting from the corresponding structures and depositing the probe atop one of the hydroxyls directly adjacent to the vacancy. Note that **Equation 95** yields only a proxy measure of Brønsted acidity as this approach does not model the exchange of a proton between an acid and a base. The overall bonding between the probe molecule and the framework is additionally dependent on steric factors as well as secondary dispersion and electrostatic interactions. During the comparison of different acid sites, it is necessary to keep these caveats in mind.

An alternative way of quantifying acidity in solid catalysts is to use the vibrational frequency of an adsorbed CO molecule. The approach is similar to that used for binding energy: the target system is optimised, then CO is introduced onto the site of interest, followed by another optimisation. Once this is done, the signature vibrational frequency of the adsorbate is estimated by calculating the full or partial Hessian matrix. Any changes in this frequency are the result of a balance between bonding and electrostatic effects, as summarised by Hadjiivanov and Vayssilov:[28]

- σ bonding

  The highest occupied molecular orbital (HOMO) of CO is slightly antibonding. As the formation of a σ bond in which the δ negative carbon end acts as an electron donor removes density from this orbital, the overall bond order increases and the frequency is blueshifted. Conversely, such a bond originating from the oxygen will remove electron density from a bonding orbital instead.

- π bonding

  It is also possible for the cation to donate some electron density from its π-symmetry orbitals into the lowest unoccupied molecular orbital (LUMO) of the probe, which is antibonding. This therefore lowers the bond order of CO and leads to redshift.

- Electrostatics

  The electric field exerted by the cation stabilises all of the orbitals on CO, increasing frequency. In addition, given that C and O will generally experience different field strengths, the charge distribution in the molecule will also be polarised. When carbon is closer to the cation, this stabilises the HOMO (which becomes either less antibonding or slightly more bonding) and leads to blueshift, whereas with oxygen, this is reversed.

The second of these effects is unlikely to be important in UiO-66 given that the relevant cations, H^+ and Zr^{4+}, do not possess filled or partially filled orbitals of π symmetry at the right energies. The other two factors will generally both be important and will increase the frequency of adsorbed carbon monoxide. Both reflect the ease with which a bond can be formed with a given cation and thus correlate with Brønsted or Lewis acidity. Note that the shift in frequency of the adsorbed CO is again only a proxy measure of acidity.

4.2.2 Functional benchmarking

Given the use of ammonia and pyridine as probe molecules, which will form hydrogen bonds with bridging hydroxyls, some benchmarking calculations were run to validate the choice of functional and basis set. The PBE and BLYP functionals, both GGAs, and PBE0, a hybrid, were tested using double (DZVP) and triple (TZVP) ζ basis sets on the preliminary array of acid sites shown in **Figure 29**. Note that the given the low cost of these calculations, the conventional unit cell was used for work on ammonia and pyridine.

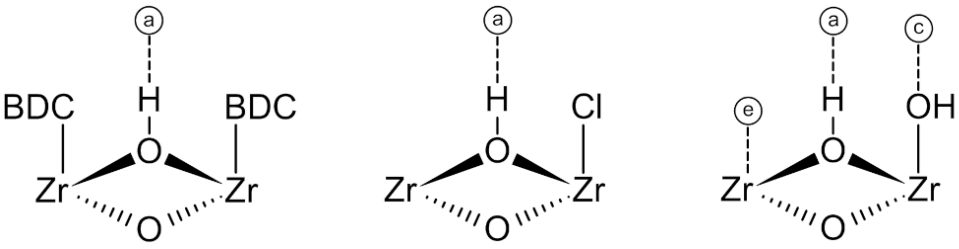


Figure 29: Acid sites used to benchmark the choice of functional and basis set. The structures depicted from left to right are respectively the pristine framework and linker vacancies terminated by Cl and OH. Sites **fcu**-a, Cl-a, OH-a, and OH-c were used for adsorption energies; **fcu**-a, OH-a, OH-c, and OH-e were used for adsorbed CO frequencies.

The corresponding binding energies for these sites, as defined in **Equation 95**, were evaluated using ammonia as the probe molecule and are shown in **Figure 30**. Note that for each functional and basis set, the results have been offset relative to **fcu**:

$$\Delta\epsilon = \Delta E_{\mathbf{fcu}} - \Delta E_{\text{site}} \tag{96}$$

$\Delta E_{\mathbf{fcu}}$ and ΔE_{site} correspond to the binding energies of ammonia on sites **fcu**-a and one of Cl-a, OH-a, or OH-c. Positive values imply stronger binding than in the pristine structure and vice versa for negative ones.

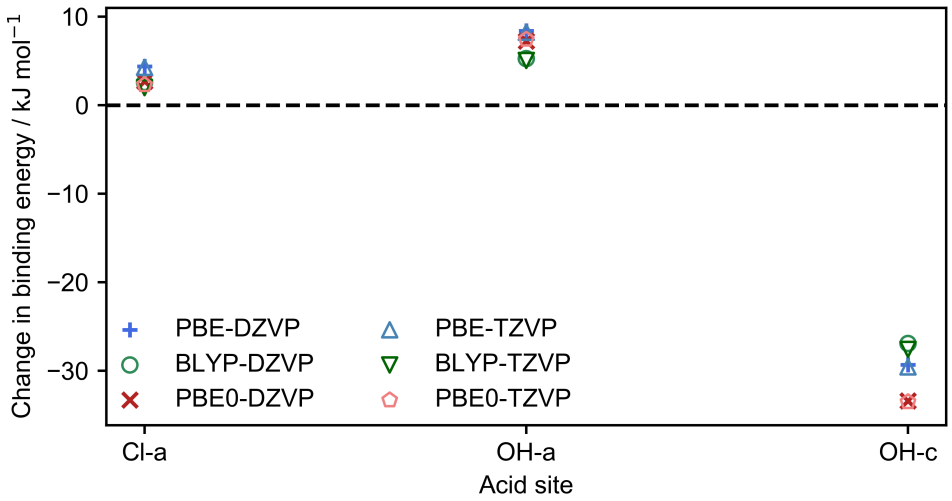


Figure 30: Benchmarking of different Hamiltonians for evaluating the binding energy between ammonia and a target acid site. The adsorption geometries correspond to those shown in **Figure 29**.

It appears that the binding energy between the probe and framework is generally insensitive to the choice of functional and basis set, with differences of only a handful of kJ mol^{-1}. There is greater spread for site c of the OH capping which suggests that for sites other than the native bridging μ_3 OH, more care may need to be taken. Nevertheless, all Hamiltonians yield the same qualitative trend for this subset of structures and the choice of PBE-DZVP is therefore suitable for this work.

Further benchmarking was carried out for the choice of functional and the choice of mobile atoms in vibrational calculations with carbon monoxide. To reduce the cost of these calculations, the half unit cell was used for work on CO rather than the conventional one. PBE and PBE0 were again tested with double and triple ζ basis sets; results for BLYP were found not to be internally consistent and are therefore not shown here. A different subset of acid sites were tested than for ammonia; these are also shown in **Figure 29**. The corresponding frequencies for CO were evaluated and are shown in **Figure 31**, along with the experimental data from Cirujano and Llabrés I Xamena.[8] The 2135 cm^{-1} peak assigned to physisorbed CO in the aforementioned study is taken here to correspond to CO adsorption on OH site c.

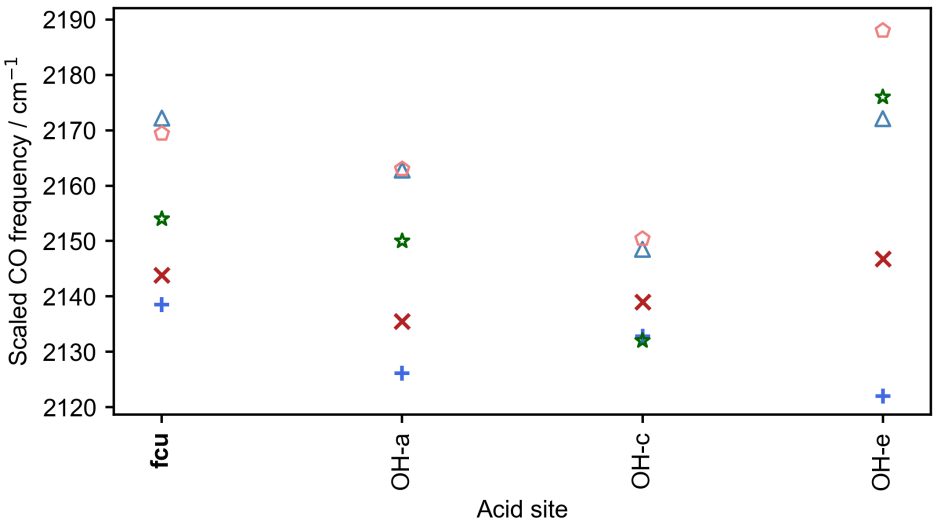


Figure 31: CO stretching frequencies with different methods on a chosen subset of acid sites as shown in **Figure 29**. PBE-DZVP, PBE-TZVP, PBE0-DZVP, PBE0-TZVP, and experimental results from [8] are respectively marked by blue pluses, light blue triangles, red crosses, pink pentagons, and green stars. Appropriate scaling factors have been computed and applied for each data series based on the frequency of isolated gas-phase CO.

This preliminary data suggests the previous choice of PBE-DZVP is not suitable for carbon monoxide frequencies. OH site c is expected to be the weakest acid site while the uncapped Zr atom of site e should be the most acidic. This is respected by the PBE0 functional for both the double and triple ζ basis sets, but only the latter also mirrors the experimental data. For this combination of functional and basis, frequencies are overestimated by about $\sim$15 cm^{-1}, but the qualitative trends are the same. PBE0-TZVP therefore appears to be a sensible choice of method for the evaluation of carbon monoxide frequencies.

For these partial vibrational calculations, minimal sets of mobile atoms were used. In any structure where adsorption took place atop a proton (via either a hydroxyl or water moiety), 4 atoms were allowed to move: carbon monoxide C and O, and acid site O and H. For adsorption directly onto Zr, 3 mobile atoms were used: carbon monoxide C and O, and the Zr to which the CO is bonded. This is the minimal set of atoms and is shown schematically in **Figure 32**.

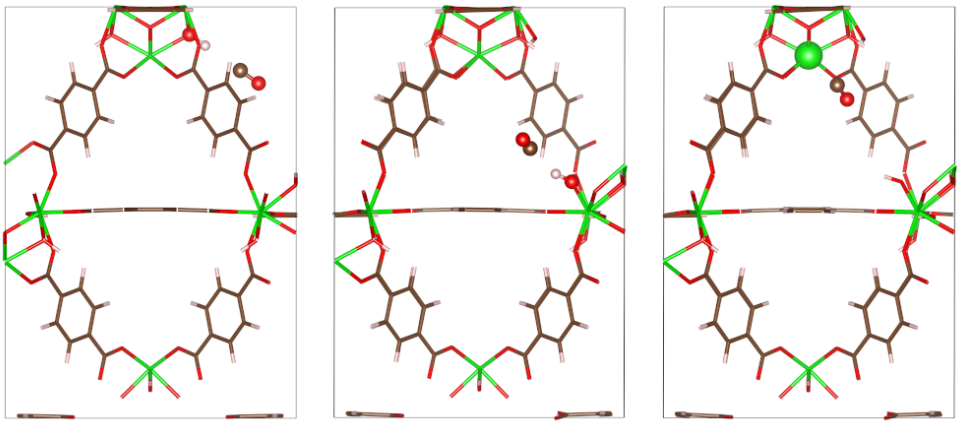


Figure 32: Mobile atoms in CO frequency calculations. From left to right: **fcu**-a, OH-c, and OH-e. Atoms free to move are drawn as spheres; for the rest of the structure, only the skeleton is shown.

This choice of mobile atoms was verified by comparing the resulting frequencies to those from calculations in which no atoms were held fixed. This was done for the **fcu**-a, OH-a, OH-c, and OH-e sites, the results of which are shown in **Table 10**.

Table 10: CO mobile atoms benchmark

Set of mobile atoms	CO scaled frequency / cm^{-1}			
	fcu-a	OH-a	OH-c	OH-e
Full	2169.9	2163.2	2150.6	2188.4
Minimal	2169.6	2163.2	2150.6	2188.2

Stretching frequency of CO on selected sites with the PBE0 functional and TZVP basis sets. Different numbers of atoms were held fixed during the vibrational calculations corresponding to those shown in **Figure 32**. A scaling factor of 0.967 was applied uniformly.

For the sites tested here, there are no differences between the minimal and full sets of mobile atoms. It appears that the stretching vibration of CO is decoupled from the framework's other normal modes. The minimal scheme will therefore be used for all partial vibrational calculations involving carbon monoxide.

4.2.3 Energy barriers

A final set of calculations was carried out in this study to determine the free energy barriers of two processes: the opening up of a Lewis acid site and the cyclisation of citronellal to isopulegol. This was done using the half unit cell in order to reduce computational cost. For each process across each structure, this involved 3 steps:

1. Low-temperature MD was used to generate a trajectory from the reactant to the product. Practically, this was done via the Plumed package by using a set of distance-based collective variables along which an energy penalty was progressively applied.[29] In the case of cyclisation, the transition step is already known from previous studies,[2,23,30] such that the trajectory was also made to pass through this extremum.

2. Frames were extracted from the MD trajectory to initialise a NEB calculation so as to identify a minimum energy pathway. This was combined with the dimer method to refine the transition state.

3. Partial vibrational calculations were then carried out on the 3 optimised extrema. The Tamkin package was then used to carry out PHVA and evaluate free energy differences between the initial, transition, and final states.[31]

The exact collective variables used in the first step are schematically detailed in **Figure 33** and **Figure 34** for a formate-terminated defect.

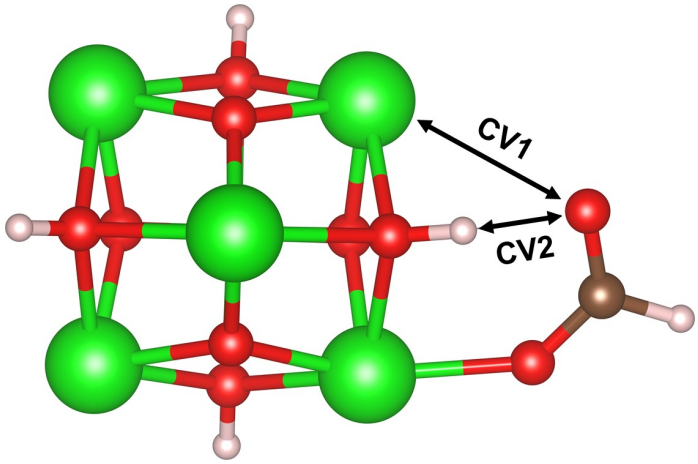


Figure 33: Schematic depiction of collective variables used in MD to generate frames for the dangling step. CVs 1 and 2 are respectively the distance between: one of the carboxylate oxygens and its nearby zirconium; and the same oxygen and the adjacent hydroxyl proton. The former was increased from its equilibrium value, while the latter was decreased.

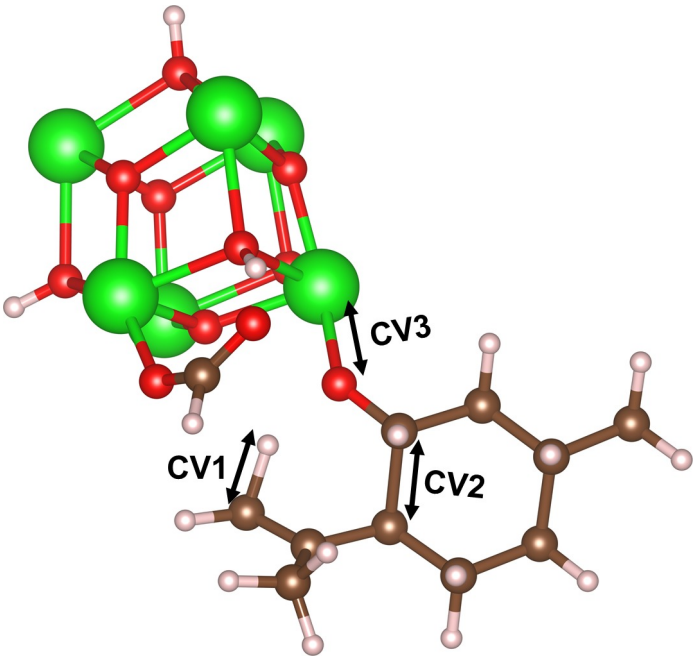


Figure 34: Schematic depiction of collective variables used in MD to generate frames for the cyclisation step. CVs 1, 2, and 3 are respectively the distances between: one of the CH_3 protons and its carbon; two of the carbon atoms in the 6-ring; and the isopulegol oxygen and its zirconium. The first two have been increased from their equilibrium values, while the third was held approximately constant.

For step 3, partial vibrational analysis was carried out on extrema from the dangling and cyclisation processes. For the first part, all atoms from the carboxylate, along with those of the adjacent μ_3 OH, were allowed to move. In addition to these, all atoms belonging to the adsorbate were also mobile during the second step. This is shown schematically for the pristine structure in **Figure 35**.

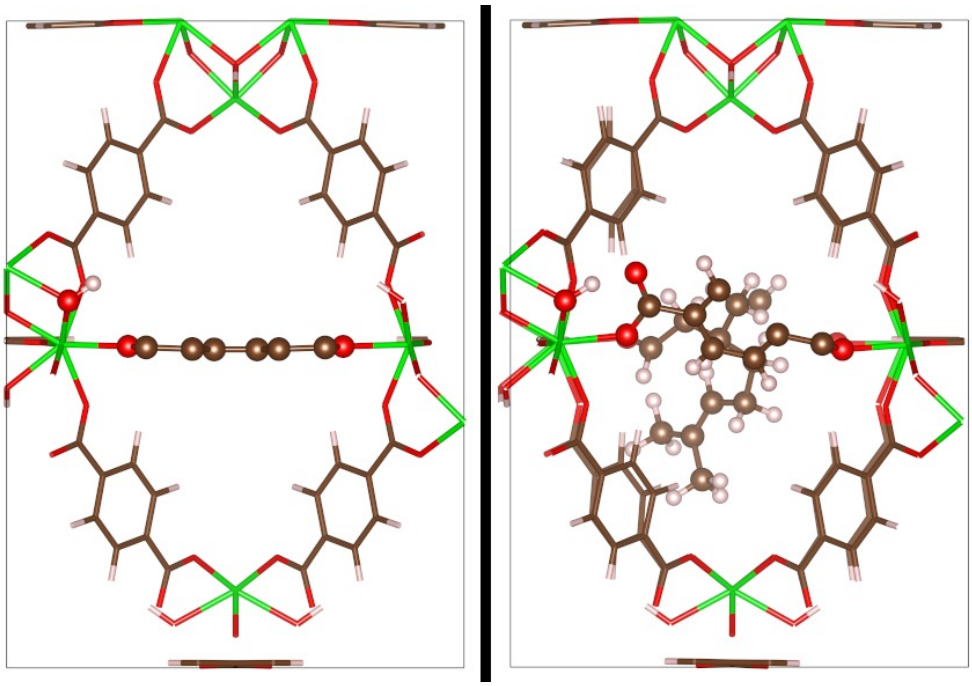


Figure 35: Schematics for the atoms left mobile during partial vibrational calculations on the **fcu** system for the dangling (left) and cyclisation (right) steps. For atoms which have been fixed during this analysis, only the framework is shown.

4.2.4 Stereoisomers

Citronellal has a chiral centre and therefore two enantiomers. Isopulegol, the cyclisation product, has four diastereoisomers, each of which has two enantiomers (three chiral centres leading to eight total stereoisomers). These are detailed in **Figure 36**.

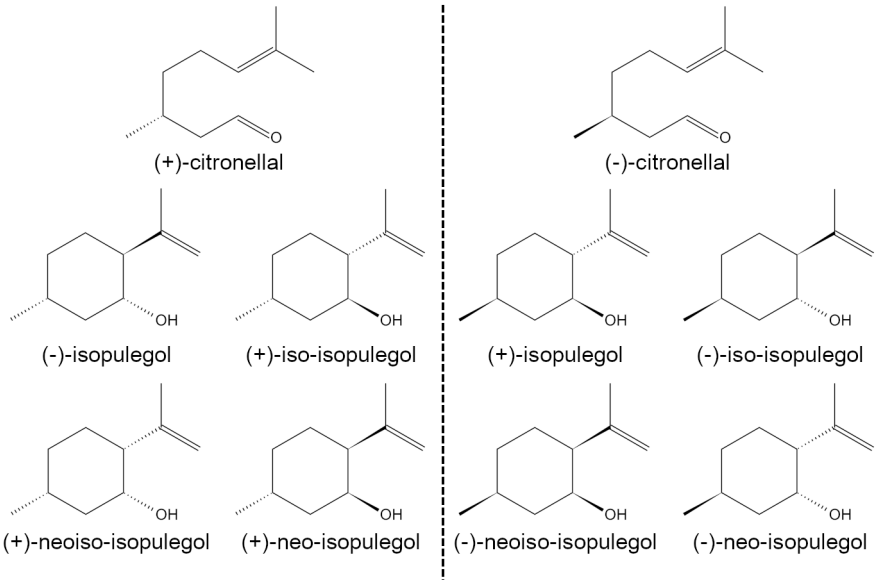


Figure 36: Stereoisomers of citronellal and isopulegol. With one chiral centre, citronellal has two enantiomers. With 3 chiral centres, isopulegol has 4 diastereoisomers, each of which has two enantiomers. The cyclisation of a given enantiomer of citronellal will yield a mixture of the corresponding 4 diastereoisomers of isopulegol.

Experimental and computational studies involving solid Lewis acids typically focus specifically on the conversion of (+)-citronellal to (-)-isopulegol. One reason for this is that the selectivity of the product diastereoisomer is an indicator of the type of acidity that the material exhibits. This cyclisation can be enabled by a proton transfer from a Brønsted acid, in which case the reaction yields an even mix of diastereoisomers.[32] When a Lewis acid is used, selectivity for both enantiomers of isopulegol is markedly higher than for the other products.[33] A further reason to focus on (-)-isopulegol specifically is that this is a precursor to commercially relevant (-)-menthol.[34] It is therefore sensible to focus in this work on the MEP between (+)-citronellal and (-)-isopulegol, which has been done notably by Vermoortele *et al.* and Vandichel *et al.*[2,23]

4.2.5 Specific computational details

The majority of calculations in this chapter were carried out using the parameters described in **Section 2.6**. For work involving ammonia and

pyridine, probe molecules were inserted atop the sites of interest in the defective unit cells optimised in **Chapter 3**. Following this, atomic positions only were relaxed. Additional corrections for BSSE were applied in a *post hoc* manner to probe binding energies following the counterpoise method of Boys and Bernardi.[35]

Significantly different computational parameters were used for calculations involving carbon monoxide. The optimal configurations of CO atop acid sites of interest were determined in the same manner as described above for ammonia and pyridine. The truncated Coulomb operator variant of the PBE0 functional was used together with D3 dispersion corrections and GTH pseudpotentials.[36–39] Higher-quality triple–ζ TZVP-MOLOPT-SR-GTH and TZVP-MOLOPT-GTH basis sets were used for Zr and all other atoms respectively,[40] together with a 1200 Ry plane wave cutoff. The Hartree-Fock exchange energy was determined through the auxiliary density matrix method (ADMM), within which Zr and other atoms were modelled via FIT11 and pFIT3 auxiliary basis sets. The same SCF and geometric convergence criteria were used except for maximum force, which was tightened to 1.0×10^{-5} Ha Bohr^{-1}. The frequency of gas-phase carbon monoxide was evaluated through a separate calculation inside a cubic box of edge length 20 Å and compared to the experimental value to generate a scaling factor. The resulting value of 0.967 for PBE0-TZVP is in line with published figures and was applied uniformly to all corresponding CO results.[41] Such uniform scaling is justified by the work of Ma *et al.*, who showed that site-specific scaling for adsorbed CO increases computational expense and complexity for only minimal gains in accuracy.[42] When evaluating the frequency of adsorbed CO, the majority of framework atoms were held fixed as described in **Section 4.2.2**

For work on the conversion of citronellal to isopulegol, the parameters described in **Section 2.6** were used for the functional, basis sets, and semi-empirical corrections. AIMD runs were carried out using a

timestep of 0.5 fs at a temperature of 5 K regulated by the CSVR thermostat (canonical sampling through velocity rescaling).[43] Nudged elastic band calculations were run initially with the improved-tangent algorithm and subsequently in climbing-image mode,[44,45] using a spring constant of 2.0×10^{-2} Ha Bohr^{-2} and either 7 or 9 images. Cell parameters remained frozen during these simulations, during which convergence criteria of 10^{-8} Ha, 4.5×10^{-4} Bohr, 4.5×10^{-4} Ha Bohr^{-1}, 4.5×10^{-4} Bohr, and 1.0×10^{-4} Ha Bohr^{-1} were respectively applied for the energy change between subsequent SCF steps, RMS displacement, RMS force, maximum displacement, and maximum force. ORCA was used with the same settings listed in **Section 2.6** to evaluate strain in twisted linkers.

4.3 Results and discussion

4.3.1 Adsorption of ammonia and pyridine

The Brønsted acid site in pristine UiO-66 is the μ-3 bridging hydroxyl group on the SBU. Removing an adjacent linker and terminating it with capping species can introduce additional Brønsted acidic moieties and alter the surrounding chemical environment. These possible geometries, which correspond to those discussed in **Chapter 3**, are shown in **Figure 37**. Ammonia has been deposited atop each Brønsted acid site (types a, b, c, and d) for the different charge-compensation schemes, following which atomic positions have been optimised. Note that the defect of type **e** has been omitted as charge compensation in this case destroys the adjacent μ_3 OH groups. The resulting binding energies, evaluated using **Equation 95**, are shown in **Figure 38**.

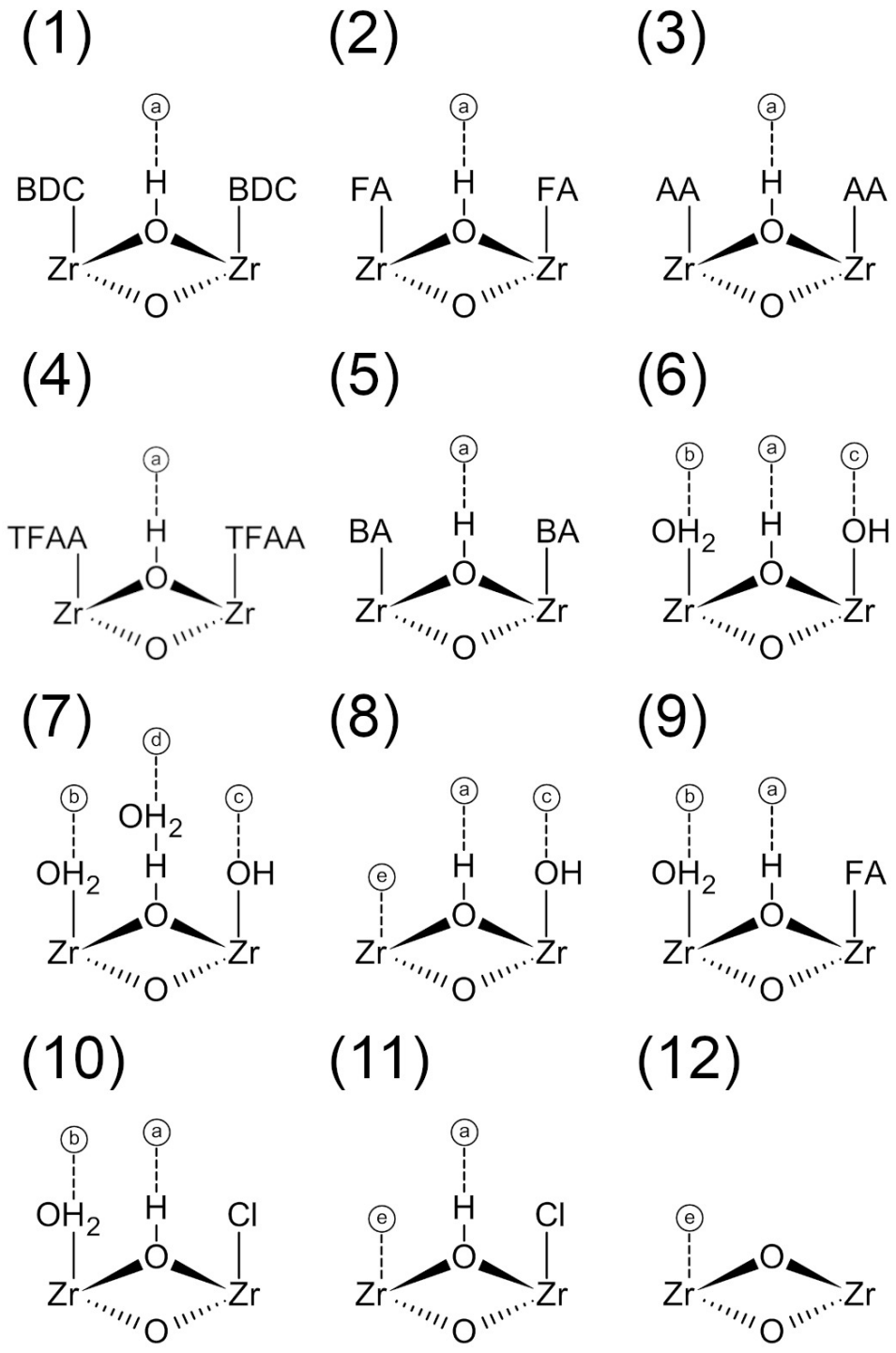


Figure 37: The different Brønsted and Lewis acidic sites available in UiO-66 depending on the capping species. Panel 1 depicts the pristine structure. The remaining panels show missing linkers terminated respectively by: FA, AA, TFAA, BA, H_2O–OH, $2\,H_2O$–OH, OH, FA–H_2O, Cl–H_2O, Cl, and nothing (charge compensation achieved by removing proton from adjacent μ_3 hydroxyl, see **Chapter 3** '**e** defect').

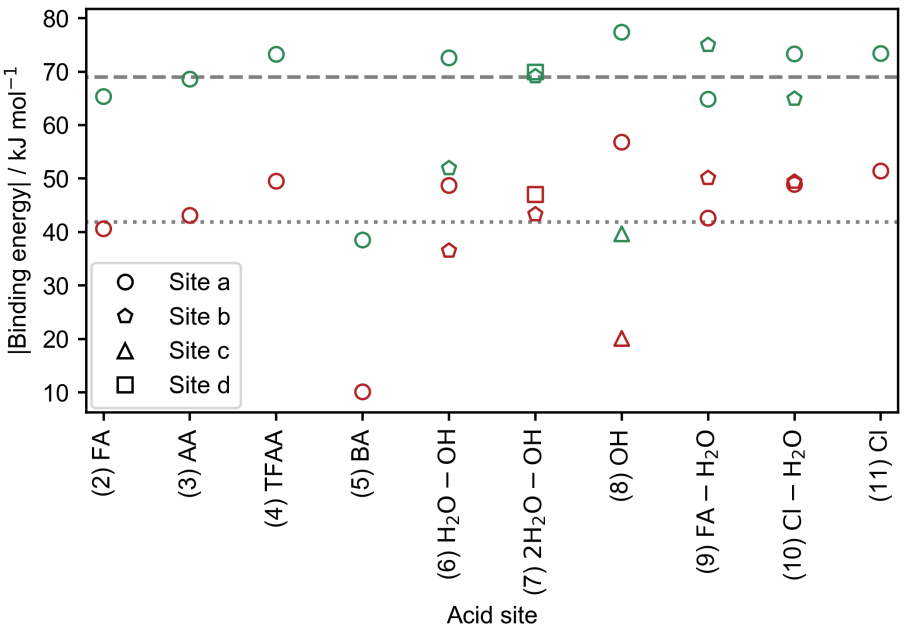


Figure 38: Binding energies of ammonia in defective UiO-66. Different acid sites are marked by different marker shapes as labelled in the legend; these correspond to those shown in **Figure 37**. Green markers correspond to the full DFT binding energy (electronic + D3 dispersion), while red ones represent only the electronic component. The **fcu** results for these datasets are respectively marked by dashed and dotted lines.

The difference in Brønsted acidity between sites leads to a range of probe adsorption energies. However, dispersion forces, steric hindrance, and secondary interactions also strongly influence the trends seen in **Figure 38**. An ammonia molecule sitting on a site of type a experiences weaker dispersion interactions when one of the nearby linkers has been replaced; this leads to weaker binding without necessarily reflecting on Brønsted acidity. The DFT binding energy can be separated into an electronic component and a dispersion part:

$$\Delta E_{\mathrm{DFT}} = \Delta E_{\mathrm{electronic}} + \Delta E_{\mathrm{D3}} \tag{97}$$

ΔE_{D3} can therefore be removed to leave behind the pure DFT binding energies, which are also shown in **Figure 38**. It can be seen in this way that while dispersion is a major contributor to the bonding between ammonia and the framework, it does not account for variation between species and instead represents a constant ∼30 kJ mol^{-1}. Taking a look at the monocarboxylates, which bond to Zr in a similar manner as BDC does, FA, AA, and TFAA are comparable to **fcu**, while BA leads to much poorer adsorption. This can be attributed to steric hindrance caused by the bulky benzoate groups, which prevents the probe from occupying a favourable position above the μ_3 hydroxyl acid site. This can be seen in the corresponding structural schematic, shown in **Figure 39** panel a.

The introduction of additional functional groups through the capping species is responsible for the variation across the remaining defects. The presence of a non-native hydroxide or a chloride anion leads to a minor enhancement of the binding energy on sites of type a, as seen for H_2O–OH, OH, Cl–H_2O, and Cl. In these cases, ammonia reorients itself so as to form a secondary hydrogen bond with either the hydroxide oxygen or the chlorine from the capping species, as shown in **Figure 40**. Non-native hydroxide anions also represent acid sites which are not present in the standard framework, but these are generally poor, as seen for site OH-c. The orientation of the O–H bond in this defect is such that ammonia is sterically hindered by the nearby linker. The hydroxide is able to re-orient to alleviate this, as shown in **Figure 39** panel b, but the end result is a strained structure in which NH_3 occupies a skewed off-axis position. While site c also exists for H_2O–OH and 2 H_2O–OH, during optimisation with ammonia, an internal proton transfer takes place such that site c and site b are exchanged, as depicted in **Figure 39** panel d. This also evidences the poor acidity of these added hydroxyl groups.

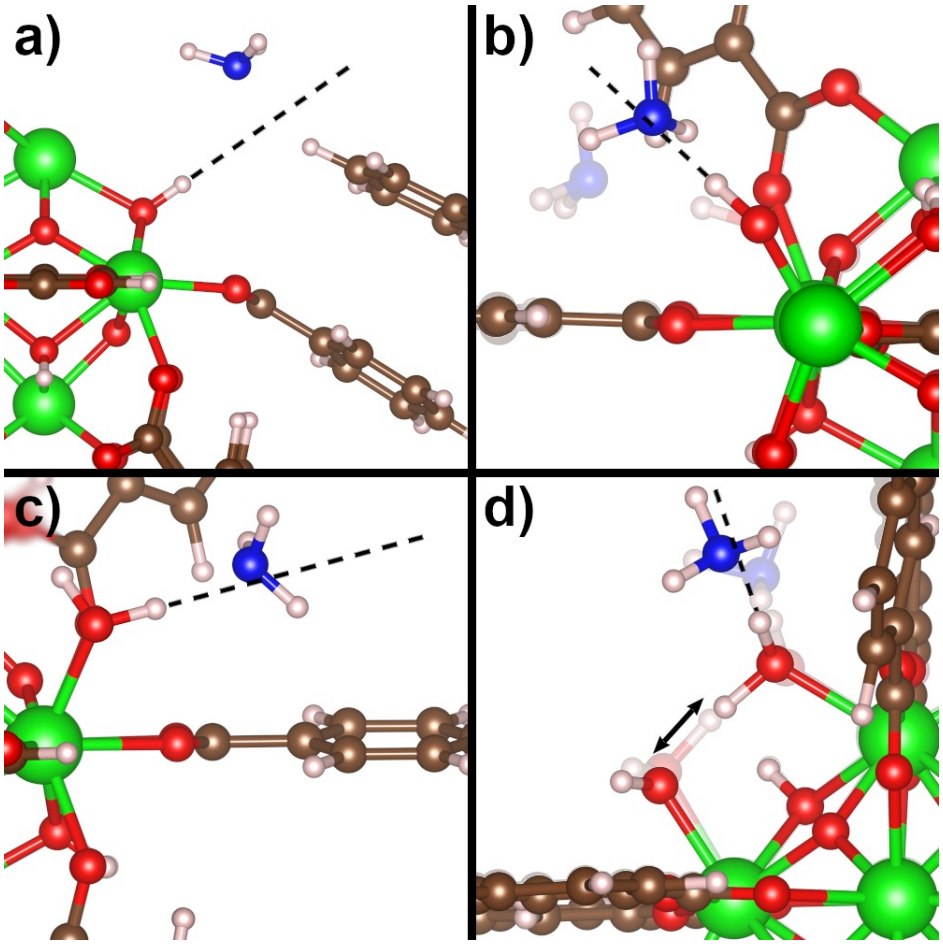


Figure 39: Irregular binding geometries of ammonia near certain defects. (a) Ammonia is forced far from site BA-a due to nearby benzoates. (b) Re-arrangement seen between starting (transparent) and final (opaque) geometries during optimisation for ammonia atop site OH-c. (c) Nearby linker becomes twisted and pushes NH_3 away from O–H axis at site H_2O–OH-b. (d) Re-arrangement seen between starting (transparent) and final (opaque) geometries during optimisation for ammonia atop sites H_2O–OH-c and 2 H_2O–OH-c. Some atoms have been obscured to enhance visibility; O–H axes have been highlighted by dashed lines.

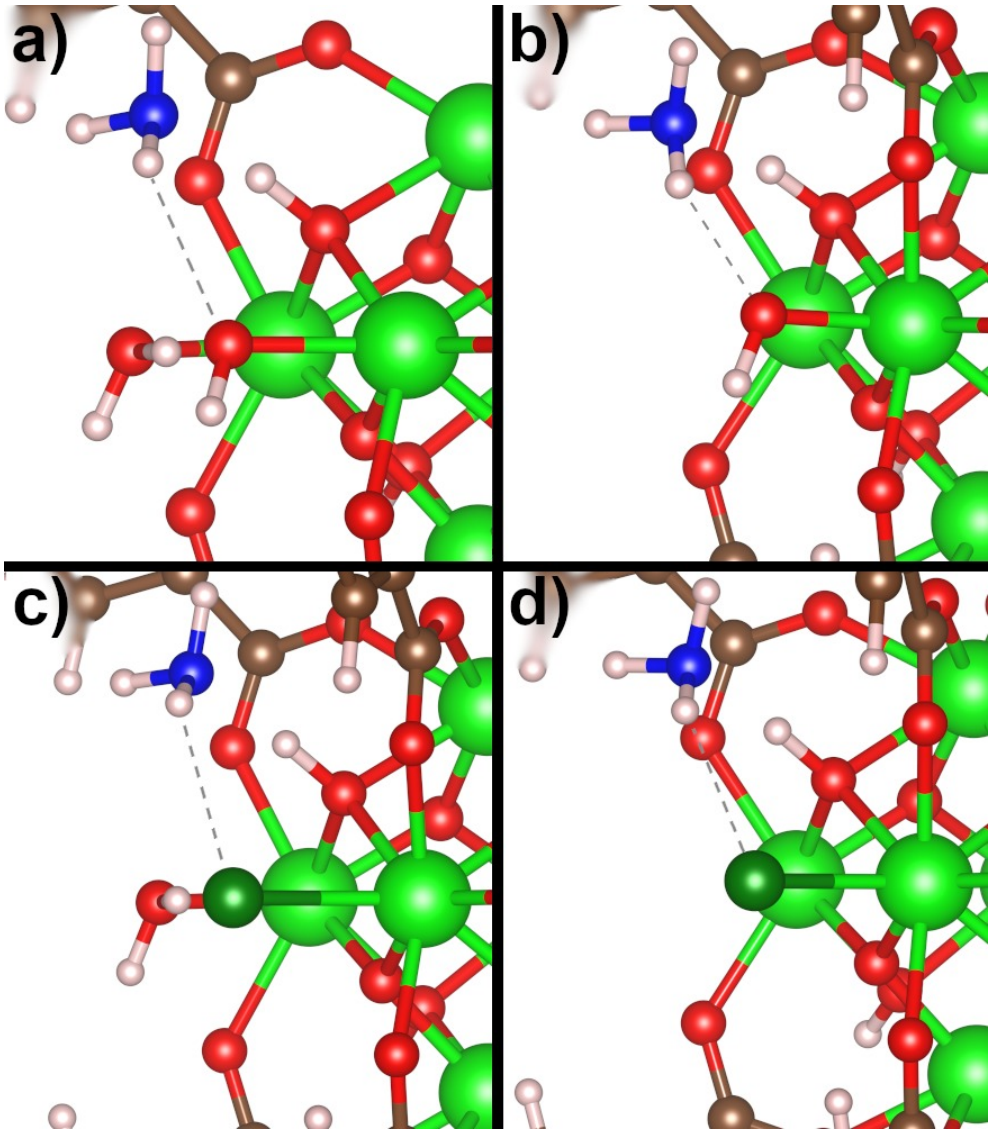


Figure 40: Secondary hydrogen bonds between ammonia and nearby capping species. Panel a): ammonia sitting atop site H_2O–OH-a with a secondary hydrogen bond to nearby hydroxide anion. b): similar case to a) but with ammonia atop site OH-a. c): NH_3 atop site Cl–H_2O-a with a secondary hydrogen bond to nearby chloride anion. d) similar case to c) but with ammonia atop site Cl-a. Hydrogen bonds have been highlighted by dashed lines; some atoms have been hidden for clarity.

Additional water molecules also present further acid sites for the capping species H_2O–OH, $2\,H_2O$–OH, FA–H_2O, and Cl–H_2O. On the first of these, the adsorption of ammonia is weak because of a similar reason to OH-c: the orientation of water is such that a nearby linker leads

to steric hindrance, as depicted in **Figure 39** panel c. For the other structures, water is differently aligned such that this issue does not arise and binding at sites of type b and d is comparable to that of the native hydroxyl. These results with water and hydroxide anions notably mean that, through a judicious choice of the synthesis and activation conditions, it should be possible to increase the density of Brønsted acidic active sites, as was found experimentally by Cirujano and Llabrés I Xamena.[8]

These experiments have been repeated using pyridine, a stronger and bulkier base than ammonia. The results are shown in **Figure 41**.

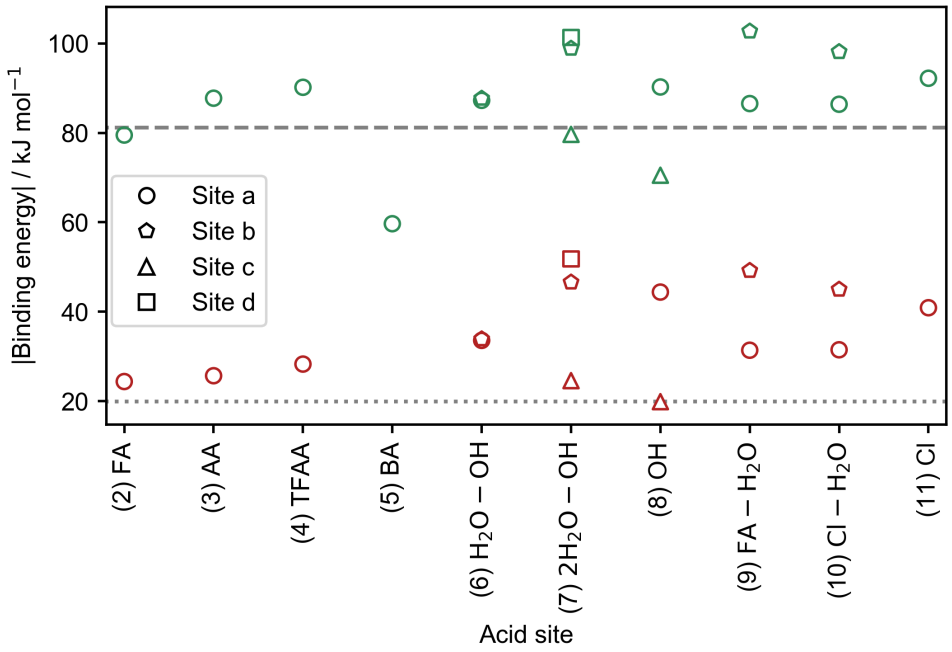


Figure 41: Adsorption of pyridine on different acid sites. Different acid sites are marked by different marker shapes as labelled in the legend; these correspond to those shown in **Figure 37**. Green markers correspond to the full DFT binding energies, while red ones represent only the electronic component. The **fcu** results for these datasets are respectively marked by dashed and dotted lines. The dispersion-corrected binding energy for BA is now positive (repulsive) and is therefore not shown here.

While this new probe is unable to form secondary hydrogen bonds with the framework, it is possible for there to be π-π overlaps between aromatic systems.[46] Indeed for almost all acid sites, pyridine orients itself such that the plane of the molecule is almost parallel to that of an adjacent terephthalate linker, as shown in **Figure 42**. This forms a major part of the binding energy for this probe, as can be seen from the uniform and significant drop between the full DFT and dispersion-corrected datasets in **Figure 41**. The pure electronic component for the pristine framework is now only 20 kJ mol^{-1}, compared to $\sim$40 kJ mol^{-1} previously with ammonia. Given that pyridine is a stronger base, this suggests that this probe is too large to optimally occupy the μ_3 OH acid sites. Examining the equilibrium geometries confirms this: the average N–H bond length increases by 0.1 Å when using pyridine as a probe molecule, which is indicative of more steric hindrance. The replacement of a linker with any species other than BA can also be seen to lead to an increase in the pure electronic component of the binding energy, which also reflects that the more open structure enables stronger adsorption.

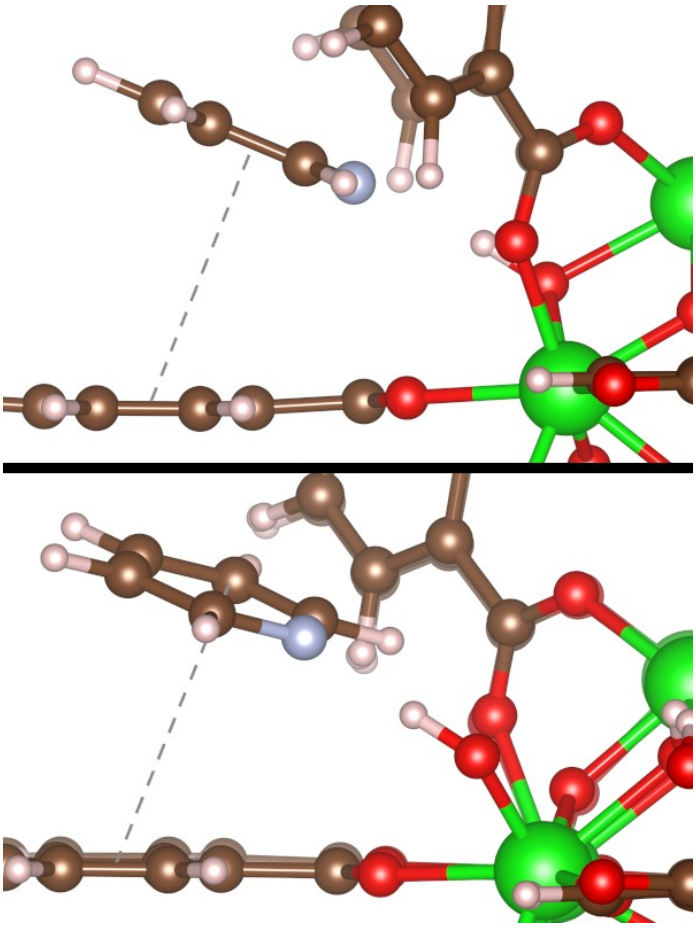


Figure 42: Alignment of pyridine with a nearby terephthalate linker. Top: pyridine adsorbed on site **fcu**-a. Bottom: pyridine adsorbed on site OH-c. The π-π interactions have been highlighted by dashed lines. In these defective structures, the distance between the molecular centroids ranges from 3 to 4 Å, which is in line with the optimal separation in benzene-pyridine dimers.[46]

Qualitatively, adsorption near defects capped by monocarboxylic acids has not changed except for BA. As a result of pyridine's large size, binding at site BA-a is now entirely dominated by dispersion, while the electronic component is now positive (repulsive). Small differences can also be seen for some caps containing hydroxide anions or water. An internal proton transfer no longer takes place when the probe is deposited atop site 2 H_2O–OH-c (though this still happens for H_2O–OH-c). This added hydroxide, along with that of site OH-c, are now almost comparable to the μ_3 OH of the pristine framework. The orientation of the O–H bond

is in these cases less of an issue as the close proximity of a nearby linker is in part compensated for by a π-π overlap, as depicted in **Figure 42**. This also contributes to the enhancement seen for all sites of type b seen in H_2O–OH, $2\,H_2O$–OH, FA–H_2O, and Cl–H_2O, for which this strong secondary interaction leads to improved adsorption.

While the method used here has shown that the presence of defects can lead to variation in binding energies at acidic protons, it is difficult to correlate this with changes in Brønsted acidity. It has been found that secondary interactions influence both the geometry and strength of adsorption, but these cannot easily be disentangled from differences in proton lability. In most cases, changes in the binding energy relative to that of the pristine framework are small and comparable to thermal energy. However, it has also been possible in this way to evidence how some capping species can contribute additional acid sites that are comparable to those present in the standard framework. Defect engineering could therefore be used to increase the density of active sites for Brønsted acid catalysis. While this was to some extent already known from the work of Cirujano and Llabrés I Xamena,[8] the analysis here has extended this to other capping species including FA–H_2O and Cl–H_2O. Additionally, these results have highlighted in a number of cases that binding is strongly influenced by secondary interactions between the probe and the framework. Dispersion forces, hydrogen bonding, and electrostatics will play an active role in catalysis by helping to immobilise molecules at the active site or by promoting certain conformations. This finding is therefore also useful for catalyst design, as activity may be indirectly tuned by modifying the defect caps.

4.3.2 Carbon monoxide frequencies

Although some inferences can be made about how Brønsted acidic a given site is using the binding energy of a probe molecule, it is difficult to disentangle contributions from different secondary interactions. An alternative means of probing acidity which has been successfully used for other porous materials is to deposit a carbon monoxide probe and instead evaluate its signature C–O stretching frequency. As this probe is less susceptible to subsidiary interactions with the framework, this method should lead to proxy acidity measures which are easier to interpret. Furthermore, CO is more suitable for binding onto unsaturated Zr atoms, so the Lewis acidic site e (panels 8, 11, and 12 in **Figure 37**) can additionally be sampled for completeness. The full set of capping species and acid sites has been evaluated using this method; the respective frequencies are presented in **Figure 43**.

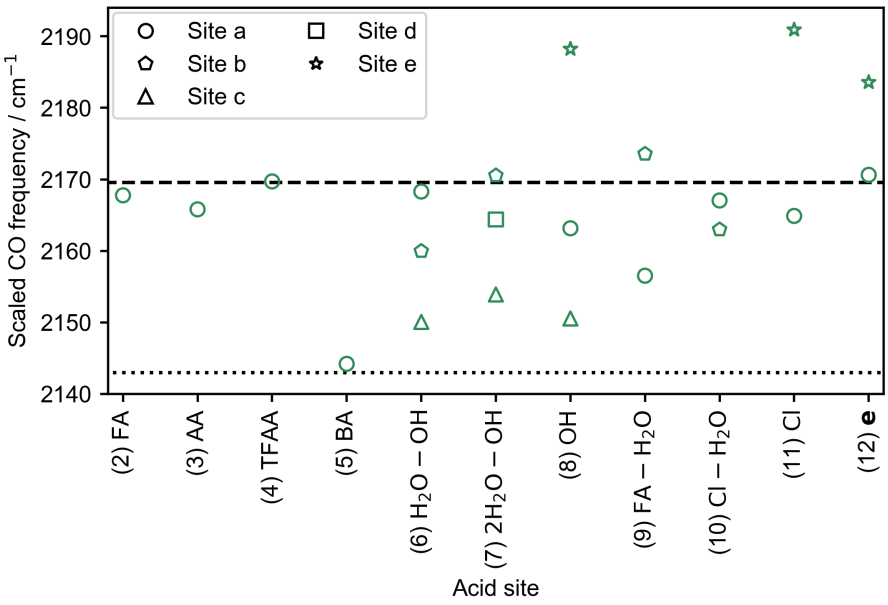


Figure 43: Stretching frequencies of CO bound on different acid sites. The frequencies of CO in vacuum and on the hydroxyl in **fcu** are denoted by dotted and dashed lines respectively. A scaling factor of 0.967 has been applied uniformly to all frequencies.

The results seen for carboxylic acids are in line with those found with ammonia and pyridine: sites FA-a, AA-a, and TFAA-a are broadly comparable to **fcu**-a. Steric hindrance remains a significant obstacle to accessing the acid site near a BA-terminated linker in spite of the small size of CO. There is essentially no binding between the probe and the μ_3 hydroxyl, as reflected by a stretching frequency is almost the same as the gas phase value. This also confirms that this metric is less sensitive to dispersion forces; with ammonia and pyridine the binding energy remained significant for this defect.

The results for the other caps are more difficult to interpret. Nevertheless, these trends can be rationalised by closely examining the structures. Some useful geometric descriptors are the distance between the CO carbon and the adsorption site oxygen, as well as the angle subtended by the CO carbon, the acidic proton it is bonded to, and the oxygen to which that proton is attached. These parameters are illustrated in **Figure 44**.

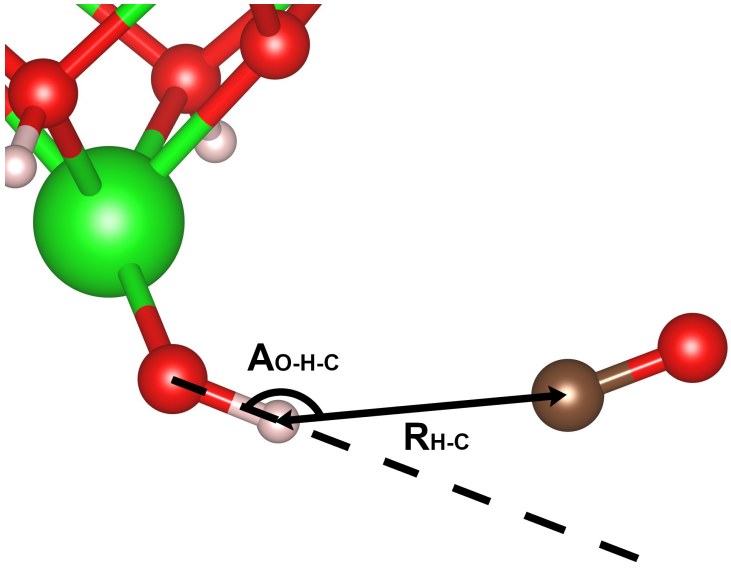


Figure 44: Schematic of geometric parameters used to describe CO adsorption. For carbon monoxide optimised in the pristine framework, the H–C separation ($R_{\text{H-C}}$) is 3.14 Å and the obtuse angle between the oxygen, proton, and carbon ($A_{\text{O-H-C}}$) is 179.9 °.

Some of this structural information is presented in **Figure 45** for the non-carboxylate capping species; tabulated values can also be found in **Appendix D**. In each case, it is more convenient to take the deviations in these parameters relative to those of the pristine framework. Close proximity to the adsorption site proton and good alignment with the O–H bond generally leads to blueshifts close to that of **fcu**, as is the case for sites H_2O–OH-a, $2\,H_2O$–OH-b, $2\,H_2O$–OH-d, FA–H_2O-b, Cl–H_2O-a, and Cl-a. Larger deviations in either $R_{\text{H-C}}$ or $A_{\text{O-H-C}}$ are associated with lower frequencies, as can be seen for sites H_2O–OH-c, $2\,H_2O$–OH-c, and OH-c. It is likely that steric hindrance from nearby linkers, which was invoked to rationalise the poor binding of ammonia for the same sites, reduces the accessibility of these protons. In addition to this, electrostatic interactions are possible between the negative charge on benzene or carboxylate groups and the δ^+ carbon monoxide oxygen atom when these are too close. Such charge effects are known to be destabilising to the bond in CO.[47] This is likely to also be important for sites OH-a and FA–H_2O-a, for which steric hindrance is minimal and the deviations in bond length and angle are modest, but which nevertheless exhibit lower frequencies. Finally CO adsorbs poorly on sites Cl–H_2O-b and H_2O–OH-b, as marked by large changes in $A_{\text{O-H-C}}$. Here also it is likely that a combination of steric hindrance and electrostatic interactions are responsible for the drop in frequency relative to site **fcu**-a.

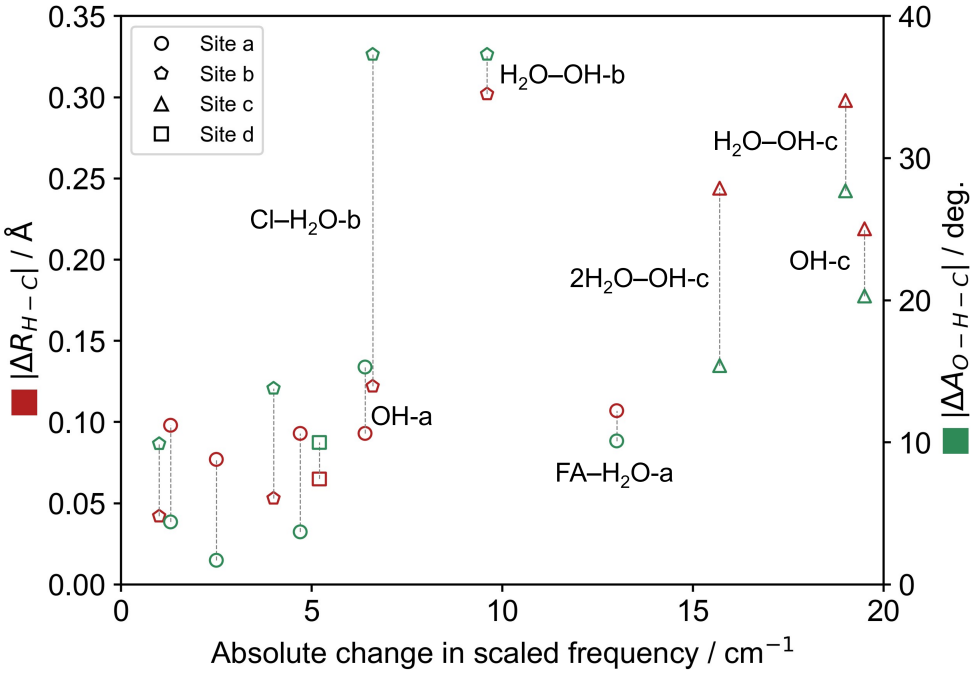


Figure 45: Correlation between CO adsorption geometry and stretching frequency. Deviations in the values of $R_{\text{H-C}}$ are marked in red and scaled by the left axis while those in $A_{\text{O-H-C}}$ are marked in green and scaled by the right axis. Site **fcu**-a has been used as the reference from which these differences have been calculated. Each structure has a pair of points on this plot (a bond and an angle); these have been connected together by dashed lines. Some sites of interest have been individually labelled.

Finally, sites of type e can be briefly commented on. The undercoordinated zirconium sites OH-e, Cl-e, and **e**-e all induce a significant blueshift of about 20 cm^{-1} in the CO frequency, which can be attributed to the strong electric field around these 4+ cations. In contrast to the red-shifting interactions described earlier, this effect strengthens the C–O bond.[47] This is consistent with the experimental data from Cirujano and Llabrés I Xamena and confirms the importance that these sites will play in Lewis acid catalysed reactions.[8] It should also be pointed out that the frequencies for these 3 structures follow the ordering Cl-e > OH-e > **e**-e. While covering only a few wavenumbers, this trend nevertheless reflects the Mulliken charge on the zirconium to which CO is bonded to: 0.952, 0.935, and 0.923 respectively. This is in agreement with the findings of

Vermoortele *et al.* who found that electron-withdrawing side groups on the linker in UiO-66 increase the Lewis acidity of zirconium atoms as reflected by increases in catalytic activity.[2]

Overall, the data suggests that the scope for tuning the Brønsted acidity of UiO-66 via defects may be limited. This analysis is complicated by secondary interactions: the proxy measures used here are also sensitive to steric hindrance as well as dispersion and electrostatic interactions. Nevertheless, adsorption sites can broadly be collected into 3 categories: highly Lewis acidic (OH-e, Cl-e, and **e**-e), poorly Brønsted acidic (BA-a, H_2O–OH-c, 2 H_2O–OH-c, OH-c, and FA–H_2O-a), and moderately Brønsted acidic (all other sites). It is possible to introduce additional labile protons to the structure, which are at best comparable with those of the framework, by capping defects with water. This can be controlled through the synthesis conditions, though care would need to be taken during activation so as to avoid the OH termination at the expense of H_2O–OH and 2 H_2O–OH.[4] The results for OH-e and Cl-e also show that Lewis acidic sites can be generated without necessarily destroying nearby Brønsted acidic protons, as would be the case for example during framework dehydroxylation. This is of particular interest for reactions which require a cooperative effect between both types of sites, such as the esterification of levulinic acid.[3,14] Nevertheless, if the acidity of protons needs to be increased for some given catalytic reaction, this cannot be done through defects and instead requires the functionalisation of the linkers or nodes.

4.3.3 Cyclisation of citronellal

The analysis of adsorbed CO frequencies presented previously suggests that there is limited potential for tuning the Brønsted acid strength of native sites in UiO-66 via defects. However, every cap involving a bare Zr atom exhibited a greatly enhanced stretching frequency, providing a way through which at least Lewis acidity may be enabled through linker

vacancies. Indeed, this is in line with numerous experiments which have reported trends between Lewis acid catalytic activity and defectivity in this material.[13,15,48] This has typically been ascribed to undercoordinated zirconium atoms generated through harsh activation conditions, leading to active sites similar to those of type e (**Figure 37** panels 8, 11, and 12) studied here. However, the framework can remain active even when it is not heated at sufficiently high temperatures to remove all capping species, suggesting that zirconiums atoms remain accessible even when fully coordinated.[8] It is also expected that over the course of a catalytic cycle, Zr coordination spheres may be transiently completed by solvent molecules when not occupied by the substrate. Therefore, it would be useful to know how the identity of the capping species influences the accessibility of the active sites along with its direct effects on Lewis acidity for designing more effective catalysts.

Different vacancies and capping species can be compared for Lewis acid catalysis by evaluating the energy barriers for a model reaction across a set of defective structures. The cyclisation of citronellal to isopulegol is a good choice in this regard as it is often used experimentally to probe the Lewis acidity of UiO-66.[2,8] Additionally, the transition state in the conversion of (+)-citronellal to (-)-isopulegol on a MOF Lewis acid site is already known from previous computational studies.[30] The structures of these molecules, along with that of the transition state, are shown schematically in **Figure 46**. The minimum energy pathway between these two specific isomers will be examined in 7 different systems: the **fcu** unit cell; linker and cluster vacancies terminated by FA and TFAA; and the **e** defect for both the missing linker and cluster. This should cover differences in steric hindrance and electronic structure.

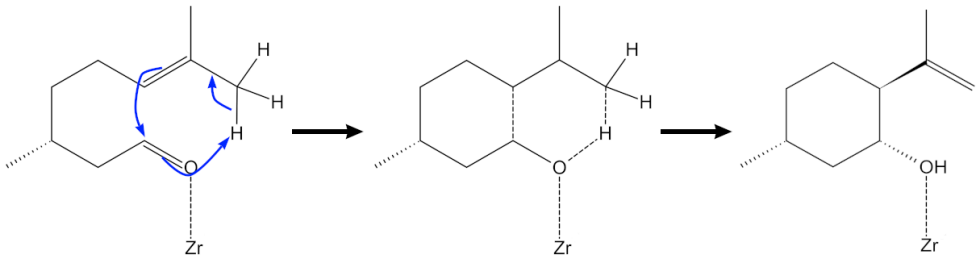


Figure 46: Scheme for the conversion of (+)-citronellal (left) to (-)-isopulegol (right) on a solid Lewis acid. The transition state reported by Vandichel *et al.* for this process is shown in the centre.[30] The movement of electrons has been marked by blue arrows and the Lewis acid centre to which the oxygen is bound has been denoted by a zirconium atom.

While **e** defects already have undercoordinated zirconium atoms to which citronellal can bond, this is not the case for vacancies terminated by FA and TFAA. For these structures, it is necessary to first free up space in the coordination sphere of Zr for the reactant. It is proposed that this could happen through one of the monocarboxylates moving from a bidentate to a dangling bonding mode in which the other end sits atop the adjacent μ_3 OH group, as shown in **Figure 47**. Such a distortion was noted by Vermoortele *et al.* as being necessary to enable citronellal cyclisation in formate-terminated cluster models.[2] In order to verify the Lewis acidity of these dangling states, an additional set of carbon monoxide frequency calculations has been carried out. The results are shown in **Table 11**.

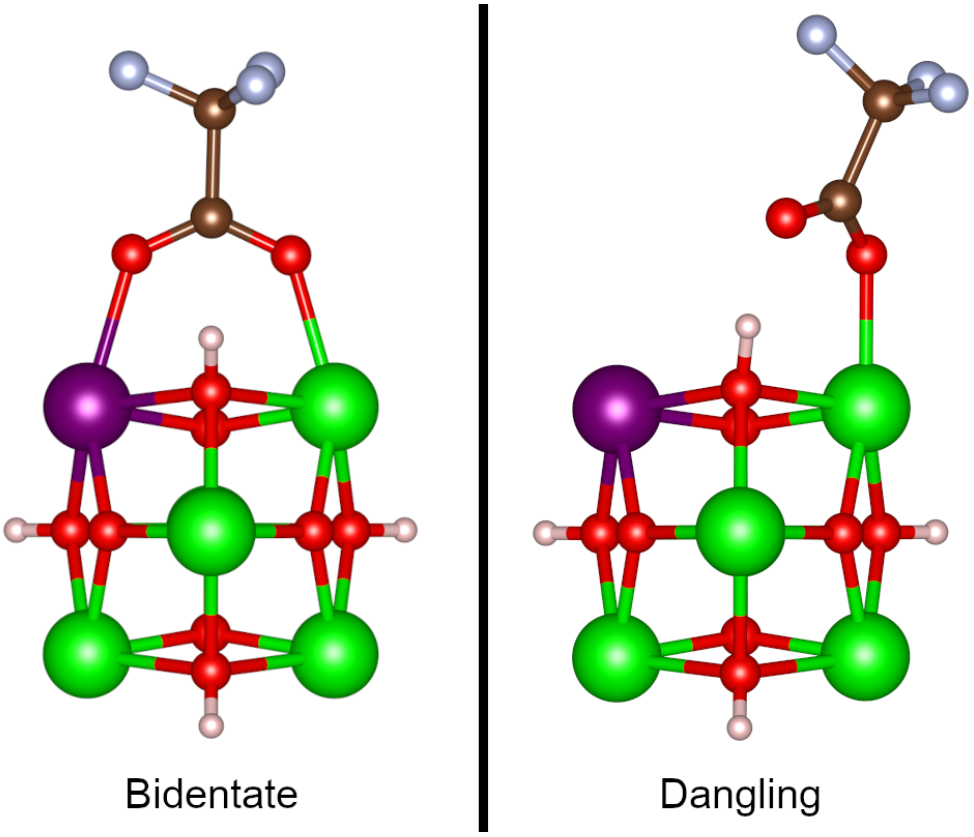


Figure 47: Schema for how coordinatively saturated zirconium can be made accessible by moving a defect-terminating monocarboxylate from a bidentate to a unidentate configuration. The zirconium which is freed up has been drawn purple. Some atoms have been obscured for clarity.

These frequencies are comparable to those of sites OH-e, Cl-e, and **e**-e seen in **Section 4.3.2**, confirming that this structural change opens up a strong Lewis acid site. The variation between capping species and defect type (linker and cluster vacancies) is minimal, which implies only minor variation in Lewis acidity. It therefore appears possible that differences in activity observed experimentally for various defective UiO-66 materials may arise from the initial step which opens the coordination sphere around zirconium. The barrier associated with this process should depend on the defect termination and type due to differences in O–Zr bond strength and steric hindrance. For the particular case of **e** defects, zirconium is already undercoordinated so it is not necessary to move from a bidentate to a dangling state. The approach described in **Section 4.2.3** has been used here to evaluate the free energies required to reach this

activated state for defective structures terminated by FA and TFAA. The results are shown in **Figure 48** and enumerated in **Table 12**. A full set of schematics for the structures in the bidentate, transition, and dangling states can also be found in **Appendix E**.

Table 11: CO frequencies on coordinatively unsaturated Zr

System		Scaled carbon monoxide frequency / cm^{-1}
fcu		2189.7
single defect	FA	2190.1
	TFAA	2200.7
	e	2183.6
reo	FA	2192.7
	TFAA	2202.4
	e	2185.4

Frequencies of carbon monoxide on dangling carboxylate structures. Stretching frequencies are strongly enhanced on all sites relative to gas-phase CO, which confirms the Lewis acidic potential of Zr in these partially decoordinated systems. There is also little variation between missing linkers and **reo**, suggesting the SBUs have similar electronic structures. All frequencies have been scaled by a factor of 0.967.

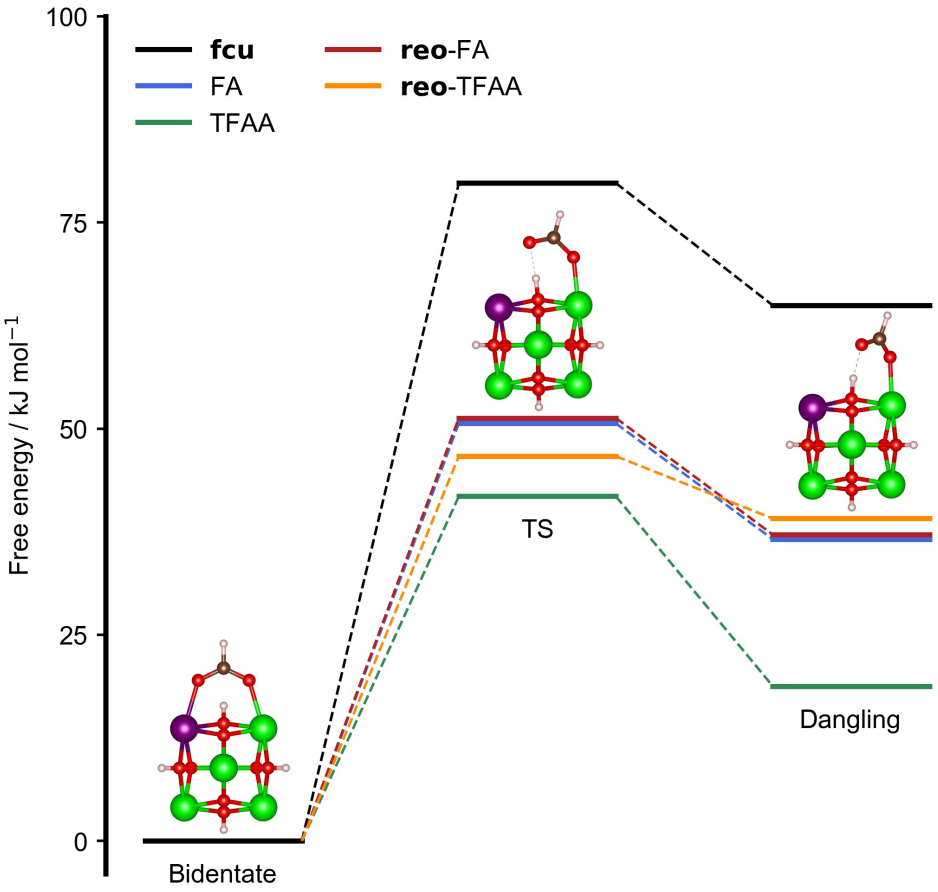


Figure 48: Free energy barriers for moving a carboxylate into the dangling position and opening up a Lewis acid site. During this first step, citronellal was not present in the system so as to avoid any drift towards the open site. The inset schematics show the geometry of a formate cap in the bidentate, transition state, and dangling configurations. The zirconium atom which becomes available is drawn in purple.

Table 12: Free energy barriers for accessing Lewis acid sites

System	Free energy barrier / kJ mol^{-1}	
	Forward	Backward
fcu	80.1	15.2
FA	50.7	14.5
TFAA	41.6	23.2
reo-FA	51.3	14.5
reo-TFAA	46.5	7.5

Free energy barriers for the dangling step. For all structures, the entropic contribution to the barrier is of the order of ∼5 kJ mol^{-1}, such that the difference in internal energy is the dominant contributor. The asymmetry between the forward and backward barriers implies that the latter process will occur more frequently and that dangling states are therefore only metastable.

Broadly, this initial step is characterised by modest forward barriers and smaller reverse ones, as illustrated by **Figure 48**. This implies that these dangling states should be accessible at reaction conditions, but that they are only metastable. The material should therefore be effective for catalysis without the need for the harsh activation procedure or dehydroxylation needed to generate **e** defects. For the **fcu** system, the forward barrier is ∼30-40 kJ mol^{-1} higher than that for the defective structures. This suggests that vacancies still make Lewis acid sites more accessible and therefore improve catalytic activity. This is borne out by experiment, wherein pristine UiO-66 typically performs poorly, unlike highly-modulated samples for which activity is correlated with defect content.[5,6,14,15]

More attention can be paid to specific structural elements within the different structures to explain the variations seen in energy barriers. This is the largest for **fcu**; the native BDC linkers are bound to 2 SBUs via 4 Zr–O bonds, such that the dangling position adopted here introduces significant strain. This configuration for the terephthalate is $\sim$ 15 kJ mol $^{-1}$ less stable than that for the standard material, meaning this deformation makes up a significant part of the forward barrier. In contrast, the monocarboxylates are connected to a single inorganic cluster via only 2 bonds, making them more flexible for this rearrangement. With FA the forward barriers are lower by 30 kJ mol^{-1}, with almost identical energetics for both the missing linker and missing cluster. These figures fall by a further 10 kJ mol^{-1} respectively when TFAA is used. In a unit cell with correlated SBUs, the CF_3 groups across a missing linker sterically hinder each other and are unable to adopt optimal orientations. Moving to the dangling state relieves this strain, hence why this defect geometry has the lowest relative energy and smallest associated barrier. This is not a concern for **reo**-TFAA, where the decrease compared to **reo**-FA must instead reflect that a weaker Zr–O bond is being broken. Indeed, examining both structures reveals that this bond is on average shorter in the formate structure, at 2.254 Å, compared to 2.269 Å with the trifluoroacetate.

Once space around zirconium has been opened up, citronellal can approach and occupy the site by bonding through its carbonyl group without further energy barriers. The cyclisation then consists of the formation of a C–C bond and proton hop from a CH_3 group to the oxygen in a concerted step. The free energies of the starting, transition, and product stages have been evaluated and are shown in **Figure 49** and **Table 13**. A full set of the corresponding structural schematics for these states can also be found in **Appendix E**.

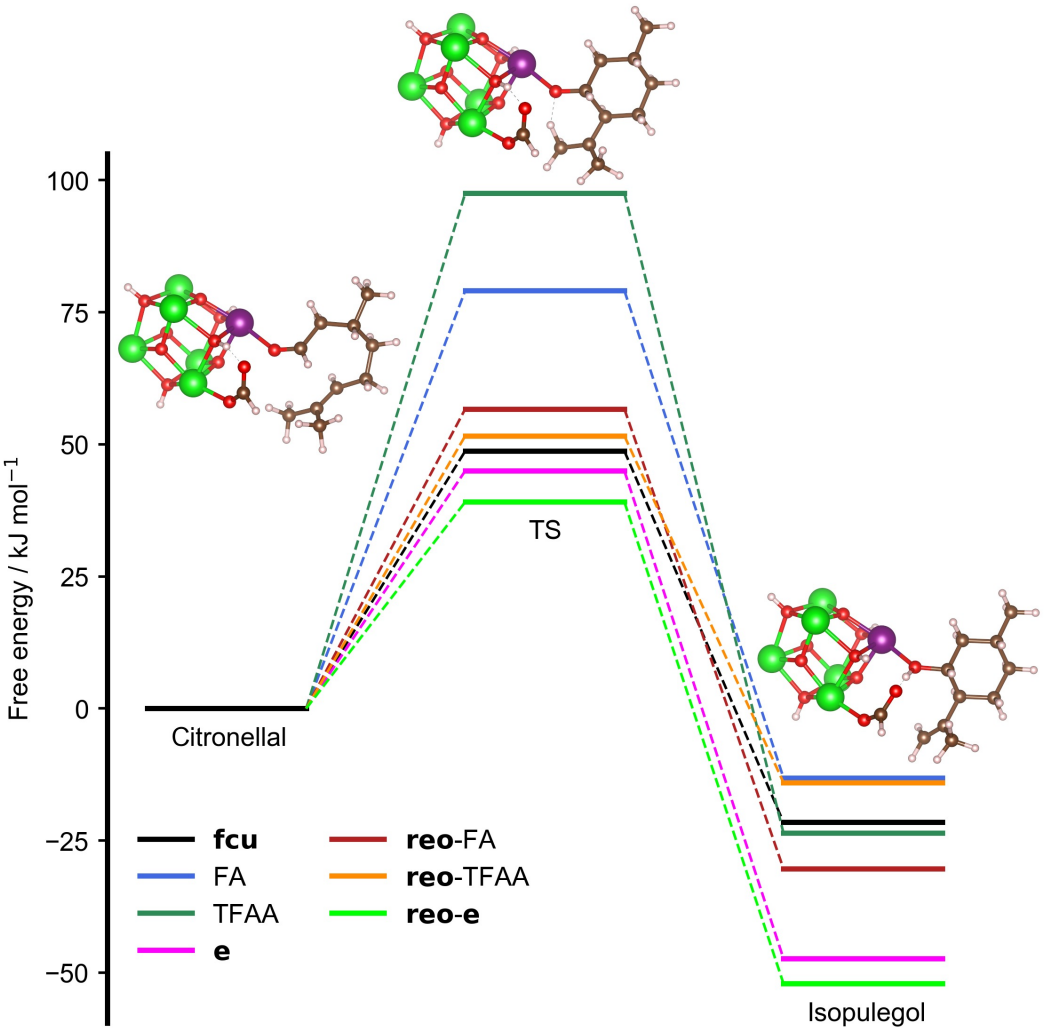


Figure 49: Free energy profile of the cyclisation of (+)-citronellal to (-)-isopulegol. The energy scale is similar to that of the previous step, but the product is significantly stabilised compared to the reactant. The inset schematics show the structure at the citronellal, transition, and isopulegol stages of the reaction at a formate-capped SBU. The Lewis acid active site has been drawn purple; some atoms have been obscured for clarity.

Table 13: Free energy barriers for citronellal cyclisation

System	Free energy barrier / kJ mol^{-1}	
	Forward	Rate-limiting
fcu	48.7	79.8
FA	79.1	79.1
TFAA	97.6	97.6
e	45.0	-
reo-FA	56.7	56.7
reo-TFAA	51.6	51.6
reo-e	39.1	-

Free energy barriers for the cyclisation of (+)-citronellal to (-)-isopulegol. The overall rate-limiting step is the one with the largest forward barrier between the dangling and cyclisation processes. Missing clusters and fully unsaturated zirconium exhibit the smallest barriers, as expected from experimental results.

It was shown earlier in **Table 11** that the Lewis acidity of the site investigated here does not vary much with the capping species. It is therefore surprising to see such large differences in free energy barriers between the different structures. For unit cells with a missing linker, the cavity space available to citronellal is limited, which likely restricts the conformation that it can adopt. This is compounded by the presence of additional charge-compensating species nearby in the cases of FA and TFAA. The forward barriers seen for these caps, 79.1 and 97.6 kJ mol^{-1}, reflect that the active site is congested and the system is destabilised by significant strain. These fall to 56.7 and 51.6 kJ mol^{-1} respectively for the missing cluster cells, highlighting the negative effect that nearby linkers and capping species have on the cyclisation step. However, this does not then

explain why **fcu**, which should be the most hindered system, has an even lower forward barrier at 48.7 kJ mol^{-1}. Examining this specific structure reveals that the dangling linker is naturally oriented in such a way that dispersion interactions with the citronellal are maximised, as can be seen in **Figure 50**. This alignment is not possible with the monocarboxylic acids. The transition state therefore benefits from additional stabilisation, leading to only a modest energy cost for cyclisation. Finally, the **e** defects perform the best for this catalytic step, with the barrier dropping as low as 39.1 kJ mol^{-1} for **reo-e**, which is to be expected given that zirconium is completely unhindered in these structures.

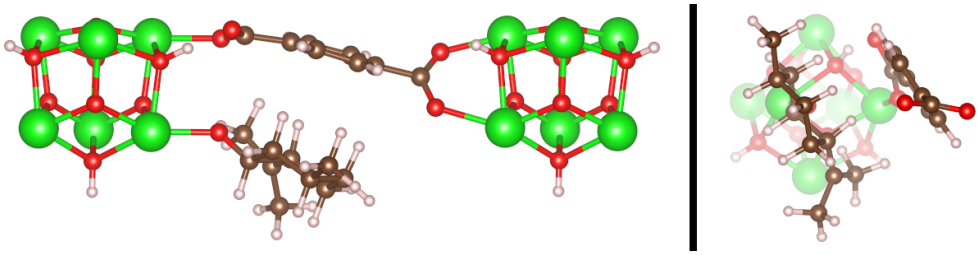


Figure 50: Schematics of the transition state and the dangling linker during the cyclisation of (+)-citronellal to (-)-isopulegol in the pristine system. It can be seen in the front (left panel) and side (right panel) views that the transition state and the terephthalate are in close proximity to each other and their molecular planes are well-aligned. This maximises dispersion interactions but is not possible in defective unit cells.

Combining thermodynamic information about both the dangling and cyclisation steps offers a more complete picture of the overall reaction. The largest barrier across both chemical events, also tabulated in **Table 13** will dictate the rate of conversion of citronellal to isopulegol. With this metric, the best kinds of defects are **e** linker and cluster vacancies, which is consistent with reported increases in activity following harsh activation procedures. However, the material should remain active even if zirconium is coordinatively saturated by species such as FA or TFAA, though this specifically requires regions of **reo** structure. This agrees with experiments which have generally shown missing clusters to be more effective

for catalysis.[5,17,18] We also note that the pristine framework and capped missing linkers should not participate much in the reaction as it is either difficult to access active sites or the cyclisation has a high barrier. Finally, it is informative to compare the results here with those of Vandichel *et al.*, who used cluster models in which the majority of linkers were replaced with formate and in which an **e** defect was used to enable access to zirconium.[23] They reported cyclisation barriers of $\sim$ 70 kJ mol^{-1} with 2 BDCs remaining and $\sim$ 61-67 kJ mol^{-1} for an inorganic node with just 1 linker left, compared to $\sim$40-45 kJ mol^{-1} seen here. The use of formate-capped aperiodic models may be responsible for this discrepancy: this affects the electronic structure of zirconium and removes some dispersion interactions between citronellal and the framework, which have been found in this study to play a role. Some differences in the choice of functional and mobile atoms in vibrational calculations are also likely to have contributed to the smaller barriers seen here.

4.4 Conclusions

The experiments in this chapter were carried out in order to better understand the relationship between defects and acidity in UiO-66. The material nominally contains Brønsted acidic μ_3 hydroxyl groups and Lewis acidic Zr^{4+} cations, both in the SBUs. Via vacancies, it is possible to alter the local structure to either introduce new moieties or to reduce steric hindrance, through which acid sites may be affected. A set of adsorption calculations has been carried out to evaluate the lability of protons, while the cyclisation of citronellal has been modelled to examine the Lewis acidity of zirconium.

Initially, the probe molecules ammonia and pyridine were used to gauge the Brønsted acidity of the bridging hydroxyl protons in different defective structures. This analysis quickly revealed that the largest contributors to variation in binding energy were secondary interactions with the framework, such as dispersion forces and hydrogen bonding. Ultimately,

these effects cannot be fully disentangled from differences in how the probe binds to a given site. However, this also means that the choice of capping species, over which some control can be exerted via synthesis conditions, will affect the conformation which a reactant will adopt on the hydroxyl. In this way, the material can still be tuned via defects to optimise the catalysis of a reaction, but not through affecting the Brønsted acidity of the active site. Another possibility is that of introducing additional labile protons through H_2O and OH moieties, though the latter of these is a weaker acid. While this approach may not be as powerful as the functionalisation of linkers, it remains interesting as it would make use of a one-pot synthesis and relatively simple building blocks.

The use of carbon monoxide via its vibrational frequency, less sensitive to the aforementioned secondary interactions, yielded further information on the different acid sites. Notably, no enhancement of the μ_3 OH was seen for any capping species, though protons on added water were confirmed to be comparable in acidity. Those on the non-native hydroxyl groups were also found to be much less labile, confirming previous findings. This probe also enabled the study of Lewis acid sites for caps containing undercoordinated zirconium; these all led to strong blueshifts, in agreement with experimental reports. A further finding is that opening these sites does not require destroying the nearby OH groups, thereby suggesting a way in which UiO-66 might behave as a dual Brønsted/Lewis acid catalyst. This functionality can be introduced simply by controlling synthesis parameters, such that the rest of the MOF design space (for example, modifying the linkers) can be used for other purposes such as tuning the zirconium cations.

A different set of simulations was then carried out to investigate the role that the capping species plays during a Lewis acid catalysed reaction. The cyclisation of citronellal to isopulegol was chosen for this given that

it is often used experimentally as a yardstick for this purpose. This was broken down into 2 steps: one in which open sites are made available, and cyclisation. The barriers for individual parts revealed that of the structures tested, missing clusters are generally more effective for catalysis, with the **e** defect termination yielding the lowest overall barriers. This is consistent with some reports of increased activity in samples containing predominantly cluster defects. Citronellal's large size, along with the limited space inside UiO-66's cavities, prevents it from adopting its optimal conformation in the transition state, making the reaction more difficult.

There are nevertheless some shortcomings to the approach adopted. A dangling carboxylate was used to make zirconium available for bonding, but it is known that there are a number of ways in which linkers may decoordinate in the presence of solvent.[24] This was neglected here because of the associated computational expense, especially for the larger **reo** cells, though metadynamics would offer a natural way to search for transition states when solvent molecules are present. The differences with Vandichel *et al.*'s work also point to the importance of the method; the PBE functional used here may not be sufficiently accurate, but upgrading to the hybrid they utilised would prove too costly. Some compromise may be possible in that the transition states and intermediate structures are now known, so some refinement at a higher level of theory may be possible with minimal compute time. A further element which may be investigated is how differences in porosity brought about by vacancies affect the transport properties of citronellal. In spite of these limitations, the results seen here suggest that a systematic experimental study comparing different structures for this catalytic reaction would be quite useful. The results seen for **reo** structures indicate the material can be effectively tuned for this cyclisation by focusing on this type of defect.

References

(1) A. Dhakshinamoorthy, A. Santiago-Portillo, A. M. Asiri and H. Garcia, *ChemCatChem*, 2019, **11**, 899–923.

(2) F. Vermoortele, M. Vandichel, B. Van De Voorde, R. Ameloot, M. Waroquier, V. Van Speybroeck and D. E. De Vos, *Angew. Chemie - Int. Ed.*, 2012, **51**, 4887–4890.

(3) Y. Kuwahara, H. Kango and H. Yamashita, *ACS Sustain. Chem. Eng.*, 2017, **5**, 1141–1152.

(4) G. Fu, F. G. Cirujano, A. Krajnc, G. Mali, M. Henrion, S. Smolders and D. E. De Vos, *Catal. Sci. Technol.*, 2020, **10**, 4002–4009.

(5) L. Liu, Z. Chen, J. Wang, D. Zhang, Y. Zhu, S. Ling, K. W. Huang, Y. Belmabkhout, K. Adil, Y. Zhang, B. Slater, M. Eddaoudi and Y. Han, *Nat. Chem.*, 2019, **11**, 622–628.

(6) X. Feng, H. S. Jena, C. Krishnaraj, K. Leus, G. Wang, H. Chen, C. Jia and P. Van Der Voort, *ACS Appl. Mater. Interfaces*, 2021, **13**, 60715–60735.

(7) R. C. Klet, Y. Liu, T. C. Wang, J. T. Hupp and O. K. Farha, *J. Mater. Chem. A*, 2016, **4**, 1479–1485.

(8) F. G. Cirujano and F. X. Llabrés I Xamena, *J. Phys. Chem. Lett.*, 2020, **11**, 4879–4890.

(9) H. Chen, *Chem. Phys. Lett.*, 2022, **800**, 139658.

(10) V. R. Bakuru, S. R. Churipard, S. P. Maradur and S. B. Kalidindi, *Dalt. Trans.*, 2019, **48**, 843–847.

(11) J. Speight, N. A. Lange and J. A. Dean, in *Lange's Handbook of Chemistry*, McGraw Hill, New York, United States, 16th edn., 2005, ch. 1.7, pp. 1.160–1.171.

(12) L. Shen, R. Liang, M. Luo, F. Jing and L. Wu, *Phys. Chem. Chem. Phys.*, 2015, **17**, 117–121.

(13) F. Vermoortele, B. Bueken, G. Le Bars, B. Van De Voorde, M. Vandichel, K. Houthoofd, A. Vimont, M. Daturi, M. Waroquier, V. Van Speybroeck, C. Kirschhock and D. E. De Vos, *J. Am. Chem. Soc.*, 2013, **135**, 11465–11468.

(14) R. Wei, J. Fan, X. Qu, L. Gao, Y. Wu, Z. Zhang, F. Hu and G. Xiao, *Eur. J. Inorg. Chem.*, 2020, **2020**, 833–840.

(15) X. Feng, J. Hajek, H. S. Jena, G. Wang, S. K. Veerapandian, R. Morent, N. De Geyter, K. Leyssens, A. E. Hoffman, V. Meynen, C. Marquez, D. E. De Vos, V. Van Speybroeck, K. Leus and P. Van Der Voort, *J. Am. Chem. Soc.*, 2020, **142**, 3174–3183.

(16) W. Xiang, J. Ren, S. Chen, C. Shen, Y. Chen, M. Zhang and C. jun Liu, *Appl. Energy*, 2020, **277**, 115560.

(17) P. Chammingkwan, G. Y. Shangkum, L. T. T. Mai, P. Mohan, A. Thakur, T. Wada and T. Taniike, *RSC Adv.*, 2020, **10**, 28180–28185.

(18) J. Wang, L. Liu, C. Chen, X. Dong, Q. Wang, L. Alfilfil, M. R. Alalouni, K. Yao, J. Huang, D. Zhang and Y. Han, *J. Mater. Chem. A*, 2020, **8**, 4464–4472.

(19) G. C. Shearer, S. Chavan, S. Bordiga, S. Svelle, U. Olsbye and K. P. Lillerud, *Chem. Mater.*, 2016, **28**, 3749–3761.

(20) F. G. Xi, Y. Yang, H. Liu, H. F. Yao and E. Q. Gao, *RSC Adv.*, 2015, **5**, 79216–79223.

(21) D. M. Driscoll, D. Troya, P. M. Usov, A. J. Maynes, A. J. Morris and J. R. Morris, *J. Phys. Chem. C*, 2018, **122**, 14582–14589.

(22) G. C. Shearer, S. Forselv, S. Chavan, S. Bordiga, K. Mathisen, M. Bjørgen, S. Svelle and K. P. Lillerud, *Top. Catal.*, 2013, **56**, 770–782.

(23) M. Vandichel, J. Hajek, F. Vermoortele, M. Waroquier, D. E. De Vos and V. Van Speybroeck, *CrystEngComm*, 2015, **17**, 395–406.

(24) C. Caratelli, J. Hajek, E. J. Meijer, M. Waroquier and V. Van Speybroeck, *Chem. Eur. J.*, 2019, **25**, 15315–15325.

(25) S. Ling and B. Slater, *Chem. Sci.*, 2016, **7**, 4706–4712.

(26) K. Suzuki, G. Sastre, N. Katada and M. Niwa, *Chem. Lett.*, 2007, **36**, 1034–1035.

(27) K. Suzuki, T. Nishio, N. Katada, G. Sastre and M. Niwa, *Phys. Chem. Chem. Phys.*, 2011, **13**, 3311–3318.

(28) K. I. Hadjiivanov and G. N. Vayssilov, *Adv. Catal.*, 2002, **47**, 307–511.

(29) G. A. Tribello, M. Bonomi, D. Branduardi, C. Camilloni and G. Bussi, *Comput. Phys. Commun.*, 2014, **185**, 604–613.

(30) M. Vandichel, F. Vermoortele, S. Cottenie, D. E. De Vos, M. Waroquier and V. Van Speybroeck, *J. Catal.*, 2013, **305**, 118–129.

(31) A. Ghysels, T. Verstraelen, K. Hemelsoet, M. Waroquier and V. Van Speybroeck, *J. Chem. Inf. Model.*, 2010, **50**, 1736–1750.

(32) J. Jiang, F. Gándara, Y. B. Zhang, K. Na, O. M. Yaghi and W. G. Klemperer, *J. Am. Chem. Soc.*, 2014, **136**, 12844–12847.

(33) L. Alaerts, E. Séguin, H. Poelman, F. Thibault-Starzyk, P. A. Jacobs and D. E. De Vos, *Chem. Eur. J.*, 2006, **12**, 7353–7363.

(34) D. Dylong, P. J. Hausoul, R. Palkovits and M. Eisenacher, *Flavour and Fragrance Journal*, 2022, **37**, 195–209.

(35) S. F. Boys and F. Bernardi, *Mol. Phys.*, 1970, **19**, 553–566.

(36) J. Spencer and A. Alavi, *Phys. Rev. B - Condens. Matter Mater. Phys.*, 2008, **77**, 1–4.

(37) M. Guidon, J. Hutter and J. VandeVondele, *J. Chem. Theory Comput.*, 2009, **5**, 3010–3021.

(38) S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.

(39) S. Goedecker and M. Teter, *Phys. Rev. B - Condens. Matter Mater. Phys.*, 1996, **54**, 1703–1710.

(40) J. VandeVondele and J. Hutter, *J. Chem. Phys.*, 2007, **127**, 114105.

(41) NIST, *Computational Chemistry Comparison and Benchmark DataBase*, 2022.

(42) X. Ma, A. Genest, L. Spanu and N. Rösch, *Computational and Theoretical Chemistry*, 2015, **1069**, 147–154.

(43) G. Bussi, D. Donadio and M. Parrinello, *J. Chem. Phys.*, 2007, **126**, 014101.

(44) G. Henkelman and H. Jónsson, *J. Chem. Phys.*, 2000, **113**, 9978–9985.

(45) G. Henkelman, B. P. Uberuaga and H. Jónsson, *J. Chem. Phys.*, 2000, **113**, 9901–9904.

(46) E. G. Hohenstein and C. D. Sherrill, *J. Phys. Chem. A*, 2009, **113**, 878–886.

(47) S. Beran, *J. Phys. Chem.*, 1983, **87**, 55–58.

(48) C. Caratelli, J. Hajek, F. G. Cirujano, M. Waroquier, F. X. Llabrés i Xamena and V. Van Speybroeck, *J. Catal.*, 2017, **352**, 401–414.

5 Mixed Ce-Zr UiO-66

5.1 Motivation

The architecture of UiO-66 is compatible with other metals than zirconium. Other isoreticular frameworks, containing similar M_6 SBUs, have been synthesised using other transition and f-block metals.[1–4] Post-synthetic exchange (PSE) can be used to replace Zr^{4+} in the host structure, most notably with Ti^{4+},[5–7] while heterometallic forms of the material can be directly assembled using a suitable mix of precursors.[8,9] This latter approach is particularly interesting as it can be used to introduce additional functionality, such as photochemical activity or redox chemistry, with minimal impact on the exceptional stability of UiO-66.[8,10,11]

One such version of the framework which has received significant attention is the cerium substituted one. CeO_2 has previously been widely studied for its catalytic properties, with applications in automotive catalysts, alcohol and hydrocarbon oxidation, and the water gas shift reaction.[12] Ceria, either alone or in solid solutions with zirconia, is also known for its redox activity.[13] While pure Ce-UiO-66 can be synthesised, it suffers from lower thermal stability and is not as well-studied as the zirconium material.[14] The interest in the mixed framework is therefore to build an effective catalyst based on a robust and extensively characterised structure.

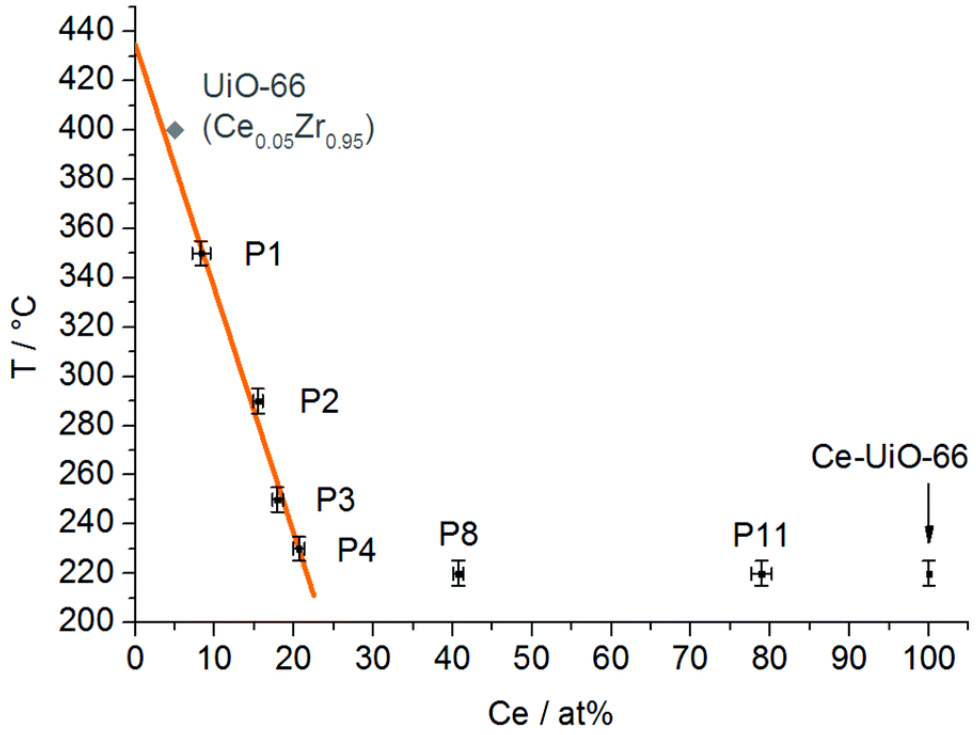


Figure 51: Variation in decomposition temperature, as determined by variable-temperature PXRD, with cerium content in UiO-66. Thermal stability drops sharply from the y-intercept (which coincides with the value for pure UiO-66 as determined by TGA) until ~20% Ce, after which it appears to stabilise. Figure adapted from ref. [14].

The synthesis of Ce-Zr-UiO-66 is largely the same as for the standard material. Zirconium and cerium precursors (typically cerium ammonium nitrate, CAN) are dissolved in DMF and mixed with BDC and a modulator, usually formic or benzoic acid, before heating, washing, and activation. Work by Lammert *et al.* suggests this procedure leads to a homogeneous product in which cerium is well-dispersed, as indicated by good adherence to Vegard's law.[14] However, uptake of this metal is sub-stoichiometric relative to the ratio of precursors used, suggesting some difficulty in including it in the structure. Any cerium not incorporated into the framework eventually forms cerium (III) formate byproduct, especially when the reaction time is long.[15,16] This is accompanied by an unusual trend in decomposition temperature, shown in **Figure 51**, which exhibits two regimes of thermal stability.

Attempts at explaining Lammert *et al.*'s results have revealed a specific distribution of SBUs may be responsible.[14] Lomachenko *et al.* carried out EXAFS experiments in which they saw changes in Zr and Ce K-edges with increasing cerium content.[17] They found good agreement between this data and a structural model in which only Zr_6, Ce_6, and Zr_5Ce_1 SBUs are allowed. As more cerium is added, the number of Zr_6 building blocks decreases, until at ∼20% only Zr_5Ce_1 and Ce_6 are present, coinciding with the change in regimes seen in **Figure 51**. The authors therefore suggested that the loss of highly ordered and stable Zr_6 units is responsible for the drop in thermal stability. Subsequent work by the same groups has revealed that theoretical O–H stretching frequencies from this so-called three-cluster model (TCM) are consistent with experimental IR spectra from samples with a range of cerium contents.[18] However, it is also noted that the data does not rule out the presence of small concentrations of other $Zr_{6-x}Ce_x$ units.

The driving force behind the TCM remains unknown, but the mismatch between Zr and Ce likely contributes to it. Zr^{4+} and Ce^{4+} have radii of 0.84 and 0.97 Å when eight-fold coordinated,[19] which is reflected in the cell parameters of the respective pure UiO-66 materials, 20.700 and 21.473 Å.[15,20] Mixing metals within the unit cell and even within individual SBUs therefore likely results in significant strain. Additionally, it has been found by Trousselet *et al.* and later confirmed through further calculations by Lomachenko *et al.* that Ce-Zr bonds are unfavourable relative to homometallic bonds.[17,21] Therefore, the prevalence of Zr_6 and Ce_6 SBUs may simply be the result of minimising bimetallic bonds and structural distortions, though this does not explain why Zr_5Ce_1 building blocks (without the similar presence of Zr_1Ce_5) are so abundant.

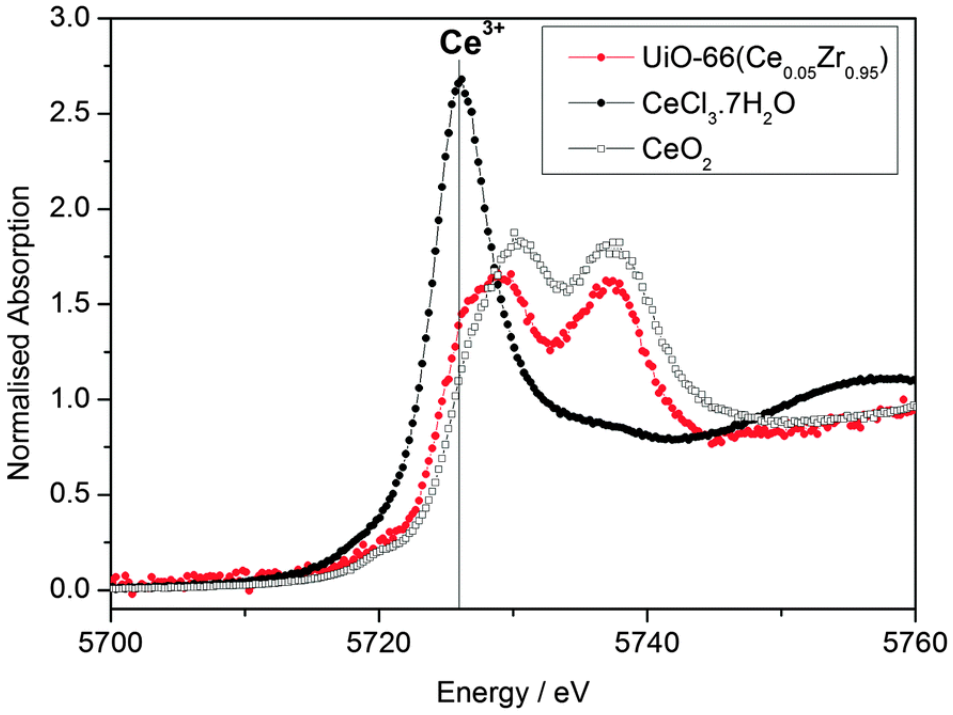


Figure 52: XANES spectrum of mixed Ce-Zr UiO-66 and reference materials for Ce^{3+} and Ce^{4+} ions, at the Ce L_{III} edge. Analysis using the edge positions suggests a ratio of 20/80 % for the respective oxidation states. Figure taken from ref. [8].

An additional factor which should be taken into account in attempting to rationalise the TCM is that some cerium may also exist in the 3+ state. Such species may result from the use of a Ce(III) precursor such as cerium chloride, or from the *in situ* reduction of Ce^{4+} by DMF or another solvent.[22,23] Studies have confirmed via XANES that Ce^{3+} can persist inside UiO-66 following synthesis.[8] **Figure 52** evidences a shift in the cerium edge consistent with a mixture of oxidation states. This ion's even larger ionic radius (1.143 Å)[19] and unpaired 4f electron will exacerbate issues of strain and unfavourable bonding. Additionally, Ce^{3+} is known to have different solution chemistry than Ce^{4+}; the latter assembles into the same M_6O_8 clusters as zirconium, while the former favours mononuclear complexes.[24] In this state, cerium may therefore be unable to form the required building blocks to participate in the growth of the framework, which would also explains low cerium uptake.

The work in this chapter was carried out with the objective of determining why Ce-Zr UiO-66 grows following the TCM. Understanding why this happens would lead to better control over the design and structure of the material in two ways: with knowledge of the factors that lead preferentially to Zr_5Ce_1 SBUs, it may be possible to control the synthesis so as to promote diversity among heterometallic building blocks; and given that the issue of low cerium uptake is likely related, this could enable syntheses in which the ratio of Ce to Zr can be better tuned. Another key question concerns defects and their role in the distribution of cerium in the mixed material. These objectives are addressed using DFT by considering the thermodynamics of UiO-66 at two levels: for isolated $Zr_{6-x}Ce_x$ clusters as they would exist in solution, and for the periodic framework in both pristine and defective forms. Given the possible importance of Ce^{3+}, both oxidation states of cerium are considered in this context.

5.2 Approach

5.2.1 Modelling cerium

The methods used in this section are broadly similar to those described in **Section 3.2** and **Section 4.2**. One major difference is the presence of cerium, for which relativistic effects present additional modelling challenges.[25] The non-relativistic Schrödinger equation introduced in **Section 2.1** made use of the classical momentum and mass for the kinetic energy operator. For heavy nuclei, in which electrons have kinetic energies corresponding to appreciable fractions of the speed of light, this simplification is no longer appropriate and relativity needs to be properly treated. A convenient way to do this is by using a pseudopotential which has been fitted to a wavefunction determined by a method which explicitly accounts for these effects.[26] Cerium can be modelled in this way thanks to the work of Lu *et al.*[27] The authors fitted GTH pseudopotentials and accompanying double-ζ basis sets for all lanthanides using relativistic all-electron calculations at the PBE level.

Another issue which must be addressed with cerium is that of its strongly-correlated 4f states. In CeO_2 and Ce_2O_3, these are poorly modelled by local and semi-local functionals, leading to over-delocalisation.[28,29] A commonly used approach to mitigate this is to apply the semi-empirical correction of Hubbard, also known as the DFT+U method.[30–32] This approximately tabulates and compensates for self-interaction error, and penalises orbitals which are less than half-empty while stabilising those which are not, the results of which are to discourage delocalisation.[33,34] One drawback with this method for ceria is that different values of the U parameter are required to optimally model various properties, and a suitable choice for Ce^{4+} is often unsuitable for Ce^{3+}. There is additionally no convenient empirical data to which this parameter can be fitted to for mixed UiO-66. An alternative is to include a fraction of the HF exchange energy via a hybrid functional, which can lead to good structural results compared to DFT+U for ceria and reduced ceria.[35]

In order to determine how best to model cerium's 4f electrons, some preliminary calculations were run using different methods. There is a large body of work covering the optimal choice of parameters for DFT+U when modelling cerium oxide and its vacancies. From this, we pinpoint the work of Castleton *et al.*, which suggests the choice U_{eff}=7 eV, and that of Lu and Liu, from which values of 5 and 6 eV appear sensible for Ce^{4+} and Ce^{3+} respectively. Additionally, the PBE0 functional was also tested. A subset of aperiodic clusters containing either Ce^{3+} or Ce^{4+} was initially evaluated using both PBE+U and PBE0. Relative energies were then defined using the general formula below for Ce^{4+}:

$$\Delta E = E_{\mathrm{Zr}_{n-x}\mathrm{Ce}_x} - \frac{(n-x)}{n} E_{\mathrm{Zr}_{\mathrm{ref}}} - \frac{x}{n} E_{\mathrm{Ce}_{\mathrm{ref}}} \tag{98}$$

The first term on the right-hand side corresponds to the mixed system which contains a total of n metal atoms, x of which have been substituted from zirconium to cerium. The subsequent terms are the energies

of Zr and Ce reference states, which in this case are Zr_6 and Ce_6 clusters given the known propensity for these 4+ ions to form such octahedral units.[38,39] A different equation was used for Ce^{3+}:

$$\Delta E = E_{Zr_4Ce_2} + E_{Zr_6} - 2E_{Zr_5Ce_1} \tag{99}$$

Terms here correspond to the doubly substituted, pure zirconium, and singly substituted M_6 clusters. This latter expression provides a simple way of comparing energetics between systems containing Ce^{3+} without needing to define a reference state. The resulting energies for aperiodic clusters containing Ce^{4+} and Ce^{3+} are shown in **Figure 53** and **Table 14** respectively.

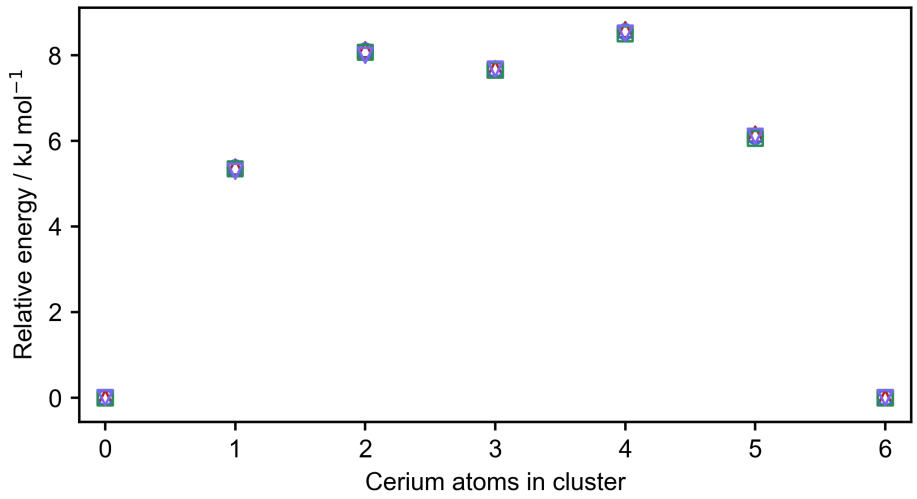


Figure 53: Benchmark of FA-terminated clusters containing Ce^{4+}. Relative energies have been evaluated using **Equation 98** with n=6 at the PBE+U and PBE0 levels. It is clear that the choice of method makes no qualitative difference and has only minimal quantitative impact for this oxidation state. Red triangles, blue circles, and green squares correspond to PBE+U-DZVP with U_{eff} values of 5, 6, and 7 eV; purple inverted triangles denote PBE0-TZVP results.

Table 14: DFT+U benchmarking - Ce^{3+}

System	Relative energy / kJ mol^{-1}			
	U_{eff}=5	U_{eff}=6	U_{eff}=7	PBE0
Zr_4Ce_2-adjacent FA	28.7	36.3	43.5	-21.7
Zr_4Ce_2-opposite FA	51.7	52.0	51.1	12.0
Zr_4Ce_2-adjacent BA	-6.2	18.1	-22.2	-14.8
Zr_4Ce_2-opposite BA	26.6	38.2	17.6	11.3

Benchmark of FA and BA-terminated aperiodic clusters containing 2 Ce^{3+} ions. U_{eff} is given here in eV. Replacing a 4+ metal with a 3+ one introduces a charge imbalance; this has been compensated for by introducing an additional proton to a neighbouring μ_3 oxygen for each cerium. Relative energies have been evaluated using **Equation 99**. Positive values imply that for these building blocks, Zr_5Ce_1 is more stable than Zr_4Ce_2.

It is clear from **Figure 53** that the choice of value for U_{eff} will have little influence on energetics when only Ce^{4+} is present, while PBE0 leads to identical results. Therefore, the balanced choice from Castleton *et al.* of $U_{eff} = 7$ will be used for this cerium ion.[36] Interpreting **Table 14** is less straightforward, as the data is noisier and variations are not systematic with U_{eff}. This may be in part because DFT+U can lead to local minima in the energy landscape,[40] as has notably been reported for CeO_2 containing vacancies.[41] Across all parameter choices, the opposite configuration is less stable than the adjacent, for both acids. However, whether this structure is also relatively more stable than Zr_5Ce_1 appears to depend on both the cap and the method. Based on these results, there is no universally satisfying choice of method, as Castleton *et al.* noted also for CeO_2 and Ce_2O_3.[36] The carboxylates used here bond in similar fashion to Zr and Ce, so based on the separation of +3 ions in the opposite

structure, relative energy would not be expected to vary strongly with U_{eff} or the cap. In this regard, PBE0 is a sensible choice because it obeys this criterion. This exercise does highlight the difficulty of modelling materials containing cerium, which has been made more complicated here by the high cost of slowly-converging open-shell calculations. Therefore, the hybrid functional was used for systems containing Ce^{3+}.

5.2.2 Computational details

The majority of calculations in this chapter were carried out using the parameters described in **Section 2.6**. For systems containing Ce^{4+}, the DZV-MOLOPT basis and GTH pseudopotentials of Lu *et al.* were used for those atoms.[27] A Hubbard correction was additionally applied to the 4f orbitals of cerium using a single effective parameter U_{eff} equal to 7 eV except where specified. For simulations involving aperiodic clusters, an empty box of dimensions 25×25×25 Å was used. Linker strain has been evaluated using ORCA with the parameters described in **Section 2.6**.

Different settings were used for systems containing Ce^{3+}. PBE was used during benchmarking with the same basis sets and settings described above. For this functional, the open-shell nature of cerium's extra electrons was treated using the ROKS scheme as this yielded better electronic solutions than UKS. The Ce^{3+} results presented in **Section 5.3.3** were generated using the truncated Coulomb operator variant of the PBE0 functional with D3 dispersion corrections and GTH pseudopotentials.[42–45] The following basis sets were used: the same DZV basis as described above for Ce; DZVP-MOLOPT for Zr; and TZVP-MOLOPT for H, C, N, and O atoms.[46] The ADMM method was used to calculate the HF exchange energy together with the following auxiliary basis sets: pFIT3 for H, C, N, and O atoms; FIT11 for Zr; and for Ce, a basis set developed by Hahn *et al.* for use on CeO_2.[47] The UKS scheme was used with the PBE0 functional only as the ROKS alternative is not currently implemented for hybrid functionals in CP2K. With both PBE and PBE0, the

convergence criteria for maximum step size during geometry optimisations was increased to 10^{-3} Ha Bohr^{-1} for all systems containing Ce^{3+}. Additionally, difficulties were encountered for some systems in the SCF procedure in the form of slow convergence and oscillation in the energy. In such cases, the SCF convergence criterion was relaxed to 10^{-6} Ha.

5.3 Results and discussion

5.3.1 Isolated Ce-Zr building blocks

In order to uncover why certain SBUs are favoured in mixed UiO-66, the Ce^{4+} solution-state building blocks were first considered. Aperiodic models for different cerium contents were generated in 2 sets: one capped by FA, and another by BA, both of which are common modulators in the synthesis of the mixed material. In both cases, relative energies per mole of cluster can readily be defined using **Equation 98** with $n = 6$ and M_6 reference states terminated using the corresponding carboxylic acids. These have been evaluated are shown in **Figure 54**.

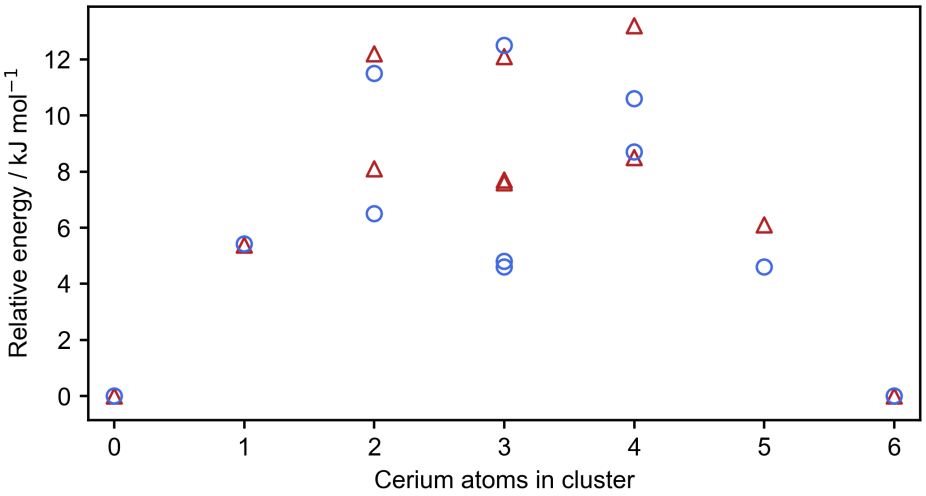


Figure 54: Relative stabilities of $Zr_{6-x}Ce_x$ building blocks capped by FA (blue circles) and BA (red triangles) found at the PBE-DZVP level. There are respectively 2, 3, and 2 possible configurations for 2, 3, and 4 cerium atoms. The scale of relative energies is comparable to thermal energy, which suggests all clusters should be accessible during synthesis.

It is clear that this simple approach does not explain the prevalence of Zr_5Ce_1 units. For both capping species, the relative energy does not increase significantly when moving away from either reference state, with a total range of 12 kJ mol^{-1}. All possible clusters should therefore be thermally accessible, and one would expect this to be reflected in the distribution of SBUs for the finished material. Furthermore, the lower-energy set of results for x=2, 3, 4 corresponds to configurations in which cerium atoms are adjacent. This implies heterometallic bonds are unfavourable, as found by Trousselet *et al.*,[21] which is also inconsistent with experimental trends since it implies a preference for separate Ce and Zr phases. It is likely that, in these aperiodic units, low rigidity means the larger radius of Ce^{4+} can easily be accommodated through structural distortions. Therefore, repeating this experiment using periodic models, in which SBUs are more geometrically constrained, is a sensible next step.

5.3.2 Metal-mixing in the unit cell

In moving to the conventional unit cell of UiO-66, the number of metal atoms increases from 6 to 24. This leads to a combinatorial explosion such that different cerium contents can no longer be exhaustively sampled. However, the results seen with clusters suggest that the minimum and maximum energy distributions can be found by respectively minimising and maximising heterometallic bonds. In this way, the configurational space is reduced such that it becomes tractable with DFT. The periodic systems have thus been optimised and relative energies (per mole of unit cell contents) have again been defined according to **Equation 98**; the periodic mixed and pure metal systems have been used as reference states, along with $n = 24$. Additionally, the sequence of unit cells $Zr_6 \rightarrow Zr_6 - Zr_5Ce_1 \rightarrow Zr_5Ce_1 \rightarrow Zr_5Ce_1 - Ce_6 \rightarrow Ce_6$ corresponding to experimental trends has also been evaluated. The results are shown in **Figure 55**.

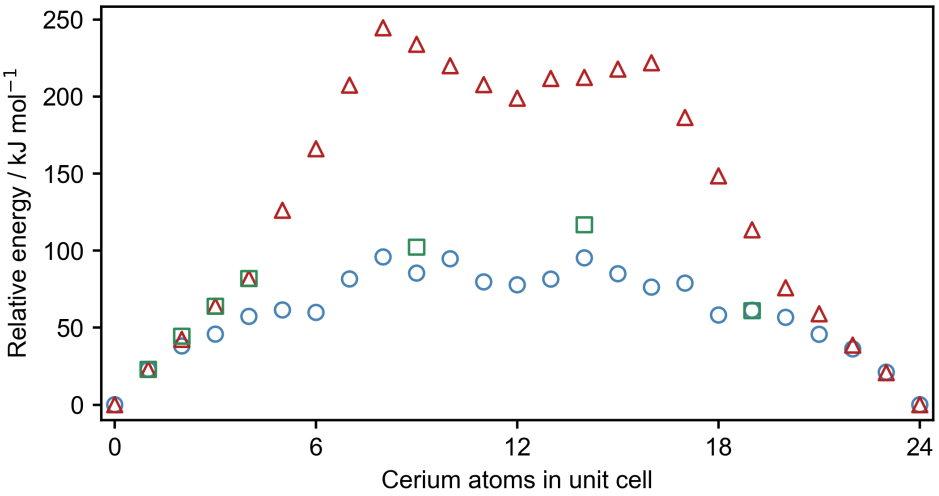


Figure 55: Stability of different cerium distributions in the pristine conventional unit cell of UiO-66. Relative energies have been evaluated using **Equation 98** with n=24 at the PBE-DZVP level. Configurations in which the numbers of heterometallic bonds have been minimised and maximised are respectively shown by blue circles and red triangles. Green squares correspond to systems built only using Zr_6, Ce_6, and Zr_5Ce_1 clusters.

It can be seen in **Figure 55** that in the fully connected framework, it is more difficult to mix the metals. The range of relative energies has increased an order of magnitude compared to cluster models, now reaching nearly 250 kJ mol^{-1} for the least stable configurations. The gap between favourable and unfavourable distributions has also been magnified to ∼100 kJ mol^{-1}, whereas it was previously comparable to thermal energy (∼3 kJ mol^{-1} at synthesis conditions). This can be attributed to the difference in ionic radii between Zr^{4+} and Ce^{4+}; the optimal metal-linker oxygen bond lengths of 2.233 and 2.400 Å of the respective pure-metal UiO-66 cannot be simultaneously satisfied.[15,20] As shown by Trousselet *et al.*, this is most pronounced when oxygens on the same end of a linker are bound to different metals (from Zr and Ce), such that this strain is more significant when cerium is more distributed.[21] The consequences of this can be seen in **Figure 56**, which compares some terephthalate geometries from the pristine Zr-UiO-66 and Zr-Ce-UiO-66 systems. The

mismatch in metal ionic radii can lead to both bowing and tilting of the linker's benzene backbone. Such distortions are more widespread when cerium is distributed and when an intermediate number of zirconium atoms have been substituted. **Figure 55** also shows that the sequence of structures made from Zr_6, Ce_6, and Zr_5Ce_1 is at best competitive with those in which cerium is maximally adjacent with itself. The data therefore shows again that there is no reason why Zr_5Ce_1 should be predominant in the mixed material, only that mixing and heterometallic bonds are unfavourable.

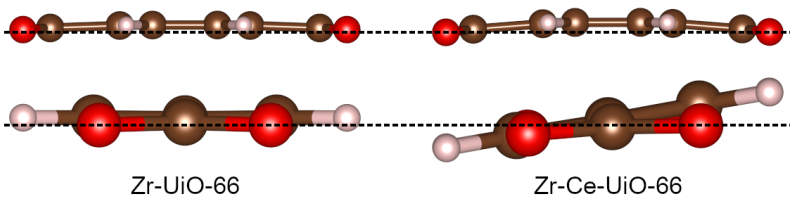


Figure 56: Schematics of some linker distortions which can be found in optimised unit cells of the Zr-Ce-UiO-66 system. While these appear minor, each one is associated with an energy penalty up to ~12 kJ mol^{-1}. In the pristine unit cell with 24 linkers, this adds up towards significant strain. The schematics here correspond to different linkers taken from unit cells with intermediate cerium content. Dotted lines have been drawn to emphasise the distortions.

An additional element which should be considered is missing linker defects, well-studied and ubiquitous in Zr UiO-66, given that these have also been reported in the mixed and pure Ce forms.[8,15] Introducing vacancies into the structure reduces rigidity,[48] which should thereby alleviate the issue of strain highlighted previously. As before, it is not possible to evaluate a range of defect contents or their positions, as the configurational space becomes problematically large. Instead, a single defective arrangement missing 8 linkers, which corresponds to the **bcu** structure reported by Liu *et al.*, has been considered.[49] This represents a high defect content and a specific topology but should provide an upper bound for how

vacancies may alleviate strain in mixed UiO-66. Vacancies have been introduced in the previously-optimised periodic systems, oriented in such a way that as many ceriums as possible lie adjacent to one. Formic acid has been used as the capping species as it leads to straightforward defect geometries and no care needs to be taken with regards to the orientations of SBUs. **Equation 98** has been used with $n = 24$ and **bcu** unit cells of Zr and Ce UiO-66 for the reference states to evaluate relative energies, which are shown in **Figure 57**

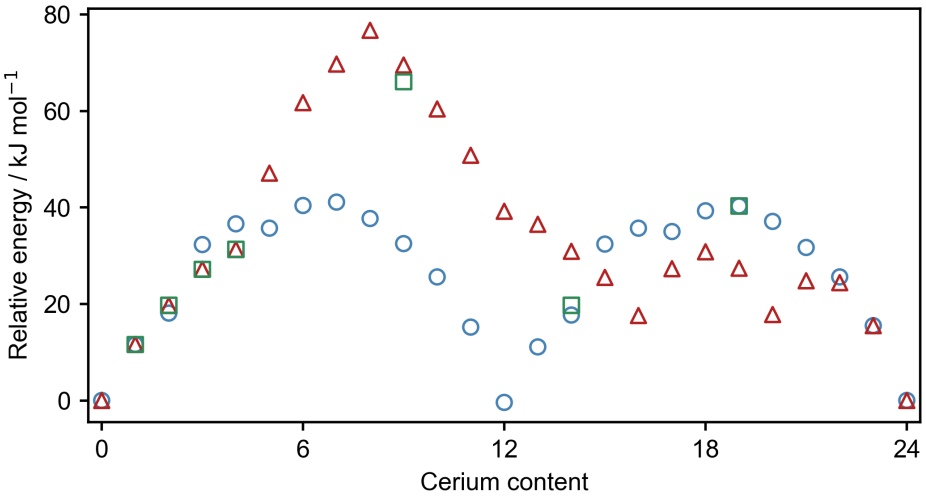


Figure 57: Stability of different cerium distributions in the **bcu** unit cell of UiO-66. Relative energies have been evaluated using **Equation 98** with n=24 at the PBE-DZVP level. Configurations in which the numbers of heterometallic bonds have been minimised and maximised are respectively shown by blue circles and red triangles. Green squares correspond to systems built only using Zr_6, Ce_6, and Zr_5Ce_1 clusters.

It is clear from **Figure 57** that the extra structural flexibility introduced by defects is beneficial to the inclusion of cerium. The range of relative energies has decreased by a factor of 3 to a maximum of 80 kJ mol^{-1}, and the difference between favourable and unfavourable configurations is at most 40 kJ mol^{-1}. With reduced framework connectivity, it is easier for the structure to relax so as to accommodate the larger radius of Ce^{4+} without buckling the terephthalates. For a cerium content of 12 with

maximum adjacency ($2\,Zr_6 - 2\,Ce_6$), the cost of mixing falls to zero. In this specific structure, shown in **Figure 58**, each Zr SBU is connected only to Ce ones and vice versa, thereby decoupling the different ionic radii of the two metals. It is likely that this effect is also responsible for the low energy of arrangements where cerium is highly distributed at high cerium contents, in which building blocks should otherwise be highly strained. Although the specific **bcu** topology has not been reported in the mixed material, the results nevertheless suggest that the distributions of cerium and defects will be intertwined, such that it may be possible to influence one by tuning the other through changes in synthesis parameters (e.g. ratio of metals, concentrations of precursors etc.). However, as before, it is favourable for cerium to be adjacent to cerium, and the data points for $Zr_6 - Ce_6 - Zr_5Ce_1$ unit cells suggest these can be competitive, but not preferred. Once again the results indicate Zr_5Ce_1 building blocks should not be prevalent in the structure.

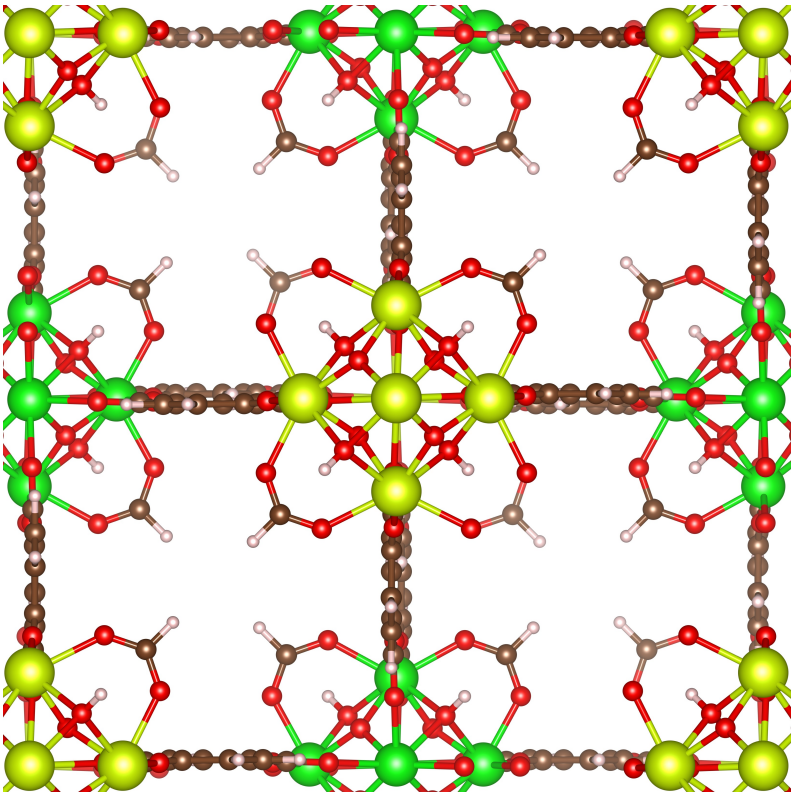


Figure 58: Schematic of the most favourable mixed unit cell with 12 cerium atoms and 8 linker vacancies. This arrangement of defects and cerium leaves only Zr SBUs coupled to Ce SBUs, such that the difference between Zr–O and Ce–O bond lengths no longer leads to any strain.

5.3.3 Building blocks containing Ce^{3+}

One further factor which may influence the distribution of SBUs in mixed UiO-66 is the presence of Ce^{3+}. This species may be present during the synthesis either through the deliberate use of a Ce(III) precursor or through the reduction of Ce^{4+} by the solvent.[50] It is also known that cerium can oscillate between the 3+ and 4+ states if suitable reducing and oxidising species are present, for example in the Belousov–Zhabotinsky reaction or as observed in nanoceria.[51,52] This ion is also known to persist in the finished framework, as found through XANES, and is important for redox-based catalysis.[15,53] One expects differences in thermodynamic trends with Ce^{3+} present as a result of its unpaired 4f electron and larger ionic radius. The former effect in particular should result in well-distributed cerium, as this will reduce repulsion between

the additional electrons, thereby leading to Zr_5Ce_1 SBUs. Taking the possible presence of this oxidation state into account may therefore be necessary in order to uncover a possible driving force behind the observed distribution of building blocks in mixed UiO-66.

Modelling Ce^{3+} requires first reviewing the way in which $Zr_{6-x}Ce_x$ clusters are capped. When all metal ions are in the 4+ state, half of this positive charge is compensated for by the carboxylic acids (modulator or BDC), and the rest by bridging oxygens and hydroxyls. Upon reducing a cerium atom, one equivalent of negative charge is no longer needed, meaning the local structure needs to be adjusted for the system to remain neutral. Unfortunately, there has been no experimental characterisation so far from which the geometry around Ce^{3+} might be inferred. Therefore, some plausible capping schemes are proposed here which maintain charge neutrality, involving species present during the synthesis. These are shown in **Figure 59**.

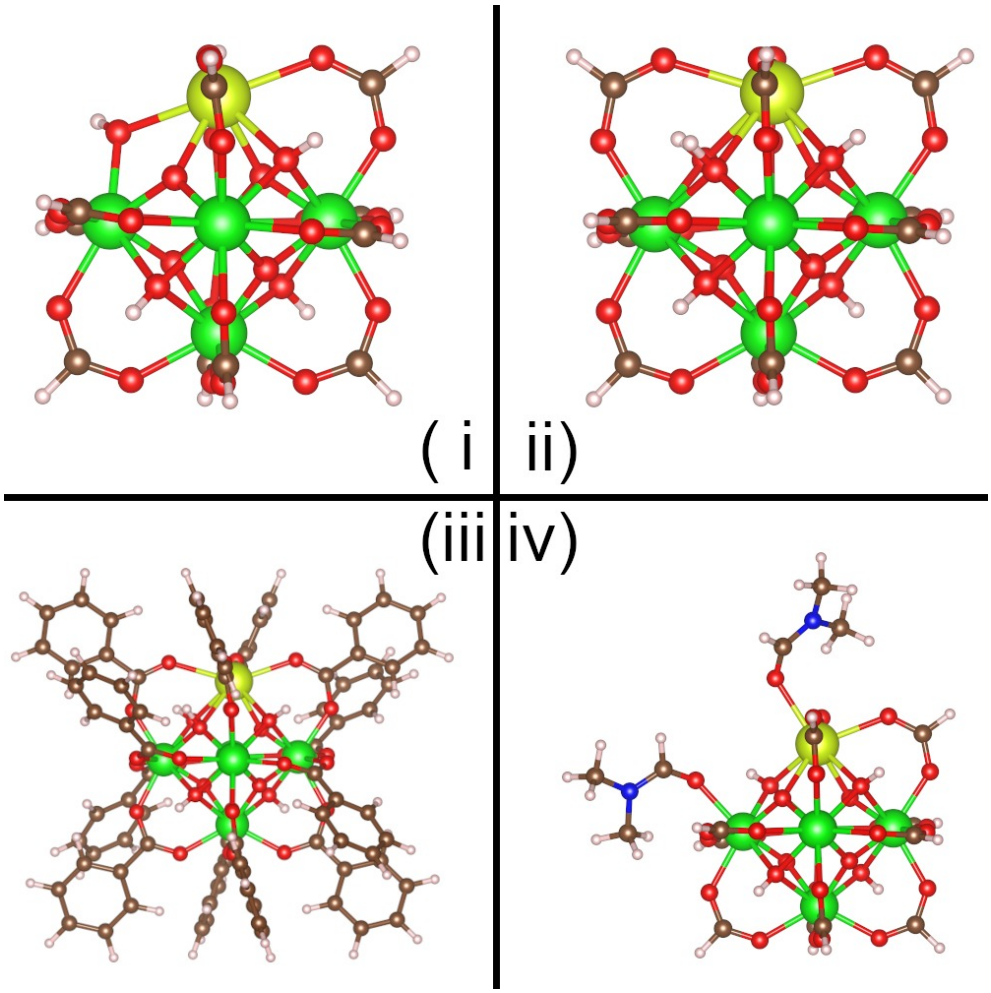


Figure 59: Proposed terminations around Ce^{3+} in a Zr_5Ce_1 cluster. Panels i-iv show, in order: no cap used, leading to the re-arrangement of a bridging hydroxyl; additional proton added to an adjacent μ_3 oxygen; same termination as ii), but with BA replacing FA; and 2 equivalents of DMF.

An additional complication when Ce^{3+} is involved is that of the reference states, the choice for which is not straightforward. Cerium (III) formate is known to form as a byproduct, but only under the right synthetic conditions following heating.[16] Hennig *et al.* have shown that in an aqueous solution of formic acid, Ce^{3+} will form monomeric complexes with average coordination numbers between 9 and 10.[24] While it is not known whether this behaviour takes place in DMF, this nonetheless appears a sensible schema for generating plausible reference states for Ce^{3+}. Candidate clusters have been generated and optimised for different combinations of

species present in solution, up to a total of 9. Relative energies have been defined according to the formula below:

$$\Delta E = E_{\text{complex}} - E_{\text{ligands}} - \min\{E_{\text{complexes}}\} \tag{100}$$

where E_{complex} is the energy of a given complex, E_{ligands} is the sum of energies of its isolated ligands, and $\min\{E_{\text{complexes}}\}$ is the minimum energy among complexes. These are shown in **Table 15**.

Table 15: Relative energies of Ce^{3+} reference states

Cluster contents					Relative energy / kJ mol^{-1}
DMF	Nitrate	Chloride	FA	BA	
6	3	0	0	0	0.0
6	2	0	1	0	36.2
6	1	0	2	0	100.8
6	0	0	3	0	107.0
6	2	0	0	1	25.5
6	1	0	0	2	56.5
6	0	0	0	3	63.3
6	0	3	0	0	31.7
6	0	2	1	0	69.5
6	0	1	2	0	110.0
6	0	2	0	1	48.4
6	0	1	0	2	64.7

Candidate reference states for Ce^{3+}. Each complex is charge neutral and provides a coordination sphere of 9 oxygens for cerium. The carboxylates are bonded in monodentate fashion. Relative energies have been defined according to **Equation 100** at the PBE0-TZVP level.

Encouragingly, the most stable reference states are those involving 0 or 1 carboxylic acids, in agreement with the work of Hennig *et al.*[24] Although the solvent in that study was water, we recover here the same qualitative trend, such that we can be confident in the proposed schema with DMF. Choosing the combination of 3 nitrates + 6 DMF, we can describe the formation of a $Zr_{6-x}Ce_x$ cluster as a reaction between one or more Ce^{3+} complexes, a monocarboxylic acid, water, and a fraction of the (monocarboxylate-terminated) Zr_6 cluster. DMF and nitric acid are released in the process. This is illustrated for the Zr_5Ce_1 building block capped by formic acid in **Equation 101**:

$$\begin{aligned} &\text{complex}(Ce^{3+}) + \frac{4}{3}\ \text{water} \\ &+ \frac{5}{6}\ \text{cluster}(Zr_6) + 2\ \text{formic acid} \end{aligned} \rightarrow \begin{aligned} &\text{cluster}(Zr_5Ce_1) + 3\ \text{nitric acid} \\ &+ 6\ \text{DMF} \end{aligned}$$

$$\begin{aligned} &Ce(NO_3)_3(C_3H_7NO)_6 + \frac{4}{3}(H_2O) \\ &+ \frac{5}{6}(Zr_6O_4(OH)_4)(HCOO)_{12} \\ &+ 2\,(HCOOH) \end{aligned} \rightarrow \begin{aligned} &Zr_5CeO_3(OH)_5(HCOO)_{12} + 3(HNO_3) \\ &+ 6(C_3H_7NO) \end{aligned} \tag{101}$$

Equation 101 can be used to define a relative energy for the mixed building block:

$$\begin{aligned} \Delta E = E_{Zr_5Ce} + 3E_{HNO_3} + 6E_{DMF} - E_{Ce^{3+}} - \frac{4}{3}E_{H_2O} \\ - \frac{5}{6}E_{Zr_6} - 2E_{FA} \end{aligned} \tag{102}$$

The terms in **Equation 102** correspond respectively to the energies of: the mixed building block; nitric acid; DMF; the Ce^{3+} complex; water; the formate-terminated Zr_6 cluster; and formic acid. This can be generalised to any monocarboxylate capping species for multiple cerium substitutions:

$$\begin{aligned} \Delta E = E_{Zr_{6-x}Ce_x} + 3xE_{HNO_3} + 6xE_{DMF} - xE_{Ce^{3+}} \\ - \frac{4x}{3}E_{H_2O} - \frac{6-x}{6}E_{Zr_6} - 2xE_{Acid} \end{aligned} \tag{103}$$

E_{acid} now refers to the energy of the monocarboxylic acid of choice, which is also used to terminate the Zr_6 and $Zr_{6-x}Ce_x$ clusters. This covers cappings ii) and iii) in **Figure 59**. For i) and iv), minor modifications are required to maintain chemical balance:

$$\begin{aligned}\Delta E = E_{Zr_{6-x}Ce_x} + 3xE_{HNO_3} + 6xE_{DMF} - xE_{Ce^{3+}} \\ - \frac{4x}{3}E_{H_2O} - \frac{6-x}{6}E_{Zr_6} - xE_{FA}\end{aligned} \tag{104}$$

$$\begin{aligned}\Delta E = E_{Zr_{6-x}Ce_x} + 3xE_{HNO_3} + 4xE_{DMF} - xE_{Ce^{3+}} \\ - \frac{4x}{3}E_{H_2O} - \frac{6-x}{6}E_{Zr_6} - xE_{FA}\end{aligned} \tag{105}$$

The full set of equations for both chemical reactions and relative energies can be found in **Appendix F**.

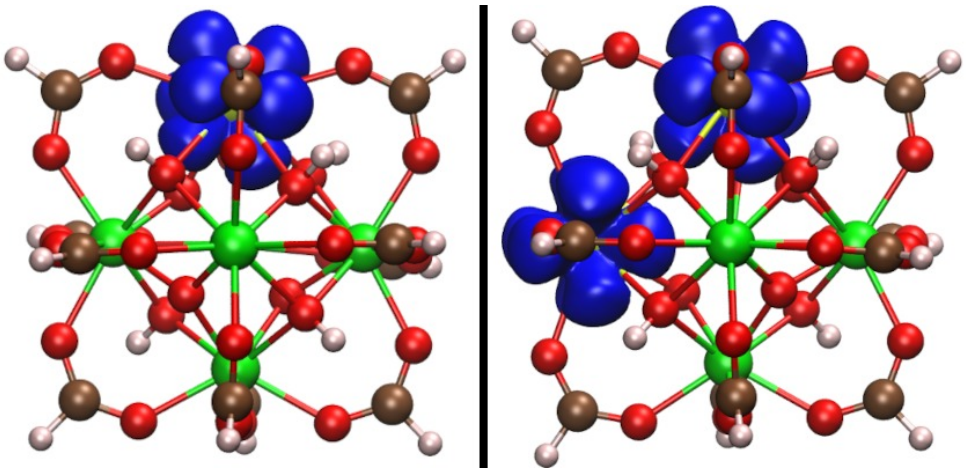


Figure 60: Spin density in Zr_5Ce_1 (left) and Zr_4Ce_2 (right) building blocks. The highly localised 4f orbitals in which the unpaired electrons reside can be clearly seen in their cubic representations. Isosurfaces have been drawn in blue using a value of 2×10^{-4} electrons Bohr^{-3}.

M_6 clusters were generated for cerium contents with $x = 1, 2$ for the cappings shown in **Figure 59**. When there are two Ce^{3+} in the same building block, the system may be modelled as a singlet or a triplet, both of which were initially tested. However, it was found that the former is prohibitively high in energy. Therefore, the results for Zr_5Ce_1 and Zr_4Ce_2, shown in **Table 16**, correspond to the triplet. For both systems, the unpaired electrons occupy well-defined localised 4f orbitals, as shown

in **Figure 60**. This suggests that the choice of method is appropriate as no delocalisation of the Ce^{3+} electron is observed, which is otherwise a common problem when modelling vacancies in materials such as CeO_2.[29] Higher Ce contents were not modelled due to computational cost.

Table 16: Relative energies of building blocks containing Ce^{3+}

Cerium distribution		Relative energy / kJ mol^{-1}			
		Uncapped	FA	BA	2DMF
Zr_5Ce_1		429.6	303.7	287.2	261.4
Zr_4Ce_2	adjacent	848.9	585.7	560.1	558.8
	opposite	830.1	619.4	586.1	566.3

Relative stabilities of open-shell clusters with Ce^{3+}. Relative energies have been defined using **Equation 103** at the PBE0-TZVP level. These are all positive, implying that the formation of these building blocks is not favourable. The different cappings correspond to those presented in **Figure 59**; clusters containing 2 cerium atoms have been modelled in the triplet spin state.

It is evident that it is not favourable for Ce^{3+} to be part of M_6 building blocks. The very high relative energies of all clusters and terminations in **Table 16** imply that this ion is much more stable in a monomeric complex, in agreement with Hennig *et al.*[24] This is also consistent with experimental reports of this oxidation state: Lammert *et al.* generated Ce^{3+} sites only through the use of a reducing species, while Nouar *et al.* reported that only 20% of ceriums in their mixed framework were in this oxidation state; these are therefore uncommon.[8,15] However, when they become part of the material, it is expected that this should be almost exclusively as part of Zr_5Ce_1 SBUS. Doubly-substituted clusters have relative energies slightly under twice as large as those containing a single

cerium; they are therefore favourable on a per ion basis, but thermodynamically less accessible given that the associated barriers will be very large. Therefore, unless there exists a mechanism through which Ce is incorporated stepwise at lower cost, this building block should be very rare. Unsurprisingly, this is exacerbated when these sites are uncapped, while FA, BA, and DMF provide increasing stabilisation. This latter species should be the majority cap for Ce^{3+}, with the acids not far behind. It should also be noted that the scheme described by **Equation 103** (and the related equations for uncapped and DMF-capped defects) involve the release of multiple equivalents of nitric acid and solvent molecules. Therefore, solvation and entropic effects should favour the formation of these mixed building blocks, such that the relative energies in **Table 16** are overestimates. These systems are therefore likely more thermodynamically accessible than presented here.

It should also be possible for cerium to change oxidation states once it becomes a part of these building blocks. The conversion of Ce^{4+} to Ce^{3+} could account for the presence of this ion in the finished framework without it necessarily being involved during synthesis. Conversely, the reverse process could explain why the majority cerium species in the material is Ce^{4+} if the assembly of building blocks is to proceed via the 3+ oxidation state. The energy required for this redox can be estimated for the formate and benzoate terminated clusters by considering the 3+ cluster as resulting from the addition of hydrogen atoms to the 4+ one:

$$Zr_{6-x}Ce_x(Ce^{4+}) + \frac{x}{2}H_2 \rightarrow Zr_{6-x}Ce_x(Ce^{3+})$$

$$Zr_{6-x}Ce_xO_4(OH)_4(acid)_{12} + \frac{x}{2}H_2 \rightarrow Zr_{6-x}Ce_xO_{4-x}(OH)_{4+x}(acid)_{12} \quad (106)$$

The associated reduction energy is then given by:

$$\Delta E_{\text{redox}} = \frac{1}{x}(E_{Zr_{6-x}Ce_x}(Ce^{3+}) - E_{Zr_{6-x}Ce_x}(Ce^{4+}) - \frac{x}{2}E_{H_2}) \quad (107)$$

where the terms on the right hand side correspond respectively to the

energies of the cluster containing Ce^{3+}, the cluster containing Ce^{4+}, and an isolated hydrogen molecule. x is the number of Ce^{3+} ions, such that the energy is normalised per cerium.

The energies of the corresponding Ce^{4+} clusters and H_2 have been evaluated using the same level of theory as the Ce^{3+} systems in order to estimate the energy required for cerium to change oxidation states. The results are shown in **Table 17**.

Table 17: Redox energies of cerium clusters

System		Reduction energy per cerium / kJ mol^{-1}
FA	Zr_5Ce_1	-98.6
	Zr_4Ce_2-adjacent	-113.2
	Zr_4Ce_2-opposite	-98.7
BA	Zr_5Ce_1	-119.7
	Zr_4Ce_2-adjacent	-131.5
	Zr_4Ce_2-opposite	-117.3

Energies required to reduce cerium from its 4+ to its 3+ state in a mixed Zr-Ce building block. These have been evaluated at the PBE0-TZVP level according to **Equation 107**. Only the formate and benzoate terminations have been considered since the 4+ and 3+ structures are straightforwardly related by one or more protons.

It can be seen in **Table 17** that the redox of cerium within the $Zr_{6-x}Ce_x$ building block should be thermodynamically feasible given relative energies for both oxidation states within ∼100 kJ mol^{-1}. The 3+ state is also more accessible when the cerium atoms are adjacent rather than isolated, suggesting it is easier to accommodate the unpaired electrons in this configuration. One can therefore expect changes in cerium's oxidation state after assembly of the cluster, such that Ce^{3+} building blocks

may be the majority species before the framework begins to grow. It is also possible that the 3+ ions evidenced through the XANES work of Nouar *et al.* were not involved in the synthesis and were generated after the growth of the framework.[8] However, this aperiodic model is more flexible than the periodic material such that the change in ionic radius between 3+ and 4+ does not lead to significant strain. Additionally, the thermodynamic metric presented here does not provide any information about the kinetics of this redox nor about the species which might provide the hydrogen atoms required. Ultimately this data only points out that post-assembly and post-synthesis changes in oxidation state should be possible and that these should therefore be considered in any explanation for the distribution of $Zr_{6-x}Ce_x$ building blocks.

While this data is encouraging in some regards, it does not conclusively explain the experimental trends in SBUs and sub-stoichiometric cerium uptake. It is after all possible to synthesise pure Ce UiO-66 using a Ce(III) precursor, while the heterometallic framework usually uses Ce(IV). However, the balance of Ce^{3+} and Ce^{4+} in solution should depend on the precursor, solvent, and any additives, which all affect the starting oxidation state as well as the rates of oxidation and reduction of cerium. The use of DMF, known to reduce Ce^{4+} to Ce^{3+}, may make the latter the majority species at high zirconium ratios, leading primarily to Zr_5Ce_1 SBUs. As the ratio of cerium increases, eventually enough Ce^{4+} persists to form Ce_6 and a minority of other mixed clusters, such that the progression of inorganic cornerstones follows the TCM. This explanation would be difficult to reconcile with the easy synthesis of pure-Ce UiO-66, and the scale of energies in **Table 16** instead suggests that Zr-UiO-66 should instead form with no cerium uptake, this metal instead remaining in solution. Changes in the cerium oxidation state following the assembly of building blocks or in the framework provides an additional means through which the majority cerium species can change during the course and after the synthesis, but this also does not explain how easily the pure

cerium material can be made. Therefore, fully explaining the propensity for Zr_5Ce_1 units will likely require taking into account other factors which have not been considered in this study.

5.4 Conclusions

This chapter has attempted to rationalise the distribution of SBUs observed in mixed cerium-zirconium UiO-66. The material seems to contain limited types of building blocks and exhibits structural and stability properties in line with the progression $Zr_6 \rightarrow Zr_6 - Zr_5Ce_1 \rightarrow Zr_5Ce_1 - Ce_6 \rightarrow Ce_6$ as the ratio of Ce precursor is increased. To this end, a series of static DFT experiments have been carried out on both aperiodic cluster models and periodic unit cells with different cerium contents and oxidation states.

Initial calculations were aimed at examining different distributions of Ce^{4+} with the aim of uncovering possible thermodynamic trends. Results for aperiodic building blocks agreed with the previous findings of Trousselet *et al.*, namely that heterometallic bonds are unfavourable and any amount of mixing lowers stability.[21] Nevertheless for clusters this penalty is small, for both FA and BA, with the possible configurations of the $Zr_{6-x}Ce_x$ hexamer representing a range only slightly larger than thermal energy. The low cost of mixing can in this case be attributed to the unconstrained flexibility enjoyed by the aperiodic model, which allows for the larger radius of Ce^{4+} to be straightforwardly accommodated for. It is expected from this data that the finished framework should contain primarily Zr_6 and Ce_6, along with an even spread of mixed SBUs no matter the ratio of cerium to zirconium precursor, unlike what is seen experimentally.

Subsequently, this analysis was extended to the periodic material. Given the highly connected **fcu** topology, the framework is stiffer and cerium's larger radius is not as easily adapted to. This manifests itself as a dif-

ference in the optimal metal-linker bond length, which cannot simultaneously be satisfied for both Zr and Ce in the pristine structure. This resulted in much larger energy costs for the introduction of cerium, which reached values of 100 to 250 kJ mol^{-1} for intermediate contents. Again these calculations showed that these metals prefer not to mix, with configurations in which cerium is aggregated incurring lower penalties. The progression of Zr_6, Zr_5Ce_1, and Ce_6 SBUs seen experimentally was also evaluated, which revealed that these are thermodynamically comparable to but not preferred over the most favourable arrangements. Repeating these calculations with 8 defects in the unit cell decreased the cost of mixing threefold, reflecting an increase in framework flexibility that benefits cerium. While this did not alter the performance of Zr_5Ce_1 compared to other building blocks, it did reveal a structure with a relative energy of zero, highlighting that the distributions of cerium and vacancies are coupled. This suggests that either could be tuned by altering synthetic parameters that affect the other.

Given that Ce^{3+} can be present during synthesis and has been identified in the finished product, this oxidation state was taken into account in a second set of DFT calculations. This species has a larger ionic radius and additional electron density, both of which should impact how it would be distributed in UiO-66 compared to its 4+ counterpart. Two additional questions were considered here: can Ce^{3+} be included in the framework (and what would be the cost of doing so?), and would its presence account for the predominance of Zr_5Ce_1 SBUs. Different plausible ways in which the coordination sphere of a Ce^{3+} ion might be filled were proposed and evaluated for a subset of aperiodic clusters. This revealed that this ion can only be included in the material at very high cost, a finding in line with experimental reports of sub-stoichiometric uptake and the low prevalence of this ion in the mixed material. Additionally, this analysis showed Zr_5Ce_1 units to be thermodynamically easier to access than Zr_4Ce_2, albeit on such an energy scale that neither should be present

in large concentrations. The energy required to reduce Ce^{4+} to Ce^{3+} in these systems was also estimated. This revealed that this process should be feasible post-synthesis, providing an additional mechanism through which the presence of the 3+ ion might be explained in the framework.

Overall, although a range of possible thermodynamic roots have been explored for the abundance of Zr_5Ce_1 bricks in mixed UiO-66, none conclusively explain the trends borne out by experiment. The results involving Ce^{3+} are promising but not without shortcomings: benchmarking data is not unequivocal about the choice of functional, while a number of assumptions have been made regarding charge compensation and reference states. In particular, a more thorough understanding of the behaviour of Zr^{4+}, Ce^{4+}, and Ce^{3+} in solution, including their mixture, would enable more robust computational experiments on UiO-66 building blocks. While a range of clusters has been reported for zirconium in solution,[54] less is known when it comes to cerium. This is especially important given the results involving the defective material and the 3+ ion, for which further study would involve changing the composition of the synthesis solution. The use of additional experimental techniques would also benefit modelling efforts by providing more information on the mixed material's electronic structure and geometry, particularly when Ce^{3+} is expected to be present. It would for example be possible to confirm whether the framework contains unpaired electrons through electron paramagnetic resonance spectroscopy, which would nicely complement the XANES results of Nouar *et al.*[8]

References

(1) S. Jakobsen, D. Gianolio, D. S. Wragg, M. H. Nilsen, H. Emerich, S. Bordiga, C. Lamberti, U. Olsbye, M. Tilset and K. P. Lillerud, *Phys. Rev. B - Condens. Matter Mater. Phys.*, 2012, **86**, 125429.

(2) C. Falaise, C. Volkringer, J. F. Vigier, N. Henry, A. Beaurain and T. Loiseau, *Chem. Eur. J.*, 2013, **19**, 5324–5331.

(3) C. Falaise, J. S. Charles, C. Volkringer and T. Loiseau, *Inorg. Chem.*, 2015, **54**, 2235–2242.

(4) A. M. Hastings, D. Ray, W. Jeong, L. Gagliardi, O. K. Farha and A. E. Hixon, *J. Am. Chem. Soc.*, 2020, **142**, 9363–9371.

(5) H. G. T. Nguyen, L. Mao, A. W. Peters, C. O. Audu, Z. J. Brown, O. K. Farha, J. T. Hupp and S. T. Nguyen, *Catal. Sci. Technol.*, 2015, **5**, 4444–4451.

(6) K. Wang, C. Li, Y. Liang, T. Han, H. Huang, Q. Yang, D. Liu and C. Zhong, *Chem. Eng. J.*, 2016, **289**, 486–493.

(7) J. G. Santaclara, A. I. Olivos-Suarez, A. Gonzalez-Nelson, D. Osadchii, M. A. Nasalevich, M. A. Van Der Veen, F. Kapteijn, A. M. Sheveleva, S. L. Veber, M. V. Fedin, A. T. Murray, C. H. Hendon, A. Walsh and J. Gascon, *Chem. Mater.*, 2017, **29**, 8963–8967.

(8) F. Nouar, M. I. Breeze, B. C. Campo, A. Vimont, G. Clet, M. Daturi, T. Devic, R. I. Walton and C. Serre, *Chem. Commun.*, 2015, **51**, 14458–14461.

(9) A. M. Ebrahim and T. J. Bandosz, *ACS Appl. Mater. Interfaces*, 2013, **5**, 10565–10573.

(10) Y. Lee, S. Kim, J. K. Kang and S. M. Cohen, *Chem. Commun.*, 2015, **51**, 5735–5738.

(11) J. Tu, X. Zeng, F. Xu, X. Wu, Y. Tian, X. Hou and Z. Long, *Chem. Commun.*, 2017, **53**, 3361–3364.

(12) *Catalysis by Ceria and Related Materials*, ed. A. Trovalleri and P. Fornasiero, Imperial College Press, London, 2nd edn., 2013.

(13) R. Di Monte and J. Kašpar, *J. Mater. Chem.*, 2005, **15**, 633–648.

(14) M. Lammert, C. Glißmann and N. Stock, *Dalt. Trans.*, 2017, **46**, 2425–2429.

(15) M. Lammert, M. T. Wharmby, S. Smolders, B. Bueken, A. Lieb, K. A. Lomachenko, D. De Vos and N. Stock, *Chem. Commun.*, 2015, **51**, 12578–12581.

(16) M. Stawowy, M. Róziewicz, E. Szczepańska, J. Silvestre-Albero, M. Zawadzki, M. Musioł, R. Łuzny, J. Kaczmarczyk, J. Trawczyński and A. Łamacz, *Catalysts*, 2019, **9**, 309.

(17) K. A. Lomachenko, J. Jacobsen, A. L. Bugaev, C. Atzori, F. Bonino, S. Bordiga, N. Stock and C. Lamberti, *J. Am. Chem. Soc.*, 2018, **140**, 17379–17383.

(18) C. Atzori, K. A. Lomachenko, J. Jacobsen, N. Stock, A. Damin, F. Bonino and S. Bordiga, *Dalt. Trans.*, 2020, **49**, 5794–5797.

(19) R. D. Shannon, *Acta Crystallogr. Sect. A*, 1976, **32**, 751–767.

(20) J. H. Cavka, S. Jakobsen, U. Olsbye, N. Guillou, C. Lamberti, S. Bordiga and K. P. Lillerud, *J. Am. Chem. Soc.*, 2008, **130**, 13850–13851.

(21) F. Trousselet, A. Archereau, A. Boutin and F. X. Coudert, *J. Phys. Chem. C*, 2016, **120**, 24885–24894.

(22) J. Jacobsen, A. Ienco, R. D'Amato, F. Costantino and N. Stock, *Dalt. Trans.*, 2020, **49**, 16551–16586.

(23) Z. Hu, Y. Wang and D. Zhao, *Chem. Soc. Rev.*, 2021, **50**, 4629–4683.

(24) C. Hennig, A. Ikeda-Ohno, W. Kraus, S. Weiss, P. Pattison, H. Emerich, P. M. Abdala and A. C. Scheinost, *Inorg. Chem.*, 2013, **52**, 11734–11743.

(25) J. Autschbach, *J. Chem. Phys.*, 2012, **136**, 150902.

(26) M. Dolg and X. Cao, *Chem. Rev.*, 2012, **112**, 403–480.

(27) J. B. Lu, D. C. Cantu, M. T. Nguyen, J. Li, V. A. Glezakou and R. Rousseau, *J. Chem. Theory Comput.*, 2019, **15**, 5987–5997.

(28) N. V. Skorodumova, R. Ahuja, S. I. Simak, I. A. Abrikosov, B. Johansson and B. I. Lundqvist, *Phys. Rev. B - Condens. Matter Mater. Phys.*, 2001, **64**, 1151081–1151089.

(29) M. Nolan, S. Grigoleit, D. C. Sayle, S. C. Parker and G. W. Watson, *Surf. Sci.*, 2005, **576**, 217–229.

(30) S. Fabris, G. Vicario, G. Balducci, S. De Gironcoli and S. Baroni, *J. Phys. Chem. B*, 2005, **109**, 22860–22867.

(31) M. D. Krcha and M. J. Janik, *Int. J. Quantum Chem.*, 2014, **114**, 8–13.

(32) M. Huang and S. Fabris, *J. Phys. Chem. C*, 2008, **112**, 8643–8648.

(33) B. Himmetoglu, A. Floris, S. De Gironcoli and M. Cococcioni, *Int. J. Quantum Chem.*, 2014, **114**, 14–49.

(34) H. J. Kulik, *J. Chem. Phys.*, 2015, **142**, 240901.

(35) J. L. Da Silva, M. V. Ganduglia-Pirovano, J. Sauer, V. Bayer and G. Kresse, *Phys. Rev. B - Condens. Matter Mater. Phys.*, 2007, **75**, 19–24.

(36) C. W. Castleton, J. Kullgren and K. Hermansson, *J. Chem. Phys.*, 2007, **127**, 244704.

(37) D. Lu and P. Liu, *J. Chem. Phys.*, 2014, **140**, 084101.

(38) R. Das, R. Sarma and J. B. Baruah, *Inorg. Chem. Commun.*, 2010, **13**, 793–795.

(39) H. Xu, S. Sommer, N. L. N. Broge, J. Gao and B. B. Iversen, *Chem. Eur. J.*, 2019, **25**, 2051–2058.

(40) D. Koch, S. Manzhos and M. Chaker, *J. Phys. Chem. A*, 2022, **126**, 3604–3611.

(41) J. P. Allen and G. W. Watson, *Phys. Chem. Chem. Phys.*, 2014, **16**, 21016–21031.

(42) J. Spencer and A. Alavi, *Phys. Rev. B - Condens. Matter Mater. Phys.*, 2008, **77**, 1–4.

(43) M. Guidon, J. Hutter and J. VandeVondele, *J. Chem. Theory Comput.*, 2009, **5**, 3010–3021.

(44) S. Grimme, J. Antony, S. Ehrlich and H. Krieg, *J. Chem. Phys.*, 2010, **132**, 154104.

(45) S. Goedecker and M. Teter, *Phys. Rev. B - Condens. Matter Mater. Phys.*, 1996, **54**, 1703–1710.

(46) J. VandeVondele and J. Hutter, *J. Chem. Phys.*, 2007, **127**, 114105.

(47) K. R. Hahn, M. Iannuzzi, A. P. Seitsonen and J. Hutter, *J. Phys. Chem. C*, 2013, **117**, 1701–1711.

(48) A. W. Thornton, R. Babarao, A. Jain, F. Trousselet and F. X. Coudert, *Dalt. Trans.*, 2016, **45**, 4352–4359.

(49) L. Liu, Z. Chen, J. Wang, D. Zhang, Y. Zhu, S. Ling, K. W. Huang, Y. Belmabkhout, K. Adil, Y. Zhang, B. Slater, M. Eddaoudi and Y. Han, *Nat. Chem.*, 2019, **11**, 622–628.

(50) L. D'Arras, C. Sassoye, L. Rozes, C. Sanchez, J. Marrot, S. Marre and C. Aymonier, *New J. Chem.*, 2014, **38**, 1477–1483.

(51) R. M. Noyes and J. J. Jwo, *J. Am. Chem. Soc.*, 1975, **97**, 5431–5433.

(52) Y. Malyukin, V. Klochkov, P. Maksimchuk, V. Seminko and N. Spivak, *Nanoscale Res. Lett.*, 2017, **12**, 1–7.

(53) S. Smolders, K. A. Lomachenko, B. Bueken, A. Struyf, A. L. Bugaev, C. Atzori, N. Stock, C. Lamberti, M. B. Roeffaers and D. E. De Vos, *ChemPhysChem*, 2018, **19**, 373–378.

(54) Y. Zhang, F. de Azambuja and T. N. Parac-Vogt, *Coord. Chem. Rev.*, 2021, **438**, 213886.

6 Conclusions and Outlook

The objective of this work was to progress the understanding of defects in UiO-66. This was done through DFT experiments directed along 3 main axes: the correlation and aggregation of vacancies into nanodomains; the impact of missing linkers and clusters on acidic properties; and the distribution of metals in the cerium-substituted material. While some of the questions presented in each section have remained unanswered, this research has produced useful conclusions and recommendations for future studies, including some considerations for the modelling of UiO-66. These key results are recounted and the outlook for future work is discussed in this chapter.

The aim of **Chapter 3** was the rationalisation of defective nanodomains in UiO-66. Initially, simple missing linker, **bcu** and **reo** structures were considered across a range of charge-compensation schemes. DFT calculations revealed that the energetics of caps are controlled in part by how they bond to the SBU, and in part by dispersion interactions with the rest of the framework. These factors supplement the commonly used arguments of modulator pKA and steric size used to justify trends seen experimentally for carboxylic acids, notably in the work of Shearer *et al.*[1–3] Nevertheless, the data implies that neither **bcu** nor **reo** are thermodynamically preferable no matter how the defects are terminated, and that there is no low-energy pathway to their formation. This contrasts with the findings of Bristow *et al.*, who previously reported that such a route exists for at least AA and Cl–H_2O.[4] The known importance of dispersion forces, modelled via the D3 semi-empirical correction in this study (compared with a Lennard-Jones potential fit to *ab initio* data in Bristow *et al.*), may be responsible for this disagreement. This highlights the potential need for a comparison of different van der Waals schemes applied to this material, though existing benchmarking on the breathing MOF MIL-53 suggests that the method used here should be adequate.[5]

This initial dataset was refined by applying the corrections necessary to convert the electronic results to free energies, allowing for a more quantitatively sound comparison of different charge-compensation schemes. With these updated results, BA was established to be the best species for terminating defects, as found by Atzori *et al.*,[3] while the other carboxylic acids did not follow experimental trends. It was also found that BA and TFAA favoured missing linkers over clusters, contrary to experimental findings; the possibility of a large energy barrier for adopting the optimal structure at a linker vacancy capped by these species was proposed to explain this discrepancy. It was also confirmed that geometries involving undercoordinated Zr are highly unfavourable, but they can be stabilised by the addition of neutral water. This is in line with both the high temperatures needed for activation or dehydroxylation and the reversible nature of these processes when the material is exposed to moisture.[6–8] In addition the Cl–H_2O and FA–H_2O caps were shown to be competitive with carboxylic acids for the formation of both missing linkers and clusters, a result consistent with previous computational studies.[4,9] The lack of experimental reports of **bcu** and **reo** with these caps was explained here in two ways: the synthesis may lack sufficient water for the formation of both SBUs and these defect terminations, and the bimolecular nature of these caps implies the need for multiple molecular collisions. Finally, it was found across all species that the **reo** structure is not favourable over isolated missing linkers, implying that taking into account differences in vibrational behaviour between these structures does not change the overall thermodynamic picture.

The interfaces between different regions were also investigated through supercell calculations. For FA, AA, and BA the associated energy penalties for mixing **fcu** and **reo** cells is close to zero, while it is of the same order of magnitude as thermal energy for TFAA. This implies that a range of defect distributions should be observed rather than nanodomains, which minimise interfacial areas. The additional case of **bcu** unit cells

capped by FA was also considered, which revealed energy penalties comparable to those seen for **reo**. The impact of removing multiple linkers from the same SBU was also investigated by enumerating the energies of FA-capped linker vacancies around a single inorganic node in the periodic system. Only a minor increase in energy cost was found when removing consecutive terephthalates, implying minimal spatial correlation between formate-terminated defects. This analysis was repeated for aperiodic building blocks using FA, AA, TFAA, and BA. It was found in this way that when linkers are allowed to dangle, they drift towards each other to form favourable slipped-triangle trimers. The resulting balance of structural distortions and stabilising dispersion interactions leads to a decrease in defect energy as more terephthalates are removed. This is especially notable for BA and TFAA, for which this penalty rapidly becomes comparable to thermal energy. Nevertheless, this enumeration did not yield any preference for the type of SBU seen in the **bcu** or **reo** structures.

A number of questions remain open regarding the work carried out in **Chapter 3**, presenting possibilities for future work. Among these is the role of interfaces between defective nanoregions. $2 \times 2 \times 2$ supercells are relatively small compared to the domains observed by Liu *et al.* and Johnstone *et al.*,[10,11] yet with 3500 atoms, they are already large for the use of DFT. An alternative way forward is to use force fields, of which Rogge *et al.*'s QuickFF parametrisation for UiO-66 is the most promising.[12,13] A collaboration remains underway with this group within which systems up to $5 \times 5 \times 5$ are being examined with the use of the GPU-accelerated OpenMM package.[14] With this method, more realistic distributions of supercells can be simulated to evaluate differences in free energies, some examples of which are shown in **Figure 61**. This study should provide more robust conclusions to the influence of domain interfaces on the distribution of defects in UiO-66.

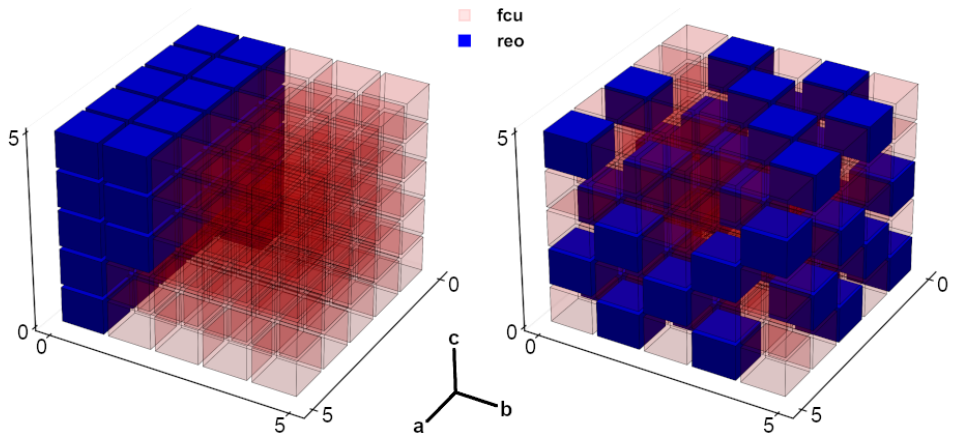


Figure 61: Schematic representations of the types of defect distributions which can be studied by increasing the system size to a $5 \times 5 \times 5$ supercell. The left and right panels respectively depict two extremes: a lamellar **reo** nanodomain, and fully distributed missing cluster defects. Even with over 50,000 atoms, calculation runtimes are only of the order of a few hours.

Another aspect of this work which may be followed up on is that of dispersion corrections, the importance of which was discussed repeatedly in **Chapter 3**. The D3 scheme was used throughout this work, and although it is known to be robust and transferable, it can underperform when used in conditions which deviate from those in which coefficients were fitted.[15,16] Discrepancies between some of the results in **Section 3.3.1** and those of Bristow *et al.*,[4] who used a different scheme, also suggest that future studies on defects would benefit from thorough benchmarking specific to UiO-66. This should be carried out over a representative subset of structures involving both small and large capping species as well as both periodic and aperiodic models. A method worth testing that naturally expands over D3 is the latest iteration, D4; this calculates pairwise coefficients on-the-fly using tight-binding derived partial atomic charges.[17] A number of functionals may also be evaluated together with the vdW kernels of Dion *et al.* or Vydrov and Van Voorhis, including the re-parametrised optPBE and optB88 variants of Klimeš *et al.*[18–20] While

it would be difficult to generate reliable and accurate data against which other methods can be compared, the work of Wieme *et al.* suggests that results found using the random-phase approximation to the correlation energy would be suitable.[5] Given the computational cost of this method, it might prove necessary to carry out the benchmarking on smaller fragments of the material or on the primitive unit cell.

Finally, the building blocks of UiO-66 can be studied in more depth. When aperiodic models involving dangling linkers were optimised in **Section 3.3.4**, these were found to distort so as to form dispersion-dominated slipped triangular trimers. The concerns raised previously regarding the choice of van der Waals correction scheme are therefore also valid here, but it is also important to consider whether these distortions would be possible under experimental conditions. Solvent molecules might act to both sterically hinder the motion of terephthalates and screen the long-range interactions that pull them towards each other. An initial MD simulation involving DMF solvent molecules could be carried out to verify whether this is the case. Follow-up experiments could then be run to evaluate whether solvation significantly changes the thermodynamic picture already drawn for clusters in vacuum. Quantum-mechanics/molecular mechanics (QMMM) would be a suitable way to offset the increase in computational cost since a number of solvent molecules will occupy corners of the simulation box in a spectator role. Larger building blocks may also be considered: Xu *et al.* have reported that Zr_6 octahedra can assemble into octahedral oligomers prior to any heating,[21] as shown in **Figure 62**. It is interesting to note in particular that the layer-wise addition of such structural elements would leave regular gaps in the framework, resulting in the **reo** topology.

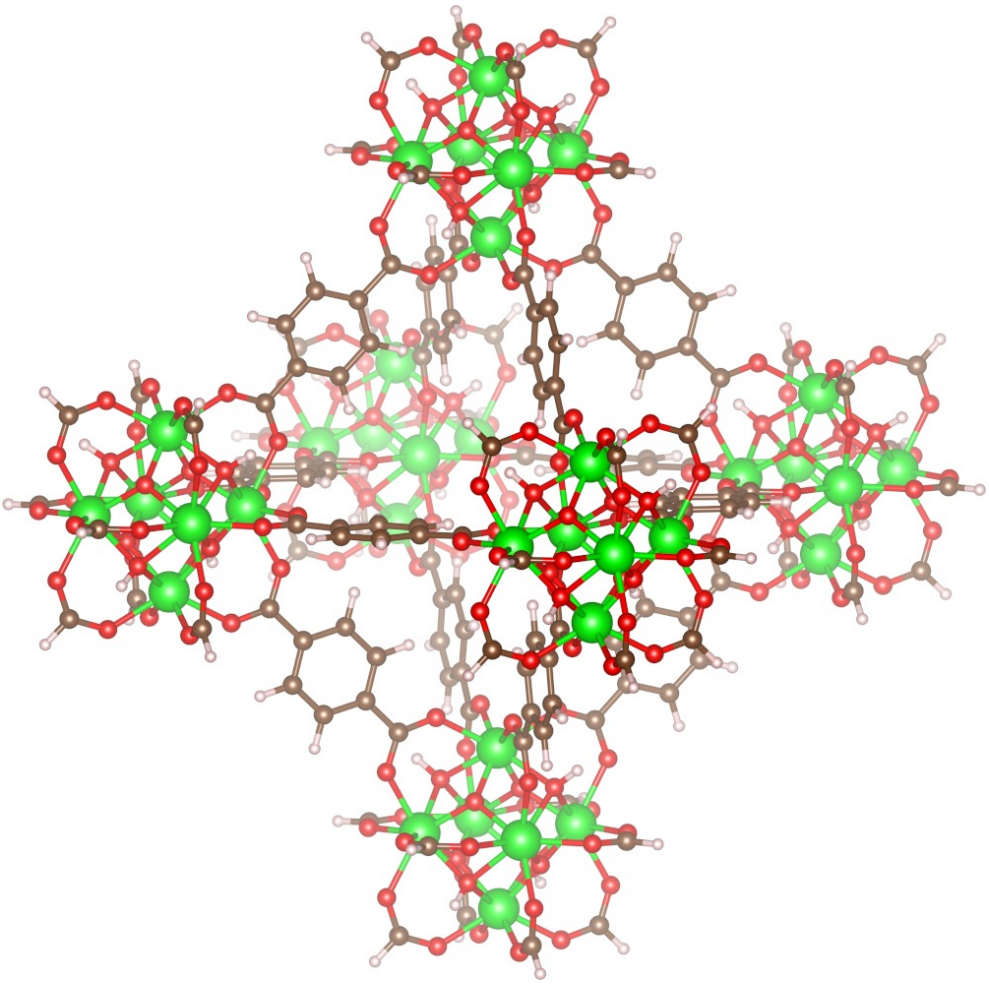


Figure 62: Octahedral oligomer of UiO-66 FA-terminated building blocks proposed by Xu *et al.*[21] This model has been fitted to X-ray pair distribution data collected *in situ* after mixing the reagents but prior to heating in the synthesis of UiO-66. Other oligomers match the data well, but all are derived from the octahedral geometry shown here.

These techniques would also prove useful for periodic calculations on the free energies of vacancies. Running MD simulations in which the unit cells have been filled with solvent would allow differences in solvation effects between capping species to be evaluated. In this way, the vibrational, rotational, and translational degrees of freedom of the framework and capping species molecules would also be properly damped by the solvent, providing a better representation of the experimental system than with gas-phase reference states. Again, given the size of the unit cell of UiO-66, this may require the use of QMMM to mitigate the increase in

computational cost associated with the larger total number of atoms.

In **Chapter 4**, the influence of defects on the acidic properties of the framework was investigated. Efforts were initially directed towards Brønsted acid sites, present in both the pristine and defective materials as μ_3 hydroxyl groups. Probe molecules, ammonia and pyridine, were deposited atop these sites and their binding energies were evaluated for a range of missing linker structures. While results initially suggested that the presence of certain moieties enhanced acidity, closer inspection revealed that increased binding strength could instead be attributed to secondary interactions between the probe and the framework. Dispersion forces contributed to differences across capping species for both probe molecules; ammonia also benefitted from additional hydrogen bonding, and pyridine from π-π overlaps with adjacent linkers. While deconvoluting these effects from changes in the acidity of the bridging hydroxyl proton is not straightforward, this analysis at least highlighted the role that charge-compensating species might play in catalysis, in which they enhance binding near the active site and promote certain conformations.

Similar experiments were carried out using carbon monoxide, during which changes to its signature stretching frequency were tracked. It was found that adsorption atop the native hydroxyl groups is generally not very sensitive to the identity of the nearby capping species. However, the additional water and hydroxide moieties which can be added to the framework through certain defects can provide additional Brønsted acidic protons. Therefore, while opportunities for tuning acid strength are limited, the density of active sites for catalysis can be increased via vacancies, the presence of which can be influenced via synthetic parameters. In addition, it was also confirmed that both Brønsted and Lewis sites can co-exist in close proximity to each other without any significant modifications to the synthesis. This has implications for catalytic reactions in which a cooperative effect between both types of acid is

beneficial, such as the esterification of levulinic acid.[22,23] Catalyst design becomes simpler since the linker no longer has to be functionalised to introduce Brønsted sites, and this part of the design space can be used for other purposes. Notably, Zhang *et al.* have instead modified the linker to make UiO-66 a stronger Lewis acid for the conversion of glucose to 5-hydromethylfurfural, which also requires Brønsted sites, improving in the process upon previous multivariate MOFs.[24]

The Lewis acidic properties of UiO-66 were investigated further by modelling the cyclisation of citronellal to isopulegol within different defective structures. This was broken down into two steps, wherein a carboxylate group is first moved to a partially dangling position to free up space around zirconium, which is then followed by cyclisation. The overall barrier to this process was found to be large for **fcu** and missing linkers capped by FA or TFAA, but small for **reo** and **e** defects. For the pristine material, high stiffness discourages terephthalates from decoordinating, whereas this is easier with monocarboxylates. However, limited space around Zr when charge-compensating species are present generally makes the second step unfavourable except for the more open missing cluster and **e** structures. These results are consistent with experimental reports of improved performance with these types of vacancies and the use of harsh activation procedures to remove terminating carboxylates. While concerns of steric hindrance will be less significant for reactants smaller than citronellal, the results nevertheless confirm that efforts to tune catalytic performance should be focused on the **reo** topology. In this regard, some of the approaches described in **Section 1.4** and **Section 3.1** for generating missing clusters may prove particularly useful.

There are a number of ways in which the current work on Lewis acidity can be extended. The dangling state examined here is but one of a number of ways in which open sites might be generated: Vandichel *et al.* and Marreiros *et al.* have proposed alternate geometries that might

be adopted upon heating or during the catalysis of other reactions.[25,26] The possibilities increase in the presence of solvent, with Caratelli *et al.* having reported a number of coordination states during metadynamics calculations involving water.[27] This latter approach is the most interesting, as it naturally includes both vibrational effects and solvation to explore the free energies of different Zr environments. Notably, it is also possible in this way to study caps containing multiple molecules, such as Cl–H_2O and FA–H_2O, of interest due to their low defect free energies, for which the method used in **Section 4.3.3** was not suitable. The solvent in such a study would have to be chosen carefully: in order to significantly affect the barriers in a catalytic reaction, it would need to be polar so that it can solvate Zr and dangling carboxylate oxygens. This rules out the cyclisation of citronellal, since this works best in apolar media,[28] but other test reactions for Lewis acidity such as the esterification of levulinic acid and the isomerisation of glucose to fructose can take place in alcohol.[29,30] Populating the unit cells with solvent molecules will increase the computational cost, but again some compromise might be reached through the use of QMMM.

Another way in which this study may be expanded is through the effect that vacancies have on transport properties. While not directly related to acidity, the ease with which reactants can diffuse through the material can have important consequences on catalytic activity. Missing linkers and clusters both increase how open the framework is, but the extent to which this is influenced by the capping species is not known. It is expected that a reactant will be less sterically hindered when moving through defects terminated FA than bulkier TFAA or BA, while specific interactions with species like H_2O–OH may inhibit movement. Such a study would have to be carried out using force fields given the long MD durations needed to estimate diffusion coefficients, but a blueprint can be taken from the work of Wardzala *et al.*[31] The authors studied the diffusion of acetone in pristine UiO-66 and concluded that modelling the

framework as flexible is necessary, with Rogge *et al.*'s QuickFF proving the best choice.[12] This methodology's automatic fitting procedure could be used to fit parameters for the defective material capped by various species. Force field terms could be constructed for the chosen reactant using the TraPPE united-atom model.[32] The resulting diffusion coefficients across different structures would allow for a more granular interpretation of experimental data on catalytic activity by separating the influences of porosity and framework acidity.

The final chapter of this work focused on the distribution of SBUs in the cerium-substituted form of UiO-66. The objective was to determine why the progressive incorporation of this metal in the material happens almost exclusively via Zr_5Ce_1 and eventually Ce_6 inorganic nodes. To do this, isolated $Zr_{6-x}Ce_x$ building blocks were initially investigated, which showed Ce-Zr bonds to be unfavourable, as was found by Trousselet *et al.*[33] The associated energy penalties for metal mixing were also found to be comparable to thermal energy; a range of mixed SBUs would therefore be expected based on thermodynamic grounds, rather than only those comprising the TCM. These calculations were repeated for the periodic unit cell, for which the cost of mixing increased by over an order of magnitude, reflecting the difficulty of accommodating Ce^{4+}'s larger ionic radius in the fully-connected material. The resulting strain in the structure could be relaxed by introducing formate-capped linker vacancies adjacent to cerium. It was additionally found that with defects present, metal configurations involving primarily Zr_6 and Ce_6 SBUs could be realised with almost no energy penalty by decoupling the different ionic radii. This implies an interdependence between the distributions of cerium and missing linkers, which might be exploited to optimise the synthesis of the Ce-Zr framework. Nevertheless, neither set of periodic calculations revealed why the TCM should be preferred over a more disordered set of building blocks.

Given that computational efforts involving Ce^{4+} ultimately did not reveal why Zr_5Ce_1 SBUs are favoured, subsequent work focused on the role that Ce^{3+} might play. As the structure of building blocks containing cerium in the 3+ state is not known, a number of geometries were proposed which maintain overall charge neutrality. The subsequent calculations on Zr_5Ce_1 and Zr_4Ce_2 clusters showed that there is a significant thermodynamic penalty associated with incorporating this ion in UiO-66 nodes which approximately scales with the number of ceriums. This favours Zr_5Ce_1 over other building blocks and is consistent with the sub-stoichiometric Ce uptake seen experimentally. However, Ce_6 units are still observed at high cerium-to-zirconium ratios, and the synthesis of pure Ce UiO-66 is straightforward, neither of which can be reconciled with these results. The data here may therefore explain why the 3+ state is uncommon, in both the pure and mixed materials, but does not clarify why the TCM is favoured. One caveat in this analysis is that relative energies have to be defined with respect to a reference state that also contains Ce^{3+}. This was done here with mononuclear complexes containing solvent and modulator ligands, but the behaviour of this ion in solution, particularly in DMF, is not well-documented.

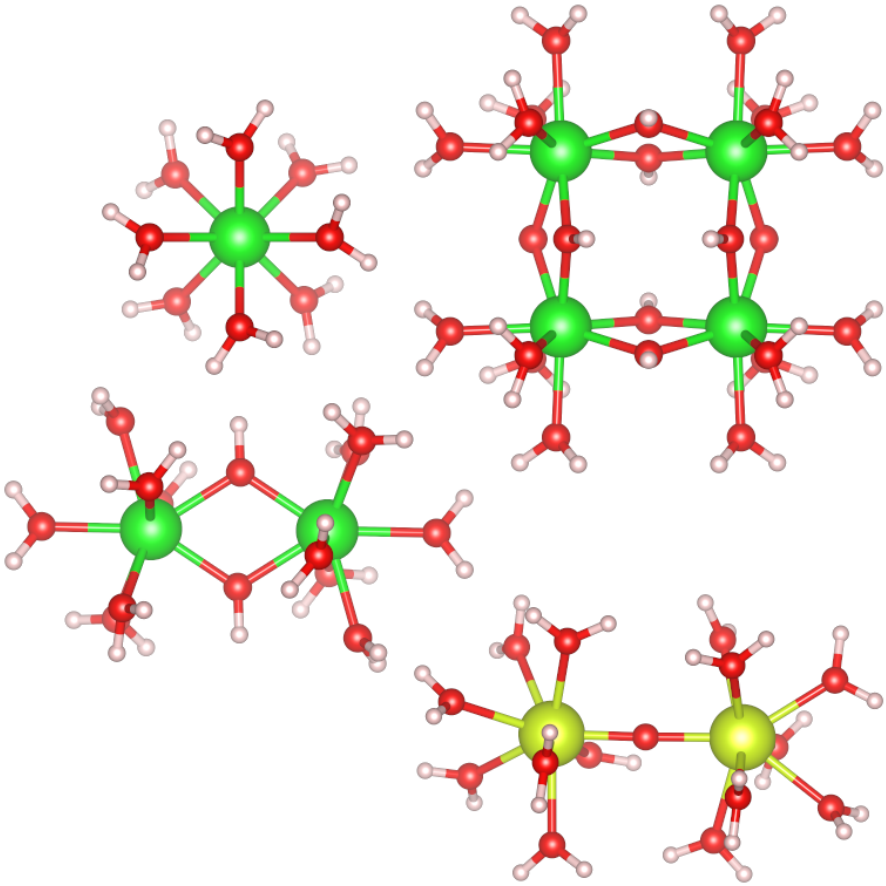


Figure 63: Some of the Zr^{4+} and Ce^{4+} complexes known to form in aqueous solutions. The clusters shown here are DFT-optimised structures from the works of Stern *et al.* (zirconium) and Ikeda-Ohno *et al.* (cerium).[34,35] It can be seen that the assembly of hexamers from dimers or tetramers would naturally limit the combinations of Ce and Zr which might be attained.

The next step in trying to determine what factors control the distribution of cerium in UiO-66 should involve a survey of what happens in solution during the synthesis. A number of assumptions underpin the results seen here with Ce^{3+}, but these might be improved on by expanding upon the work of Hennig *et al.*[36] Their use of EXAFS has already provided useful information on the behaviour of cerium in water, but this should be repeated for DMF given its known ability to reduce this metal. It is also known that in water, Zr can form a range of clusters other than Zr_6, while dinuclear complexes have also been reported for cerium, as shown

in **Figure 63**.[35,37] This suggests that the formation of $Zr_{6-x}Ce_x$ could proceed via intermediates, which would affect the distribution of Ce. By carrying out this work *in situ*, it would also be possible to track how the building blocks change before and during the synthesis, as well as to determine what materials are not incorporated into the framework. The advantages of such a study would be twofold: the data collected can be used to optimise the synthesis of the mixed framework, and it would also enable subsequent computational work of the type carried out here with a more complete view of cerium's behaviour.

To conclude, the objective of this work was to further the understanding of point defects in UiO-66. A number of DFT experiments has been carried out in order to uncover new insights into the formation of vacancies, their importance in catalysis, and the role they hold in the mixed cerium material. While some objectives have not been achieved, a number of useful results have been produced, including key considerations for the modelling of this framework and data which can be used to optimise the formation and design of linker and cluster vacancies. The work herein has also paved the way for follow-up experiments such that the current findings can be built upon with the aim of improving the performance of UiO-66 using defects.

References

(1) G. C. Shearer, S. Chavan, S. Bordiga, S. Svelle, U. Olsbye and K. P. Lillerud, *Chem. Mater.*, 2016, **28**, 3749–3761.

(2) G. C. Shearer, J. G. Vitillo, S. Bordiga, S. Svelle, U. Olsbye and K. P. Lillerud, *Chem. Mater.*, 2016, **28**, 7190–7193.

(3) C. Atzori, G. C. Shearer, L. Maschio, B. Civalleri, F. Bonino, C. Lamberti, S. Svelle, K. P. Lillerud and S. Bordiga, *J. Phys. Chem. C*, 2017, **121**, 9312–9324.

(4) J. K. Bristow, K. L. Svane, D. Tiana, J. M. Skelton, J. D. Gale and A. Walsh, *J. Phys. Chem. C*, 2016, **120**, 9276–9281.

(5) J. Wieme, K. Lejaeghere, G. Kresse and V. Van Speybroeck, *Nat. Commun.*, 2018, **9**, 4899.

(6) F. G. Xi, Y. Yang, H. Liu, H. F. Yao and E. Q. Gao, *RSC Adv.*, 2015, **5**, 79216–79223.

(7) D. M. Driscoll, D. Troya, P. M. Usov, A. J. Maynes, A. J. Morris and J. R. Morris, *J. Phys. Chem. C*, 2018, **122**, 14582–14589.

(8) G. C. Shearer, S. Forselv, S. Chavan, S. Bordiga, K. Mathisen, M. Bjørgen, S. Svelle and K. P. Lillerud, *Top. Catal.*, 2013, **56**, 770–782.

(9) K. Tan, H. Pandey, H. Wang, E. Velasco, K. Y. Wang, H. C. Zhou, J. Li and T. Thonhauser, *J. Am. Chem. Soc.*, 2021, **143**, 6328–6332.

(10) L. Liu, Z. Chen, J. Wang, D. Zhang, Y. Zhu, S. Ling, K. W. Huang, Y. Belmabkhout, K. Adil, Y. Zhang, B. Slater, M. Eddaoudi and Y. Han, *Nat. Chem.*, 2019, **11**, 622–628.

(11) D. N. Johnstone, F. C. Firth, C. P. Grey, P. A. Midgley, M. J. Cliffe, S. M. Collins and D. N. Johnstone, *J. Am. Chem. Soc.*, 2020, **142**, 13081–13089.

(12) S. M. Rogge, J. Wieme, L. Vanduyfhuys, S. Vandenbrande, G. Maurin, T. Verstraelen, M. Waroquier and V. Van Speybroeck, *Chem. Mater.*, 2016, **28**, 5721–5732.

(13) S. M. Rogge, P. G. Yot, J. Jacobsen, F. Muniz-Miranda, S. Vandenbrande, J. Gosch, V. Ortiz, I. E. Collings, S. Devautour-Vinot, G. Maurin, N. Stock and V. Van Speybroeck, *ACS Mater. Lett.*, 2020, **2**, 438–445.

(14) P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L. P. Wang, A. C. Simmonett, M. P. Harrigan, C. D. Stern, R. P. Wiewiora, B. R. Brooks and V. S. Pande, *PLoS Comput. Biol.*, 2017, **13**, 1–17.

(15) S. Ehrlich, J. Moellmann, W. Reckien, T. Bredow and S. Grimme, *ChemPhysChem*, 2011, **12**, 3414–3420.

(16) A. Hansen, C. Bannwarth, S. Grimme, P. Petrovic, C. Werlé and J. P. Djukic, *ChemistryOpen*, 2014, **3**, 177–189.

(17) E. Caldeweyher, C. Bannwarth and S. Grimme, *J. Chem. Phys.*, 2017, **147**, 034112.

(18) M. Dion, H. Rydberg, E. Schröder, D. C. Langreth and B. I. Lundqvist, *Phys. Rev. Lett.*, 2004, **92**, 22–25.

(19) O. A. Vydrov and T. Van Voorhis, *Phys. Rev. Lett.*, 2009, **103**, 7–10.

(20) J. Klimeš, D. R. Bowler and A. Michaelides, *J. Phys. Condens. Matter*, 2010, **22**, 022201.

(21) H. Xu, S. Sommer, N. L. N. Broge, J. Gao and B. B. Iversen, *Chem. Eur. J.*, 2019, **25**, 2051–2058.

(22) Y. Kuwahara, H. Kango and H. Yamashita, *ACS Sustain. Chem. Eng.*, 2017, **5**, 1141–1152.

(23) R. Wei, J. Fan, X. Qu, L. Gao, Y. Wu, Z. Zhang, F. Hu and G. Xiao, *Eur. J. Inorg. Chem.*, 2020, **2020**, 833–840.

(24) Y. Zhang, B. Zhao, S. Das, V. Degirmenci and R. I. Walton, 2022.

(25) M. Vandichel, J. Hajek, A. Ghysels, A. De Vos, M. Waroquier and V. Van Speybroeck, *CrystEngComm*, 2016, **18**, 7056–7069.

(26) J. Marreiros, C. Caratelli, J. Hajek, A. Krajnc, G. Fleury, B. Bueken, D. E. De Vos, G. Mali, M. B. Roeffaers, V. Van Speybroeck and R. Ameloot, *Chem. Mater.*, 2019, **31**, 1359–1369.

(27) C. Caratelli, J. Hajek, E. J. Meijer, M. Waroquier and V. Van Speybroeck, *Chem. Eur. J.*, 2019, **25**, 15315–15325.

(28) N. M. Bertero, A. F. Trasarti, M. C. Acevedo, A. J. Marchi and C. R. Apesteguía, *Mol. Catal.*, 2020, **481**, 110192.

(29) F. G. Cirujano and F. X. Llabrés I Xamena, *J. Phys. Chem. Lett.*, 2020, **11**, 4879–4890.

(30) M. D. de Mello and M. Tsapatsis, *ChemCatChem*, 2018, **10**, 2417–2423.

(31) J. J. Wardzala, J. P. Ruffley, I. Goodenough, A. M. Schmidt, P. B. Shukla, X. Wei, A. Bagusetty, M. De Souza, P. Das, D. J. Thompson, C. J. Karwacki, C. E. Wilmer, E. Borguet, N. L. Rosi and J. K. Johnson, *J. Phys. Chem. C*, 2020, **124**, 28469–28478.

(32) J. M. Stubbs, J. J. Potoff and J. I. Siepmann, *J. Phys. Chem. B*, 2004, **108**, 17596–17605.

(33) F. Trousselet, A. Archereau, A. Boutin and F. X. Coudert, *J. Phys. Chem. C*, 2016, **120**, 24885–24894.

(34) R. D. Stern, R. S. Kingsbury and K. A. Persson, *Inorg. Chem.*, 2021, **60**, 15456–15466.

(35) A. Ikeda-Ohno, S. Tsushima, C. Hennig, T. Yaita and G. Bernhard, *Dalt. Trans.*, 2012, **41**, 7190–7192.

(36) C. Hennig, A. Ikeda-Ohno, W. Kraus, S. Weiss, P. Pattison, H. Emerich, P. M. Abdala and A. C. Scheinost, *Inorg. Chem.*, 2013, **52**, 11734–11743.

(37) Y. Zhang, F. de Azambuja and T. N. Parac-Vogt, *Coord. Chem. Rev.*, 2021, **438**, 213886.

Appendices

A Chemical equations for defects

This appendix contains equations for each of the terminations considered for missing linkers in this work. Each one is given as a schematic on the top and a chemical reaction on the bottom.

A.1 FA - formic acid

$$\textbf{fcu} + 2 \text{ formic acid} \rightarrow \text{missing linker defect} + H_2BDC$$

$$\begin{array}{l} (Zr_6O_4(OH)_4)_4(C_8O_4H_4)_{24} \\ + 2\,HCOOH \end{array} \rightarrow \begin{array}{l} (Zr_6O_4(OH)_4)_4(C_8O_4H_4)_{23}(HCOO)_2 \\ + H_2C_8O_4H_4 \end{array}$$

A.2 AA - acetic acid

$$\textbf{fcu} + 2 \text{ acetic acid} \rightarrow \text{missing linker defect} + H_2BDC$$

$$\begin{array}{l} (Zr_6O_4(OH)_4)_4(C_8O_4H_4)_{24} \\ + 2\,CH_3COOH \end{array} \rightarrow \begin{array}{l} (Zr_6O_4(OH)_4)_4(C_8O_4H_4)_{23}(CH_3COO)_2 \\ + H_2C_8O_4H_4 \end{array}$$

A.3 TFAA - trifluoroacetic acid

$$\textbf{fcu} + 2 \text{ trifluoroacetic acid} \rightarrow \text{missing linker defect} + H_2BDC$$

$$\begin{array}{l} (Zr_6O_4(OH)_4)_4(C_8O_4H_4)_{24} \\ + 2\,CF_3COOH \end{array} \rightarrow \begin{array}{l} (Zr_6O_4(OH)_4)_4(C_8O_4H_4)_{23}(CF_3COO)_2 \\ + H_2C_8O_4H_4 \end{array}$$

A.4 BA - benzoic acid

$$\textbf{fcu} + 2 \text{ benzoic acid} \rightarrow \text{missing linker defect} + H_2BDC$$

$$\begin{array}{l} (Zr_6O_4(OH)_4)_4(C_8O_4H_4)_{24} \\ + 2\,H_5C_6COOH \end{array} \rightarrow \begin{array}{l} (Zr_6O_4(OH)_4)_4(C_8O_4H_4)_{23}(H_5C_6COO)_2 \\ + H_2C_8O_4H_4 \end{array}$$

A.5 2 H_2O–OH

$$\textbf{fcu} + 2(3\,H_2O) \rightarrow \text{missing linker defect} + H_2BDC$$

$$\begin{array}{l} (Zr_6O_4(OH)_4)_4(C_8O_4H_4)_{24} \\ + 6\,H_2O \end{array} \rightarrow \begin{array}{l} (Zr_6O_4(OH)_4)_4(C_8O_4H_4)_{23}((H_2O)_2OH)_2 \\ + H_2C_8O_4H_4 \end{array}$$

A.6 H_2O–OH

$$\mathbf{fcu} + 2\ 2\,H_2O \rightarrow \text{missing linker defect} + H_2BDC$$

$$\begin{array}{l}(Zr_6O_4(OH)_4)_4(C_8O_4H_4)_{24} \\ +\,4\,H_2O\end{array} \rightarrow \begin{array}{l}(Zr_6O_4(OH)_4)_4(C_8O_4H_4)_{23}((H_2O)(OH))_2 \\ +\,H_2C_8O_4H_4\end{array}$$

A.7 Cl–OH

$$\mathbf{fcu} + 2\ (HCl + H_2O) \rightarrow \text{missing linker defect} + H_2BDC$$

$$\begin{array}{l}(Zr_6O_4(OH)_4)_4(C_8O_4H_4)_{24} \\ +\,2HCl + 2H_2O\end{array} \rightarrow \begin{array}{l}(Zr_6O_4(OH)_4)_4(C_8O_4H_4)_{23}((Cl)(H_2O))_2 \\ +\,H_2C_8O_4H_4\end{array}$$

A.8 FA–H_2O

$$\mathbf{fcu} + 2\ (FA + H_2O) \rightarrow \text{missing linker defect} + H_2BDC$$

$$\begin{array}{l}(Zr_6O_4(OH)_4)_4(C_8O_4H_4)_{24} \\ +\,2HCOOH + 2H_2O\end{array} \rightarrow \begin{array}{l}(Zr_6O_4(OH)_4)_4(C_8O_4H_4)_{23}((HCOO)(H_2O))_2 \\ +\,H_2C_8O_4H_4\end{array}$$

A.9 Cl

$$\mathbf{fcu} + 2HCl \rightarrow \text{missing linker defect} + H_2BDC$$

$$\begin{array}{l}(Zr_6O_4(OH)_4)_4(C_8O_4H_4)_{24} \\ +\,2HCl\end{array} \rightarrow \begin{array}{l}(Zr_6O_4(OH)_4)_4(C_8O_4H_4)_{23}Cl_2 \\ +\,H_2C_8O_4H_4\end{array}$$

A.10 OH

$$\mathbf{fcu} + 2H_2O \rightarrow \text{missing linker defect} + H_2BDC$$

$$\begin{array}{l}(Zr_6O_4(OH)_4)_4(C_8O_4H_4)_{24} \\ +\,2H_2O\end{array} \rightarrow \begin{array}{l}(Zr_6O_4(OH)_4)_4(C_8O_4H_4)_{23}(OH)_2 \\ +\,H_2C_8O_4H_4\end{array}$$

A.11 e

$$\mathbf{fcu} \rightarrow \text{missing linker defect} + H_2BDC$$

$$(Zr_6O_4(OH)_4)_4(C_8O_4H_4)_{24} \rightarrow \begin{array}{l}(Zr_6O_4(OH)_4)_2(Zr_6O_5(OH)_3)_2(C_8O_4H_4)_{23} \\ +\,H_2C_8O_4H_4\end{array}$$

B Using Python to construct defects

Sample of the Python code used to generate defective structures. Modifications were written and applied as necessary for different requirements such as supercells, uncorrelated SBUs, or aperiodic clusters

```
###############################################################################
import numpy as np
import os
import sys
import networkx as nx
from datetime import date
###############################################################################

'''
Script for generating missing linker defects in Zr6 MOFs. Takes in some command
line arguments: an input .cif file (which should have P1 symmetry), a cap
coordinate .cif file (the cap attached to a bare Zr6 cluster), and some indices
for the linkers to be replaced.
The script may be run blank to generate a .cif file from which indices can
be chosen.

python3 defect-generator.py

will generate skeleton.cif file.

python3 defect-generator.py file.cif coordinates.cif 1,2,3

will replace linkers indexed as 1, 2, and 3 in file.cif with the capping species
contained in coordinates.cif
'''

#==============================================================================
#==============================================================================
#raycov_dictionary is a dictionary containing the covalent radii of
#all elements. data taken from the CSD at
#https://www.ccdc.cam.ac.uk/support-and-resources/
#ccdcresources/Elemental_Radii.xlsx
#at 18:10 on 27/11/2018
#Two further dummy radii, for linkers and clusters, are defined.
raycov_dict={
'H':0.23,'He':1.5,'Li':1.28,'Be':0.96,'B':0.83,'C':0.68,
'N':0.68,'O':0.68,'F':0.64,'Ne':1.5,'Na':1.66,'Mg':1.41,'Al':1.21,'Si':1.2,
'P':1.05,'S':1.02,'Cl':0.99,'Ar':1.51,'K':2.03,'Ca':1.76,'Sc':1.7,'Ti':1.6,
'Va':1.53,'Cr':1.39,'Mn':1.61,'Fe':1.52,'Co':1.26,'Ni':1.24,'Cu':1.32,
'Zn':1.22,'Ga':1.22,'Ge':1.17,'As':1.21,'Se':1.22,'Br':1.21,'Kr':1.5,'Rb':2.2,
'Sr':1.95,'Y':1.9,'Zr':1.75,'Nb':1.64,'Mo':1.54,'Tc':1.47,'Ru':1.46,'Rh':1.42,
'Pd':1.39,'Ag':1.45,'Cd':1.54,'In':1.42,'Sn':1.39,'Sb':1.39,'Te':1.47,'I':1.4,
'Xe':1.5,'Cs':2.44,'Ba':2.15,'La':2.07,'Ce':2.04,'Pr':2.03,'Nd':2.01,'Pm':1.99,
'Sm':1.98,'Eu':1.98,'Gd':1.96,'Tb':1.94,'Dy':1.92,'Ho':1.92,'Er':1.89,'Tm':1.9,
'Yb':1.87,'Lu':1.87,'Hf':1.75,'Ta':1.7,'W':1.62,'Re':1.51,'Os':1.44,'Ir':1.41,
'Pt':1.36,'Au':1.36,'Hg':1.32,'Tl':1.45,'Pb':1.46,'Bi':1.48,'Po':1.4,'At':1.21,
'Rn':1.5,'Fr':2.6,'Ra':2.21,'Ac':2.15,'Th':2.06,'Pa':2,'U':1.96,'Np':1.9,
'Pu':1.87,'Am':1.8,'Cm':1.69,'Bk':1.54,'Cf':1.83,'Es':1.5,'Fm':1.5,'Md':1.5,
'No':1.5,'Lr':1.5,'Rf':1.5,'Db':1.5,'Sg':1.5,'Bh':1.5,'Hs':1.5,
'Mt':1.5,'Ds':1.5,'Linker':0,'Cluster':0
}

#==============================================================================
#==============================================================================
#==============================================================================
#Class for storing data about atoms. Normally contains an integer label,
#the atomic symbol, the covalent radius, position coordinates, a boolean
#which describes whether the atom is real or a periodic image, and a set of
#bonded neighbours.

class Atom:
```

```
    def __init__(self,label,atom_type,position,image):

        self.L=label
        self.t=atom_type
        self.r=raycov_dict[self.t]
        self.p=np.asarray(position,dtype=np.float64)
        self.i=image
        self.n=set()

#================================================================================
#================================================================================
#Class for storing data and functions related to the unit cell. Initialised
#using 6 cell parameters (angles in degrees).

class Crystal:

    def __init__(self,cell_params,n_atoms):

        if len(cell_params)==6:
            self.cell_params=cell_params
        else:
            print('Cell parameters should contain 6 entries.')
            sys.exit()

        #Also store number of atoms and supercell status.
        self.n=int(n_atoms)
        self.supercell=(1,1,1)

#====
    #Method for building a supercell from the current unit cell. Adjusts
    #positions and cell parameters while adding new atoms.

    def make_supercell(self,supercell,atoms):

        #A supercell should really be generated only once, so attributes
        #which would be needed to generate it a second time are not tracked.
        if self.supercell[0]*self.supercell[1]*self.supercell[2]!=1:
            print('Supercell should only be made once and will not be made.\n')
            return

        self.supercell=supercell
        #Edit the cell parameters.
        for i in range(3):
            self.cell_params[i]=self.cell_params[i]*self.supercell[i]

        #Variable N used to label the new atoms
        N=self.n
        #Add new atoms for each unit cell we are adding to the supercell.
        for i in range(self.supercell[0]):
            for j in range(self.supercell[1]):
                for k in range(self.supercell[2]):
                    #Skip the original unit cell.
                    if i+j+k==0:
                        continue
                    for x in range(self.n):
                        new_position=atoms[x].p+[i,j,k]
                        atoms.append(
                            Atom(x+N,atoms[x].t,new_position,False)
                            )
                        N+=1

        #Edit the coordinates of the original cell to reflect changes in cell
        #parameters.
        for q in range(self.n):
            for w in range(3):
```

```
                atoms[q].p[w]=atoms[q].p[w]/self.supercell[w]

#====
    #Builds and returns the metric tensor for the given unit cell from the
    #cell parameters.

    def make_metric_tensor(self):

        a=self.cell_params[0]
        b=self.cell_params[1]
        c=self.cell_params[2]
        #Assumes the cell parameters are in degrees
        alpha=np.radians(self.cell_params[3])
        beta=np.radians(self.cell_params[4])
        gamma=np.radians(self.cell_params[5])

        #Standard crystallographic formula for an abritrary unit cell.
        m_tensor=np.array([
            [a*a,a*b*np.cos(gamma),a*c*np.cos(beta)],
            [a*b*np.cos(gamma),b*b,b*c*np.cos(alpha)],
            [a*c*np.cos(beta),b*c*np.cos(alpha),c*c]
            ]
            )

        return m_tensor

#====
    #Builds and returns the transformation matrix for going from fractional to
    #cartesian coordinates from the cell parameters. Optional flag allows for
    #the inverse (cartesian to fractional) to be made instead.

    def make_frac2cart(self,inverse=False):

        a=self.cell_params[0]
        b=self.cell_params[1]
        c=self.cell_params[2]
        #Assumes the cell parameters are in degrees
        alpha=np.radians(self.cell_params[3])
        beta=np.radians(self.cell_params[4])
        gamma=np.radians(self.cell_params[5])

        #Omega quantity is required to define the transformation matrix.
        omega=a*b*c*np.sqrt(
            1-np.power(np.cos(alpha),2) \
            -np.power(np.cos(beta),2) \
            -np.power(np.cos(gamma),2) \
            +2*np.cos(alpha)*np.cos(beta)*np.cos(gamma)
            )

        #Standard crystallographic formula for an arbitrary unit cell.
        t_matrix=np.array(
            [
            [a,b*np.cos(gamma),c*np.cos(beta)],
            [0,b*np.sin(gamma),c*(np.cos(alpha) \
                -np.cos(beta)*np.cos(gamma))/np.sin(gamma)],
            [0,0,(omega)/(a*b*np.sin(gamma))]
            ]
            )

        if inverse:
            return np.linalg.inv(t_matrix)
        else:
            return t_matrix

#====
    #Writes a cif file for the current structure given a set of atom objects.
```

```
    def makecif(self,outfile,atoms):

        with open(outfile,'w') as file:

            file.write('#File created on {}\n\n'.format(date.today()))

            #Pre-position block.
            file.write('data_image0\n'\
            +'_symmetry_space_group_name_H-M    "P 1"\n'\
            +'_symmetry_int_tables_number       1\n\n'\
            +'loop_\n'\
            +' _symmetry_equiv_pos_as_xyz\n'\
            +" 'x, y, z'\n\n")

            #Cell parameter block.
            file.write(
            '_cell_length_a {:.4f}\n'.format(self.cell_params[0])\
            +'_cell_length_b {:.4f}\n'.format(self.cell_params[1])\
            +'_cell_length_c {:.4f}\n'.format(self.cell_params[2])\
            +'_cell_angle_alpha {:.4f}\n'.format(self.cell_params[3])\
            +'_cell_angle_beta {:.4f}\n'.format(self.cell_params[4])\
            +'_cell_angle_gamma {:.4f}\n\n'.format(self.cell_params[5]))

            #Position block.
            file.write('loop_\n'\
            +' _atom_site_label\n'\
            +' _atom_site_type_symbol\n'\
            +' _atom_site_fract_x\n'\
            +' _atom_site_fract_y\n'\
            +' _atom_site_fract_z\n')

            for n,a in enumerate(atoms):
                file.write('{} {} {:.5f} {:.5f} {:.5f}\n'.format(a.t+str(n),a.t
                        ,a.p[0],a.p[1],a.p[2]))

        print('New .cif file {} was successfully generated.'.format(outfile))

#============================================================================
#============================================================================
#Function to read lines until the target string is found. Line is returned
#if found, else False returned if EOF is reached.

def readuntil(fin,target):

    while True:
        line=fin.readline()
        if target in line:
            return line
        #Evaluates to False for empty string which EOF returns
        elif not line:
            return False

#============================================================================
#============================================================================
#Function for reading a cif file and pulling cell and atom information. Works
#only for cifs of P1 symmetry, though this functionality could be added using
#pymatgen.

def cifreader(filename):

    with open(filename,'r') as fin:

        #Read cif file until cell parameters
        a=float(readuntil(fin,'_cell_length_a').split()[-1])
        b=float(fin.readline().split()[-1])
```

```
        c=float(fin.readline().split()[-1])
        alpha=float(fin.readline().split()[-1])
        beta=float(fin.readline().split()[-1])
        gamma=float(fin.readline().split()[-1])

        #Read until coordinate and atom type headers and collect them
        headers=[]
        headerstring='_atom_site_'
        headers.append(readuntil(fin,headerstring).strip())
        while True:
            line=fin.readline()
            if headerstring in line:
                headers.append(line.strip())
            else:
                break

        #Once headers have been collected, move back to start of atoms block
        fin.seek(0)
        readuntil(fin,headers[-1])

        atoms=[]
        i=0
        while True:
            line=fin.readline()
            if line:
                try:
                    line=line.split()
                    #Using the index of the needed headers in headers array
                    #ensures that, even if other fields are present, only the
                    #atom type, x, y, and z coordinates will be pulled.
                    atoms.append(Atom(
                        i,line[headers.index('_atom_site_type_symbol')],
                        [line[headers.index('_atom_site_fract_x')],
                        line[headers.index('_atom_site_fract_y')],
                        line[headers.index('_atom_site_fract_z')]],
                        False
                        ))
                    i+=1
                except IndexError:
                    break
            else:
                break

    crystal=Crystal(
        np.array([a,b,c,alpha,beta,gamma],dtype=np.float64),len(atoms)
            )

    return atoms,crystal

#==============================================================================
#==============================================================================
#This function finds all atoms too close to the far edge of the unit cell
#(as controlled by cutoff) and appends their periodic images at the
#appropriate position (i.e. current position -1) to atoms list to mimic
#translational symmetry.

def t_symm(atoms,crystal,cutoff=3):

    for tt in range(0,len(atoms)):
        #first, make sure all atoms are inside the unit cell
        for s in range(3):
            while atoms[tt].p[s]<0.0:
                atoms[tt].p[s]=atoms[tt].p[s]+1.0
            while atoms[tt].p[s]>1.0:
                atoms[tt].p[s]=atoms[tt].p[s]-1.0
```

```
        #track adjacency to a cell boundary with 3 flags (x,y,z)
        edge_flags=[False,False,False]
        for ttt in range(0,3):
            if (1-atoms[tt].p[ttt])*crystal.cell_params[ttt]<cutoff:
                edge_flags[ttt]=True
                #for each far edge that an atom is near, set flag to true

        #for each combination of True flags, add an image atom in the
        #corresponding position
        #e.g. [False,True,True] will add atoms at p-[0,0,1], p-[0,1,0], and
        #p-[0,1,1]
        for i in range(int(edge_flags[0])+1):
            for j in range(int(edge_flags[1])+1):
                for k in range(int(edge_flags[2])+1):
                    if i+j+k!=0:
                        vector=[i,j,k]
                        atoms.append(
                            Atom(
                                atoms[tt].L,atoms[tt].t,
                                atoms[tt].p-vector,True
                                )
                            )
                        atoms[tt].edge=edge_flags
                        #copy all attributes of the original atom, except
                        #the image label which is True

    return atoms

#============================================================================
#============================================================================
#Function for finding bonds between the given set of atoms. Requires a crystal
#object containing some cell parameters in order to generate the metric
#tensor, which is used to calculate distances valid for any crystal system.
#Default tolerance of 0.4 angstrom is a standard value.

def build_bonds(atoms,crystal,tol=0.4):

    #Metric tensor object is a 3x3 matrix.
    m_tensor=crystal.make_metric_tensor()

    for i in range(0,len(atoms)):
        for j in range(i+1,len(atoms)):
            vector=atoms[i].p-atoms[j].p
            #Generalised Pythagoras formula works as long as m_tensor is correct
            mag=np.sqrt(vector@m_tensor@vector)
            if atoms[i].r+atoms[j].r-tol < mag < atoms[i].r+atoms[j].r+tol:
                #Due to method used for translational symmetry, image atoms are
                #also present. Hence add the labels of the ith and jth atoms
                #instead (since these were copied from their counterparts).
                atoms[atoms[i].L].n.add(atoms[j].L)
                atoms[atoms[j].L].n.add(atoms[i].L)

#============================================================================
#============================================================================
#Function for finding the centroid of a molecular fragment. Takes into account
#possible disconnects across unit cell boundaries

def get_centroid(atoms,fragment):

    #Extract the atom positions to avoid editing atom positions.
    positions=[atoms[i].p for i in fragment]
    #Set one of the atoms as the initial geometric centre to compare to.
    gc=positions[0]


    #For each atom, if it is farther away than 0.75 for any coordinate from
```

```
    #the centroid, edit its position to place it in a closer image position
    for p in positions:
        for i in range(3):
            if gc[i]-p[i]>0.75:
                p[i]+=1
            elif gc[i]-p[i]<-0.75:
                p[i]-=1

    #Geometric centroid simply the average of all corrected coordinates.
    gc=np.mean(positions,axis=0)

    return gc

#==========================================================================
#==========================================================================
#Function for generating a local Cartesian basis set given a pair of oxygens
#corresponding to linker-SBU bonds

def basis_generator(atoms,ends,crystal):

    metals=['Zr']
    vectors=[]

    #Find O-Zr vectors
    for oxygen in ends:
        for j in atoms[oxygen].n:
            if atoms[j].t in metals:
                vectors.append(atoms[oxygen].p-atoms[j].p)

    #Check that vector isn't cut across a unit cell boundary
    for v in vectors:
        for i in range(3):
            if v[i]>0.5:
                v[i]-=1
            elif v[i]<-0.5:
                v[i]+=1

    #Transform vectors to Cartesian
    t_matrix=crystal.make_frac2cart()
    vectors=[t_matrix@v for v in vectors]

    #Basis vector 1 defined using mean of O-Zr bonds, is normalised.
    basis1=np.mean([vectors],axis=1)[0]
    basis1/=np.linalg.norm(basis1)

    #Basis vector 2 defined using cross product of O-Zr bonds, is normalised.
    basis2=np.cross(vectors[0],vectors[1])
    basis2/=np.linalg.norm(basis2)

    #Basis vector 3 defined using cross product of other 2 basis vectors and is
    #already normalised.
    basis3=np.cross(basis1,basis2)
    basis_set=np.array([basis1,basis2,basis3],dtype=np.float64)

    return basis_set

#==========================================================================
#==========================================================================
#Function for generating capping species coordinates given a coordinate system
#and an origin corresponding to the specific part of the linker being replaced.

def capping_generator(filename,mof_basis,mof_crystal,origin):

    #Read capping species coordinates and separate into linker+SBU
    cap_atoms,cap_crystal=cifreader(filename)
    build_bonds(cap_atoms,cap_crystal,tol=0.4)
```

```
    cap,cluster=find_fragments(cap_atoms)

    #Find the capping species' end oxygens in order to generate a local basis
    #set.
    ends=[]
    for a in cap_atoms:
        ntypes=set([cap_atoms[i].t for i in a.n])
        if ntypes.intersection(set(['O','C'])) and len(a.n)==3:
            for j in a.n:
                if cap_atoms[j].t=='O':
                        ends.append(j)

    #Generate cap basis set and fractional-to-cartesian matrix
    cap_basis=basis_generator(cap_atoms,ends,cap_crystal)
    cap_t_matrix=cap_crystal.make_frac2cart()

    #Generate a set of transform matrices from the MOF coordinate system.
    inverse=mof_basis.transpose()
    t_matrix=mof_crystal.make_frac2cart(inverse=True)


    cap_origin=np.mean([cap_atoms[k].p for k in ends],axis=0)

    '''In order:
    1. set the capping species' origin to be the centroid of the 2 end oxygens
    2. transform cap coordinates to Cartesian, then to local basis
    3. transform edited cap coordinates to MOF basis, then to MOF fractional,
    then add the origin of the linker end in the MOF'''
    for a in cap_atoms:
        a.p-=cap_origin
        a.p=cap_basis@cap_t_matrix@a.p
        a.p=t_matrix@inverse@a.p+origin

    #Keep only the atoms corresponding to the cap atoms
    cluster_atoms=[cap_atoms[l] for l in cluster[0].members]
    cap_atoms=[a for a in cap_atoms if a not in cluster_atoms]

    return cap_atoms

#==============================================================================
#==============================================================================
#Function for generating missing linker defects. Take a set of target indices
#from the array of identified linkers, as well as an external coordinate file
#which should contain the cap attached to the SBU.

def defect_maker(targets,linker_array,atoms,crystal,coordfile):


    linker_ends=[]
    removal=[]

    #Identify all sets of linker end oxygens.
    for linker in targets:
        ends=[]
        for member in linker_array[linker].members:
            #Track all atoms from the original linkers so they can be removed.
            removal.append(atoms[member])

            #Simple criteria for finding the ending oxygens
            ntypes=set([atoms[i].t for i in atoms[member].n])
            if ntypes.intersection(set(['O','C'])) and len(atoms[member].n)==3:
                for j in atoms[member].n:
                    if atoms[j].t=='O':
                        ends.append(j)

                #Append ends only if it contains items; empty list evaluates
```

```
                #to False.
                if ends:
                    linker_ends.append(ends)
                    ends=[]

    #Generate a set of coordinates for the capping species for each end of each
    #linker that was given.
    new_atoms=[]
    for ends in linker_ends:
        origin=np.mean([atoms[k].p for k in ends],axis=0)
        mof_basis=basis_generator(atoms,ends,crystal)
        new_atoms.append(capping_generator(
                coordfile,mof_basis,
                crystal,origin
                ))
    #Remove original linker atoms
    atoms=[a for a in atoms if a not in removal]
    for new in new_atoms:
        for k in new:
            atoms.append(k)

    return atoms

#================================================================================
#================================================================================
#Function to isolate building blocks in the given MOF. Builds a graph object
#by initialising every bond as an edge. Then removes edges corresponding to
#linker-SBU bonds, and finally finds connected subgraphs.

def find_fragments(atoms):

    metals=set(['Zr'])
    removebonds=[]

    #For Zr6 MOFs, this set of criteria is sufficient for identifying the
    #bonds which should be removed from the graph.
    for a in atoms:
        if len(a.n)==2 and a.t=='O':
            for b in a.n:
                if atoms[b].t in metals:
                    removebonds.append((a.L,b))

    #Initialise the graph object, build the edges, then remove SBU-linker
    #bonds
    graph=nx.Graph()
    for a in atoms:
        for b in a.n:
            graph.add_edge(a.L,b)
    for rb in removebonds:
        graph.remove_edge(rb[0],rb[1])

    #Find molecular fragments by identifying connected subgraphs
    subgraphs=[]
    for g in nx.connected_components(graph):
        subgraphs.append(list(graph.subgraph(g).nodes))

    #Each subgraph is stored as an atom object with a label and position.
    clusters=[]
    linkers=[]
    for g in range(len(subgraphs)):
        fragment=[i for i in subgraphs[g]]
        atomtypes=set([atoms[j].t for j in fragment])
        #A set with elements in it evaluates to True, hence if there is an
        #overlap with the metals set, must be an SBU
        if atomtypes.intersection(metals):
            cluster=Atom(0,'Cluster',get_centroid(atoms,fragment),False)
```

```
            cluster.members=fragment #Members attribute not present by default
            #for atoms but useful to add to molecular fragments.
            clusters.append(cluster)
        else:
            linker=Atom(0,'Linker',get_centroid(atoms,fragment),False)
            linker.members=fragment
            linkers.append(linker)

    #Relabel fragments for convenience.
    for q in range(len(clusters)):
        clusters[q].L=q
    for q in range(len(linkers)):
        linkers[q].L=q

    return linkers,clusters

#===============================================================================
#===============================================================================
#Wrapper calls the necessary functions sequentially.

def _wrapper(filename,coordfile,indices,skeleton=False):

    atoms,crystal=cifreader(filename) #Read input data
    t_symm(atoms,crystal) #Apply periodic boundary conditions using image atoms

    build_bonds(atoms,crystal,tol=0.4) #Build bonds using covalent radii
    atoms=[gg for gg in atoms if gg.i==False] #Remove image atoms

    linker_array,cluster_array=find_fragments(atoms) #Use graph theory to find
    #molecular fragments

    #Make skeleton cif file if requested; used to identify indices of linkers
    if skeleton:
        for c in cluster_array:
            c.t='Xe'
        for l in linker_array:
            l.t='La'
        crystal.makecif('skeleton.cif',cluster_array+linker_array)
        sys.exit()


    foutname=filename[:-4]+'-{}-defects.cif'.format(len(indices))
    #Generate defects by replacing some chosen defects with capping species
    new_atoms=defect_maker(indices,linker_array,atoms,crystal,coordfile)
    crystal.makecif(foutname,new_atoms)

#===============================================================================
#===============================================================================
#Run the code via the wrapper function. Receive and format some command-line
#arguments for file name, cap coordinates, and defect indices.

if __name__=="__main__":

    if 1 < len(sys.argv)<4:
        print('No coordinate file or indices given; skeleton.cif file will'+\
            ' be generated instead \n')
        filename=sys.argv[1]
        coordfile='dummy'
        indices=[]
        _wrapper(sys.argv[1],coordfile,indices,True)

    elif len(sys.argv)>3:
        filename=sys.argv[1]
        coordfile=sys.argv[2]
        indices=[int(i) for i in sys.argv[3].split(',')]
```

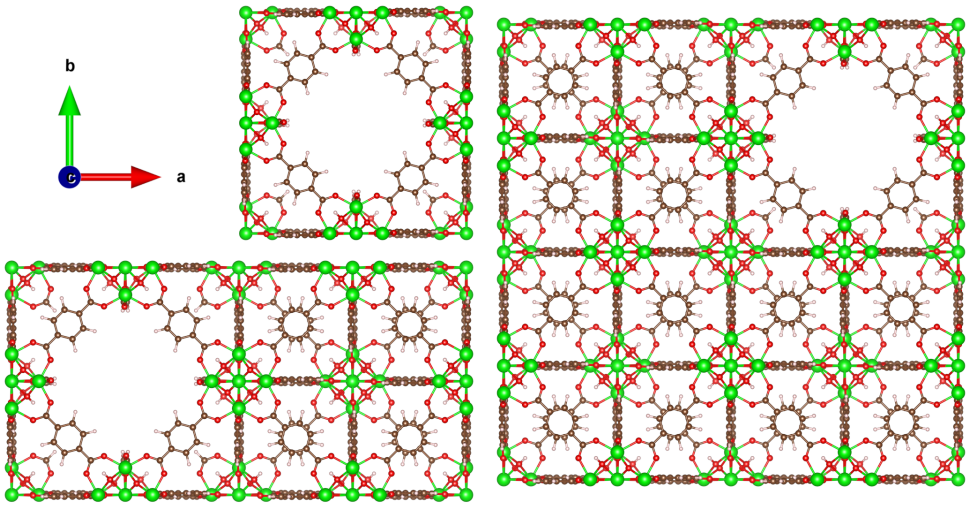
```
    print('Arguments received for input  file: {}'.format(filename))
    print('Argument received for coordinate file: {}'.format(coordfile))
    print('Arguments received for linker indices: {}\n'.format(sys.argv[3]))

    _wrapper(sys.argv[1],sys.argv[2],indices,False)
else:
    print('Insufficient arguments received; no defects will be generated.')
    sys.exit()
```

C Supercell schematics

C.1 fcu-reo supercells

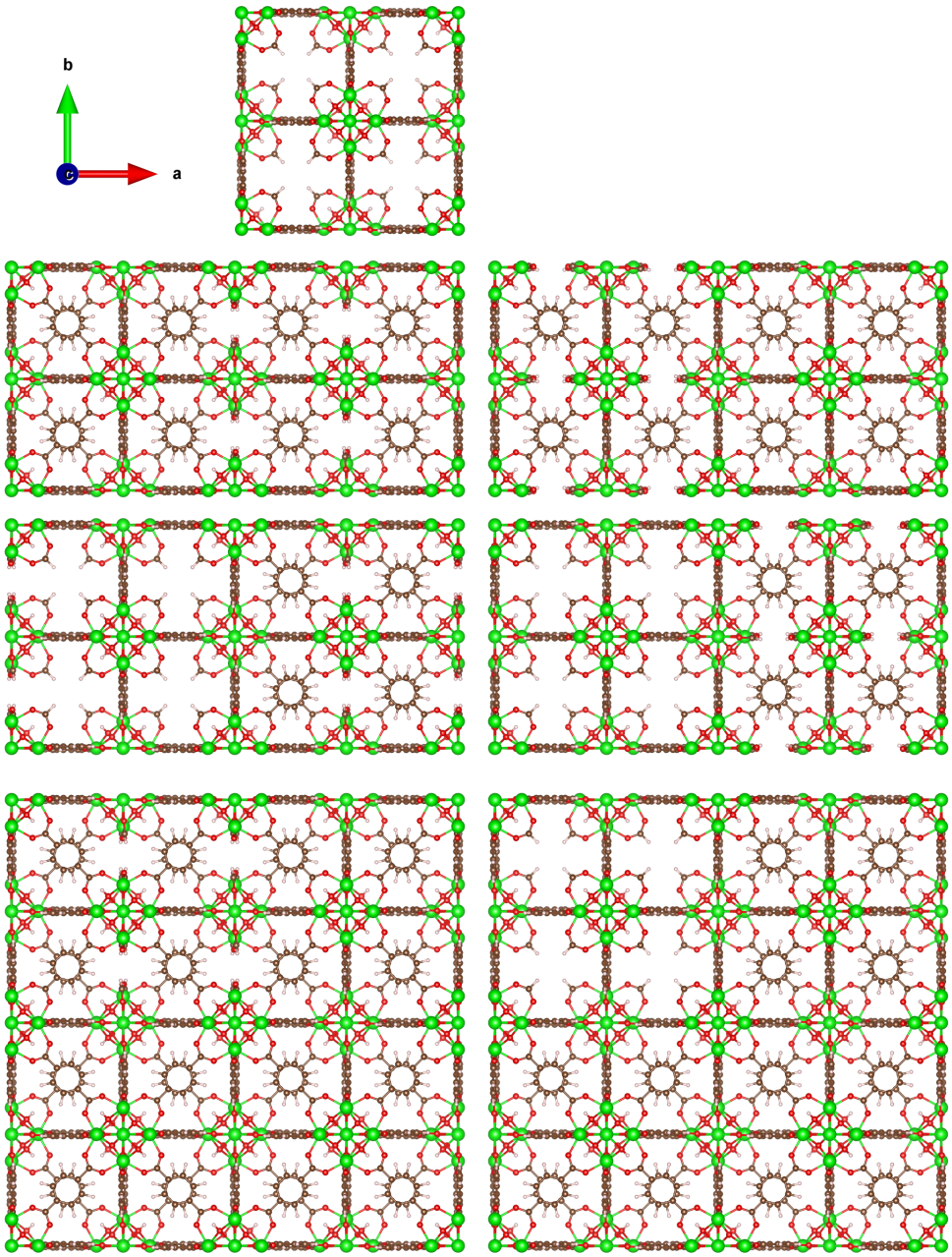
Schematics of the mixed **fcu-reo** supercells discussed in **Section 3.3.3**, shown here with FA as the capping species. 2×2×2 systems have been omitted for clarity.



Schematics of **fcu-reo** supercells. Left column, top to bottom: 1×1×1-**reo**, 2×1×1-**fcu-reo**. Right column: 2×2×1-**fcu-reo**

C.2 fcu-bcu supercells

Schematics of the mixed **fcu-bcu** supercells discussed in **Section 3.3.3**, shown here with FA as the capping species. 2×2×2 systems have been omitted for clarity. Note that as the **bcu** system has an orientation, additional supercell configurations are possible compared to those for the **reo** system.



Schematics of **fcu-bcu** supercells. Left column, top to bottom: 1×1×1-**bcu**, 2×1×1-**fcu-bcu**-a, 2×1×1-**bcu-bcu**-ac, 2×2×1-**fcu-bcu**-a. Right column, top to bottom: 2×1×1-**fcu-bcu**-b, 2×1×1-**bcu-bcu**-bc, 2×2×1-**fcu-bcu**-c.

D CO adsorption geometric data

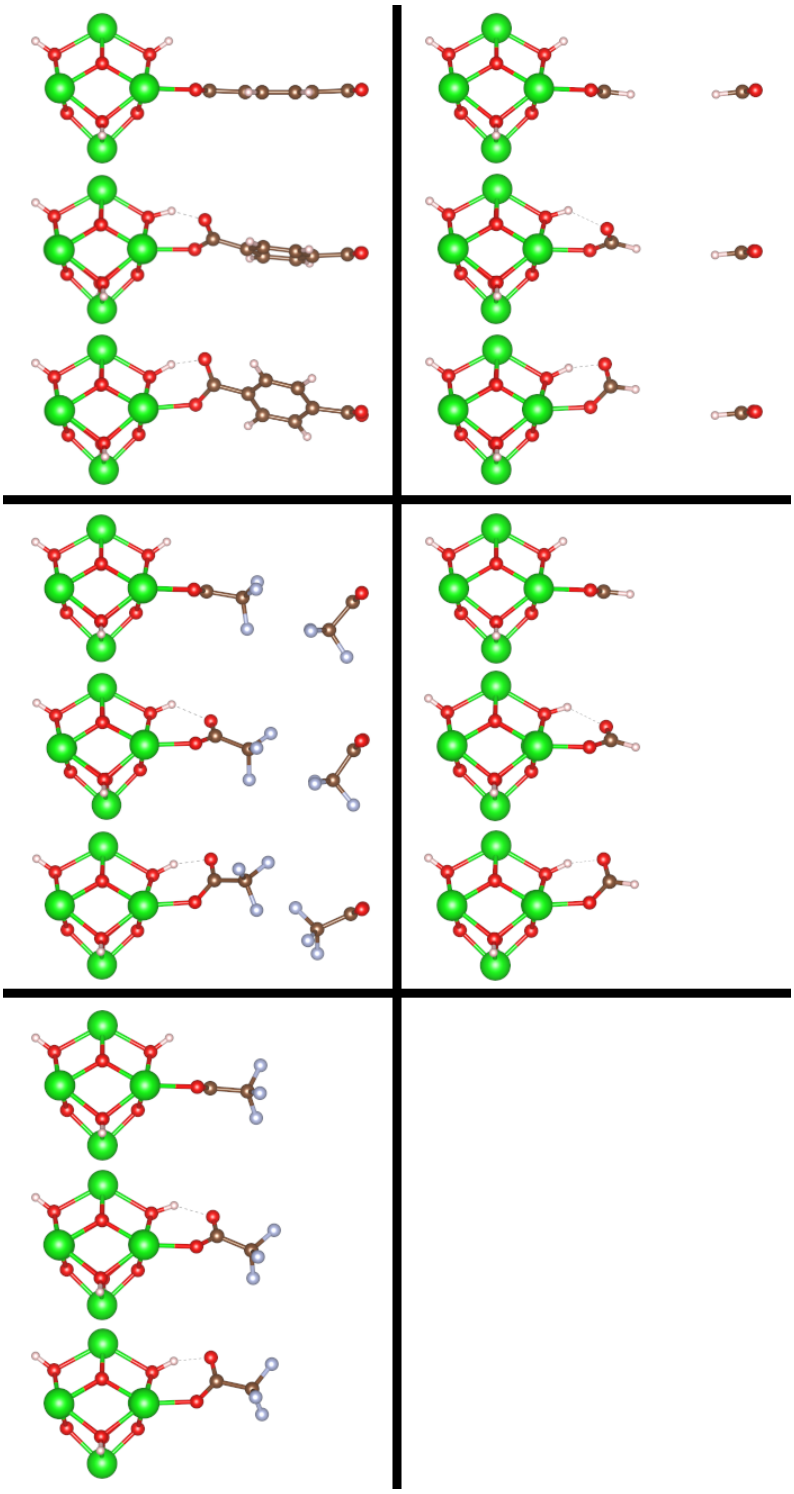
This appendix contains geometric information about the adsorption geometries of CO on Brønsted acid sites discussed in **Section 4.3.2**. The atomic separations and angles described in the table correspond to those shown schematically in **Figure 44**.

Adsorption site	$R_{\text{H-C}}$ / Å	$A_{\text{O-H-C}}$ / deg.
1-a (**fcu**)	3.14	179.9
6-a	3.04	175.5
6-b	3.29	142.6
6-c	3.31	159.6
7-b	3.09	170.1
7-c	3.35	164.5
7-d	3.19	170.0
8-a	3.02	164.6
8-c	3.35	152.2
9-a	3.23	169.8
9-b	3.07	166.1
10-a	3.06	178.2
10-b	3.12	142.7
11-a	3.05	176.2

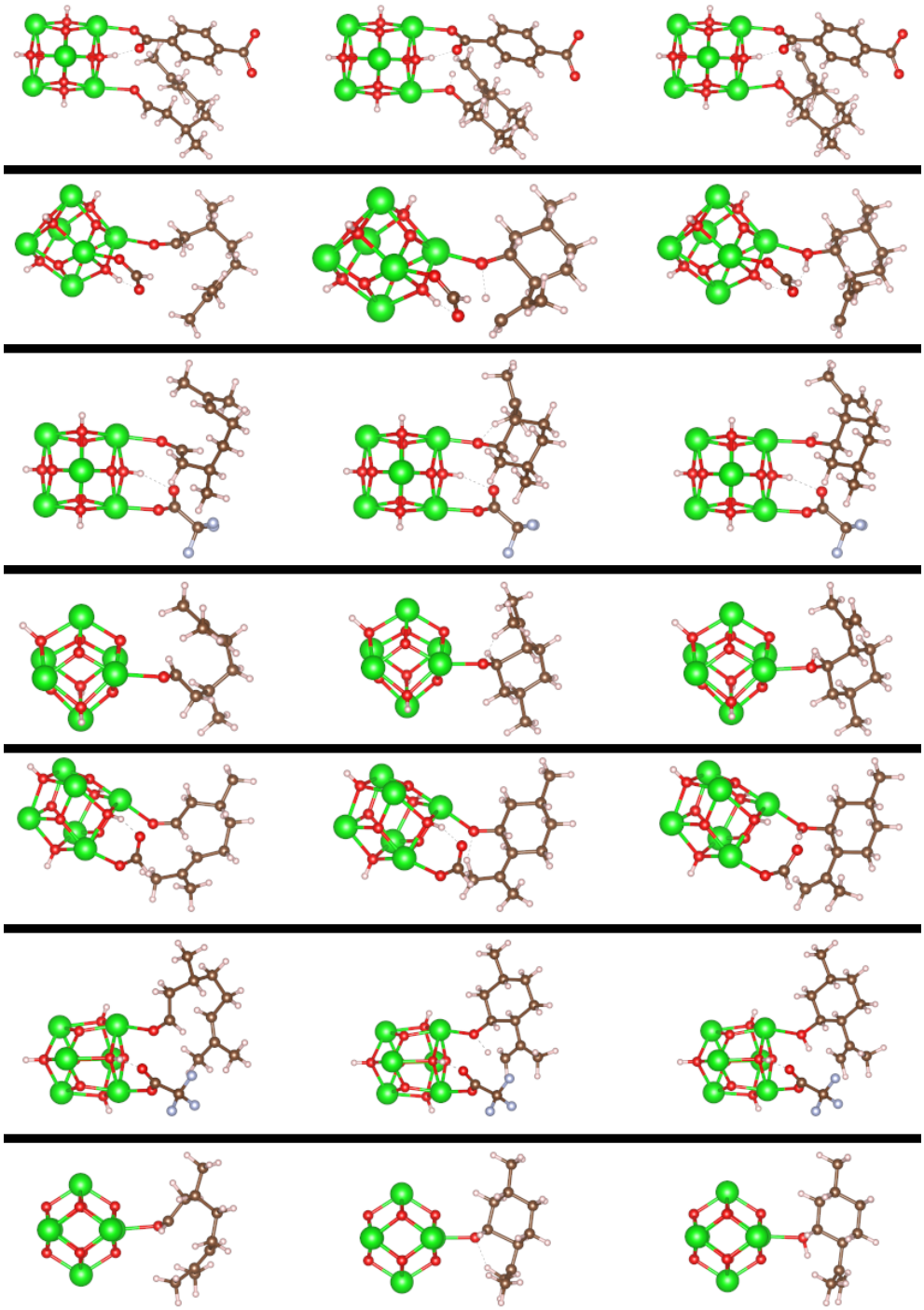
Geometric parameters of carbon monoxide adsorption on a selection of Brønsted acid sites. The separation denotes the distance between the probe carbon and the adsorption site oxygen; the angle is subtended between the probe carbon, the acidic proton, and the adsorption site oxygen. These are detailed in **Figure 44**.

E Cyclisation of citronellal transition states

The two figures in this appendix are schematic depictions of the converged minima and transition states for the cyclisation of citronellal to isopulegol that is discussed in **Section 4.3.3**. The majority of framework atoms have been obscured to improve visibility.



Minima and transition states for the dangling step. Each set of 3 schematics show the carboxylic acid in its normal, transition, and dangling states from top to bottom respectively. Left column, from top to bottom: **fcu**, TFAA missing linker, TFAA **reo**. Right column, from top to bottom: FA missing linker and FA **reo**.



Minima and transition states for the cyclisation step. Each set of 3 schematics show the citronellal, transition, and isopulegol states from left to right respectively. From top to bottom: **fcu**, FA missing linker, TFAA missing linker, **e** missing linker, FA **reo**, TFAA **reo**, and **e reo**.

F Equations for systems containing Ce^{3+}

This appendix contains all the possible equations for the reactions which describe the formation of the various $Zr_{6-x}Ce_x$ building blocks containing Ce_{3+} discussed in this work. These are accompanied by the corresponding equations for relative energy.

F.1 Uncapped

The equations in this subsection correspond to the defect termination in panel i) of **Figure 59**.

$$\begin{aligned} &x \text{ complex}(Ce^{3+}) + \frac{4x}{3} \text{ water} \\ &+ \frac{5x}{6} \text{ cluster}(Zr_6) + x \text{ formic acid} \end{aligned} \rightarrow \begin{aligned} &\text{cluster}(Zr_{6-x}Ce_x) + 3x \text{ nitric acid} \\ &+ 6x \text{ DMF} \end{aligned}$$

$$\begin{aligned} &x(Ce(NO_3)_3(C_3H_7NO)_6) + \frac{4x}{3}(H_2O) \\ &+ \frac{6-x}{6}(Zr_6O_4(OH)_4)(HCOO)_{12} \\ &+ x(HCOOH) \end{aligned} \rightarrow \begin{aligned} &Zr_{6-x}Ce_xO_4(OH)_4(HCOO)_{12-x} + 3x(HNO_3) \\ &+ 6x(C_3H_7NO) \end{aligned}$$

$$\begin{aligned} \Delta E = E_{Zr_{6-x}Ce_x} + 3xE_{HNO_3} + 6xE_{DMF} - xE_{Ce^{3+}} \\ - \frac{4x}{3}E_{H_2O} - \frac{6-x}{6}E_{Zr_6} - xE_{FA} \end{aligned}$$

F.2 FA

The equations in this subsection correspond to the defect termination in panel ii) of **Figure 59**.

$$\begin{aligned} &x \text{ complex}(Ce^{3+}) + \frac{4x}{3} \text{ water} \\ &+ \frac{5x}{6} \text{ cluster}(Zr_6) + 2x \text{ formic acid} \end{aligned} \rightarrow \begin{aligned} &\text{cluster}(Zr_{6-x}Ce_x) + 3x \text{ nitric acid} \\ &+ 6x \text{ DMF} \end{aligned}$$

$$\begin{aligned} &x(Ce(NO_3)_3(C_3H_7NO)_6) + \frac{4x}{3}(H_2O) \\ &+ \frac{6-x}{6}(Zr_6O_4(OH)_4)(HCOO)_{12} \\ &+ 2x(HCOOH) \end{aligned} \rightarrow \begin{aligned} &Zr_{6-x}Ce_xO_4(OH)_4(HCOO)_{12} + 3x(HNO_3) \\ &+ 6x(C_3H_7NO) \end{aligned}$$

$$\Delta E = E_{\mathrm{Zr_{6-x}Ce_x}} + 3xE_{\mathrm{HNO_3}} + 6xE_{\mathrm{DMF}} - xE_{\mathrm{Ce^{3+}}} - \frac{4x}{3}E_{\mathrm{H_2O}} - \frac{6-x}{6}E_{\mathrm{Zr_6}} - 2xE_{\mathrm{FA}}$$

F.3 BA

The equations in this subsection correspond to the defect termination in panel iii) of **Figure 59**.

$$x\ \mathrm{complex(Ce^{3+})} + \frac{4x}{3}\ \mathrm{water} + \frac{5x}{6}\ \mathrm{cluster(Zr_6)} + 2x\ \mathrm{benzoic\ acid} \rightarrow \mathrm{cluster(Zr_{6-x}Ce_x)} + 3x\ \mathrm{nitric\ acid} + 6x\ \mathrm{DMF}$$

$$x(\mathrm{Ce(NO_3)_3(C_3H_7NO)_6}) + \frac{4x}{3}(\mathrm{H_2O}) + \frac{6-x}{6}(\mathrm{Zr_6O_4(OH)_4})(\mathrm{H_5C_6COO})_{12} + 2x(\mathrm{H_5C_6COOH}) \rightarrow \mathrm{Zr_{6-x}Ce_xO_4(OH)_4(H_5C_6COO)_{12}} + 3x(\mathrm{HNO_3}) + 6x(\mathrm{C_3H_7NO})$$

$$\Delta E = E_{\mathrm{Zr_{6-x}Ce_x}} + 3xE_{\mathrm{HNO_3}} + 6xE_{\mathrm{DMF}} - xE_{\mathrm{Ce^{3+}}} - \frac{4x}{3}E_{\mathrm{H_2O}} - \frac{6-x}{6}E_{\mathrm{Zr_6}} - 2xE_{\mathrm{BA}}$$

F.4 DMF

The equations in this subsection correspond to the defect termination in panel iv) of **Figure 59**.

$$x\ \mathrm{complex(Ce^{3+})} + \frac{4x}{3}\ \mathrm{water} + \frac{5x}{6}\ \mathrm{cluster(Zr_6)} + x\ \mathrm{formic\ acid} \rightarrow \mathrm{cluster(Zr_{6-x}Ce_x)} + 3x\ \mathrm{nitric\ acid} + 4x\ \mathrm{DMF}$$

$$x(\mathrm{Ce(NO_3)_3(C_3H_7NO)_6}) + \frac{4x}{3}(\mathrm{H_2O}) + \frac{6-x}{6}(\mathrm{Zr_6O_4(OH)_4})(\mathrm{HCOO})_{12} + x(\mathrm{HCOOH}) \rightarrow \mathrm{Zr_{6-x}Ce_xO_4(OH)_4(HCOO)_{12-x}} + 3x(\mathrm{HNO_3}) + 4x(\mathrm{C_3H_7NO})$$

$$\Delta E = E_{\mathrm{Zr_{6-x}Ce_x}} + 3xE_{\mathrm{HNO_3}} + 4xE_{\mathrm{DMF}} - xE_{\mathrm{Ce^{3+}}} - \frac{4x}{3}E_{\mathrm{H_2O}} - \frac{6-x}{6}E_{\mathrm{Zr_6}} - xE_{\mathrm{FA}}$$