# Computational Screening and Analysis of Functional Porous Materials

**Andrew Tarzia**

A thesis submitted in fulfilment of the requirements for the degree of

Doctor of Philosophy

January 2019


Department of Chemistry

The University of Adelaide

North Terrace Campus

Adelaide, South Australia 5005

Supervisors:

Prof. Christian J. Doonan

A/Prof. David M. Huang

Dr. Aaron Thornton

# Contents

# Abstract

Functional porous materials are a class of materials that have found use in many industrial applications. In particular, extended framework materials, such as metal–organic frameworks (MOFs) and porous aromatic frameworks (PAFs), which are the subject of this thesis, show significant promise for applications including gas storage and separations, catalysis, drug delivery, microelectronics and sensing. This broad scope of applications stems from the immense chemical diversity afforded by their modular bottom-up synthesis and design. Additionally, the rational choice of building blocks allows for the precise control of the properties of the pore networks of crystalline extended porous materials. However, the process of finding optimal porous materials for emerging applications is slow due to arduous trial-and-error experimental approaches. The application of computational methods to analyze porous materials allows for the development of design principles, which can guide experimental endeavors. Furthermore, high-throughput screening can be used to expand on experimental findings by efficiently exploring chemical space for the best candidates for a given application.

This thesis reports several studies in which novel computational protocols are developed and applied to more rapidly screen porous functional materials for applications. A coarse-grained molecular dynamics model was developed to investigate the formation mechanism of PAFs and the role of structural and dynamics factors in determining their highly porous, amorphous networks. PAF formation, which is kinetically controlled, was found to robustly lead to a high degree of defects and porosity, and that relatively weak dispersion interactions are responsible for inducing porosity-reducing interpenetration. The simulations suggest that bulky reaction intermediates or building blocks with diminished dispersion interactions can be used to eliminate interpenetration and increase material porosity.

Highly-ordered MOF thin films with macroscale in-plane and out-of-plane alignment have many potential applications, but only a handful of examples have been reported to date. Therefore, a high-throughput screening process was developed to suggest

new MOFs that are likely to undergo aligned growth. The screening process was parameterized from a set of experimental observations of the aligned growth of copper-based MOFs from copper(II) hydroxide ($Cu(OH)_2$) and allows for the screening of thousands of MOF structures in a few days. Importantly, the number of known MOFs that are likely to grow aligned from $Cu(OH)_2$ was expanded and some design principles were uncovered. In particular, it was found that the substrate imparts a directing effect on the MOFs able to grow aligned, but does not limit the possible pore network properties of aligned MOFs.

The biomimetic mineralization of MOFs around biomacromolecules was investigated in two joint experimental and computational studies. Biomimetic mineralization is a general and facile method for encapsulating biological entities to, for example, enhance their stability in harsh conditions. Systematic experimental studies of the encapsulation of proteins and carbohydrates by zeolitic imidazolate framework-8 (ZIF-8) found that the electrostatic properties of the biomacromolecule govern biomimetic mineralization and showed that chemical functionalization can be used to control this process. Computational modelling verified the role of the negative charge on biomacromolecules in inducing ZIF growth as a result of enhancement of the surrounding zinc ion concentration. Furthermore, calculations of the surface electrostatic potential and pI of a protein were shown to accurately and efficiently predict whether a biomacromolecule seeds MOF growth.

Finally, a high-throughput screening process was developed to explore enzymatic reaction space to discover candidate reactions for MOF-encapsulated enzymes. This screening process uses the molecular size of the components of a reaction to predict whether the reaction can occur inside MOFs. The number of possible enzymatic reactions that have been carried out inside ZIF-8 is very small, and many of those reactions were found to have components that are likely too big to diffuse through ZIF-8. Therefore, the screening process was applied to suggest reactions that can investigate the relationship between the size of reaction components and enzymatic activity inside ZIF-8. In this process, a reaction of significant commercial value was identified that should occur in a MOF-encapsulated enzyme.

# Declaration

I certify that this work contains no material which has been accepted for the award of any other degree or diploma in my name, in any university or other tertiary institution and, to the best of my knowledge and belief, contains no material previously published or written by another person, except where due reference has been made in the text. In addition, I certify that no part of this work will, in the future, be used in a submission in my name, for any other degree or diploma in any university or other tertiary institution without the prior approval of the University of Adelaide and where applicable, any partner institution responsible for the joint-award of this degree.

I acknowledge that copyright of published works contained within this thesis resides with the copyright holder(s) of those works.

I also give permission for the digital version of my thesis to be made available on the web, via the University's digital research repository, the Library Search and also through web search engines, unless permission has been granted by the University to restrict access for a period of time.

I acknowledge the support I have received for my research through the provision of an Australian Government Research Training Program Scholarship

Andrew Tarzia

January 2019

# Acknowledgements

I have many people to thank for their support during the last 4 years of PhD studies, 8 years in total of studies at Adelaide University and 26 years of life.

Firstly, I am incredibly grateful to my supervisory team who provided me with many insights for the research contained in this thesis, but also the remainder of my academic career and life. Thank you, Aaron Thornton, David Huang and Christian Doonan, for your time, support and advice. In particular, the opportunities given to me to travel to conferences, collaborate with other world-class researchers and take on leadership roles within your research groups, all of which have assisted in my development in many ways. I believe I have come a long way as a scientist under your supervision these last four years and I hope to continue developing based on the advice you have all given me.

A special mention to academics in the department, Richard Keene, Christopher Sumby and Stephen Bell, for offering their guidance and support on many occasions. I would also like to thank Masahide Takahashi for hosting me in Osaka, Japan for three weeks, and his students Ken Ikigaki and Takaaki Hara for their support and hospitality. Finally, I would like to show immense gratitude for the support provided to me through CSIRO and the welcoming nature of the whole research group at Clayton.

I must thank the entirety of the Chemistry department at the University of Adelaide. Especially Matt Bull for his technical support and demonstrator support. I would also like to thank the current and past members of the Sumby-Doonan and Huang research groups. In particular, Alex Burgen, Jesse Teo and Witold Bloch for teaching me so much and always supporting me. Thank you to Dr. Patrick Tapping for this LaTeX template and constant support throughout my PhD and undergraduate studies. Thank you to Michael Huxley, Kate Flint, Alex Stuart, Stella Child, Harley Betts, Oliver Linder-Patton, Natasha Maddigan (**co-pick of the week!**), Belinda Boehme and Huong Nguyen for being part of this process with me, you have all made it more enjoyable in so many ways. Finally, thank you to Kenji Sumida for so many

discussions about work and life.

To Jack Evans, Carla Toop, Justin Spence, Sophie Arnerić, Noby Leong, and Philip Gibson, each of you prepared me for this process, supported me throughout it and taught me so much (not limited to the consumption of beer and coffee, where to have good lunches, and how to get on the ABC). Extra thanks to Noby and Philip, for staying in Adelaide until the end of my PhD — It was crucial at so many points *flower emoji*.

Outside of academia, I have been supported by many close friends and family. Thank you to Dom, Iryna, Alex, Ali, George, Jacqueline Lee and "Squad". Thank you for always being a distraction from my work and for helping me celebrate our victories in life. To my second family, Pash, Josh, Michael, and Raquel (I hope this order need not change), thank you all for so much over the last four years (and more). You have helped me survive.

Finally, I would like to acknowledge my family for their constant and unwavering support. Thank you to my extended family for your support and love. Thank you to my brother Nick and my sisters, Lisa and Natalie for your encouragement, your advice and how you have shaped me. A special thank you to Natalie and Lisa for drafting parts of this thesis! Thank you to my parents, my entire life you have offered me guidance, inspired me and supported me. I could not have made it here successfully without you by my side.

# Contextual statement

This thesis consists of four research chapters, three of which are in manuscript style. Chapters 3, 4 and 5 have been published (or submitted for publication) in peer-reviewed international journals.[1–4] Statements of authorship can be found at the beginning of each chapter. Chapter 6 is in the style of a manuscript, which we intend to submit to a peer-reviewed international journal. Chapter 5 collates the computational analysis that was carried out by the candidate (and the relevant experimental observations) in two collaborative projects. This chapter is in a traditional thesis style. The candidate completed all of the work presented in this thesis during the doctoral candidature.

Chapters 1 and 2 introduce the porous materials that are the focus of this thesis, the concept of screening materials to explore chemical space, and the computational methods that were applied to do so. Chapter 1 provides context for the significance of the two classes of materials (porous aromatic frameworks (PAFs) and metal–organic frameworks (MOFs)) studied in this thesis. In particular, two emerging applications of MOFs are introduced, and the need for computational exploration is highlighted. A review of the literature on the use of computational screening processes in materials science is also provided. Chapter 2 serves as an introduction to the broader methods that underpin the analysis applied in this thesis.

Chapters 3 and 5 highlight computational methods as a means to investigate molecular processes in cases for which an experimental investigation is difficult. Chapter 3 details the use of coarse-grained molecular simulations to study the formation mechanism of amorphous PAFs as well as the role of monomer–monomer interactions and monomer structure on their porosity.[1] Chapter 5 details the application of theory and computation to verify the role of the electrostatic properties of biomacromolecules in the biomimetic mineralization of MOFs. In this case, the computational analysis was performed to support a series of experiments. Importantly, this chapter provides methods to predict whether a biomacromolecule seeds MOF growth, which enables the efficient selection of proteins and carbohydrates for future experiments.[3,4]

Chapters 4 and 6 report the development of high-throughput screening processes to identify promising candidates for two different applications. Chapter 4 describes screening of MOFs for aligned heteroepitaxial growth using geometrical descriptions of MOF–substrate interfaces.[2] Chapter 6 details an exploration of the known enzymatic reaction space for viable reactions that can occur inside ZIF-8. In each case, a narrow set of MOFs and enzymes have been tested experimentally to date, and our screening processes efficiently explore viable experiments to help expand these relatively new fields.

# Publications

This thesis is based on the following publications that detail some of the research undertaken during the period of candidature:

1. Tarzia, A.; Thornton, A. W.; Doonan, C. J.; Huang, D. M. Molecular Insight into Assembly Mechanisms of Porous Aromatic Frameworks. *J. Phys. Chem. C* **2017** *121*, 16381–16392. DOI: 10.1021/acs.jpcc.7b04860

2. Tarzia, A.; Takahashi, M.; Falcaro, P.; Thornton, A. W.; Doonan, C. J.; Huang, D. M. High-Throughput Screening of Metal–Organic Frameworks for Macroscale Heteroepitaxial Alignment. *ACS Appl. Mater. Interfaces* **2018**, *10*, 40938–40950. DOI: 10.1021/acsami.8b14756

3. Maddigan, N. K.; Tarzia, A.; Huang, D. M.; Sumby, C. J.; Bell, S. G.; Falcaro, P.; Doonan, C. J. Protein Surface Functionalisation as a General Strategy for Facilitating Biomimetic Mineralisation of ZIF-8. *Chem. Sci.* **2018**, *9*, 4217-4223. DOI: 10.1039/C8SC00825F

4. Astria, E.; Thonhofer, M.; Ricco, R.; Liang, W.; Chemelli, A.; Tarzia, A.; Alt, K.; Hagemeyer, C. E.; Rattenberger, J.; Schroettner, H; Wrodnigg, T.; Amenitsch, H.; Huang, D. M.; Doonan, C. J.; Falcaro, P. Carbohydrates@MOFs *Mater. Horiz.* **2019**, *accepted.* DOI: 10.1039/C8MH01611A

Additionally, contributions have been made to the following publications:

5. Huxley, M.; Coghlan, C. J.; Burgun, A.; Tarzia, A.; Sumida, K.; Sumby, C. J.; Doonan, C. J. Site-specific Metal and Ligand Substitutions in a Microporous $Mn^{2+}$-based Metal–organic Framework. *Dalton Trans.* **2016**, *45*, 4431–4438. DOI: 10.1039/C5DT05023E

6. Morshedi, M.; Thomas, M.; Tarzia, A.; Doonan, C. J.; White, N. G. Supramolecular Anion Recognition in Water: Synthesis of Hydrogen-bonded Supramolecular Frameworks. *Chem. Sci.* **2017**, *8*, 3019–3025. DOI: 10.1039/C7SC00201G

7. Sun, B.; Southam, H. M.; Butler, J. A.; Poole, R. K.; Burgun, A.; Tarzia, A.; Keene, F. R.; Collins, J. G. Synthesis, Isomerisation and Biological Properties of Mononuclear Ruthenium Complexes Containing the bis[4(4′-methyl-2,2′-bipyridyl)]-1,7-heptane Ligand. *Dalton Trans.* **2018**, *47*, 2422–2434. DOI: 10.1039/C7DT04595F

8. Liang, W.; Xu, H.; Carraro, F.; Maddigan, N. K.; Li, Q.; Bell, S. G.; Huang, D. M.; Tarzia, A.; Solomon, M. B.; Amenitsch, H.; Vaccari, L.; Sumby, C. J.; Falcaro, P.; Doonan, C. J. Enhanced Activity of Enzymes Encapsulated in Hydrophilic Metal–Organic Frameworks, *J. Am. Chem. Soc.* **2019**, *141*, 2348–2355. DOI: 10.1021/jacs.8b10302

# Abbreviations

**BET** Brunauer–Emmett–Teller

**CG** coarse-grained

**CIF** crystallographic information file

**CSD** Cambridge Structural Database

**DBSCAN** Density-based spatial clustering of applications with noise

**DFT** density functional theory

**EC** Enzyme Commission

**GB** Gay-Berne

**LD** Langevin dynamics

**LJ** Lennard-Jones

**MC** Monte Carlo

**MD** molecular dynamics

**MOF** metal–organic framework

**MSD** mean squared displacement

**PAF** porous aromatic framework

**PB** Poisson–Boltzmann

**PDB** Protein Data Bank

**PXRD** powder X-ray diffraction

**RDF** radial distribution function

**SALE** solvent-assisted linker exchange

**SAM** self-assembled monolayer

**SBU** secondary building unit

**SURMOF** surface-mounted metal–organic framework

**UFF** universal force field

**vdW** van der Waals

**XRD** X-ray diffraction

**ZIF** zeolitic imidazolate framework

**ZIF-8** zeolitic imidazolate framework-8

# CHAPTER 1

# Introduction

## 1.1 Extended porous materials

Porous materials are an important class of materials with a long history of industrial application for heterogeneous catalysis, gas storage and separation, and ion exchange.[9] These materials are defined by the properties of their pore network. In particular, microporous materials, which are the focus of this thesis, are defined as having pores smaller than 2 nm.[10] This makes them ideal for the adsorption of gases and other small molecules. Furthermore, porous materials can be classified as extended, network materials or discrete, molecular materials. This thesis is specifically concerned with the properties of extended porous materials formed from a bottom-up, modular synthetic approach, whereby distinct building blocks come together through covalent or metal-ligand bonds to form a contiguous network that supports some pore architecture.[11]

The chemistry that governs the connection between building blocks determines the structural properties of the resulting framework. For example, reversible framework growth (e.g. when the building blocks bond through labile metal–ligand bonds) allows for the formation of ordered, crystalline materials.[12] Alternatively, irreversible bonding during framework growth leads to the formation of disordered, amorphous materials (Figure 1.1a).[13,14] Furthermore, the modular synthetic approach allows for the systematic modification of the building blocks toward materials that are finely tuned, with respect to their pore chemistry and structure, for target applications (Figure 1.1b).[12,15,16] Importantly, the number of potential building blocks is vast, and so too are the libraries of existing and hypothetical extended porous materials.[17] Hence, the chemical space that has been and may be covered by porous materials is immense.

In the following sections, we introduce the two subclasses of porous materials and computational strategies that are the subject of this thesis. In the remaining chapters,

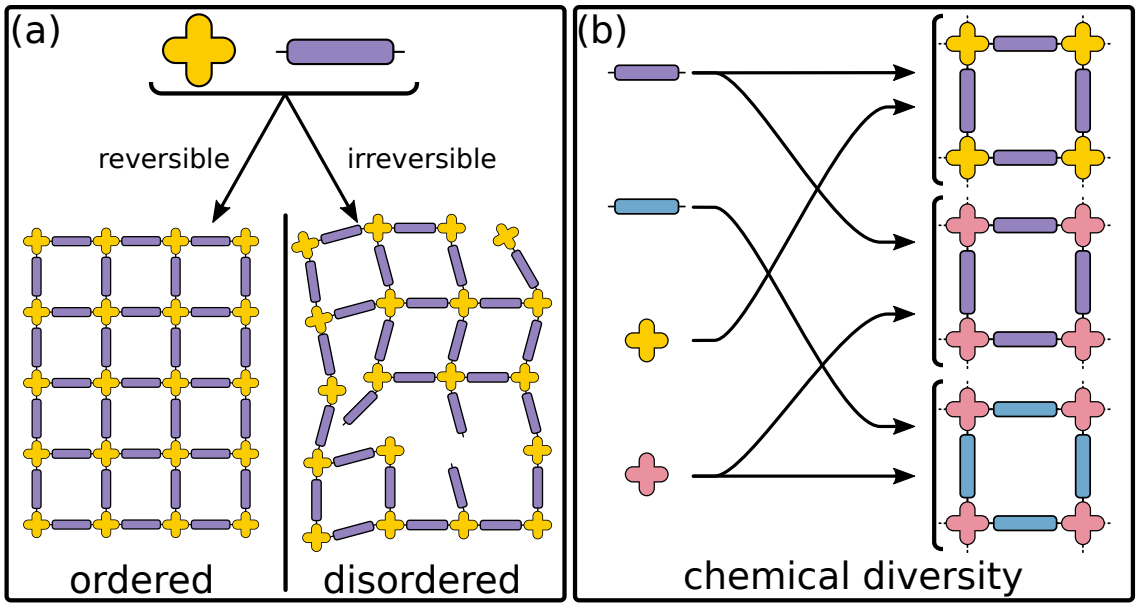



Figure 1.1: (a) Schematic of the impact of the framework formation mechanism on the extent of order in extended framework materials. (b) Schematic of the chemical diversity that arises from a bottom-up and modular synthetic approach.

we apply computation and theory to understand material assembly phenomena and their impact on material properties and to explore chemical space toward new and advanced materials efficiently.

## 1.2 Porous amorphous polymers

Porous amorphous polymers constitute a broad subclass of extended porous materials formed through irreversible covalent linkages of organic building blocks.[13,14] There are several different classes of porous amorphous polymers, such as conjugated microporous polymers (CMPs),[18,19] hyper-cross-linked polymers (HCPs),[20,21] polymers of intrinsic microporosity (PIMs)[22,23] and porous aromatic frameworks (PAFs).[24,25] There exists a diverse set of chemical reactions and building blocks that have been developed to form porous amorphous polymers, which has led to large libraries of synthesized materials with distinct properties.[14,15] Material growth using irreversible reactions leads to the formation of amorphous materials, i.e. materials without long-range order (shown schematically in Figure 1.1a). The amorphous nature of these extended networks makes them difficult to characterize on the molecular or nano scales using experimental techniques, such as X-ray diffraction (XRD).[13,26] Nonetheless, many of these materials have

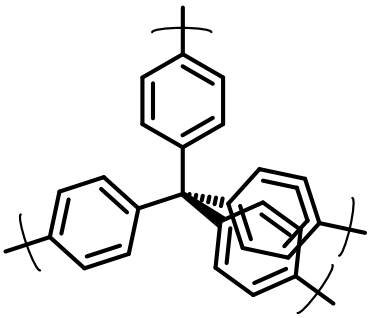


Figure 1.2: Chemical structure of the PAF-1 monomer.[27]

been shown to be porous and have useful catalytic and electronic properties.[14] PAFs are a class of porous amorphous polymers formed via cross-coupling reactions of tetrahedral building blocks containing aromatic substituents (Figure 1.2).[24,25] PAFs have some of the highest recorded Brunauer–Emmett–Teller (BET) surface areas of microporous materials (e.g. PAF-1: $5640\,\mathrm{m^2\,g^{-1}}$,[27] PPN-4: $6461\,\mathrm{m^2\,g^{-1}}$[28]). Importantly, the use of covalent linkages between building blocks provides these materials with high physiochemical stability.[27,28] With exceptional porosity and stability, PAF materials have found use in many applications, including gas capture and separation,[25,29] and have been shown to improve the performance and long-term stability of membrane materials.[30,31] Nevertheless, PAFs exhibit a wide range of porosities depending on the building block used, which is not well understood. Chapter 3 presents a computational study into the formation mechanism of PAFs to investigate the molecular-level processes that underpin their porosity. This work highlights the need for computational strategies when experimental elucidation of a materials molecular-level structure is not possible.

## 1.3 Metal–organic frameworks

A significant challenge with porous amorphous polymers is the experimental characterization of their structure due to the high degree of disorder. In contrast to this, metal–organic frameworks (MOFs), which form via reversible coordination bonds, have ordered structures (Figure 1.1a) that can be characterized using techniques such as single-crystal XRD. Since the discovery of permanent microporosity in the archetypal MOF, MOF-5,[32] these materials have been the subject of significant research effort. MOFs are formed through a modular approach, in which distinct building blocks (or

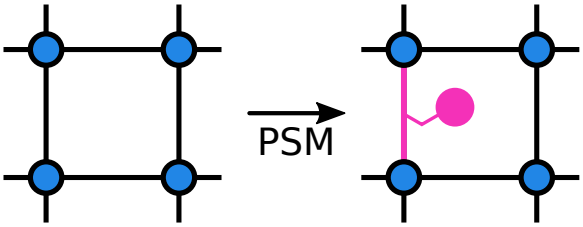



Figure 1.3: Schematic of the post-synthetic modification (PSM) of MOFs to introduce new chemistry into the pore network.

secondary building units (SBUs)) come together to form a 1D, 2D or 3D framework.[33] In contrast to other microporous ordered materials (e.g. zeolites[9]), many thousands of new MOFs have been discovered as a result of the modular synthetic approach[11,16] and relative ease of the synthesis.[34] A unique property of MOFs is that the pore network and chemistry can be precisely controlled and finely tuned toward a target application using multiple strategies.[33] Furthermore, due to the crystalline nature of MOFs, this pore network is well defined throughout a single crystal. Firstly, the chemical and structural properties of the SBUs can be treated independently from the chemistry that connects them. Hence, by using SBUs of different sizes and different connectivities many frameworks (with tunable pore sizes and shapes, as well as diverse topologies) can be achieved in a process termed "reticular chemistry".[16] Additionally, chemical functionalization of the SBUs allows for modification of the chemistry of the pore network. Different pore chemistries can be introduced by the synthesis of isoreticular MOFs with functionalized ligands or MOFs with multiple ligands with distinct functionalities in the same crystal.[35,36] Post-synthetic modification (PSM) of a MOF through chemical functionalization of a linker or the inclusion of functional guests can be used to control pore chemistry and introduce new functionality (Figure 1.3).[37–39] In summary, the chemical and structural diversity of the building blocks of MOFs, the modular synthetic approach, and the ability to post-synthetically modify MOFs has resulted in an immense library of MOF materials over the past two decades.

The experimental approaches described above indicate that MOFs can be finely tuned toward applications rationally. Indeed, MOFs have found use in a broad range of applications (including gas storage and separations,[40,41] sensing,[42] catalysis[43] and many more)[44] for which the ability to finely tune their properties is crucial. The ability to control the chemistry of MOFs allows for their integration with other materials. For

example, the inclusion of MOFs into mixed-matrix membranes as filler agents can improve membrane performance (i.e. permeability and selectivity).[45] In this thesis, we study two emerging applications of MOFs: in Chapter 4, we study the integration of MOFs into macroscopic structures and in Chapters 5 and 6 we study the MOF-based encapsulation of biological entities. We briefly introduce these two applications of MOFs in the following sections.

### 1.3.1 Integration of MOFs into thin films

As described above, MOF chemists have substantial control over the nature of the pore network at the single-crystal scale. Recently, the growth of MOF structures from macroscopic substrates has led to the translation of MOF properties to the macroscale.[46–51] In particular, thin-film fabrication of MOFs, through layer-by-layer synthesis, liquid-phase epitaxy,[52] and chemical-vapor or atomic-layer deposition,[53,54] allows for the formation of uniform MOF layers. The ability to extend the well-defined pore structure of MOF crystallites to the macroscale is crucial for the application of MOFs in fields such as optoelectronics, gas separations, and sensing.[47,52,54]

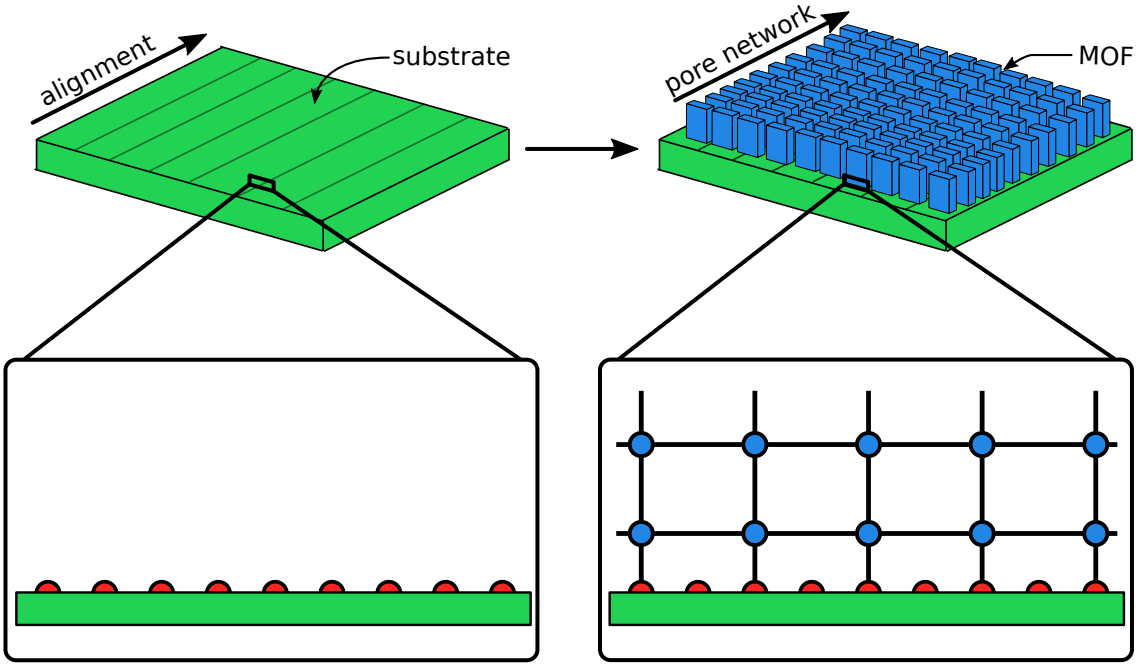



Figure 1.4: Schematic of the macroscale alignment of MOF crystallites (blue rectangles) grown from a substrate (green surface) via heteroepitaxy.[55] Matching of the MOF lattice with the binding sites of the substrate (red semi-circles) leads to heteroepitaxial growth and MOF alignment, which results in macroscale orientation of the MOF pore network.

A growing class of MOF thin films are surface-mounted metal–organic frameworks (SURMOFs),[52] in which MOF growth occurs from self-assembled monolayers (SAMs) in a layer-by-layer process. SURMOFs have garnered significant interest because of their homogenous morphology, the ability to precisely control their thickness, their low defect densities and their potential for scale up to industrial settings.[56] SURMOFs are potentially useful in many applications that take advantage of their precise structures such as gas adsorption and separations, sensing and electronic applications.[52,56] The out-of-plane orientation of SURMOFs with respect to the substrate can be controlled by changing the chemistry of the SAM.[57–59] The precise control over the thickness and orientation of SURMOFs offers a well-defined model system for studying formation mechanisms,[60] chiral separations,[57] and other intrinsic properties of MOFs.[52] However, in-plane alignment is also required to achieve a precise translation of the single-crystal scale pore network to the macroscale. Recently, Falcaro and co-workers achieved macroscale in-plane and out-of-plane alignment of MOF crystallites via epitaxial MOF growth from a well-ordered substrate in a facile, one-pot method (Figure 1.4).[55] Thorough XRD analysis confirmed macroscale alignment of a series MOFs, where epitaxial matching between the substrate and MOF was found to be necessary for aligned growth. However, there are very few examples of MOF thin films with controlled in-plane and out-of-plane orientation. Therefore, in Chapter 4, we develop and apply a computational screening process to rapidly identify candidate MOFs for aligned heteroepitaxial growth in order to expand the range of MOFs for which aligned thin films can be produced.[55]

### 1.3.2 Encapsulation of biomacromolecules inside MOFs

The immense diversity of MOF structures has assisted with their integration into biological fields. For example, the ability to tailor the structure and chemistry of the pore network toward a particular guest has been employed in drug delivery and biosensing studies.[61–63] Recently, the encapsulation of biomacromolecules inside MOFs has become of significant interest because of the potential of MOFs to protect biomacromolecules from harsh environments.[64–66] The encapsulation of biomacromolecules has been achieved in a variety of ways, including through their infiltration into pre-assembled MOF networks (Figure 1.5a).[64–66] Alternatively, MOF crystallization can be induced

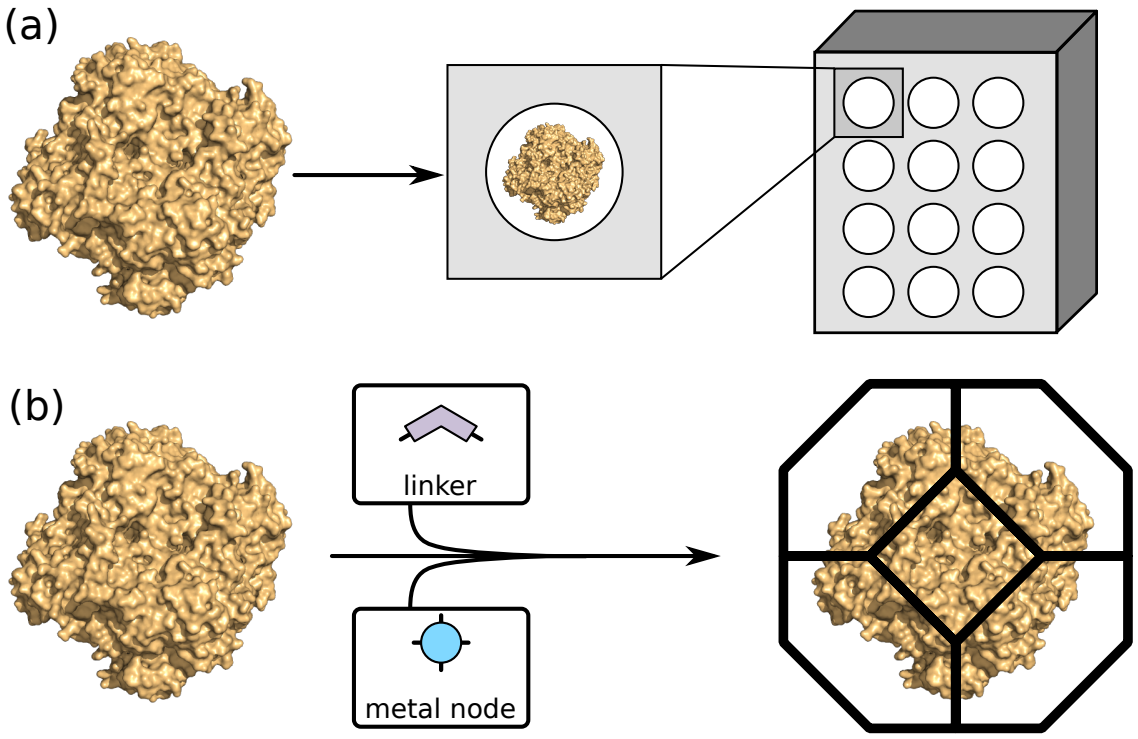



Figure 1.5: Schematic of enzyme encapsulation into a MOF lattice via (a) the diffusion of the enzyme (shown in gold) into the MOF pores[64] or (b) the biomimetic mineralization of ZIFs, which is shown schematically as a sodalite cage surrounding the enzyme.[67]

in the presence of a biomacromolecule in a process termed "biomimetic mineralization", which leads to the formation of a MOF shell around the biomacromolecule (Figure 1.5b).[67] Zeolitic imidazolate framework-8 (ZIF-8),[68,69] which is formed from $Zn^{2+}$ metal-ion nodes and 2-methylimidazole linkers, is the most widely explored MOF for biomimetic mineralization. ZIF-8 is chemically and thermally stable and can be synthesized in biologically compatible conditions.[70–72] Note that biomimetic mineralization has been achieved for other zeolitic imidazolate framework (ZIF) materials.[8]

Biomimetic mineralization (Figure 1.5b) is a facile process that has been used to encapsulate biomacromolecules, including proteins, enzymes, DNA, and living cells and viruses.[67,73–76] The encapsulation of biomacromolecules inside MOFs is a very new field with many open questions about the accessibility, stability and activity of encapsulated enzymes.[66] Nonetheless, there are many potential applications for this technology. For example, the protection of therapeutic agents from harsh conditions by encapsulating them in MOFs could help avoid degradation during transport, which is a significant challenge for the administration of vaccines and therapies.[66] Another potential application for this technology is the encapsulation and protection of enzymes.

Enzymes are proteins that catalyze chemical reactions with remarkable efficiency and selectivity, but they are generally not stable to harsh industrial conditions and favor homogeneous reaction conditions.[64] Therefore, their application in industry is limited. However, enzymes encapsulated inside MOFs (via biomimetic mineralization or infiltration) have been shown to have improved stability to temperature and denaturing conditions.[64,66,67,77] Furthermore, the MOF matrix can act as a gate for selective transport of molecules to the enzyme and simplifies the collection and reuse of the catalyst (compared to the use of enzymes in a homogeneous solution).

These potential applications of MOF-encapsulated biomacromolecules has stimulated significant research in this field, but there remain many open questions that need to be answered before the widespread application of these materials.[66] In particular, the impact of encapsulation on enzymatic activity remains unclear.[8] Only a small number of enzymes and MOFs have been studied experimentally for biomimetic mineralization. By better understanding the mechanism of biomimetic mineralization by MOFs and by developing high-throughput tools for screening MOFs and/or enzymes, the range of materials and reactions can be rapidly expanded. To this end, the research presented in Chapter 5 details studies into the mechanism that drives biomimetic mineralization and how to apply it in general to protein and carbohydrate systems. Chapter 5 shows that the number of enzymes that have a propensity to seed MOF growth is substantial. Thus, in Chapter 6 we provide simplified ways to predict the viability of an enzymatic reaction occurring inside MOFs to assist in the selection of new experiments.

## 1.4 Materials design using computational screening

As highlighted above, porous materials cover a broad range of chemical space and have immense potential for many applications. It remains intractable to synthesise all possible combinations of building blocks to determine the optimal material.[78] Therefore, it is crucial to be able to efficiently explore chemical space when designing new materials and new experiments. Computational screening processes that efficiently determine the viability of a target molecule have been commonplace in the field of drug discovery for many years.[79] In this case, large databases of potential candidates are screened computationally, where candidates are filtered and ranked in a step-wise approach using

simple descriptors. Experimental trials are then conducted only on the remaining and top-ranked candidates, thereby avoiding arduous experimental effort. Figure 1.6 shows an idealized screening process that takes a library of candidates and simplified descriptors that are defined using theory and experiments and produces future experiments, advances on the existing theory, and useful materials. In the example of drug discovery, the library of candidates are computationally screened using simplified descriptors to predict the properties of interest (e.g. solubility) to determine their viability for a specific application. This type of approach to drug discovery relies on large amounts of available input data for small molecules (e.g. data available in the PubChem,[80] and ZINC[81] databases), as well as their interactions with biological entities (e.g. available in the ChEMBL[82] database). The field of high-throughput computational drug discovery has been highly successful due to the ability to describe properties of drug-like compounds using easily calculated descriptors. For example, Lipinski's rule of five provides five simple molecular properties that can be used to efficiently screen drug molecules.[83] Similarly, simplified models and descriptors of the relevant chemical and physical processes are required for the use of any computational screening process.

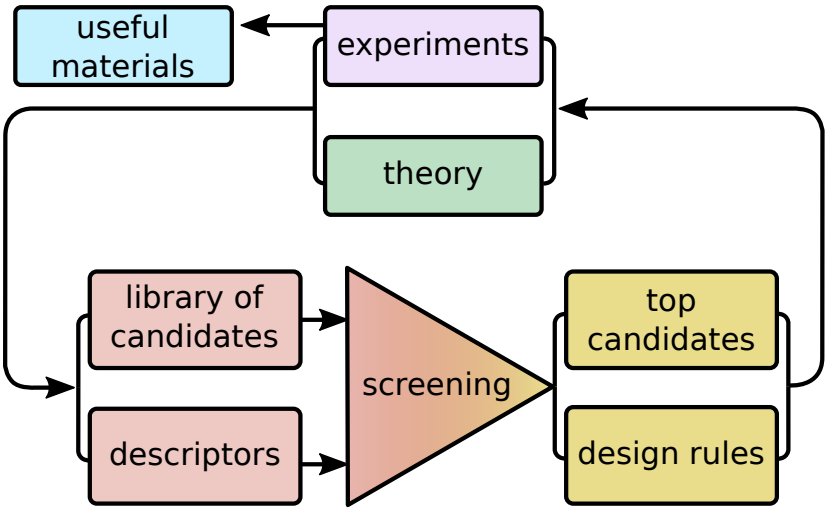



Figure 1.6: Flowchart of a computational screening process.

Recently, there has been a push toward the collation of large amounts of data on functional materials, including porous materials, to allow computational screening.[78,84,85] An example of this is the Materials Genome Initiative (akin to the Human Genome project), which attempts to bring together and make available experimental and computational data on a large and diverse range of materials to enhance the ma-

terials discovery process.[86] Recent studies have used multiple approaches to exploit the size of the Materials Genome Initiative to discover design principles (structure–property relationships) as well as porous materials that could be optimal for hydrogen[17] and methane[87] storage and many other applications.[88–92] Recently, Moghadam and co-workers used a data set of experimentally known MOFs to computationally screen for materials for oxygen storage. Their process resembled that shown schematically in Figure 1.6, in which large-scale computational screening led to the synthesis of a top candidate MOF that performed as predicted during experiments for oxygen storage.[93]

Crucial to these processes is the development of databases of known[94] and hypothetical porous materials.[95,96] In particular, the modular nature of MOFs has allowed for many in silico MOF generation algorithms.[97–100] Wilmer and co-workers first reported the generation of 138 000 hypothetical MOF structures by connecting SBUs at predefined binding sites based on known MOF chemistry.[101] However, the a priori prediction of the structure of crystalline porous materials remains a daunting task, and many computationally designed materials are not synthetically feasible. One approach to overcome this deficiency is to model material formation, but this is not a trivial task and has only been carried out a handful of times for MOFs[102,103] and covalent organic frameworks (COFs).[104,105] Furthermore, an understanding of the formation mechanisms of materials can help elucidate the relationship between local structure and global material properties.[104] More widespread calculations of the assembly of porous materials will require the development of more efficient simulation methods and models. Chapter 3 shows how a simplified model can be used to investigate the formation mechanism of PAFs efficiently, which we use to understand the main factors underlying their impressive porosity.

Simplified descriptors of target properties are required to be able to predict the viability of a particular material efficiently. The vast majority of screening studies of porous materials have focussed on their adsorption properties.[78,84] This is partially because of the gas storage and separation applications that have dominated much research on porous functional materials, and partially because simple geometrical descriptors of the pore network (discussed in detail in Section 2.2) and accurate and efficient methods for the simulation of adsorption processes exist. Furthermore, the properties of interest

to specific adsorption or separation applications are well understood and can be clearly defined.[78,84] Many other properties of MOFs are calculable, e.g. electronic[106–109] and mechanical[110–112] properties, using computationally expensive techniques (e.g. quantum mechanical methods).[84,113] However, these methods are not generally amenable to high-throughput screening of large data sets and computationally cheap descriptors for these properties are difficult to identify.[84] Therefore, very few high-throughput screening studies for applications other than adsorption exist, with a handful of small-scale computational screening studies of MOFs addressing applications in catalysis,[114] drug delivery,[63] and thin-film growth[115,116] have been performed.

A focus of this thesis is to develop and identify simplified models and descriptors that can be used to study and predict the properties of porous materials. In Chapters 3, 4 and 5 we simplify several assembly phenomena in porous materials to their essential components to facilitate their study. In Chapter 3, we use a coarse-grained model to simulate the kinetically controlled formation of PAFs on large length and time scales. Chapter 3 is unique from the remaining chapters because we have used a simplified simulation model to investigate the dynamics and kinetics of material formation. In Chapter 5, we use computation and theory to verify that simple descriptors of the electrostatic properties of a protein can predict whether the protein will seed ZIF growth or not. This result indicates that computational filtering of biomacromolecules for encapsulation inside MOFs can be done accurately and efficiently. Finally, in the studies discussed in Chapters 4 and 6 we use novel simplified descriptors and high-throughput screening to identify candidate materials/molecules for applications in two relatively new fields of MOF chemistry. In both cases, the number of promising candidate materials/molecules is increased significantly over the number already studied experimentally in a few days of computation time, highlighting the importance of low-cost, high-throughput computational strategies.

# CHAPTER 2

# Computational methods

The following chapter broadly introduces the computational methods and theories applied throughout this thesis. Each following research chapter includes detailed methodology sections that describe the work carried out by the candidate.

## 2.1 Classical molecular simulations

Classical molecular simulations, including molecular dynamics (MD), Monte Carlo (MC) methods and molecular mechanics (MM), are computational tools that describe the interactions in molecular systems using classical mechanics. MD and MC approaches are efficient and allow for the sampling of large systems ($10^5$ or more particles) for long times (up to $\sim$1 ms).[117] In the realm of porous materials, a range of molecular simulation methods are commonly used to accurately and efficiently model adsorption[78] and diffusion[118] processes as well as MOF flexibility.[84] In this thesis we have used molecular simulations specifically to model the formation of porous materials (Chapter 3), and for computationally efficient energy minimizations of molecular and framework structures (Chapters 4 and 6).

Classical mechanics offers a low-cost strategy to define the energy of a chemical system based on the configuration of the particles (e.g. atoms) and the bonds between them. In this case, the energy of a system is defined by a force field, which is typically a sum of several terms, each of which is a function of the particle (atom) positions:

$$U = U_{\text{vdW}} + U_{\text{electrostatic}} + U_{\text{bond}} + U_{\text{angle}} + U_{\text{dihedral}} + U_{\text{inversion}}, \tag{2.1}$$

where $U_{\text{vdW}}$ and $U_{\text{electrostatic}}$ represent nonbonded van der Waals (vdW) and electrostatic (Coulombic) interactions, respectively, $U_{\text{bond}}$, $U_{\text{angle}}$, $U_{\text{dihedral}}$ and $U_{\text{inversion}}$ are the bond stretch, bond angle, dihedral angle and inversion potentials, respectively. The terms

$U_{\text{bond}}$, $U_{\text{angle}}$ and $U_{\text{dihedral}}$ represent the bonded potentials that are functions of bond lengths, bond angles and dihedral angles, respectively, within a system. The term $U_{\text{inversion}}$ describes the inversion potential, or improper dihedral potential, of a bonded system. Force fields are parameterized to reproduce thermodynamic or geometrical properties of a system based on ab initio calculations or experiments. The universal force field (UFF),[119] which is used extensively in this thesis, has been parameterized to describe the interactions between all atoms on the periodic table, and therefore can be generally transferred to any system. Universality often results in a decreased accuracy, but it is useful for studies of a broad range of chemical systems. Importantly, Addicoat and co-workers have extended UFF to accurately describe MOF systems.[120,121]

Classical MD simulations provide a means to model the evolution of systems of particles as a function of time.[122] When considering the dynamics of condensed-phase molecular systems, classical MD is often more accurate and physically appropriate than quantum mechanical methods. The motion of particles in MD is determined by numerically solving Newton's equations of motion,

$$\boldsymbol{F}_i(\boldsymbol{r}^N) = m_i \frac{\mathrm{d}^2 \boldsymbol{r}_i}{\mathrm{d}t^2} = -\nabla_i \boldsymbol{U}(\boldsymbol{r}^N), \tag{2.2}$$

where $t$ is the time, $\boldsymbol{F}_i$ is the force vector on particle $i$, $m_i$ is the mass of particle $i$, $\boldsymbol{r}_i$ is the position of particle $i$, $\boldsymbol{r}^N$ is the configuration of the N-particle system, and $\boldsymbol{U}$ is the potential energy described by the force field. Note that Equation 2.2 applies to constant energy simulations and to consider other thermodynamic ensembles additional terms are necessary, as discussed further below. During the simulation of molecular systems, periodic boundary conditions and the minimum image convention are applied (Figure 2.1). These state that particles that cross the periodic boundaries appear on the other side of the simulation box and that only the energy due to interactions between a particle and the nearest image of each of its neighbours is included. Furthermore, simulation efficiency is improved by implementing a cutoff distance, whereby the interactions between pairs of particles further apart than this cutoff distance are not considered.

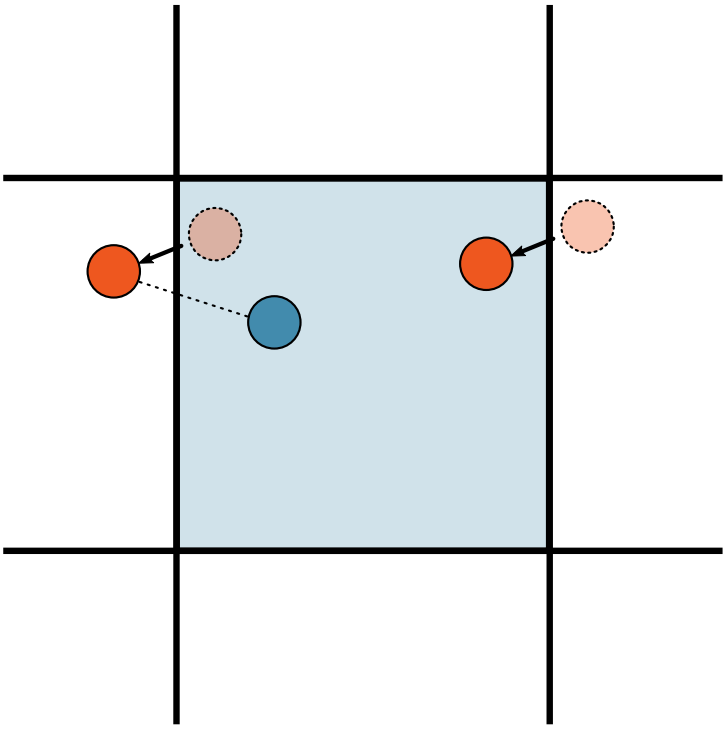


Figure 2.1: Schematic illustrating the concepts of periodic boundary conditions and the minimum-image convention in MD simulations. The primary simulation cell that is simulated is shaded blue. The particle in red translates across the boundary and appears on the other side, but due to the minimum image convention, the blue particle interacts with the image of the red particle in the neighbouring simulation box after the move (the dashed line represents the interaction).

### 2.1.1 Coarse-grained molecular dynamics

Simulations of atomistic models are useful because they are accurate and allow for the analysis of atomistic levels of detail. However, because all atoms are explicitly specified, all-atomistic systems are not computationally efficient for simulations of large systems over long times. Therefore, to study slow molecular processes, such as self-assembly, it is necessary to improve simulation efficiency. The efficiency of molecular simulations can be significantly improved by rationally agglomerating atoms into a smaller number of particles or treating them as a continuous medium, i.e. coarse-graining the system. Coarse-graining aims to simplify an atomistic system while maintaining its overall physical and thermodynamic properties. The first step to significantly reducing the number of atoms in a system is to replace the solvent atoms with an implicit solvent, e.g. by using Langevin dynamics (LD).[123] LD represents the collisions of molecular bodies with solvent atoms by augmenting Newton's equations of motion (Equation 2.2)

with a frictional force and a fluctuating random force to give (for spherical particles)

$$F(t) = -\nabla U(x) - \xi_{\mathrm{t}} m \boldsymbol{v}(t) + f_{\mathrm{R}}(t), \tag{2.3}$$

where $F(t)$ is the total force on a body, $\nabla U(x)$ is the conservative force due to the interaction potentials acting on the body, $\xi_{\mathrm{t}}$ is the translational friction coefficient, $m$ and $\boldsymbol{v}(t)$ are the mass and translational velocity of the body at time $t$, respectively, and $f_{\mathrm{R}}(t)$ is the random force applied to the bodies, which is assumed to have zero mean and be $\delta$-correlated in time and is related to the frictional force by the fluctuation-dissipation theorem,[123,124] which gives $\langle f_{\mathrm{R}}(t) \rangle = 0$ and $\langle f_{\mathrm{R}}(t) f_{\mathrm{R}}(t') \rangle = 2\xi_{\mathrm{t}} k_{\mathrm{B}} T m \delta(t - t')$.[123,125] To further reduce the number of atoms in a model, multiple atoms in a molecule can be mapped into a single interaction site to produce a coarse-grained (CG) model. The design of CG models is often based on unifying atoms in such a way that freezes fast degrees of freedom, such that the integration of the equations of motions can be done over longer time steps. For example, unifying C–H groups into a single site removes the need to model the very fast C–H bond stretch.

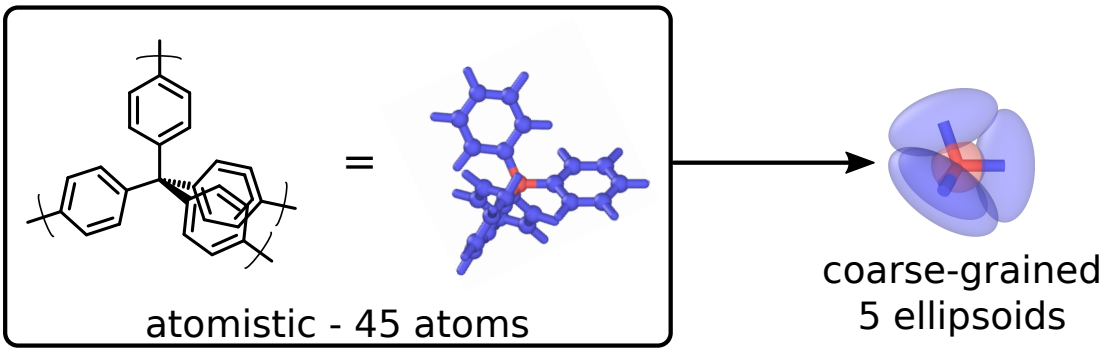



Figure 2.2: Coarse-graining scheme applied to porous aromatic frameworks (PAFs) in Chapter 3. The atomistic structure on the left (of 45 atoms) is coarse-grained to a model containing five ellipsoids. The all-atom model is colored to match the segments replaced by the ellipsoid of the same color.

Figure 2.2 shows the coarse-graining scheme applied in Chapter 3 for our model of PAFs. In this CG MD model, we applied LD to simulate an implicit solvent at the desired temperature, we replaced effectively rigid groups (i.e. benzene rings) by a single, ellipsoidal interaction site and we froze intramolecular degrees of freedom. We used an ellipsoidal CG model because they can correctly capture the anisotropic properties of the interaction sites. For example, ellipsoidal models of DNA can capture the

anisotropic effects of the stereochemistry of the bases.[126] In Chapter 3, the ellipsoidal model allows for the description of anisotropic interactions between PAF monomers.[1] Note that an additional equation besides Equation 2.3 needs to be solved to describe the evolution of the orientational degrees of freedom of the ellipsoids, as described in Chapter 3. The use of CG models simplifies the physics of a given process to as few variables as is necessary, which allows for the study of complex processes in an efficient way. In particular, the study of assembly processes in biological systems[126,127] and porous materials[102] can be simplified using the strategies discussed above. Indeed, in Chapter 3 a CG model is applied to simulate the assembly of PAFs using large systems and long time scales, which were not accessible using atomistic approaches.[1]

## 2.2 Descriptors of porosity

The pore network defines key properties of porous materials. Especially when considering the adsorption of gases and other small molecules, i.e. for gas separations and storage. Geometrical pore characteristics, such as surface area and pore size distributions, can be determined indirectly from experimental adsorption isotherms. In general, the surface area of porous materials are determined by applying the Brunauer–Emmett–Teller (BET) theory to a $N_2$ adsorption isotherm obtained at 77 K.[128] Use of the BET theory assumes that adsorption occurs on a homogeneous surface, that there is no interaction between multilayers, and that Langmuir theory applies to each layer. Similarly, pore size distributions can be determined indirectly from adsorption data of $N_2$ obtained at 77 K (other conditions may also be used), under the assumptions of a reference pore geometry (i.e. cylindrical, slit-like, or spherical pores) and approximate chemical composition.[129] In both cases, approximations are made that may not hold for MOFs and other microporous materials, due to their complex pore architectures, and because atom-level details are ignored.[84] Furthermore, there is variability in experimental adsorption data among similar materials.[130] Nonetheless, these methods represent standardized characterization techniques that allow for the comparison of different materials.

For crystalline porous materials (e.g. MOFs), many computational methods have been developed to directly determine pore characteristics from an experimental or

calculated crystal structure.[84,131] Geometrical descriptions of the solvent accessible surface of the pore network (Figure 2.3a) can be used to calculate a surface area that agrees with BET theory efficiently.[132,133] Figure 2.3a also shows the van der Waals and Connolly surfaces, which do not accurately consider the accessibility of a pore surface to probe molecules. The Connolly surface is defined as the solvent excluded surface, which is determined by tracing a probe molecule with a radius of 0 Å along the van der Waals surface of all of the atoms.[133] The surface area of a crystal structure can be determined using an MC approach, in which probe particles are placed at random positions inside the framework and their overlap with framework atoms is determined. From the overlap of many probes, a surface area, as well as pore volume, can be calculated. Importantly, the size of the probe particle can be set to match the diameter of the adsorbate used experimentally. However, this type of approach does not consider the connectivity of the pore network. For example, the surface area of a void that is only accessible via channels smaller than the size of the probe should not be included in estimating total surface area of that material because the probe cannot diffuse into that void. Algorithms to calculate surface areas and exclude inaccessible pore space have been implemented in software such as RASPA2, which can also extract pore size distributions.[131]

Voronoi decomposition of the void space in periodic networks of atoms is an alternative and efficient way to analyze the entirety of a pore network.[134,135] Voronoi decomposition represents a mathematical partitioning of the space surrounding $n$ points into $n$ polygons such that the faces of each polygon is a plane equidistant from the two points sharing the face (illustrated in 2D in Figure 2.3b).[134] The nodes of the Voronoi network of a porous material represent the void space in the pores. The algorithm implemented in the open-source software Zeo++,[134] which is used throughout this thesis, calculates advanced geometrical descriptors that take into account the connectivity of the pore network from a periodic Voronoi network.[134,135] In particular, the diameter of the largest included sphere ($D_\mathrm{I}$), the diameter of the largest free sphere ($D_\mathrm{F}$) and the diameter of the largest included sphere along the free sphere path ($D_\mathrm{IF}$) (shown in Figure 2.3c) can be used to better describe the bulk properties of a network of pores. For example, $D_\mathrm{F}$ represents the largest sphere able to freely diffuse through the pore network, which can help understand the adsorption and separation of molecules based

on their size.[134] Furthermore, the Voronoi network of a system can be used as a fingerprint to describe porous materials. For example, Lee and co-workers used topological analysis of the pore-space fingerprints obtained by Zeo++ to identify similar materials, based purely on the properties of their pore network.[136] This approach accurately finds pairs of zeolites with similar pore networks and gas storage performances in a high-throughput way.[137]

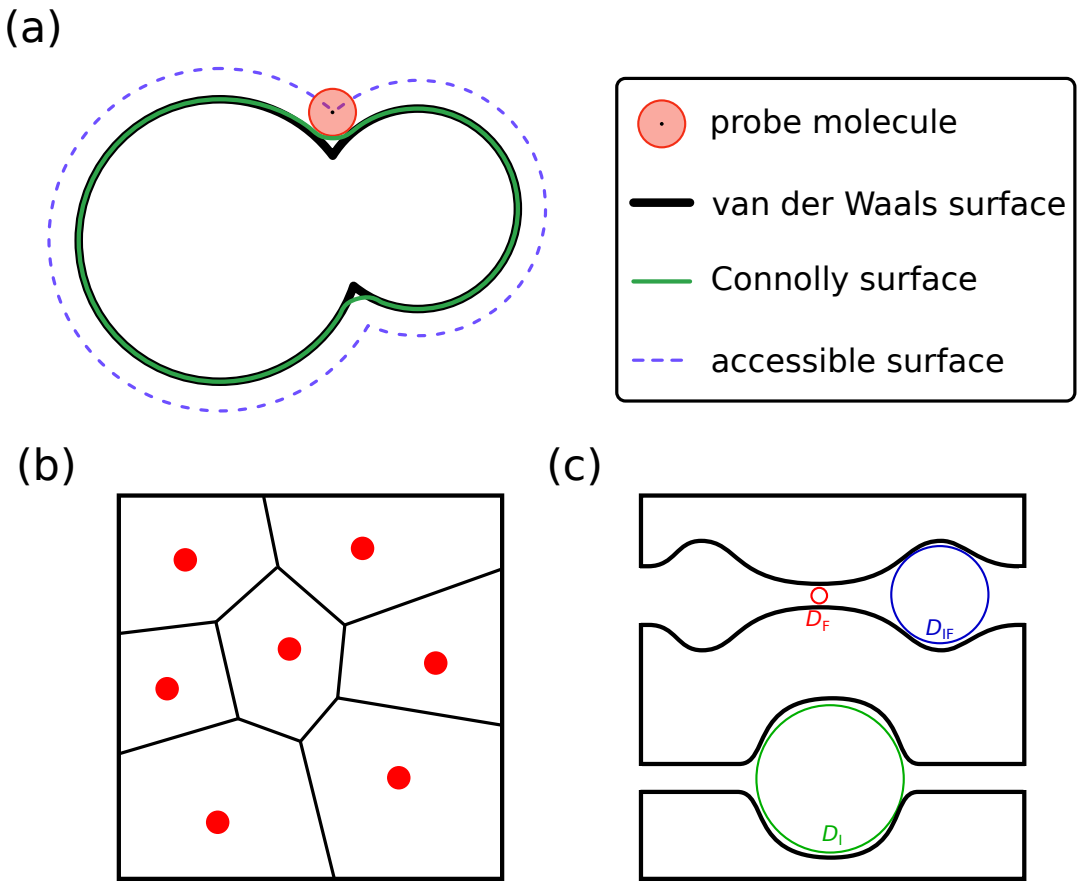



Figure 2.3: (a) Schematic of the van der Waals surface (black line), Connolly surface (green line) and accessible surface (dashed purple line) for a given probe molecule (red circle) used to calculate geometric porosities of porous materials. (b) A representative depiction of a Voronoi decomposition of 2D space. (c) Definition of the diameter of the largest included sphere ($D_I$), the diameter of the largest free sphere ($D_F$) and the diameter of the largest included sphere along the free sphere path ($D_{IF}$).

The methods described above specifically analyze rigid periodic structures. Importantly, the porosity of dynamic and amorphous systems has also been successfully captured using a combination of MD simulations to build structural models and Voronoi decomposition to analyze the pore space of each transient structure.[138–140] In Chapter 3, CG MD simulations were used to simulate the early-time formation of discrete,

amorphous PAF clusters, for which the porosity was approximated using Zeo++ and the techniques described above.[134]

## 2.3 Modelling electrostatic properties of biomacromolecules in solution

Electrostatic interactions play vital roles in biological systems, which is evidenced by the number of atoms carrying partial charges present in biological systems.[141,142] Electrostatic complementarity between contacting surfaces is known to drive enzyme–substrate binding and protein–protein interactions.[143–145] The presence of ionizable groups (on the side chains of amino acids) in a polypeptide chain results in electrostatic interactions that are crucial in defining their structure, which inevitably defines their function.[145] Furthermore, because amino acids are ionizable, their charge state is a function of pH and their environment. Hence, the electrostatic properties of biomacromolecules are influenced by the solution surrounding them. Computational and theoretical methods can be used to model electrostatic interactions in biological systems.[142,143,145] However, modelling biological systems explicitly, i.e. by considering each atom, is computationally expensive.[142] Therefore, simplified models based on mean-field approximations are often used to implicitly model the components (e.g. the electrolyte solution) of these systems.

Continuum models of electrolytes, based on a mean-field approximation of ions in solution, have been widely applied to study a range of electrolyte phenomena accurately.[146] In such models, the relationship between the electrostatic potential and the spatial distribution of charges is typically described by the Poisson–Boltzmann (PB) equation,

$$\nabla \cdot [\varepsilon(r)\nabla\psi(r)] + e\left[\sum_i c_{i,\mathrm{bulk}} z_i \exp\left(\frac{-z_i e\psi}{k_\mathrm{B}T}\right)\right] = -\rho(r), \tag{2.4}$$

where $\psi$, $\rho(r)$ and $\varepsilon(r)$ are the electrostatic potential, fixed (i.e. non-electrolyte) charge density and the spatially varying dielectric permittivity at position $r$, respectively, $c_{i,\mathrm{bulk}}$ and $z_i$ are the bulk concentration and valency of ions of type $i$, respectively, $e$ is the elementary charge, $k_\mathrm{B}$ is the Boltzmann constant and $T$ is temperature. Note that the sum is over all types of ions in the electrolyte solution. Equation 2.4 is referred to as

the nonlinear PB equation. Analytical solutions to the nonlinear PB equation exist in only a few simple cases (e.g. near a charged plate for a constant dielectric permittivity, shown in Figure 2.4a),[146] but by solving Equation 2.4 numerically it is possible to obtain ionic distributions surrounding complex systems, such as proteins.[147] To model the electrostatic properties of proteins in Chapter 5, two simplifications were applied to Equation 2.4. Firstly, the system was split into two separate media (Figure 2.4b) with different $\varepsilon(r)$: the solute (the protein) with a low dielectric constant ($\varepsilon_{\text{int}}$), and the solvent (or electrolyte) with a high dielectric constant ($\varepsilon_{\text{ext}}$). In this case, solvent ions are excluded from the solute and the charge density of the solute is constant.

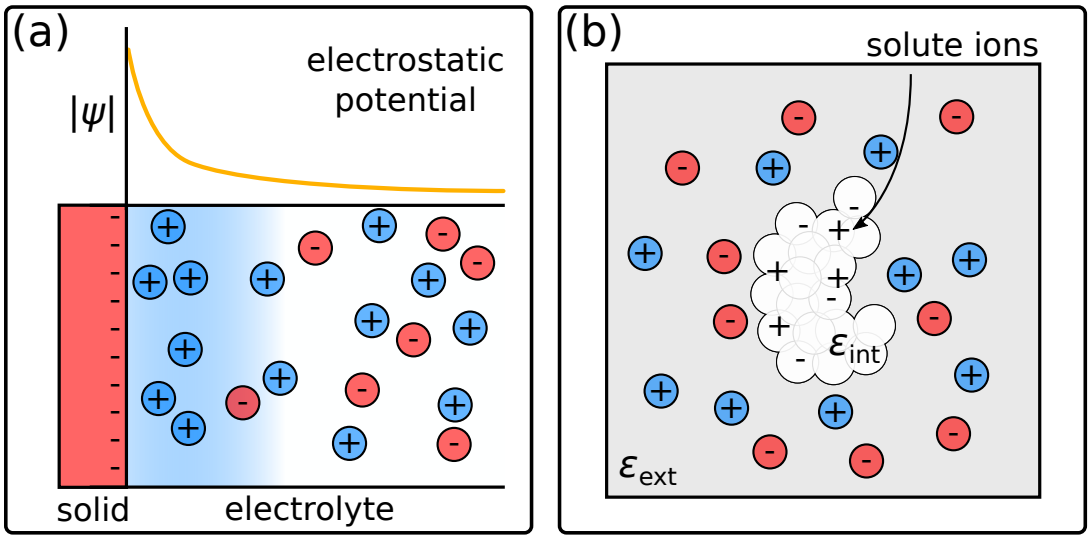



Figure 2.4: (a) Representative ion distribution and electrostatic potential (represented by the yellow line) near a charged plate. (b) Schematic of a two-phase system with an internal dielectric medium representing the solute ($\varepsilon_{\text{int}}$) and an external dielectric medium representing the solvent ($\varepsilon_{\text{ext}}$).

Solving the nonlinear PB equation for large systems can be computationally demanding. Therefore the second simplification was to use the linearized PB equation. In the low-potential regime ($z_i e\psi \ll k_{\text{B}}T, \forall z_i$),[142] Equation 2.4 can be linearized using the first two terms of the series expansion

$$\sum_i c_{i,\text{bulk}} z_i e \exp\left(\frac{-z_i e\psi}{k_{\text{B}}T}\right) = \sum_i c_{i,\text{bulk}} z_i e \left(1 - \frac{z_i e\psi}{k_{\text{B}}T} + \cdots\right) \tag{2.5}$$

$$= \sum_i c_{i,\text{bulk}} z_i e - \sum_i \frac{(z_i e)^2 c_{i,\text{bulk}}\psi}{k_{\text{B}}T} \tag{2.6}$$

$$= -\sum_i \frac{(z_i e)^2 c_{i,\text{bulk}}\psi}{k_{\text{B}}T}, \tag{2.7}$$

since $\sum_i c_{i,\text{bulk}} z_i e = 0$ due to charge neutrality of the bulk electrolyte, to give the linearized PB equation,[148]

$$\nabla \cdot [\varepsilon(r) \nabla \psi(r)] - \varepsilon_\text{r} \varepsilon_0 [\kappa(r)]^2 \psi(r) = -\rho(r), \tag{2.8}$$

where $\varepsilon_0$ is the vacuum permittivity, $\varepsilon_\text{r}$ is the relative permittivity of the solution (which is water (80) in all cases in this thesis), $\kappa(r)$ is the Debye screening parameter, given by

$$\kappa = \left( \frac{\varepsilon_\text{r} \varepsilon_0 k_\text{B} T}{2 e^2 I} \right)^{-1/2}, \tag{2.9}$$

for values of $r$ in the solution and taken to be zero for $r$ inside of the solute. $\kappa$ in the solution is the inverse of the Debye screening length $\lambda_\text{D}$, i.e. $\lambda_\text{D} = \kappa^{-1}$. In Equation 2.9, $I$ is the ionic strength of the electrolyte solution, given by $I = \frac{1}{2} \sum_i c_{\text{i,bulk}} z_i^2$. Finite-difference methods can be used to solve Equation 2.8 numerically, by mapping all physical quantities onto a grid.[148] Indeed, this approach has been implemented in software, such as APBS[149] and DelPhi.[147,150]

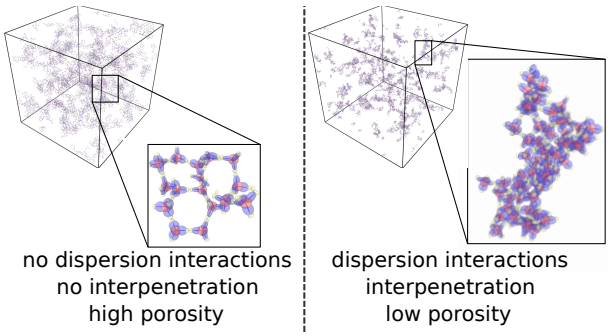
Similarly, the nonlinear PB equation (Equation 2.4) in the solution can be simplified to

$$\varepsilon_0 \varepsilon_\text{r} \nabla^2 \psi + e \left[ \sum_i c_{i,\text{bulk}} z_i \exp\left(\frac{-z_i e \psi}{k_\text{B} T}\right) \right] = -\rho(r), \tag{2.10}$$

by assuming a constant dielectric permittivity, i.e. $\varepsilon(r) = \varepsilon_0 \varepsilon_\text{r}$ for all $r$. This is useful for a given model where distinctions between solute and solvent are physically unclear. In each of the cases above, the underlying physics being modelled is similar while the models and approximations used to do so differ. In Chapter 5, the linear (Equation 2.8) and nonlinear (Equation 2.10) PB equations are successfully used to study the interactions of zinc ions (precursors to zeolitic imidazolate framework-8 (ZIF-8)) with charged proteins and carbohydrates to verify the role of electrostatic properties in the biomimetic mineralization process.

# CHAPTER 3

# Molecular insight into assembly mechanisms of porous aromatic frameworks




*Published*

This work has appeared in the following publication:

Tarzia, A.; Thornton, A. W.; Doonan, C. J; Huang, D. M. Molecular Insight into Assembly Mechanisms of Porous Aromatic Frameworks. *J. Phys. Chem. C* **2017**, *121*, 16381–16392. DOI: 10.1021/acs.jpcc.7b04860

**Principal author (candidate):** Andrew Tarzia

**Contribution:** 85%

Design of coarse-grained model and carried out simulations and analysis. Construction of figures. Conception, writing and drafting of manuscript.

This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.

Signature: Date: 11/1/19

## Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i the candidate's stated contribution to the publication is accurate (as detailed above);

ii permission is granted for the candidate in include the publication in the thesis; and

iii the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

**Co-Author:** Aaron W. Thronton

**Contribution:**

Supervised the development of computational model, assisting in the interpretation of results and revision of the manuscript.

Signature: Date: 14 Jan 2019

**Co-Author:** Christian J. Doonan

**Contribution:**

Conceived project, supervised the interpretation of results and the conception and revision of manuscript. Acted as corresponding author.

Signature: Date: 21/1/2019

**Co-Author:** David M. Huang

**Contribution:**

Supervised the model development, simulations, interpretation of results and the conception and revision of manuscript. Acted as corresponding author.

Signature: Date: 21/01/2019

## 3.1 Abstract

The structural and dynamics factors governing porosity in porous aromatic frameworks (PAFs) are investigated using coarse-grained molecular dynamics simulations. PAFs form amorphous, porous networks with potential for gas storage and separation applications. We focus on a series of four PAFs – PAF-1, PPN-1, PPN-2 and PPN-3 – which exhibit an unexpected trend in porosity as the structure of the PAF monomer is varied. The simulations suggest that nonbonding dispersion interactions that stabilize misbound monomer configurations play an essential role in the formation of porosity-reducing interpenetrated frameworks in PAFs comprising the larger PPN-1 and PPN-2 monomers; on the other hand, the simulations indicate that the steric bulk of a key reaction intermediate acts to limit interpenetration in PAFs made up of the smaller PAF-1 and PPN-3 monomers. The simulations also show that the rate of cluster growth, which depends largely on the monomer concentration used in the experimental synthesis, is significantly higher for PPN-1 and PPN-2, which would exacerbate the kinetic trapping of interpenetrated misbound configurations. This work provides design rules for synthesizing highly porous amorphous networks through the choice of monomer structure and reaction conditions that limit framework interpenetration.

## 3.2 Introduction

Porous amorphous polymers, such as conjugated microporous polymers (CMPs),[18] hyper-cross-linked polymers (HCPs)[20] and porous aromatic frameworks (PAFs),[24,25] are a class of porous solids that are synthesized via irreversible chemical reactions. Their chemically robust structures and permanent porosity make them promising candidates for application to the fields of catalysis and gas storage and separation. In particular, porous aromatic frameworks (PAFs) have garnered significant interest[29–31,151–158] due to their high physiochemical stability and remarkably high porosity. For example, the archetypal PAF, PAF-1,[27] has one of the highest nitrogen BET surface areas reported to date ($5600\,m^2\,g^{-1}$). Other PAF materials have been synthesized, namely PPN-1, PPN-2, and PPN-3, which are based upon larger tetrahedral building units than PAF-1. However, in each case, their empirical surface areas are significantly

and unexpectedly lower than PAF-1.[28,159] The structural similarity of the monomers used to construct this family of materials is illustrated in Figure 3.1a. PPN-3 differs from PAF-1 by the presence of an adamantane moiety instead of a carbon atom at the monomer core, while PPN-1 and PPN-2 are analogues of PAF-1 and PPN-3, respectively, in which the phenyl groups of the monomer have been extended by the addition of an ethynyl group. Several factors could give rise to the lower porosity observed for PPN-1, PPN-2, and PPN-3 compared with PAF-1. The most likely of these are framework interpenetration[160,161] and pore blocking as a result of structural defects or reaction side-products.[162] Furthermore, previous computer simulations performed on related materials suggest that factors such as monomer flexibility have a more pronounced effect on surface area than increasing the size of the monomer.[163] Although these studies have provided some basic strategies for obtaining high surface area amorphous polymers,[161] the origin of the disparity in porosity between PAF-1 and its analogues PPN-1, PPN-2, and PPN-3 remains unclear. Developing a detailed physical understanding of these systems will facilitate the design of novel PAF materials with tailored porosity.

Structural characterization of amorphous solids is challenging as their lack of long-range order limits the information that can be obtained from diffraction techniques. Accordingly, computational methods have been widely employed to understand the relationship between local structure and global properties of a material.[26,163–166] A key challenge in this area is to develop methods that produce hypothetical structures that accurately represent the amorphous network. This requires a fundamental understanding of their synthesis chemistry and the noncovalent interactions that determine their 3D architectures. Molecular simulations, such as molecular dynamics (MD) or Monte Carlo (MC) approaches, can provide insight into these processes, but their application is computationally demanding due to the large length and time scales involved. To alleviate this issue coarse-grained (CG) models can be employed as they enhance the efficiency of MD and MC simulations. Indeed, CGMD/CGMC simulations have been used to study assembly processes in a range of model systems, such as DNA and proteins.[126,127,167–170] Furthermore, simulation algorithms developed to model the formation of amorphous microporous polymers have previously been able to replicate experimental data and offer insight into key factors governing a materials

microporosity.[26,165,166,171–174]

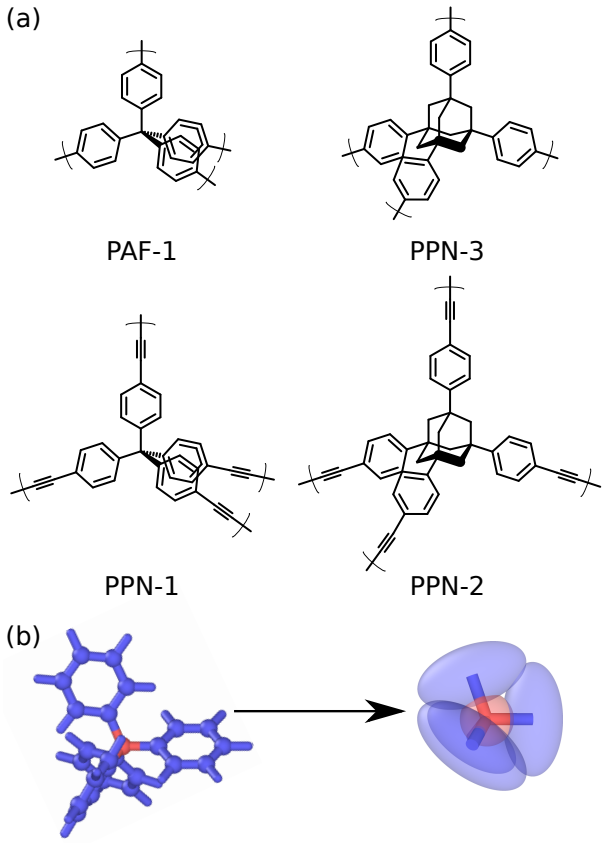



Figure 3.1: (a) Four PAF monomers studied in this work. (b) Schematic showing the coarse-graining of the PAF-1 monomer to the coarse-grained (CG) representation, which consists of ellipsoidal sites. The all-atom model is colored to match the segments replaced by the ellipsoid of the same color.

PAF-1 has been the focus of a series of modeling papers that employed molecular simulations to capture the growth of the polymer network.[165,166] These studies suggest potential mechanisms that prevent framework interpenetration and give rise to the high porosity of PAF-1: solvent templating, which is the result of framework–solvent or solvent–solvent interactions,[165] and a "snap-out" mechanism, whereby the rigidity of the PAF monomers and the large change in the relative orientation of monomer pairs during bond formation drive the formation of an open network.[166] However, in the aforementioned work, the influence of noncovalent $\pi \cdots \pi$ interactions on the 3D structure of the polymer were not investigated. We contend that such interactions are important as they have been shown to engender interpenetrated structures in other amorphous porous polymers.[164]

In this work, we sought to develop a model for the synthesis of PAFs that captures

the essential features of the assembly process—namely the monomer size and shape, monomer concentration, rate of diffusion of reacting species, and strength of intermolecular interactions between these species—to explain the origins of the large variations in porosity observed for different PAF materials. The four PAF structures we focused on were those based on "short-arm" monomers, PAF-1 and PPN-3,[175,176] and "long-arm" monomers, PPN-1 and PPN-2 (Figure 3.1a).[177] Although the short-arm and long-arm PAFs are formed via different synthetic procedures (see Appendix A.1 for details) they nevertheless share common features: both reactions are reported to go to completion with high yields and the proposed mechanisms proceed via a monomer–catalyst intermediate with a very fast subsequent carbon–carbon bond-forming reaction step. Both reactions are irreversible; thus, to a good approximation, the assembly process is kinetically rather than thermodynamically controlled. As a result, the kinetics of the assembly mechanism is expected to be a key factor in determining the hitherto unexplained differences in the properties of these PAFs. We employed coarse-grained molecular dynamics (CGMD) as it provides an efficient means of simulating the kinetics of assembly on the length and time scales necessary to produce frameworks that are representative of the bulk materials. This approach to generating hypothetical PAF structures addresses aspects of their synthesis that have not previously been considered and will thus provide new insights into their chemistry and a guide for the design of materials with tailored properties.

## 3.3 Methods

The coarse-graining process that was applied replaced groups of atoms by an ellipsoid centered at the center-of-mass of the atom group. The shape of each ellipsoid was specified by principal diameters (given in Table A.1) determined from the Connolly volume of the group of atoms that the ellipsoid replaces, obtained using the "Atom Volumes and Surfaces" Tool within Materials Studio,[178] where the Connolly volume is an approximation for the van der Waals volume.[179] The shape of the bond-forming ellipsoid was modified in the binding direction to yield the correct bond length upon binding; the other principal axes remained unchanged. The mapping from the atomistic to the CG representation is shown in Figure 3.1b for PAF-1 and in Figure A.2 for all the PAF

monomers studied. Each CG PAF monomer was treated as a rigid body, which means intramonomer structural fluctuations were quenched. This approximation is expected to be reasonably accurate, given that the only significant intramonomer motions are hindered rotations of the phenyl groups, which do not produce large conformational changes of the monomer. Interactions between ellipsoids were described by the Gay-Berne (GB) potential,[180,181] which accounts for short-range repulsions and long-range attractions between anisotropic particles (see Appendix A.3 for details). We used a modular approach to distinguish three general types of interactions: primary bonding (1°), secondary nonbonding (2°) and steric. The highly directional primary interactions represent the irreversible coupling reaction between monomers, which for simplicity was assumed to occur via an energetically barrierless process (Figure 3.2a). For all four PAFs the maximum well depth of the primary interactions was assumed to be $100k_{\mathrm{B}}T$ to give effectively irreversible bonding. The secondary interactions represent $\pi\cdots\pi$, alkyne$\cdots\pi$, or alkyne$\cdots$alkyne nonbonded interactions. The strength of these interactions was set to values between $\sim$3 kJ mol$^{-1}$ (for alkyne$\cdots$alkyne interactions) and $\sim$9 kJ mol$^{-1}$ (for $\pi\cdots\pi$ interactions), which is typical for such interactions[182,183] (see Appendix A.3 for details). To model the formation of the monomer–catalyst–monomer complex that forms prior to monomer–monomer coupling in the reaction mechanism for PAF-1 and PPN-3 (Figure 3.2b,c), a potential energy well at a larger distance than the primary binding interaction, which we will call the "steric potential", was added (Figure 3.2a). The energy of the steric potential well was chosen to match approximately the energy of a typical nickel–carbon bond (see Appendix A.3 for details).[175] The position of the energy minimum in the steric potential was chosen to match the approximate distance between a monomer core site and an incoming phenyl group in the presence of the catalyst complex (Figure 3.2a).

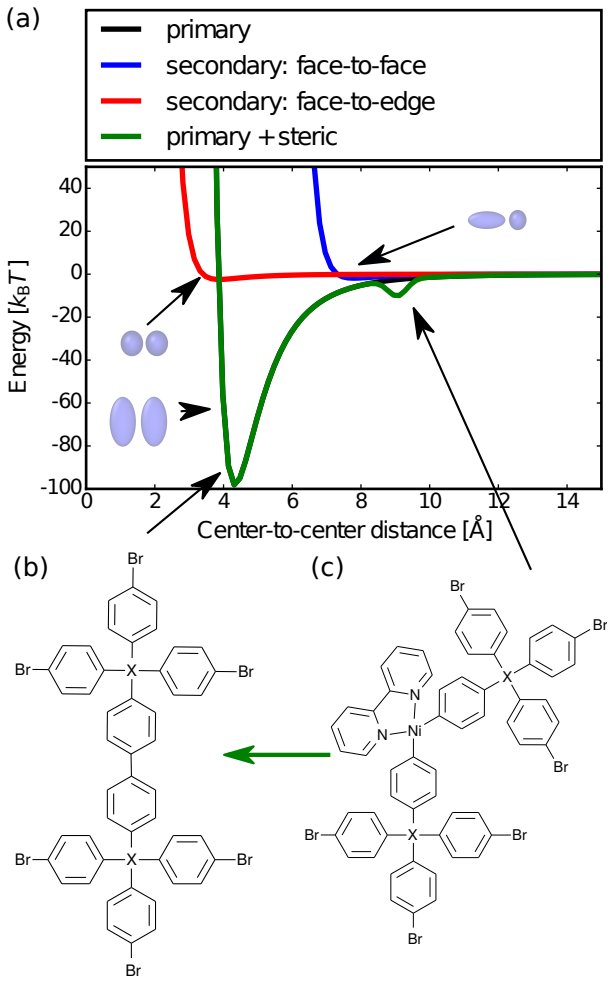



Figure 3.2: (a) Pair potential energy as a function of the center-to-center distance between two CG phenyl groups (shown in inset as blue ellipsoids) for the various interaction potentials in our model. (Note that the curve for the primary interaction overlaps that for the primary + steric interactions except at the small well around 9.3 Å, that these interaction potentials are angular dependent and that the potential curves in the figure are for the specific relative site orientations shown.) (b, c) Chemical structures of the (b) product and (c) proposed catalyst-bound intermediate of the Yamamoto reaction used to synthesize PAF-1 and PPN-3.

The motion of the PAF monomers in implicit solvent was simulated using GPU-accelerated Langevin dynamics with LAMMPS.[123,184,185] The temperature and friction coefficients used in the Langevin dynamics were chosen to yield diffusion coefficients that matched predicted values for Brownian motion of particles the size of the PAF monomers at the experimental solvent viscosity and temperature, as described in the Appendix A.3. Using PACKMOL,[186] an initial configuration of 8000 randomly positioned monomers was specified in a cubic simulation box at the experimental monomer concentration (given in Table A.3). To push apart overlapping atoms, the system

was evolved with soft nonbonded potentials using a small integration time step before carrying out simulations with interactions specified by the GB potential with an integration time step (given in Table A.4) defined as one tenth of the Langevin dynamics relaxation time, which is inversely proportional to the predicted diffusion coefficient. Constant-volume simulations at the temperatures used in the experimental synthesis of the PAFs (see Appendix A.3 for details) were run for at least one microsecond or until the growth of clusters had halted. Simulations were carried out with only primary interactions and with primary and secondary interactions for all four PAFs, as well as with the additional presence of the steric potential for PAF-1 and PPN-3. Simulations of all four PAFs were also carried out with the full set of interactions (primary + secondary (+ steric for PAF-1 and PPN-3)) with isotropic compression of the simulation box at a constant rate (see Appendix A.4 for details). Compression was stopped when the system reached a cut-off pressure (Figure A.5) indicating that monomers were starting to overlap.

All images of simulation configurations were rendered using Ovito.[187] In-house Python code was used to identify clusters of bonded monomers in the simulation trajectories, which were distinguished by distance and density criteria via the DBSCAN clustering algorithm.[188] The clusters were then reverse mapped to an atomistic representation. Details of the clustering and reverse-mapping procedures are given in the Appendix A.5. The RASPA2 software was used to calculate pore size distributions (PSDs) and X-ray diffraction (XRD) patterns of extracted clusters.[131] The Zeo++ software was used to calculate densities, $N_2$ accessible surface areas, and accessible pore volumes of extracted clusters.[134] For clusters extracted from constant-volume simulations, an algorithm was applied that minimizes the effect of the external surface of the clusters on measured properties by taking the highest density portions of the cluster (see Appendix A.7 for details). At the end of the compression simulations, the system consisted of a single cluster; due to the high computer memory requirements of the porosity analysis of a cluster of this size, smaller samples of the system-spanning cluster were analyzed instead, which was found to be representative of the full system (see Appendix A.8 for details).

## 3.4 Results and discussion

To provide molecular-level insight into how PAFs are formed, we developed a model designed to capture the essential features of their constituent monomers as they are assembled. These include the monomer geometry and concentration, the strength and anisotropy of intermonomer interactions, and the dynamics in solution of the monomers and their aggregates. Furthermore, all simulation parameters were selected to match the experimental conditions employed for each PAF. An understanding of how these parameters affect the assembly process is an important step towards establishing design principles for these materials. We note that our model is specific to these PAF systems, but the approach should be generally applicable to the assembly of any porous polymer synthesized via kinetically controlled reactions. A salient feature of our approach is that the CG model allows simulations to be carried out using experimental concentrations of monomer units at the length and time scales required to observe mesoscale assembly processes (microseconds or longer).[126] Such conditions were not employed for previous simulations of amorphous, porous polymers,[164–166] that used fully atomistic models due to their high computational cost.

### 3.4.1 Effect of monomer structure on PAF structure

Examining the early stages of PAF synthesis in constant-volume simulations provides insight into the role of which molecular interactions influence framework growth. We performed simulations with and without secondary interactions (*vide supra*) for all four PAF monomers. Figure 3.3 shows how different intermonomer interactions affect the structures of assembled clusters for PPN-2 (see also Figures A.6–A.9 in the Appendix for typical clusters formed by all four PAFs at early times on the order of 100 ns). The simulation trajectories for all PAFs show the formation of numerous small clusters on time scales of ca. 100–200ns (the free monomer concentration had fallen to 1% of its original value by this point). These clusters then agglomerate to form larger clusters on time scales on the order of a microsecond and longer. We stopped our simulations after at least one microsecond as cluster growth had halted due to the slow diffusion of larger clusters.

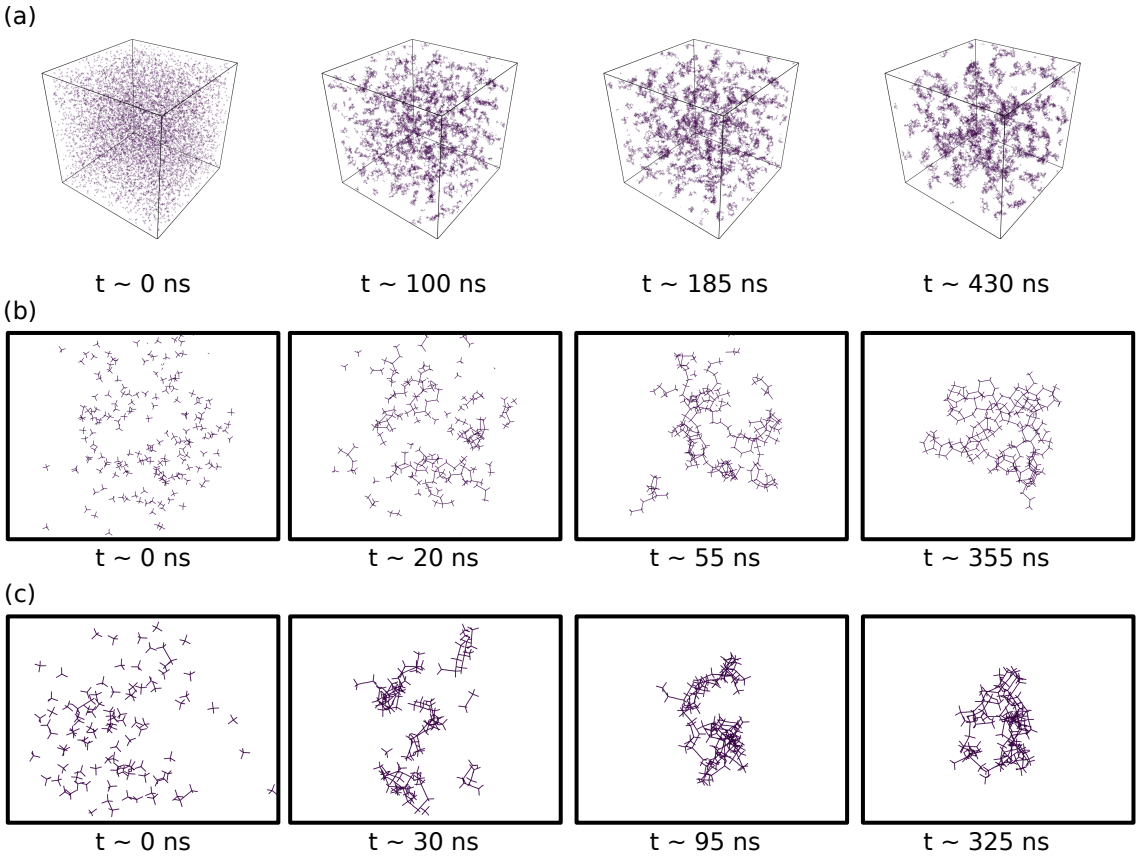



Figure 3.3: (a) Snapshots of a simulation trajectory of PPN-2 with primary and secondary interactions. (b, c) Snapshots of the growth of a typical cluster from a simulation trajectory of PPN-2 (b) with only primary interactions and (c) with primary and secondary interactions, illustrating the more compact and less porous structure when secondary interactions are present. Only bonds between sites (both intra- and intermonomer) and not the ellipsoidal sites themselves are shown for clarity; intermonomer bonds are defined to exist between pairs of binding sites separated by less than a threshold distance slightly larger than the position of the energy minimum for the primary binding interaction in Figure 3.2.

The radial distribution functions (RDFs) in Figure 3.4, obtained using the SMAC code,[189] of the CG core sites (core sites are the carbon or adamantane group in PAF-1 and PPN-1 or PPN-2 and PPN-3, respectively) at the end of our simulations highlight the structural differences due to the inclusion of secondary interactions and different monomer sizes. For each PAF there is a dominant peak at the distance corresponding to the irreversible binding of two monomers ($r_{\mathrm{bind}}$) and a second large peak at $\sqrt{\frac{8}{3}}r_{\mathrm{bind}}$, corresponding to two monomers indirectly bonded via a third intervening monomer. Figure 3.4 shows some less intense peaks at larger distances, but the lack of any other significant peaks at larger values indicates that only short-ranged order exists in these systems, which is consistent with experimental findings.[27,28,159] We have confirmed that the clusters extracted from our simulations are highly disordered by comparing their

XRD patterns to the XRD patterns of perfectly ordered diamondoid PAF networks (see Figures A.21 and A.22). When secondary interactions are present in the simulations, the second highest peak in the RDFs in Figure 3.4 loses intensity and the inset shows a smearing of peaks at larger distances, indicating a further decrease in order; peaks in the RDF at distances smaller than $r_{\text{bind}}$ also gain in prominence due to defect formation and framework interpenetration. Herein, we define interpenetration by any deviation from local tetrahedral ordering near a PAF monomer. Note that this definition differs from that of crystalline, frameworks with tetrahedral building blocks.[160]

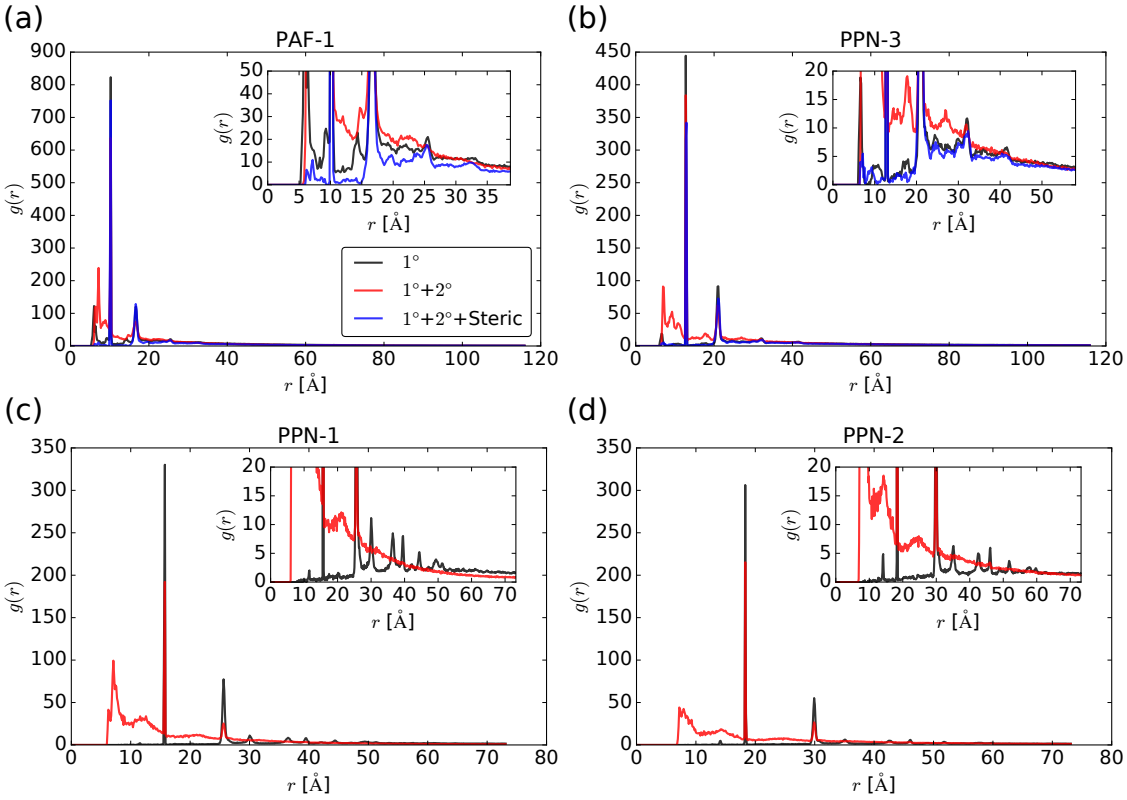



Figure 3.4: RDFs for (a) PAF-1, (b) PPN-3, (c) PPN-1, and (d) PPN-2 for all interaction types. The insets highlight peaks at small distances.

Defect formation is expected for reactions governed by kinetic control as monomers are unable to rearrange upon misbinding. We have defined a monomer pair as misbound if their core-to-core distance is smaller than $r_{\text{bind}}$. (Specifically, we took the threshold distance of misbinding to be $r_{\text{bind}} - 2\sigma_{\text{bind}}$, where $\sigma_{\text{bind}}$ is the width of this peak obtained from a Gaussian fit.) We note that this definition does not differentiate between persistently and transiently misbound monomers. Using this definition, the average number of monomers within $r_{\text{bind}} - 2\sigma_{\text{bind}}$ of any given monomer was measured

for each PAF as a function of time in the simulations and is shown in Figure 3.5. These data highlight the increased degree of interpenetration in PPN-1 and PPN-2 compared with PAF-1 and PPN-3 resulting from secondary interactions that effectively "switch on" monomer misbinding. This conclusion is in agreement with a previous model of covalent triazine frameworks (CTFs)[164] and previous experimental work on PAFs.[157,158]

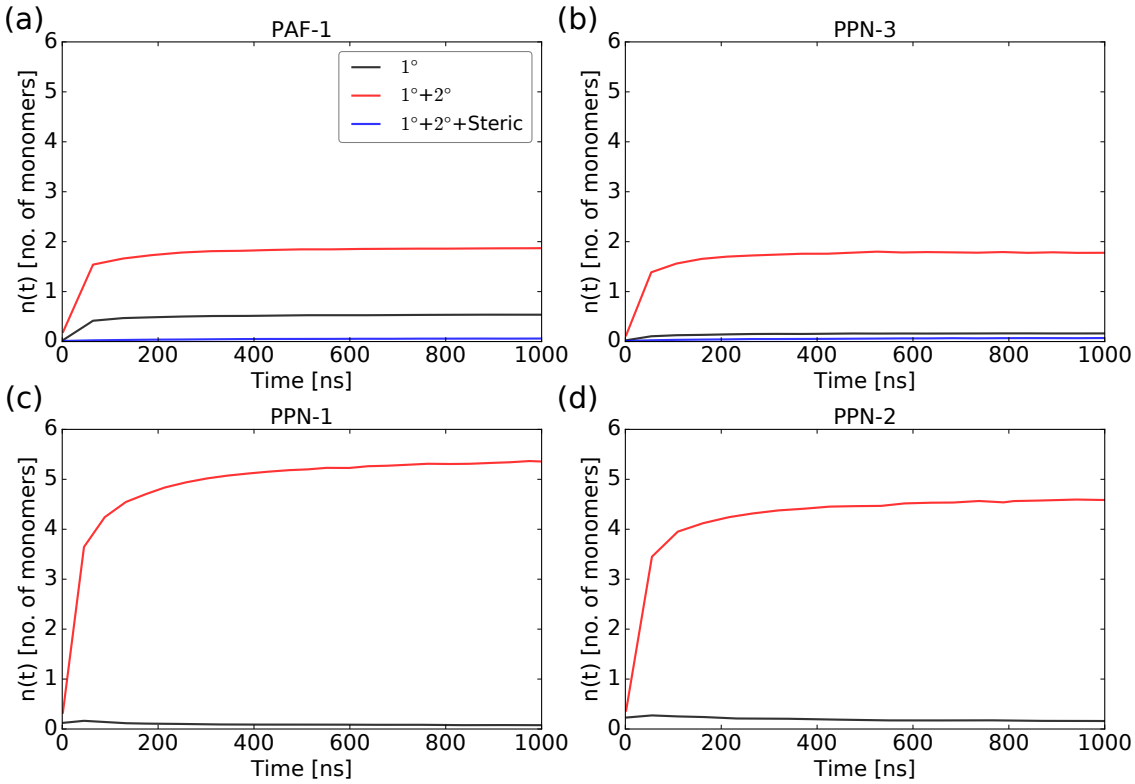



Figure 3.5: Average number of monomers that are misbound to a given monomer versus time for (a) PAF-1, (b) PPN-3, (c) PPN-1, and (d) PPN-2 for all interaction types.

To further investigate the cause of increased misbinding in PPN-1 and PPN-2, we have calculated the energetic contributions from primary and secondary interactions separately for all misbound monomer pairs at the end of simulations with both primary and secondary interactions for all four PAFs (see Appendix A.9 for details). The distribution of primary and secondary pairwise interaction energies (shown in Figure A.14) overlaps for the short-arm PAFs, PAF-1 and PPN-3, whereas the distribution of secondary pairwise interaction energies for the long-arm PAFs, PPN-1 and PPN-2, is skewed to significantly more negative energies than the distribution of primary pair interaction energies. This result suggests that stronger secondary interactions in the long-arm PAFs, PPN-1 and PPN-2, due to the additional alkyne functionality per monomer arm, contributes significantly to the higher degree of misbinding compared

with the short-arm PAFs, PAF-1 and PPN-3. Furthermore, it can be concluded that energetic stabilization of misbound monomer configurations by secondary nonbonding attractive interactions, such as $\pi \cdots \pi$, alkyne$\cdots$alkyne, and alkyne$\cdots\pi$ interactions appear to be crucial for driving interpenetration and reduced porosity (discussed below) in PAFs. These metastable configurations are sufficiently long-lived on the time scale of framework assembly that the framework can grow around the misbound monomers, preventing relaxation to the thermodynamically stable state, as discussed further below.

### 3.4.2 Role of bulky catalyst–monomer intermediate on PAF structure

Hitherto, we have considered primary and secondary interactions between PAF monomers. These data suggest that interpenetration is possible for all four PAFs, but more prevalent in the long-arm PAFs, PPN-1 and PPN-2. Experimentally determined surface areas and pore size distributions for the short-arm PAFs, PAF-1[27,29] and PPN-3,[28,159] however, are consistent with values expected for noninterpenetrated frameworks. In addition to the difference in the size of the monomers, PAF-1 and PPN-3 are formed via a Yamamoto mechanism[176] while PPN-1 and PPN-2 are synthesized via Eglinton coupling (see Appendix A.1 for the proposed reaction mechanisms).[177] In both reaction mechanisms, catalyst–monomer intermediates are formed prior to C–C bond formation. An important difference is that the intermediate formed during the Yamamoto mechanism is more sterically demanding than that in the Eglinton coupling (Figure A.1) and may impede framework interpenetration in PAF-1 and PPN-3. The synthesis of sterically bulky organic ligands is a common design strategy for minimizing the likelihood of interpenetration in metal–organic frameworks (MOFs).[190–192] Furthermore, the monomer–monomer coupling step in the proposed mechanism of the Eglinton reaction does not involve a catalyst–monomer intermediate and so the steric bulk of such intermediates is not expected to play an important role in determining the degree of framework interpenetration in PPN-1 and PPN-2.

To account for the effect of the monomer–catalyst–monomer intermediate in the Yamamoto coupling mechanism, a potential energy well (the steric potential) was added to the monomer–monomer interaction potential at a distance approximately corresponding to the monomer separation in the intermediate (Figure 3.2). The RDFs in

Figure 3.4 show that the steric potential essentially eliminates interpenetration and defect formation and thus facilitates the growth of the highly porous networks PAF-1 and PPN-3. The introduction of a steric potential affords a metastable state that acts to slow down binding and to reduce misbinding of monomers. This is represented in Figure A.15, which shows that the rate of free monomer loss is decreased when steric interactions are present. Previous modeling studies performed on PAF-1 suggested that the formation of the reaction intermediate eliminates interpenetration via a "snap-out" mechanism,[166] in which the two monomers change their orientation from noncollinear in the intermediate to collinear after bond formation. This requires that monomers binding to the edge of a cluster point away from the growing framework, leading to a relatively low-density material. Our results show that a highly directional "snap-out" mechanism is not required to explain the lack of interpenetration in PAF-1, as the steric potential we employed is isotropic. Rather, it appears that the formation of a sterically demanding intermediate prevents the monomers from approaching close enough to become kinetically trapped in a misbound configuration via secondary interactions. Solvent templating has also been advanced as a mechanism for minimizing interpenetration.[165] This current study indicates that this effect is also unnecessary for explaining the apparent lack of interpenetration in PAF-1 and PPN-3, as our model does not explicitly account for the structure of the solvent molecules.

### 3.4.3 Kinetics of PAF assembly

Figure 3.6 shows that the growth rate of the average cluster size with time depends on the specific monomer and interaction type. However, in all cases, the growth rate is approximately linear with time in the early stages (time scales of $< 200$ ns) of framework formation (see Figure A.16 for full data). The introduction of secondary nonbonding interactions in addition to the primary irreversible bonding interactions does not affect the cluster growth rates of PAF-1 and PPN-3, but these rates are approximately halved upon the introduction of the steric potential representing the bulky catalyst intermediate. In contrast the rate of cluster growth for the long-arm PAFs, PPN-1 and PPN-2, is significantly slower upon addition of secondary interactions. Furthermore, the rate of cluster growth for the long-arm PAFs is roughly double that of the short-arm PAFs in the absence of the steric potential. The faster rate of cluster growth for the long-

arm PAFs would facilitate trapping of misbound monomers in the growing framework, which, are more prevalent in these systems due to stronger secondary nonbonding interactions. In summary, the more pronounced framework interpenetration in PPN-1 and PPN-2 likely results from a combination of thermodynamic (energetic stabilization of misbound monomer pairs) and kinetic (rapid framework growth) effects.

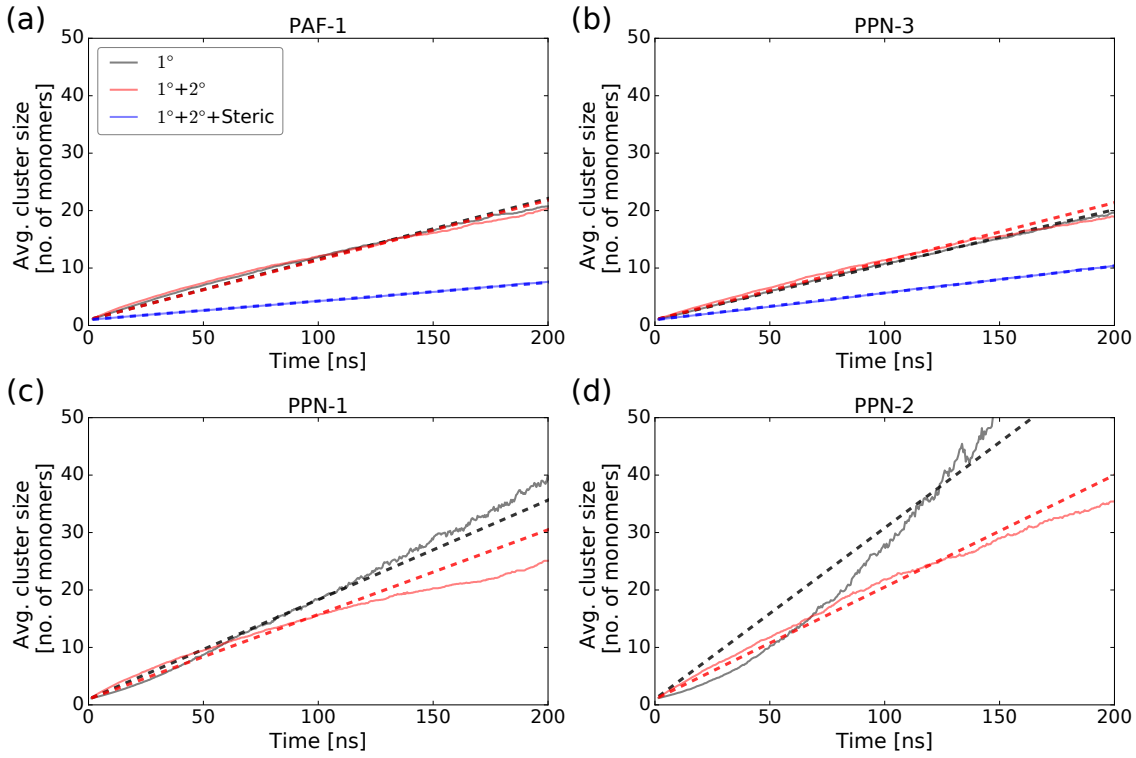



Figure 3.6: Average cluster size at early times for (a) PAF-1, (b) PPN-3, (c) PPN-1, and (d) PPN-2 for all interaction types (solid lines). The dashed lines are linear fits.

The approximately linear growth of the average cluster size with time and the different growth rates for the PAF networks can be rationalized using the Smoluchowski coagulation model,[193–195] a kinetic model that describes the size distribution of clusters formed by irreversible two-body collisions in a well-mixed system. Analytical solutions for the size distribution or its moments exist only in certain special cases of the model. In particular, when the bimolecular rate constant $K_{ij}$ for the coagulation of a cluster of $i$ units with a cluster of $j$ units is constant ($K_{ij} = K$, for all $i$, $j$), the average cluster size increases linearly with time, as observed in our simulations. Assuming that $K = \alpha k_s$ is made up of two components, where $\alpha$ is a size-independent sticking probability and $k_s$ is the rate constant for bimolecular collisions, this behavior implies that $k_s$ is independent of the size of the colliding clusters in our simulations. We have

confirmed this hypothesis by approximating $k_s$ as the rate constant of collision of two Brownian hard spheres of radius $R_i$ and $R_j$, respectively, and diffusion coefficient $D_i$ and $D_j$, respectively, which gives the analytical expression $k_s = 4\pi(R_i+R_j)(D_i+D_j)$.[195] By taking the radii of gyration and diffusion coefficients of clusters measured in our simulations as the radii and diffusion coefficients in this equation, we obtained a large spread of values of $k_s$, which, as expected, do not depend very strongly on the sizes of the colliding clusters in the linear growth regime (see Appendix A.10 for details). This behavior can be rationalized as being due to the compensatory effects of the cluster size and diffusion coefficient (which is inversely related to cluster size in the Stokes–Einstein equation, $D_i = k_B T \,/\, (6\pi\eta R_i)$) in the equation for $k_s$. From the bimolecular rate constant $K$, obtained by fitting the time dependence of the simulated average cluster size to the Smoluchowski coagulation model (slope = $KC_1 \,/\, 2$, where $C_1$ is the monomer concentration), and the calculated collision rate constants $k_s$, the sticking probability $\alpha$ was determined for each of the simulated systems. $K$, $k_s$ and $\alpha$ are given in Table 3.1.

Table 3.1: Bimolecular coagulation rate constant $K$, collision rate constant $k_s$, and sticking probability $\alpha$ obtained from simulations.

| PAF | interactions | $K$ [nm$^3$ ns$^{-1}$] | $k_s$ [nm$^3$ ns$^{-1}$] | $\alpha$ |
|---|---|---|---|---|
| PAF-1 | 1° | 26.85 | 69.0 | 0.39 |
| | 1° + 2° | 26.48 | 60.0 | 0.44 |
| | 1° + 2° + steric | 8.28 | 87.0 | 0.10 |
| PPN-3 | 1° | 24.40 | 80.0 | 0.31 |
| | 1° + 2° | 25.93 | 69.0 | 0.38 |
| | 1° + 2° + steric | 11.89 | 99.0 | 0.12 |
| PPN-1 | 1° | 26.10 | 57.0 | 0.46 |
| | 1° + 2° | 22.20 | 30.1 | 0.74 |
| PPN-2 | 1° | 44.90 | 89.3 | 0.50 |
| | 1° + 2° | 29.37 | 49.0 | 0.60 |

The coagulation rate constants for the same interaction type do not depend strongly on monomer type, further supporting the assumption of cluster-size-independent coagulation. This uniformity in coagulation rates across all monomer types is partly due to the specific conditions of temperature $T$ and solvent viscosity $\eta$ used in the experiments and simulations, which lead to similar size-dependent diffusion coefficients for all four PAF systems according to the Stokes–Einstein relation. Thus, the 2-fold higher rate

of cluster growth in the long-arm PAFs compared with the short-arm PAFs in Figure 3.6 is accounted for by the 2-fold higher monomer concentration used to synthesize the long-arm PAFs. Hence, monomer concentration is the key parameter controlling framework growth rates in these systems. Temperature or solvent viscosity can also act as control parameters, but would produce smaller changes in growth rates for the typical range of temperatures and viscosities that are feasible experimentally.

As expected, the sticking probability increases (slightly) with the addition of secondary nonbonding attractive interactions and decreases (dramatically) with the addition of the steric potential, which introduces a barrier to bond formation. Despite the higher sticking probability when secondary interactions are present, cluster growth is slower for the long-arm PAFs with these interactions because they lead to significantly smaller and denser clusters for a given number of monomers due to substantial monomer misbinding. This result is supported by a shift in the cluster radius of gyration distribution to smaller values as a function of number of monomers when secondary interactions are added; on the other hand, the radius of gyration distribution for the short-arm PAFs is less sensitive to interaction type due to the lower level of monomer misbinding (see Figure A.18). The significantly smaller clusters in the presence of secondary interactions for the long-arm PAFs results in a smaller cluster collision rate constant $k_s$ that reduces the cluster growth rate.

### 3.4.4 Cluster shape anisotropy and unbound terminal sites

The clusters formed at the early stages of framework growth in the simulations are typically anisotropic for all PAFs and all interaction types (see Figures A.6–A.9). Cluster shape can be quantified in terms of the relative shape anisotropy (see Appendix A.11 for details), which was found to have a broad range of values for all PAFs, independent of the interaction type. The results also show that a slower reaction rate due to the addition of the steric potential does not influence the anisotropy of the clusters formed. An implication of the formation of highly anisotropic clusters during framework growth and nonreversible bonding is that large voids (mesopores) are formed as a result of inefficient packing when these clusters join together at later stages. Indeed, mesoporosity has been measured experimentally in PAF-1 using positron annihilation lifetime spectroscopy, which showed mesopores as big as 5 nm in size.[29] These mesopores have

proven useful in applications involving impregnation of PAF-1 with polymers[30,31,151] and electrolytes.[29]

With regards to unbound terminal sites, in the experimental systems these represent an unreacted bromine in PAF-1 or PPN-3 or an unreacted alkyne in PPN-1 or PPN-2. The number of unbound terminal sites in the simulations versus time plateaus above at least 30% for all four PAFs for all types of interactions (Figure A.23). This contrasts with experimental evidence of negligible unbound terminal sites present in all four PAFs. Nevertheless, our results are consistent with previous atomistic simulations describing PAFs.[165,166] For example, Trewin and co-workers[166] reported large variations in the dependence of the number of unbound terminal sites on reaction conditions in atomistic molecular dynamics simulations with an artificial distance- and angle-based bonding step and only found small values (down to $\approx 5\%$) when very weak constraints on the directionality of bonding (100° from collinear) were used. Similarly, simulations of amorphous silicon (topologically similar to PAF monomers[26,165]) show that without a high-temperature annealing step the number of unbound terminal sites remains high.[196] Our model suggests that a high proportion of unbound terminal sites is intrinsic to the assembly mechanism under experimental conditions. At the experimental monomer concentration, the rate at which monomers diffuse and irreversibly bind together produces small, anisotropic clusters at the early stages of assembly that then agglomerate with irregular binding interfaces, which leaves many unbound terminal sites. Furthermore, slowing down the rate of irreversible bond formation by a factor of two, by adding the steric potential for PAF-1 and PPN-3, does not appear to reduce the fraction of unbound terminal sites. In fact, our simulations suggest it increases the number of unbound terminal sites, as shown in Figure A.23. Our results suggest that further experiments designed to assess the presence of unbound terminal sites are needed. Cooper and co-workers[161] have suggested that discrepancies in elemental analyses of PPN-1 (formed via palladium-catalyzed homocoupling in their case) is a result of a high degree of unbound terminal sites. Similarly, the presence of residual nickel in PAF-1 and PPN-3 and residual copper in PPN-1 and PPN-2 suggests the possible formation of unbound terminal sites.[27,28,159] Previous work suggests that unbound terminal sites may be difficult to discern using standard techniques such as solid-state NMR and FTIR.[161,162]

### 3.4.5 Framework porosity

The properties of the bulk PAF materials were estimated from measurements of the discrete clusters formed in the constant-volume simulations by using a method devised to minimize the effect of the external surface of the clusters (see Appendix A.7 for details). The calculated average density and nitrogen-accessible surface areas are reported in Table 3.2. The standard deviation of the calculated properties is large relative to the mean, implying a high degree of variability among the simulated clusters. Note that experimental porosity data of these types of materials are highly variable, which is evident when comparing reported PAF-1 BET surface areas in two different publications in which identical synthetic procedures were used (5600 versus 3639 $m^2\,g^{-1}$).[27,29] Table 3.2 also provides experimental BET surface areas for PPN-3[28,161] and PPN-1[161] produced via different synthetic methods. For PPN-3, Yamamoto coupling was employed in all cases, but the temperature and solvents were varied. In this case the material with the highest surface area (4221 $m^2\,g^{-1}$) was synthesized at room temperature in toluene and dimethylformamide.[28] The synthesis of PPN-1 followed a palladium-catalyzed procedure.[161]

Table 3.2: Mean nitrogen-accessible surface areas and densities obtained from constant-volume simulations (mean and standard deviations weighted by cluster mass) and for a perfectly ordered diamondoid (dia) network, where experimental surface areas are also given.

| | | mean density [$g\,cm^{-3}$] | | mean surface area [$m^2\,g^{-1}$] | | |
|---|---|---|---|---|---|---|
| PAF | interactions | sim. | sim. dia. | sim. | sim. dia. | expt. |
| PAF-1 | 1° | 0.20 ± 0.15 | 0.32 | 7620 ± 1240 | 5930 | 5600,[27] 3639[29] |
| | 1° + 2° | 0.35 ± 0.10 | | 5150 ± 1020 | | |
| | 1° + 2° + steric | 0.29 ± 0.00 | | 7680 ± 300 | | |
| PPN-3 | 1° | 0.18 ± 0.11 | 0.20 | 8430 ± 1600 | 6623 | 2840,[159] 3180,[161] 4221[28] |
| | 1° + 2° | 0.28 ± 0.09 | | 5990 ± 940 | | |
| | 1° + 2° + steric | 0.16 ± 0.04 | | 8960 ± 1220 | | |
| PPN-1 | 1° | 0.08 ± 0.04 | 0.07 | 10100 ± 680 | 10560 | 1249,[159] 1470[161] |
| | 1° + 2° | 0.13 ± 0.08 | | 6370 ± 670 | | |
| PPN-2 | 1° | 0.04 ± 0.04 | 0.09 | 10100 ± 600 | 8340 | 1764[159] |
| | 1° + 2° | 0.10 ± 0.06 | | 7070 ± 420 | | |

The results in Table 3.2 highlight the drastic effect of secondary nonbonding interactions on the gravimetric surface area of PPN-1 and PPN-2. The greater decrease in porosity for PPN-1 and PPN-2 compared with PAF-1 and PPN-3 when secondary interactions are introduced indicates a higher propensity for interpenetration in PPN-1

and PPN-2 due to monomer misbinding, as discussed earlier. For all PAFs except PPN-1, the gravimetric surface area obtained from our model is greater than the surface areas obtained from the ideal diamondoid framework. This is largely the result of the lower densities in clusters produced by our model. Experimental validation of the densities, using helium or mercury picometry for example, would offer much needed insight. Volumetric surface areas and pore volumes are also reported in Table A.8. In our models, interpenetration does not completely block the pores, the result of which is the formation of many small pores and a higher volumetric surface area, as shown in Table A.8, when secondary interactions are present compared with when only primary interactions are present. Our results show that the volumetric surface area increases as the density of the framework increases and pore volume decreases as the density of the framework increases (Table 3.2 and Table A.8).

Table 3.2 shows that our model can qualitatively rationalize the relative porosities of PAF-1, PPN-1, PPN-2, and PPN-3. As expected, interpenetration leads to a significant loss in porosity for all four PAFs ($1^\circ + 2^\circ$). Furthermore, the presence of sterically demanding intermediates in the synthesis of PAF-1 and PPN-3 ($1^\circ + 2^\circ$ + steric) lead to the retention of high degrees of porosity. However, the order of the surface areas of PAF-1 and PPN-3 disagree with experimental findings. We note that the simulated porosity in all cases is higher than that measured experimentally, and this behavior is replicated in the pore volumes (see Appendix A.12). The porosity analysis minimized the effect of the external surface of the clusters and is consistent with results obtained from the compression simulations (discussed below) in which there was no external surface in the system. Therefore, we suggest that the discrepancies between our results and experiments are due to an unexplored aspect of the mechanism of PAF formation that is not captured by our simulation model, which accurately accounts for the key properties of monomer shape, interaction strength, concentration, diffusion coefficients, and irreversible binding. Considering the frameworks produced by our model are consistently highly porous, despite the substantial variations in system parameters between the different PAFs studied, we suggest the possibility of side reactions that produce defect sites and pore blockages in PAF synthesis as a key factor in determining framework porosity. It has been suggested previously for other amorphous porous polymers that the kinetically controlled reactions that are employed likely pro-

duce defect-ridden frameworks that would be difficult to distinguish from the desired products by standard techniques.[161,162] Previous syntheses of each PAF have reported elemental analyses indicating the presence of residual reactants and impurities within the pores, which suggests the blocking of pores could play a role in experimental porosity loss that our model does not capture.[28,161] With respect to PPN-1 and PPN-2, the color change to brown reported in the experimental procedure suggests the presence of defects, as reported for other alkyne-bridged polymers.[197] The presence of copper in the Eglinton reaction (mostly as copper(II), although short-lived copper(I) species may be produced) suggests the possibility for nanoparticle formation, which would not only act as pore-blocking defect sites but also show reactivity towards alkyne groups (such as oxidation of the alkyne groups).[198,199] Framework collapse upon desolvation is also likely to occur based on the high degree of dangling bonds (discussed above), as computer simulations have shown that amorphous cross-linked polymers formed under conditions that lead to decreased cross-linking have a high propensity for framework collapse and our model does not account for such a process.[173] Finally, within our models, all coupling reactions were treated equally but the reaction efficiencies and mechanisms for the different PAFs likely differ in experiments, which may impact the quantitative results.

### 3.4.6 Compression simulations

The analysis up to this point has been of the early stages of PAF assembly and of properties of relatively small discrete clusters from constant-volume simulations. The extrapolation to the properties of the bulk PAF material from those of the discrete clusters was inexact, so we have also run simulations of each PAF in which the system was isotropically compressed at a constant rate to approximate the later stages of PAF assembly, during which the small clusters agglomerate to form the bulk PAF (Figure 3.7a). Porosity analysis of compressed systems obtained using two different compression rates that differed by a factor of 2 yielded similar results (see Appendix A.13), suggesting that the simulated porosity is representative of PAFs formed under conditions of much slower agglomeration that may occur experimentally. Furthermore, the results of analysis of the discrete clusters from the constant-volume simulations is comparable to that for the compressed systems, in which a single cluster spanned

the entire periodically replicated system with no external surface, indicating that the cluster analysis from the early stages of assembly provides a good representation of the bulk material. Although the systems were compressed, there remains significant void space, as highlighted by the holes running through slices of the final compressed systems shown in Figure 3.7b–e. This result further supports the suggestion that the formation of large void spaces is inherent to the irreversible bond formation in PAF materials and results from anisotropic clusters that are formed early in the assembly process agglomerating in a disordered fashion under kinetic control. Furthermore, Cooper and co-workers have reported syntheses of PAF-1, PPN-1 and PPN-3 in which only 30–60% of the experimental pore volume was due to micropores.[161] Pore size distributions (PSDs) were calculated for the compressed systems and are shown in Figure 3.7f–i. Good agreement with experiment is obtained in all cases for the position of the peak in the pore size distribution when the full set of interactions (primary + secondary (+ steric for PAF-1 and PPN-3)) is used for each PAF. Additionally, a significant shift in the distribution to smaller pore sizes is seen for PPN-1 and PPN-2 when secondary interactions leading to interpenetration are switched on. Note that the approximate size of the catalyst complex in the Yamamoto mechanism is described by an energy well (Figure 3.2a) that is only slightly smaller than the position of the peak in the simulated pore size distributions for PAF-1 and PPN-3 in Figure 3.7, suggesting why the formation of this intermediate effectively suppresses interpenetration in these PAFs.

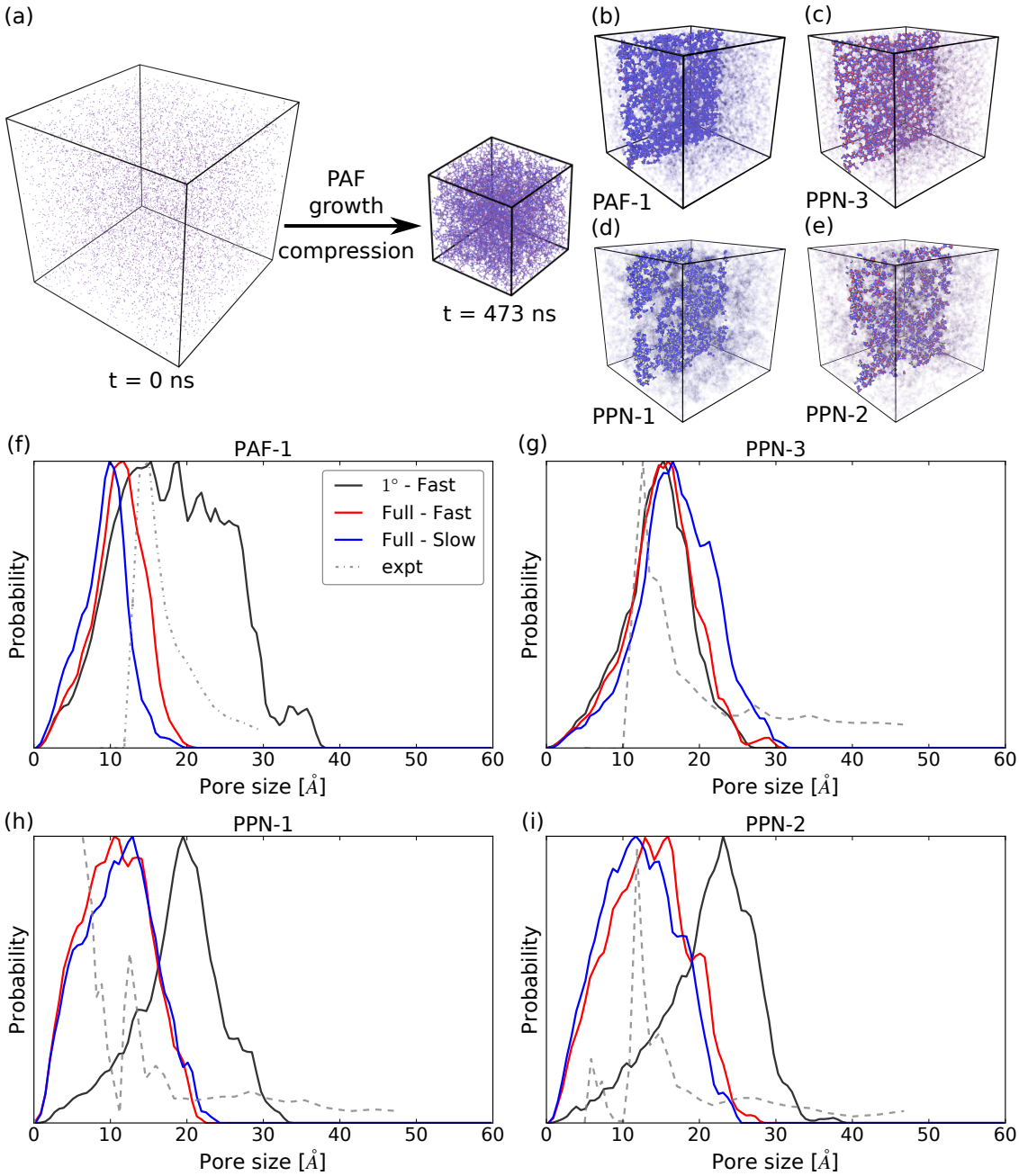



Figure 3.7: (a) Compression of PAF-1 at a constant rate toward the cutoff pressure at the "fast" rate. (b-e) Slices through the final compressed system for (b) PAF-1, (c) PPN-3, (d) PPN-1, and (e) PPN-2 in the presence of the full set of interactions. Simulated and experimental pore size distributions for (f) PAF-1, (g) PPN-3, (h) PPN-1, and (i) PPN-2 compressed at two rates with the full set of interactions and in the presence of primary interactions only.

## 3.5 Conclusions

We have developed a mesoscale CG molecular dynamics simulation model to account for the essential features of porous aromatic framework (PAF) assembly: monomer geometry, monomer concentration, energetics and anisotropy of intermonomer interactions, and dynamics of monomers and their aggregates. With experimental characterization of PAF materials limited due to their amorphous nature, the results of our model help to clarify the structural properties and assembly mechanism of a series of four PAFs; PAF-1, PPN-1, PPN-2, and PPN-3. The simulations suggest that relatively weak dispersion forces due to $\pi \cdots \pi$, alkyne $\cdots \pi$, or alkyne $\cdots$ alkyne interactions and bulky reaction intermediates are crucial in determining the structure and porosity of PAF materials. Our findings suggest that the propensity for framework interpenetration of a PAF material is substantially determined by the presence of these nonbonding dispersion interactions, which stabilize misbound monomer configurations, and the rate of cluster growth, which depends largely on the monomer concentration used in the experimental synthesis of the PAFs studied. We found that interpenetration only becomes significant in the presence of $\pi \cdots \pi$, alkyne $\cdots \pi$, or alkyne$\cdots$alkyne interactions, which implies that minimization of these interactions (for example, via the replacement of alkyne and aryl groups with rigid alkyl groups such as bicyclo[1.1.1]hexane) would reduce interpenetration. We note that quantum chemistry calculations[200,201] have shown that adding substituents to benzene always increases the strength of face-to-face binding in aromatic dimers, regardless of the electron-donating or electron-withdrawing nature of the substituent, and can increase the strength of face-to-edge binding, suggesting that adding substituents to the aromatic rings in PAFs is unlikely to be a fruitful strategy for reducing framework interpenetration. We introduced an interaction potential into the model that mimics the formation of bulky metastable reaction intermediates in the Yamamoto mechanism for synthesizing PAF-1 and PPN-3 and showed that interpenetration is effectively eliminated in PAFs formed by this mechanism, resulting in high porosity. Thus, one strategy to reduce interpenetration and improve porosity in PAFs is to employ a reaction that introduces steric hindrance using a bulky intermediate, for example by increasing the steric bulk of the ligands associated with the catalyst. In particular, our results suggest that using a catalyst complex similar in size to the

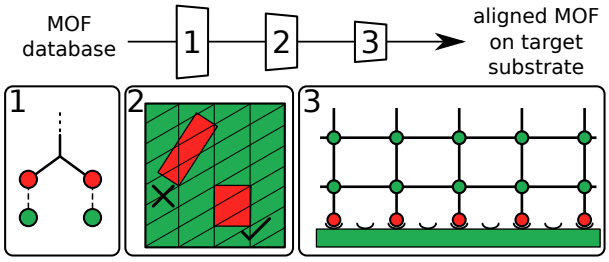
pores in the noninterpenetrated framework can effectively suppress interpenetration. However, we also note that too large a catalyst could impede framework formation. We showed that irreversible binding of monomers generally leads to highly disordered, poorly connected porous networks and a significant proportion of unbound terminal sites. Slowing down the rate of bond formation by adding a barrier to irreversible binding does not dramatically change this result. In fact, our results suggest that for typical monomer shapes, concentrations, and diffusion coefficients in experiments, frameworks produced by irreversible binding of monomers form highly porous structures very robustly. Counterintuitively, this behavior is a result of, rather than in spite of, their kinetically controlled formation mechanism and inherent disorder. Although interpenetrated structures were found for PPN-1 and PPN-2, interpenetration under the conditions simulated did not reduce porosity as much as expected based on experimental measurements. This result suggests other causes of low porosity (such as side reactions and reaction byproducts) that are not captured by the simulation model and remain poorly understood. However, if these factors could be eliminated, our findings indicate that very porous structures would be obtained. These findings should be generally applicable to other porous polymers formed through irreversible reactions.

## Acknowledgements

A.T. thanks the Australian Government for a Research Training Program Stipend Scholarship, the CSIRO Materials Science and Engineering for a Ph.D. top-up scholarship, and the CSIRO HPSC for computational resources. This research was undertaken with the assistance of resources from the National Computational Infrastructure (NCI), which is supported by the Australian Government. We would also like to thank Jack D. Evans for helpful discussions.

# CHAPTER 4

# High-throughput screening of metal–organic frameworks for macroscale heteroepitaxial alignment




*Published*

This work has appeared in the following publication:

Tarzia, A.; Takahashi, M; Falcaro, P; Thornton, A. W.; Doonan, C. J; Huang, D. M. High-Throughput Screening of Metal–Organic Frameworks for Macroscale Heteroepitaxial Alignment. *ACS Appl. Mater. Interfaces* **2018**, *10*, 40938–40950. DOI: 10.1021/acsami.8b14756

**Principal author (candidate):** Andrew Tarzia

**Contribution:** 85%

Design and implementation of the screening process. Analysis and interpretation of the data. Construction of figures. Conception, writing and drafting of manuscript.

This paper reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this paper.

Signature: Date: 11/1/19

## Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i the candidate's stated contribution to the publication is accurate (as detailed above);

ii permission is granted for the candidate in include the publication in the thesis; and

iii the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

**Co-Author:** Masahide Takahashi

**Contribution:**

Assisted the design of the high-throughput screening process, the interpretation of results and revision of the manuscript.

Signature: Date: 7 January 2019

**Co-Author:** Paolo Falcaro

**Contribution:**

Assisted the design of the high-throughput screening process, the interpretation of results and revision of the manuscript. Assisted production of figures and made supplementary video for submission.

Signature: Date: 21/01/2019

**Co-Author:** Aaron W. Thronton

**Contribution:**

Assisted the design of the high-throughput screening process, the interpretation of results and revision of the manuscript.

Signature: Date: 14 Jan 2019

**Co-Author:** Christian J. Doonan

**Contribution:**

Conceived project, assisted in interpretation of results and revision of the manuscript. Acted as corresponding author.

Signature: Date: 21/1/2019

**Co-Author:** David M. Huang

**Contribution:**

Supervised the development of screening process, interpretation of results and the conception and revision of manuscript. Acted as corresponding author.

Signature: Date: 21/01/2019

## 4.1 Abstract

The ability to align porous metal–organic frameworks (MOFs) on substrate surfaces on a macroscopic scale is a vital step toward integrating MOFs into functional devices. But macroscale surface alignment of MOF crystals has only been demonstrated in a few cases. To accelerate the materials discovery process, we have developed a high-throughput computational screening algorithm to identify MOFs that are likely to undergo macroscale aligned heteroepitaxial growth on a substrate. Screening of thousands of MOF structures by this process can be achieved in a few days on a desktop workstation. The algorithm filters MOFs based on surface chemical compatibility, lattice matching with the substrate, and interfacial bonding. Our method uses a simple new computationally efficient measure of the interfacial energy that considers both bond and defect formation at the interface. Furthermore, we show that this novel descriptor is a better predictor of aligned heteroepitaxial growth than other established interface descriptors, by testing our screening algorithm on a sample set of copper MOFs that have been grown heteroepitaxially on a copper hydroxide surface. Application of the screening process to several MOF databases reveals that the top candidates for aligned growth on copper hydroxide comprise mostly MOFs with rectangular lattice symmetry in the plane of the substrate. This result indicates a substrate-directing effect that could be exploited in targeted synthetic strategies. We also identify that MOFs likely to form aligned heterostructures have broad distributions of in-plane pore sizes and anisotropies. Accordingly, this suggests that aligned MOF thin films with a wide range of properties may be experimentally accessible.

## 4.2 Introduction

Metal–organic frameworks (MOFs) are a class of materials formed by connecting metal nodes and organic links into an extended network. In the past two decades, a focus of MOF research has been to develop synthetic strategies for the precise control of pore structure and functionality on the single-crystal scale.[33] However, more recently, extending the design principles of MOF chemistry from single-crystal to macroscopic dimensions (nm to cm) has garnered significant interest.[46–48] Approaches to

MOF thin-film fabrication include layer-by-layer synthesis, liquid-phase epitaxy, and chemical-vapor or atomic-layer deposition.[53,54] However, these techniques have yielded partially oriented thin films with out-of-plane alignment only.[46,57,202,203] To fully realize the potential of crystalline MOF thin films in applications such as optoelectronics, gas separation, and sensing, alignment in both in-plane and out-of-plane directions is required.[47,54] Recently, Falcaro and co-workers have developed a one-pot methodology for the fabrication of centimeter-scale MOF films that achieved both in-plane and out-of-plane alignment.[55] In this case, prealigned crystalline copper(II) hydroxide ($Cu(OH)_2$) nanobelts were used as a sacrificial substrate for the epitaxial growth[204] of copper-based MOFs and precise pore alignment was confirmed by comprehensive diffraction experiments. It was found that a necessary condition for epitaxial growth was a close matching between the MOF and $Cu(OH)_2$ lattice parameters. Such macroscale alignment was demonstrated only for four MOFs, which represents a very small subset of known MOF structure types and functionalities. To fully exploit this fabrication strategy, an efficient procedure for selecting MOFs that possess lattice parameters that match those of the $Cu(OH)_2$ substrate from the vast database of known structures is essential.

Materials design through high-throughput computational screening offers a low-cost strategy to efficiently guide experimental endeavors. In recent years, advancements in computer hardware and the development of open databases of experimental and theoretical material properties, which are continually growing, have led to rapid growth in this research area.[17,84,87] Indeed, high-throughput screening of candidate materials for aligned heteroepitaxial growth of MOFs would avoid an arduous and inefficient experimental approach. To this end, we have developed a high-throughput screening process to select and rank MOFs based on their likelihood to form aligned crystallites on a $Cu(OH)_2$ substrate. Although an atomic-level description of the growth mechanism and interfacial chemistry is necessary to fully characterize the kinetics and thermodynamics of these MOF films, such considerations are not amenable to rapid screening.[48,116,205–208] Instead, we have extended established methods for evaluating interfacial lattice matching[115,116,209–212] and chemical compatibility[213] to assess MOF–substrate interactions in a manner that accounts for framework porosity and allows for high-throughput screening. Geometrical lattice-matching algorithms[209,210] have been

employed in many screening studies as they offer an efficient way to exclude incompatible interfacial structures based on lattice mismatches.[115,116,211–213] However, chemical compatibility at the interface is expected to be a more discriminating measure of interfacial stability than lattice matching.[212,213] For example, Walsh and co-workers reported an efficient screening process for matching contact layers in hybrid perovskites that considered the chemical compatibility at the interface by calculating the overlap of atoms on either side of the interface.[213] Although this approach was demonstrated to be effective for nonporous solids, in this present work, we show that the atomic site overlap is not the most suitable measure of interfacial binding for highly porous materials such as MOFs. We identify a simple and easily computed proxy for the interfacial energy that considers the number of bonds formed and broken as a result of interface formation, which accurately predicts macroscale alignment of MOFs on a $Cu(OH)_2$ substrate. Application of our screening process to the CORE MOF,[94] hMOF,[101] and TOBACCO[99,214] databases shows that specific MOF topologies are favored for aligned growth from $Cu(OH)_2$, but that a wide range of pore network properties can nevertheless be obtained. Such knowledge will underpin design principles that will lead to a significant decrease in the experimental workload to develop precisely aligned MOF films.

## 4.3 Computational approach

We have applied a three-step screening process: (1) selection of candidate MOFs based on chemistry (in this study, MOF structures were searched for that possess copper-bound carboxylate functionalities); (2) refinement of the data set based on lattice matching; and (3) ranking of MOFs by interfacial bonding (see Figure 4.1). While step 3 is the most computationally intensive, high-throughput screening can be achieved with this process: e.g., ∼5000 crystal structures were screened in ∼25 h on six Intel i7-4790K CPU cores. In general, crystallographic information files (CIFs) are used as the input for the MOF layer (film) and a single crystal structure (also from a CIF) is used as input for the substrate layer ($Cu(OH)_2$ in the following work). A definition of the binding geometry across the interface between the two materials is also required. The coordination geometry within the MOF is defined by the input crystal structure,

and thus the screening process automatically accounts for multiple coordination geometries in the MOF that could arise, e.g., from multiple oxidation states of the metal nodes. Although we consider only a single interfacial binding geometry in this work, the algorithm can readily be extended to consider different types of interfacial binding atoms and/or binding geometries. This work focuses on the epitaxial growth of MOFs from a metal-hydroxide surface (specifically carboxylate-bound copper MOFs on $Cu(OH)_2$); however, the screening algorithm is not fundamentally limited to these specific systems.

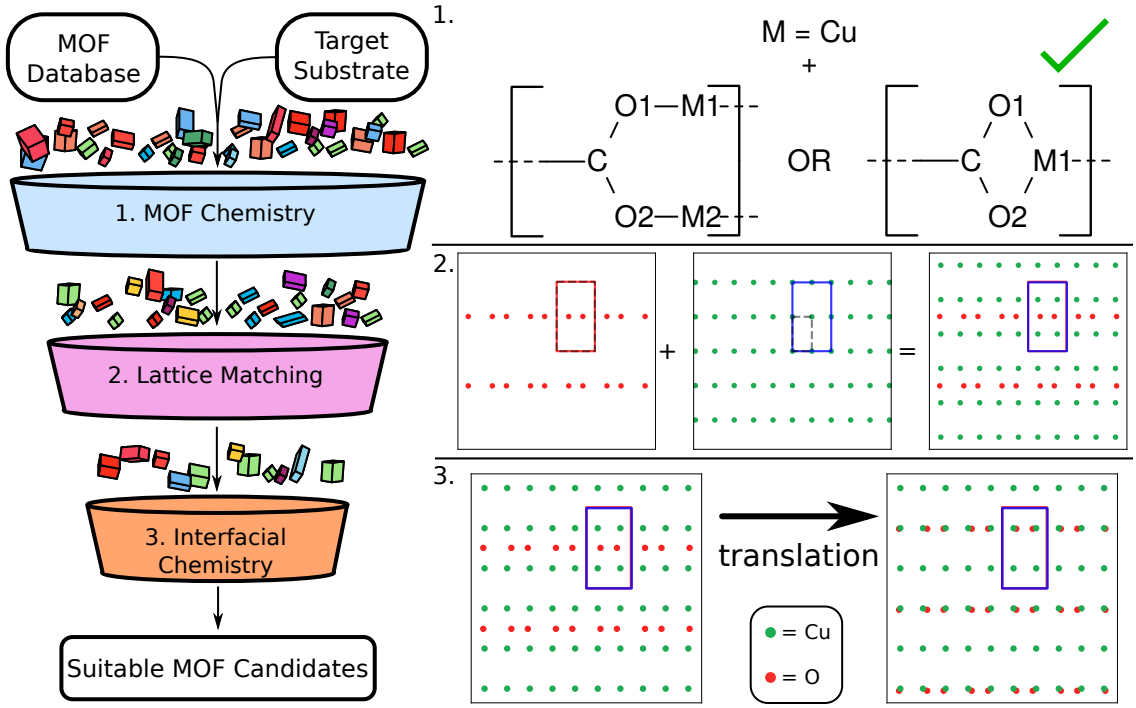



Figure 4.1: Flowchart of the screening process, in which a database of MOF crystal structures and a target substrate crystal structure act as inputs. The MOF database is filtered toward a smaller list of candidate structures for aligned heteroepitaxial growth. Each step is illustrated schematically on the right. Step 1 filters the database for MOFs that contain a carboxylate functionality bound to copper atoms and do not contain any metal other than copper. Step 2 checks if matching supercells (e.g., those outlined in solid lines, with dashed lines showing the corresponding unit cells; in this example, the red supercell is the same size and shape as the unit cell) exist for a given MOF–substrate pair. Step 3 checks all possible supercells for interfacial bonding, allowing for translations along the supercell vectors.

### 4.3.1 Identifying structures with appropriate chemistry

Step 1 of the screening process is designed to filter out all MOFs that do not possess the appropriate chemistry to bind to the substrate. This step can be generalized

to any binding geometry. The algorithm uses the Atomic Simulation Environment (ASE)[215] and pymatgen[216] Python libraries (description provided in Appendix B.1). In this study, the search was limited to MOFs containing copper nodes and carboxylate functional groups with each oxygen bound to a copper atom (see Figure 4.1). This constraint was guided by experiments by Falcaro and co-workers, in which $Cu(OH)_2$ acted as a sacrificial substrate and metal source for the MOF. However, there is no experimental evidence that heteroepitaxial growth of MOFs is limited to sacrificial substrates or to MOFs of the same metal as the substrate. Walsh and co-workers[115] suggested that the chemical similarity of both sides of the interface (defined in their work by the difference in Pearson hardness between the metals in the two environments) is a driving force for stability at the interface. According to this criterion, copper-based MOFs on a copper-based substrate are expected to produce stable interfaces.

### 4.3.2 Identifying structures with matching lattices

Step 2 of the screening process determines whether at least one Miller plane of a given MOF matches the lattice dimensions of the target substrate. In this work, the (010) Miller plane of $Cu(OH)_2$[217] was selected as it is known from experiments that this plane is exposed at the surface of this material.[55] Zur and McGill[209,210] developed an efficient geometrical lattice-matching algorithm that has been widely adopted and shown to predict heteroepitaxial interfaces consistent with experimental data.[116,211–213] The lattice-matching algorithm requires that the translational symmetry of each lattice matches at the interface within a specified tolerance. Although lattice matching is not a sufficient condition for growth, poor lattice matching can impact interfacial growth and can be used to rule out candidate structures for aligned MOF films. For example, large lattice mismatches have been found to anisotropically limit the size of hybrid perovskite crystals/grains in the direction of high mismatch.[213,218] Given a set of primitive vectors for a specified lattice on one side of the interface (the unit cell), the algorithm (implemented using the pymatgen Python library[216,219]) produces a series of supercells made up of an integer multiple of contiguous unit cells. A supercell is defined by two lattice vectors, which are reduced such that the supercell is unique.[209,210] A pair of supercells on either side of the interface, with match areas $m$ times the MOF unit cell area and $n$ times the substrate unit cell area, are deemed to match based on

four criteria: (1) the supercell match areas are below a defined threshold (set to the smaller of 90× the substrate unit cell area or 9× the film unit cell area), (2) the area ratio deviation, $\left|\frac{A_{\mathrm{MOF}}}{A_{\mathrm{sub}}} - \frac{n}{m}\right|$, where $A_{\mathrm{MOF}}$ and $A_{\mathrm{sub}}$ are the MOF and substrate unit cell areas, respectively, is below a defined threshold (set to 15%), (3) the corresponding lattice vectors are equal in length within a defined tolerance (set to 10%), and (4) the angles between the lattice vectors on either side of the interface are equal within a defined tolerance (set to 2%). (The choice of the parameters used in the matching criteria are discussed further in Section 4.4.1 and Appendix B.2). In the screening process, a MOF structure is eliminated if no Miller plane with indices $-1 \leq h, k, l \leq 1$ is found to have a matching lattice with the substrate. If at least one pair of supercells for a MOF–substrate pair passes all four tests, then that MOF continues to step 3 of the screening process.

### 4.3.3 Identifying interfaces with favorable binding

Even if MOF and substrate lattices are found to match, chemical compatibility at the interface is expected to be the major factor in determining interfacial stability.[209,212,213] Hence, the purpose of the final step of the screening algorithm is to identify interfaces with favorable binding by employing a geometric definition of interfacial binding. To do this, all possible binding planes of each MOF Miller plane that pass step 2 of the screening process are determined. We define MOF binding planes as those that include oxygens from carboxylate groups and $Cu(OH)_2$ binding planes as those that have a layer of copper atoms in the (010) Miller plane. When building the MOF binding plane, a 3D slab of the crystal structure with one face corresponding to the interface Miller plane[216,219] is selected, in which the axis bisecting the O–C–O angle of each carboxylate functionality at the surface is approximately orthogonal to the interface (see Appendix B.3). We use this definition because density functional theory calculations show that the bridging binding mode (one oxygen bound to one metal) of deprotonated carboxylates with this axis orthogonal to a metal-oxide surface (e.g., $TiO_2$) is most stable for the benzene dicarboxylate (BDC) ligand.[116,220] To test the chemical compatibility between a selected Miller plane of a MOF and the (010) Miller plane of $Cu(OH)_2$, the degree of binding at the interface of the MOF and $Cu(OH)_2$ binding plane unit cells is calculated once the unit cells are mapped onto each of their correspond-

ing supercells (identified in the lattice-matching step (step 2)).[221] By iterating over all prescribed supercells and searching interfacial configurations, interfaces with the most favorable binding for each MOF–substrate pair are identified (see Appendix B.3.4).

Several different geometric measures of interfacial binding were used to quantify the chemical compatibility of an interface. Each of these measures can take values between 0 (no binding) and 1 (perfect binding). One previously applied geometric definition of the degree of bonding at the interface is the atomic site overlap (ASO), defined as[213]

$$\mathrm{ASO} = \frac{2S_\mathrm{C}}{S_\mathrm{A} + S_\mathrm{B}}, \tag{4.1}$$

where $S_\mathrm{A}$ and $S_\mathrm{B}$ are the number of binding atoms in the substrate and film binding planes, respectively, and $S_\mathrm{C}$ is the number of coincident atom pairs at the interface. Throughout this work, two atoms on either side of the interface were defined as coincident (bonded) if their separation in the 2D binding plane was less than the mean of their ionic radii (0.96 Å for copper (0.57 Å) and oxygen (1.35 Å)).[222,223] This definition for the ASO has previously been applied to determine the chemical stability of contact layers in hybrid perovskites;[213] however, for interactions involving porous structures, such as MOFs, it yields low ASOs in general due to the disparity between the number of binding sites per unit area in the MOF and substrate layers as a result of the porous MOF structure.[115] Therefore, we have redefined the ASO to be a function of the coordination of the MOF binding sites only as

$$\mathrm{ASO} = \frac{S'_\mathrm{C}}{S_\mathrm{B}}, \tag{4.2}$$

where $S'_\mathrm{C}$ is the number of MOF (film) binding atoms that form bonds with the substrate and $S_\mathrm{B}$ remains the number of film binding atoms, as defined above. Thus, the ASO describes the proportion of MOF binding atoms that can form bonds with the substrate.

The atom positions in a given supercell are not uniquely specified, but can be varied by rigid translations of the lattice along the supercell vectors without changing the supercell itself. We use a Metropolis Monte Carlo (MC) algorithm to apply random rigid translations along the MOF supercell vectors with respect to the fixed substrate to find the maximum ASO (see Appendix B.4 for details). For computational simplicity,

we assumed rigid MOF and substrate structures (we also did not consider the possibility of surface reconstruction); however, flexibility is implicitly taken into account by the lattice-matching and binding tolerances used in screening steps 2 and 3. Assuming rigid frameworks could lead to false negatives in the screening process as structural relaxation could enhance interfacial binding. Nevertheless, we found that the rigid approximation yields sufficient agreement with experimental results while allowing efficient screening (see Section 4.4.1). Structural distortions of the MOF framework[55,206] and in-plane rotations that afford improved epitaxial matching[224] have previously been shown to relieve interfacial stress in MOF heterostructures. Accordingly, our screening process accounts—either explicitly or implicitly—for mechanisms by which interfacial stress can be reduced by considering multiple interface orientations and inherent MOF flexibility.

As MOFs are formed from metal nodes and organic linkers, slicing their crystal structures without breaking strong covalent bonds (that may be part of the organic linkers) is nontrivial.[115] We use a bond network approach to determine appropriate binding planes when slicing a MOF crystal at a particular Miller plane (see Appendix B.3), which introduces missing linker defects (i.e., removes the entire linker) when covalent bonds of the linker are broken by the slicing process. This approach is similar to that used in recent work that applied a graph network analysis to determine the most stable way to slice a crystal structure based on the assumption that the most stable slice minimizes the number of broken bonds.[225] A missing linker defect results in coordinatively unsaturated metal centers in the MOF near the binding plane, which is likely to be energetically unfavorable with respect to the pristine structure.[226] However, we note that a number of MOFs possessing defect sites are known to retain their structural integrity.[206,226–228] As the ASO does not account for the number of dangling or noncoordinated MOF atoms that are not carboxylate oxygens, thermodynamically unstable interfaces with a large proportion of dangling bonds can potentially have high ASOs and thus falsely predict macroscale aligned heteroepitaxial growth. Therefore, we propose a novel geometric measure of interfacial binding that is similar to the ASO, but also considers the number of dangling MOF bonds at the interface and thus is expected to be a better proxy for the interfacial energy. We call this quantity the change

in interfacial bonds or ΔIB and define it to be

$$\Delta\mathrm{IB} = \frac{n_{\mathrm{formed}}}{n_{\mathrm{formed}} + n_{\mathrm{broken}}}, \tag{4.3}$$

where $n_{\mathrm{formed}}$ is the number of MOF–substrate bonds that are formed and $n_{\mathrm{broken}}$ is the number of MOF metal–linker bonds that are broken to form the interface but are not re-formed as MOF–substrate bonds. ΔIB is identical to the ASO if no missing linker defects (broken bonds) are introduced when building the interface and only one MOF–substrate bond is formed per binding atom pair at the interface, but decreases relative to the ASO with increasing numbers of missing linker defects. We contend that considering the energetic cost of forming the interface using ΔIB provides a more accurate measure of interface stability than previously used geometric descriptors. Indeed, in Section 4.4.1.2 we demonstrate the superiority of the ΔIB over the ASO as a predictor for aligned heteroepitaxial growth of MOFs.

### 4.3.4 Structural characterization

We measured crystal structure porosity using Zeo++,[134] which applies a Voronoi decomposition to model the void space in a crystal structure and, due to its efficiency, has been widely employed for high-throughput screening processes.[17,229–231] Where applicable, all Zeo++ calculations used a probe radius of 1.82 Å to represent $N_2$. The pore-limiting diameter (or largest free sphere) was calculated along each of the three crystallographic axes using the resex command in Zeo++. All structures were rendered with OVITO.[187]

## 4.4 Results and discussion

### 4.4.1 Validation of screening methodology

We validated our screening algorithm and determined an optimal set of parameters for each step and substep by comparing the results to experimental data obtained for a series of aligned MOFs heteroepitaxially grown through a one-pot synthetic method on a $Cu(OH)_2$ (010) surface.[55] The experimental MOF data set (given in

Table B.2) comprised six ditopic carboxylate-linked MOFs and CuBTC (also known as HKUST-1), a well-studied 3D MOF with a tritopic carboxylate linker. Two of the dicarboxylate-linked MOFs were 3D pillared frameworks (containing 1,4-benzene dicarboxylate (BDC) and either 4,4′-bipyridine (BPY) or 1,4-diazabicyclo[2.2.2]octane (DABCO) coordinated to the axial positions of the copper paddle wheel), while the remaining four were 2D eclipsed sheet structures. Each of the MOFs possess copper paddle wheels as their metal nodes. Aside from the 2D MOF formed from the 1,4-terphenyl dicarboxylate (TDC) linker, all ditopic MOFs were shown to grow with in-plane alignment on the $Cu(OH)_2$ substrate. CuBTC did not form crystals with in-plane alignment, but out-of-plane alignment was reported. Note that our algorithm specifically identifies structures that are likely to grow with in-plane alignment on $Cu(OH)_2$.

The choice of tolerance parameters in the screening process was motivated by the values used in previous applications of the lattice-matching algorithm,[115,116,209–213] as well as by the ability to correctly categorize the MOFs in the experimental validation data set. A length mismatch tolerance between supercell vectors of 10 %, an angle mismatch tolerance of 2 %, a maximum match area of the smaller of 90× the substrate unit cell area (∼1350 Å$^2$ for $Cu(OH)_2$ (010)) or 9× the film unit cell area, and a maximum area ratio tolerance of 15 % were used. A larger length mismatch tolerance was used than generally applied in the lattice-matching literature[115,209–211,213] due to the flexible nature of MOFs.[115] This choice is supported by experimental data for synthesized MOF-on-MOF heterostructures, in which large lattice mismatches are alleviated by MOF flexibility.[206] To allow testing of MOFs with large unit cells and to ensure that all MOFs in the experimental data set were found to have at least one Miller plane with a substrate-matching lattice, the maximum match area used was much larger than values previously used for metal-oxide interfaces (∼600 Å$^2$)[209,210] or for MOF-on-metal-oxide or MOF-on-metal interfaces (20× the unit cell area for the MOF or metal/metal oxide).[115] A limit on the number of unit cells in the supercell was imposed because large-area interfaces are unlikely to form without defects. Thus, comparison of computed results for very large match areas with experiment is not expected to be meaningful. An analysis of the effect of varying the tolerances and maximum area constraints on the final results for a test database found that the chosen parameters

provide robust results (see Appendix B.2).

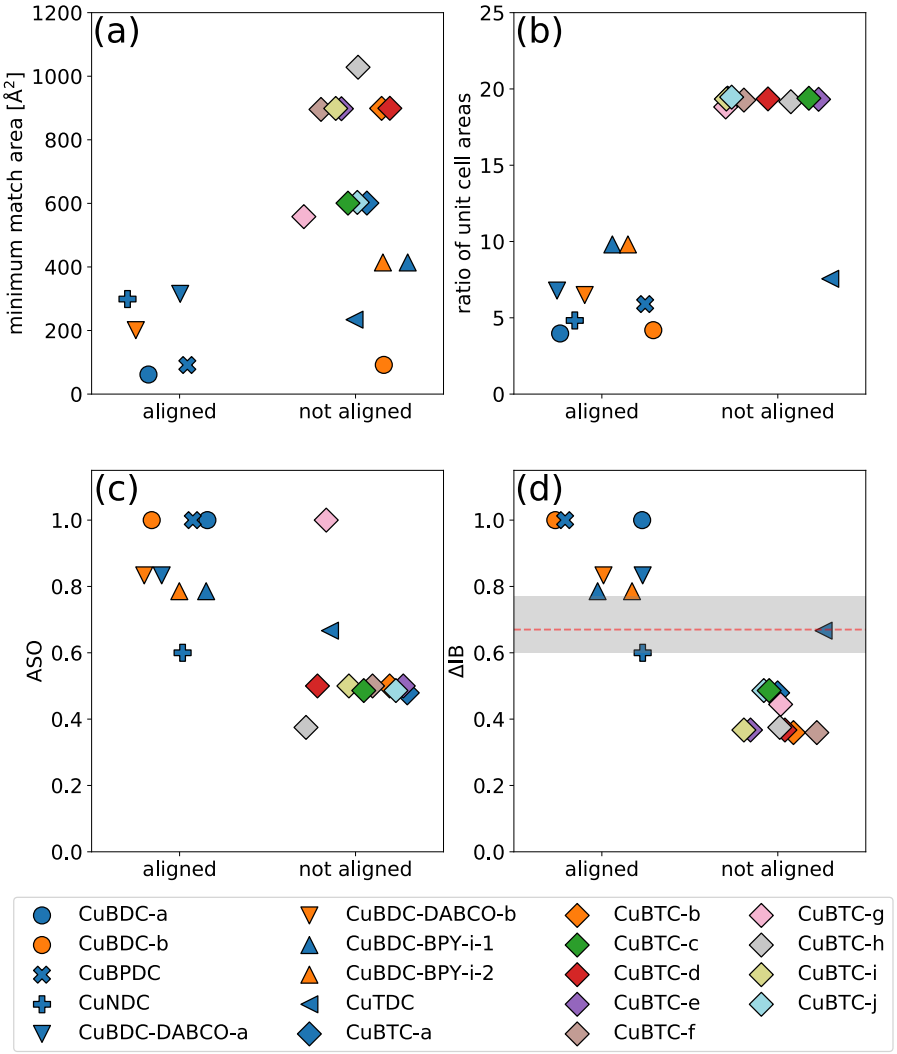



Figure 4.2: Categorical scatter plots of all crystal structures in the experimental data set (listed in Tables B.2 and B.3) showing categorization based on the (a) minimum coincident match area, (b) minimum ratio of MOF and substrate unit-cell areas, (c) maximum ASO, and (d) maximum $\Delta$IB. The shaded region indicates the approximate threshold $\Delta$IB above which in-plane heteroepitaxial alignment occurs. The red line indicates the $\Delta$IB threshold (0.67) applied throughout the remainder of this work. Separated nets of CuBDC-BPY are shown as CuBDC-BPY-i-1 and CuBDC-BPY-i-2, while the interpenetrated structure is not shown (see Appendix B.7). Categorization is based on whether the associated Miller plane grows experimentally from $Cu(OH)_2$ with in-plane alignment (determined by X-ray diffraction).[55]

Figure 4.2 shows the experimental data set[55] categorized by four different descriptors: (1) the minimum coincident match area, (2) the minimum ratio of MOF and substrate unit cell areas, (3) the maximum ASO, and (4) the maximum $\Delta$IB obtained for each crystal structure in the experimental data set (Table B.3). Screening of materials,

including MOFs, for heteroepitaxial growth using the lattice-matching algorithm[209,210] has previously shown that top candidates can be determined by ranking matching lattices by their minimal coincident match area (or the smallest unit-cell multiple that yields a matching lattice).[115,116,211] This is because interfaces with smaller match areas (and smaller mismatch) are expected to have fewer defects.[209,210,213] Figure 4.2a shows that the minimum coincident match area does not distinguish between MOFs that undergo aligned heteroepitaxial growth from those that do not; furthermore, in some cases for which such growth occurs, the MOF interface that grows on the substrate is not the one with the minimum coincident match area. On the other hand, Figure 4.2b shows that the minimum ratio of the MOF and substrate unit-cell areas for all Miller planes successfully categorizes almost all the MOFs in the experimental data set and suggests that a small ratio of MOF and substrate unit-cell areas is correlated with aligned heteroepitaxial growth.

Both the minimum coincident match area and unit-cell area ratio are geometrical descriptors that do not consider the chemistry at the interface and are significantly faster to calculate than the ASO or ΔIB. But the neglect of interface chemistry misses crucial features of interface formation. For example, in each of the cases in Figure 4.2a for which the lowest match-area interface is not the one that is known to exhibit heteroepitaxial growth, the associated MOF binding interface was found to have low ASO and ΔIB ($\approx 0.5$; not shown in Figure 4.2). Figure 4.2b suggests that aligned heteroepitaxial growth occurs for a unit-cell area ratio below about 10. However, several MOFs in the validation data set have Miller planes that satisfy this criterion but are not observed to undergo aligned growth on $Cu(OH)_2$. By contrast, the ΔIB descriptor, which does account for bonding between atoms at the interface, correctly categorizes almost all of the MOF interfaces that grow aligned on $Cu(OH)_2$, as shown in Figure 4.2d, without predicting aligned growth for Miller planes that have not been observed. On the other hand, the ASO, which also considers interfacial binding between atoms and has previously been used to categorize heteroepitaxial growth,[213] does not appear to be as good a predictor of aligned heteroepitaxy as ΔIB, as indicated by the lack of any obvious trend in Figure 4.2c.

Figure 4.2 shows that CuTDC is an outlier by both the unit-cell area ratio and ΔIB descriptors; i.e., by these measures, it would be predicted to undergo aligned het-

eroepitaxial growth that is not observed experimentally. We suggest that experimental factors associated with CuTDC that are not accounted for in the screening algorithm, such as the low solubility of the TDC linker, may inhibit aligned MOF growth. Out of the MOFs that experimentally show aligned heteroepitaxial growth on $Cu(OH)_2$, only CuNDC has a lower ASO or $\Delta$IB than CuTDC. CuNDC has been found to deform at MOF–MOF interfaces with high lattice mismatch,[206] which our screening algorithm does not consider, and so this MOF could also potentially be miscategorized by the metrics in Figure 4.2. Neglecting CuTDC, Figure 4.2 suggests that MOFs with a $\Delta$IB above a threshold somewhere between 0.60 and 0.77 undergo aligned heteroepitaxial growth, while those with $\Delta$IB below this threshold do not. We have applied a threshold value around the midpoint of this range ($\Delta$IB = 0.67) in further analysis. For all MOFs for which aligned heteroepitaxial growth has been shown to occur experimentally, only one in-plane orientation of the MOF crystallites was found. Our screening results are consistent with this observation: for example, the MOFs in Figure 4.2d with $\Delta$IB above the threshold for aligned heteroepitaxial growth ($\geq 0.67$) have no other in-plane orientations or Miller planes with a $\Delta$IB above this threshold (the method used to identify unique interfaces is described in Appendix B.5.1). The following two sections discuss the robustness of the screening algorithm and highlight the importance of considering the interfacial energy (via $\Delta$IB) to avoid false positives.

#### 4.4.1.1 Effect of crystal structure

In many cases, multiple published crystal structures exist for a given set of MOF building blocks (i.e., organic linker and metal node). For example, there are at least 17 entries of CuBTC in the CORE MOF database, which contains approximately 5000 experimental MOF structures.[94] The lattice parameters of these structures vary in many cases by a small amount ($\sim$0.1 Å). Step 3 of our screening process uses the structures of matching MOF and substrate supercells identified in step 2 to calculate the ASO and $\Delta$IB. Each supercell has a particular relative orientation of the atoms on either side of the interface, with only certain relative orientations aligning the binding sites such that the ASO (or $\Delta$IB) is high (see Appendix B.3.4). Both the lattice mismatch and the set of matching supercells depend on the crystal lattice parameters, and these affect the ASO and $\Delta$IB. We have used a set of crystal structures (Table B.3)

for the MOFs CuBDC, CuBDC-DABCO, and CuBTC in the experimental data set to investigate the impact of different crystal structures of the same MOF on the results of our screening process.

For both CuBDC and CuBDC-DABCO, slight differences between unit-cell vectors for different crystal structures yield MOF supercells with the same highest ASO that have the same orientation but different area. This is because small changes in the unit-cell vectors can lead to certain supercells failing the lattice-matching step of the screening. However, these differences do not impact the screening process, as the same ASO or $\Delta$IB is obtained for all crystal structures. On the other hand, in the case of CuBTC, Figure 4.3a shows that the particular crystal structure can affect the maximum ASO and, more importantly, the Miller plane with the highest ASO. Figure 4.3b–e shows the four different binding interfaces found for all CuBTC crystal structures. This difference between the maximum ASO for different CuBTC structures is a consequence of two factors. First, small changes to the unit-cell vectors can result in a given supercell orientation of a particular MOF Miller plane failing the lattice-matching criteria; i.e., different crystal structures yield different supercell orientations and ASOs. This explains the dramatic difference between the maximum ASO of the binding interface (Figure 4.3b) for CuBTC-g compared with those of the other crystal structures (Figure 4.3a). Second, small changes to the unit-cell vectors modify the spacing between atoms in the supercell, which can impact the coincidence of atoms at the interface. Figure 4.3f shows the orientation and lattice spacing that gives rise to the ASO of 1 for CuBTC-g and an example interface with a maximum ASO $\approx 0.33$ for the other CuBTC crystal structures. These issues could be partially resolved by using less stringent interface bonding criteria, but this would likely lead to spurious lattice and binding matches for other MOFs. On the other hand, the maximum $\Delta$IB is comparable and low for all CuBTC crystal structures (Figure 4.2d). Thus, $\Delta$IB does not show the same sensitivity to the crystal structure lattice parameters as the ASO and so appears to be a more robust descriptor for aligned heteroepitaxial growth.

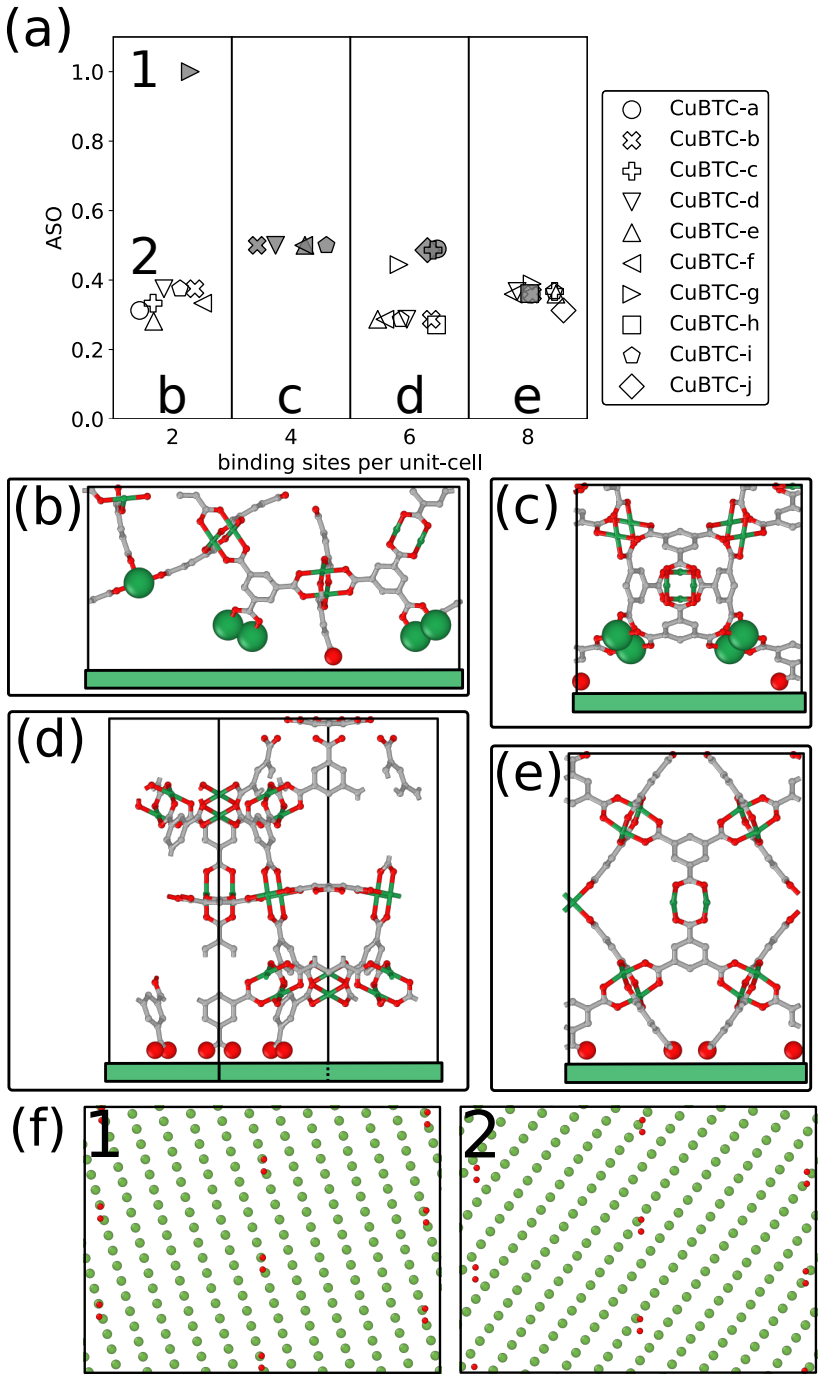



Figure 4.3: (a) Maximum ASO for all possible binding interfaces of a set of CuBTC crystal structures as a function of the number of binding oxygens per unit cell. Filled symbols are the maximum ASO obtained for each crystal structure. Points in the sections of the plot labelled b–e correspond to the same number of binding sites and interfacial structure shown in (b)–(e), respectively, in which surface-binding carboxylate oxygens and undercoordinated copper atoms are highlighted as large spheres (Cu: green; O: red; C: gray; H atoms omitted for clarity). Substrate surfaces are shown schematically as green slabs in (b)–(e). (f) Comparison of the interfaces with a maximum ASO of 1.0 and ≈0.33 for the sets of points marked "1" and "2", respectively, in (a).

#### 4.4.1.2 Cost of interface formation on chemical compatibility

Figure 4.3b shows the binding interface of the $(11\bar{1})$ Miller plane of CuBTC and highlights the carboxylate-oxygen binding sites and coordinatively unsaturated copper sites associated with missing-linker defects. This interface gives the maximum possible ASO of 1 for the CuBTC-g crystal structure. However, this surface of CuBTC has not been shown to exist experimentally on the facets of CuBTC crystals,[232] nor has it been shown to be the most energetically stable.[233] Furthermore, it has not been shown to bind to surfaces in any examples of surface-anchored CuBTC structures.[58,59] This contradiction between the computed ASO and experimental observations indicates that the ASO is not always a good predictor of favorable interface formation. In particular, the ASO ignores energetically unfavorable broken or dangling bonds. In the case of the $(11\bar{1})$ Miller plane of CuBTC, 16 bonds are broken per unit cell to form the interface in Figure 4.3b.

On the other hand, the $\Delta$IB metric that we have introduced considers both the bonds that are formed and those that are broken to form the interface, resulting in a low $\Delta$IB of 0.125 for this interface. In contrast, the (111) and (100) Miller planes, which have been experimentally observed at CuBTC surfaces, each have lower ASOs but no missing linker defects are required to form interfaces with $Cu(OH)_2$ (Figure 4.3d,e); thus, they have higher $\Delta$IBs ($\approx$0.36 and $\approx$0.28, respectively) than the $(11\bar{1})$ Miller plane. This result is consistent with previous theory and experiment, in which the (111) and (100) Miller planes have been found to be the most stable surfaces[232,233] and have been shown to form interfaces with other materials.[58,59] In contrast to the ASO, $\Delta$IB is similar and low for all Miller planes of CuBTC, indicating that in-plane heteroepitaxial alignment of this MOF on $Cu(OH)_2$ is not favored. This agrees with experimental results that show crystal growth of CuBTC with out-of-plane orientation only.[55] It is also possible that the presence of multiple interfaces with similar $\Delta$IB values could lead to competitive growth of multiple distinct orientations, which could be the cause of diminished in-plane alignment on the macroscale for CuBTC. In either case, this example highlights the superiority of the $\Delta$IB over the ASO in explaining the experimental observations.

### 4.4.2 Screening CORE MOF database

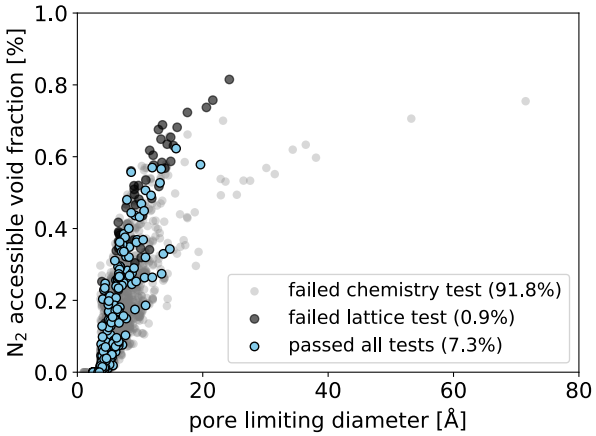



Figure 4.4: Nitrogen-accessible void fraction versus pore-limiting diameter for all crystal structures in the CORE MOF database. Points are colored by how far the MOF progressed through the screening process.

We have applied our screening algorithm to the Computation-Ready, Experimental (CORE) MOF database. The CORE MOF database comprises approximately 5000 MOF structures with pore-limiting diameters greater than 2.4 Å that have been processed by removing solvent from the pores, fixing atom position disorder, and converting each structure to its primitive unit cell.[94] Figure 4.4 shows the nitrogen-accessible void fraction as a function of the pore-limiting diameter (also known as the largest free sphere) of all structures in the CORE MOF database. For each crystal structure, the screening process was applied with the validated parameters determined from the analysis of the experimental data set in Section 4.4.1. Only 8 % of the CORE MOF database passed the chemistry test and 89 % of those MOFs passed the lattice-matching step, resulting in 7 % of the database or 336 structures whose interfacial binding was tested; 20 of these structures were interpenetrated, which led to 40 additional structures representing the separated nets of those 20 structures that were screened (the processing of interpenetrated structures is described in Appendix B.7). To emphasize the efficiency of the screening process, the first two steps took minutes on six Intel i7-4790k CPU cores for the entire CORE MOF database, while the final step for the remaining 376 structures took approximately 25 h on six CPU cores.

Figure 4.5 shows the maximum $\Delta$IB of each of the 376 structures that passed the first two screening steps as a function of the ratio of the MOF and substrate unit-

cell areas. Because macroscale aligned heteroepitaxial growth relies on the preferential growth of only one MOF crystal orientation with respect to the substrate, the points in Figure 4.5 are colored according to the second highest ΔIB ($\Delta IB_2$), which corresponds to a different Miller plane or a different in-plane orientation of the same Miller plane. The symbol shape in Figure 4.5 indicates the MOF topology, which was obtained from the TTO ToposPro topological database.[234,235] The standard cluster representation of 3D MOFs was used, which separates the atomic net into building blocks (such as metal clusters and organic linkers), which are then treated as nodes to determine the underlying network topology.[236,237] Structures listed as unknown did not have cluster representation topology entries in the database. Similar to what was found for the limited experimental data set in Figure 4.2, there is no clear correlation between the maximum ΔIB and the ratio of the unit-cell areas for the MOF structures in the CORE MOF database. Although all structures with high ΔIB values correspond to low unit-cell area ratios, the converse is not true, with structures with low ΔIB exhibiting a wide range of unit-cell area ratios. Therefore, the data suggest that a low ratio of unit cell areas does lead to favorable interfaces, as has previously been suggested,[115] but the interfacial energy (through ΔIB) should be considered to identify MOFs likely to show aligned heteroepitaxial growth.

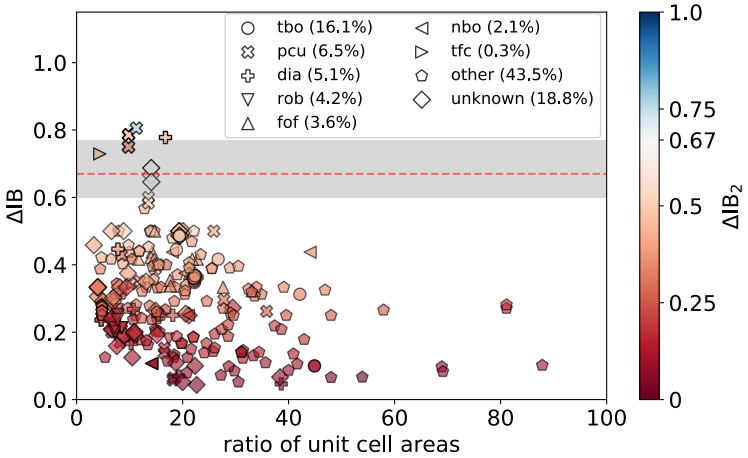



Figure 4.5: Maximum ΔIB of crystal structures in the CORE MOF database that passed screening steps 1 and 2 as a function of the ratio of the MOF and substrate unit-cell areas for the Miller plane with the maximum ΔIB. Coloring is by the value of the second highest ΔIB ($\Delta IB_2$). The symbol shape indicates the 3D MOF topology (structures with no reported topology are given as "unknown"). The percentage of MOFs that passed screening step 2 of each topology is given in the legend. The shaded region indicates the approximate ΔIB threshold for heteroepitaxial in-plane alignment determined in Section 4.4.1. The red line indicates the value of ΔIB used as the threshold for top candidates.

From Figure 4.5, the top candidates in the MOF database for aligned heteroepitaxial growth on $Cu(OH)_2$ can be identified as those with maximum $\Delta$IB above the previously defined threshold of approximately 0.67 (see Section 4.4.1) and second highest $\Delta$IB ($\Delta IB_2$) below this threshold. Figure 4.6 shows the interface structures of the top candidates (MOFs with similar structures have been grouped together). Three of the top candidates (NEJRUR, NEJSAY, and NEJSEC) correspond to different crystal structures (all of which are interpenetrated MOFs with **pcu** topology) of the MOF CuBDC-BPY (Figure 4.6a), which has been shown experimentally to grow aligned from $Cu(OH)_2$ (one of these structures, NEJRUR, was in the parameterization data set). Top candidate CEHPIP (Figure 4.6b) corresponds to an interpenetrated MOF that is isoreticular to CuBDC-BPY, in which the BDC linker has been replaced with 1,4-cyclohexanedicarboxylic acid. CEHPIP also has an underlying **pcu** topology. Top candidate ZAZBUZ (**tfc** topology) is made up of BDC linkers bound to chainlike copper nodes, rather than the paddle wheel in CuBDC (Figure 4.6c), and the top candidate ZECKID has a **dia** topology (Figure 4.6d). Two binding planes (equivalent within a 90° rotation) for the same Miller plane of ZECKID were found with the same maximum $\Delta$IB, one of which was excluded manually upon visual inspection of all top candidates. Finally, one of the top candidates corresponds to an interpenetrated MOF (UNABUH, Figure 4.6e), which has an unknown topology by the cluster representation (using the standard unsimplified representation, it has a **tfs** topology).

Figure 4.5 along with experimental data show that the $Cu(OH)_2$ (010) substrate favors the growth of MOFs with the **pcu** topology (many of the high-ranking points with **pcu** topology are overlapping in Figure 4.5). Furthermore, the Miller planes that meet the $\Delta$IB criteria for top candidates for aligned heteroepitaxial growth have approximate rectangular symmetry, as indicated by their in-plane unit-cell lattice vectors (Figure B.11a). These findings suggest that the rectangular symmetry of the (010) plane of $Cu(OH)_2$ selects out top candidates for aligned heteroepitaxial growth with similar symmetry. Note that MOFs with **pcu** and other similar topologies that commonly have at least one lattice plane with rectangular symmetry have been shown previously to exhibit in-plane alignment when grown via heteroepitaxy;[224,238] importantly, the **pcu** topology has been shown to produce superior internal surface areas compared to other MOF topologies, suggesting that the ability to target **pcu** MOFs

with heteroepitaxy could be beneficial for developing useful MOF heterostructures.[239] Note that MOFs with topologies other than **pcu** have high maximum ASOs, while their maximum ΔIB is significantly diminished due to the number of missing linker defects required to form the interface. This result suggests that certain topologies may not grow heteroepitaxially due to the difficulty of cleaving the crystal structure at a planar interface without breaking a large number of bonds.

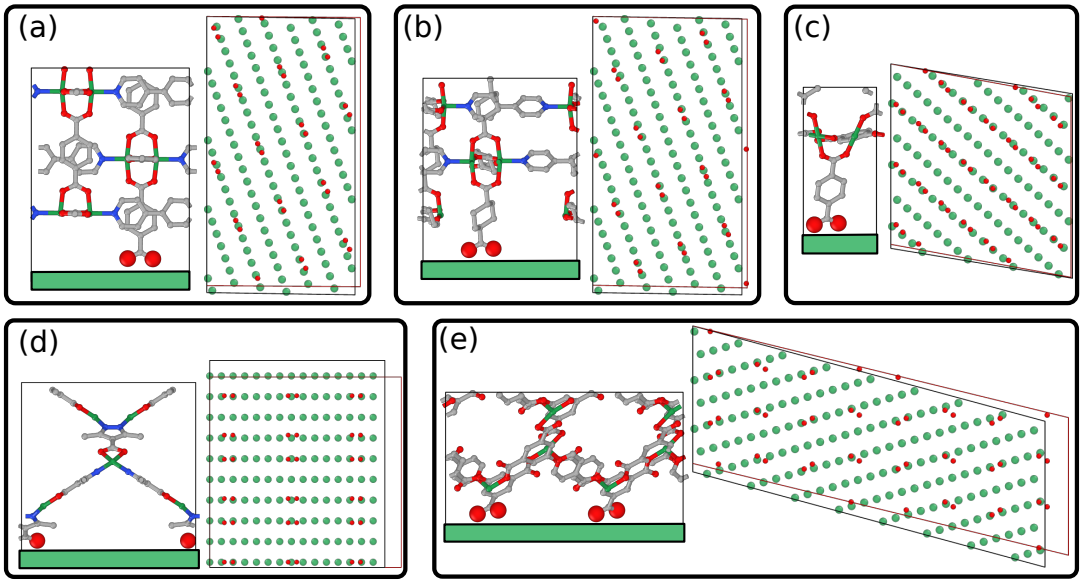



Figure 4.6: Representative interface structures of the top candidates for heteroepitaxial growth from screening the CORE MOF database: (a) NEJRUR, NEJSAY, and NEJSEC (all of which are isostructural to CuBDC-BPY), (b) CEHPIP, (c) ZAZBUZ, (d) ZECKID, and (e) UNABUH. In each case, the interface that yields the highest ΔIB is shown (Cu: green; O: red; C: gray; N: blue; H atoms are omitted for clarity; MOF supercells: red; substrate supercells: black). Interpenetrated structures are shown in (a), (b) and (e). Substrate surfaces are shown schematically as green slabs.

### 4.4.3 Selective MOF growth by topology

Experimental data[55] as well as the top candidates found using our screening process on the CORE MOF database suggest that the rectangular symmetry of the (010) Miller plane of $Cu(OH)_2$ preferentially selects MOFs with at least one Miller plane with rectangular symmetry. The majority of selected MOFs were found to also have **pcu** topology, a topology that often corresponds to structures having lattice planes of rectangular symmetry. The ability to select a specific MOF topology could greatly aid the development of MOF heterostructures and provides a design principle for further experimental study. To investigate whether selectivity by MOF topology is expected

to occur generally for $Cu(OH)_2$, we applied our screening process to two other MOF databases, the TOBACCO database[99,214] and the hMOF database,[101] which contain hypothetical MOF structures that were built using in silico structure generation. The TOBACCO database contains approximately 13 000 noninterpenetrated MOF structures constructed from a diverse range of metal nodes, organic linkers, and underlying network topologies.[99,214] Figure 4.7a shows the maximum $\Delta$IB as a function of the ratio of the MOF and substrate unit-cell areas for all crystal structures that passed screening steps 1 and 2 (10.3 % of the full database). The points are also colored by the value of the second highest $\Delta$IB. Note that none of the MOFs in this database with **pcu** topology passed the chemistry test of the screening process, as they were all zinc MOFs. Nonetheless, the screening process picks out a small subset of topologies with high maximum $\Delta$IBs, namely **lvt-b** and **nbo-b**, although only **nbo-b** MOFs are predicted to yield aligned heteroepitaxial growth on $Cu(OH)_2$. Screening of the TOBACCO database predicts again that the $Cu(OH)_2$ substrate selects MOFs with only a small set of topologies as well as Miller planes with in-plane rectangular symmetry (see Figure B.11b). Note that there are MOFs in the TOBACCO database with topologies other than **lvt-b** and **nbo-b** with high maximum ASO and low maximum $\Delta$IB, which suggests that the ability to cleave 3D MOF structures at a 2D planar interface with minimal broken bonds also drives their selection. In addition, Figure 4.7a shows that many binding planes in the TOBACCO database have large unit-cell areas compared to the $Cu(OH)_2$ substrate, which are not expected to be good candidates for aligned heteroepitaxial growth because large, pristine substrate surfaces would be required, which is unlikely to occur in real systems.

To confirm the favorability of the **pcu** topology for heteroepitaxial alignment on the $Cu(OH)_2$ substrate, we screened the hMOF database, which contains approximately 150 000 hypothetical MOF structures, approximately 90% of which have underlying **pcu** nets.[101] Only the unfunctionalized and noninterpenetrated structures were considered, resulting in only 2570 structures that were screened, 27% or 693 of which made it to the final step of the screening process. Figure 4.7b shows that out of the 693 structures, 266 have only one orientation of one Miller plane with $\Delta$IB greater than the 0.67 threshold, all of which have **pcu** topologies, which supports the assertion that the **pcu** topology is favorable for aligned growth on the $Cu(OH)_2$ substrate. As

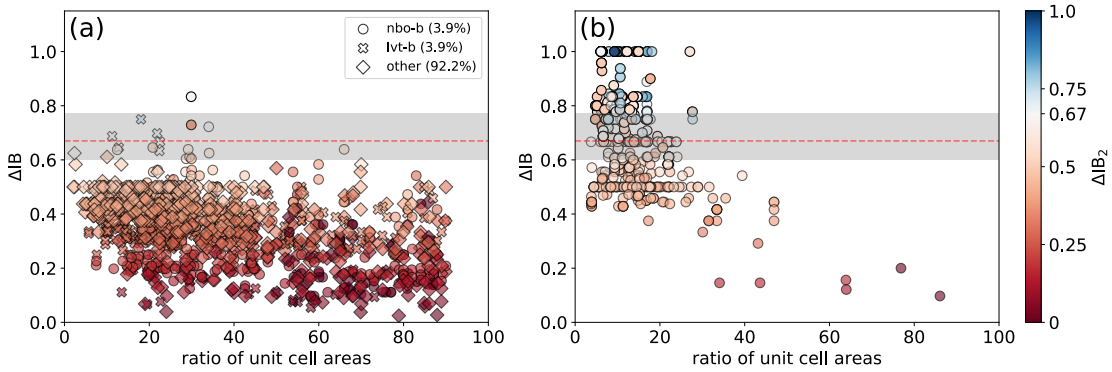



Figure 4.7: Maximum ΔIB of crystal structures in the (a) TOBACCO[99,214] and (b) hMOF database[101] that passed screening steps 1 and 2 as a function of the ratio of the MOF and substrate unit-cell areas for the Miller plane with the maximum ΔIB. Points are colored by the value of the second highest ΔIB ($\Delta IB_2$). The symbol shape in (a) indicates the 3D MOF topology (the percentages of MOFs in the entire TOBACCO database with the specified topologies are given in the legend). The shaded region indicates the approximate ΔIB threshold for heteroepitaxial in-plane alignment determined in Section 4.4.1. The red line indicates the value of ΔIB used as the threshold for top candidates.

observed for the other two MOF databases, these top candidates for aligned growth also have binding planes with approximate rectangular symmetry (see Figure B.11c). Figure 4.7b also shows that many MOFs in this database (145 in total) have multiple Miller planes with $\Delta IB \geq 0.67$, unlike the experimental data set, in which only one Miller plane was found to have a high ΔIB for each MOF. This is because each hMOF structure is built from up to three unique (but sometimes structurally similar) ligands from a ligand database, leading to more inequivalent Miller planes per MOF and often to multiple Miller planes with similar ΔIB values. Many MOFs in the hMOF database also have similar organic linkers, which results in similar structures and interfaces and helps to explain why many MOFs have a maximum ΔIB above the threshold. Both the hMOF and TOBACCO databases were compiled from mostly hypothetical structures, which limits their applicability in an experimental setting. Nevertheless, screening of the three MOF databases has identified a small number of MOF topologies as useful starting points for future experimental investigation of MOF heteroepitaxy.

### 4.4.4 Templating pore architecture through epitaxial growth

Intuitively, aligned heteroepitaxially grown MOF crystals could lead to anisotropy in the pores along the different directions with respect to the substrate. To test this hypothesis, we calculated the pore-limiting diameters in the $a$, $b$, and $c$ crystallographic

dimensions for all **pcu** MOFs in the hMOF database whose maximum ΔIB occurs for a Miller plane with $|h| + |k| + |l| = 1$ (a subset of those in Figure 4.7b). This subset of MOFs was chosen because, for the **pcu** topology, it can reasonably be assumed that the pore windows align with the three crystallographic axes of the MOF for these Miller planes, for which the pore-limiting diameter was calculated. Figures B.13 and B.14 show that high ΔIB MOFs that are predicted to undergo aligned heteroepitaxial growth have a similar distribution of in-plane pore-limiting diameters to low ΔIB MOFs, which suggests that epitaxial growth does not constrain the in-plane pore-limiting diameters as a result of the registry with the substrate lattice. This suggests that a broad range of possible pore sizes can be achieved for a given substrate and, therefore, MOFs with precise pore sizes can be chosen for target applications.

Falcaro and co-workers[55] showed that macroscale alignment of the CuBPDC MOF on a $Cu(OH)_2$ substrate could be employed to align fluorescent molecules within the MOF and yield anisotropic fluorescence from the MOF film. It follows that anisotropy of the pore network is an important design aspect for MOF thin films that could be controlled through in-plane heteroepitaxial alignment. Furthermore, the directionality of the pores with respect to the surface due to a preferred crystal orientation has been shown to impact adsorption properties of porous-coordination polymers,[240] and such properties can be predicted once the in-plane orientation is determined. To investigate this concept further, we calculated the pore network anisotropies from the pore-limiting diameters along the three crystallographic axes (described above) for each **pcu** MOF in the hMOF database with a maximum $\Delta IB \geq 0.67$ associated with a Miller plane with $|h| + |k| + |l| = 1$. Pore network anisotropies were calculated as either the ratio or the absolute difference between the maximum in-plane pore-limiting diameter ($PLD_{\text{in-plane,max}}$) and the pore-limiting diameters in the other two crystallographic dimensions (the minimum in-plane pore-limiting diameter ($PLD_{\text{in-plane,min}}$) or the out-of-plane pore-limiting diameter ($PLD_{\text{out-of-plane}}$)). Figure 4.8 shows that the top candidates from our screening process for the hMOF database cover a broad range of anisotropies and pore shapes (using either definition), including some very anisotropic pore networks. While this allows for the selection of the desired pore properties for a target application, we note that the definition of anisotropy that matters will be different for certain applications. For example, a MOF in which all three pores are

large may have high anisotropy with respect to the difference between the pore sizes, but low anisotropy with respect to their ratios. This type of pore network would not provide any notable anisotropic effect for the adsorption of relatively small molecules, but may be useful for some other application, such as alignment of large guests.

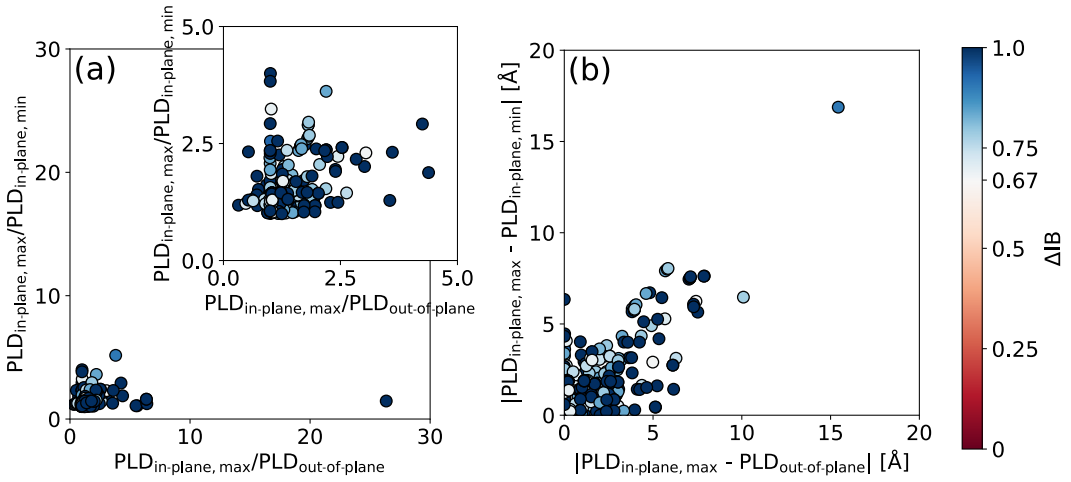



Figure 4.8: Pore network anisotropy for all MOF structures in the hMOF database predicted to grow aligned on $Cu(OH)_2$ whose maximum $\Delta$IB is associated with a Miller plane for which $|h| + |k| + |l| = 1$: (a) ratio and (b) absolute difference of the largest in-plane pore-limiting diameter ($PLD_{in\text{-}plane,max}$) and minimum in-plane pore-limiting diameter ($PLD_{in\text{-}plane,min}$) versus that of the largest in-plane pore-limiting diameter and out-of-plane pore-limiting diameter ($PLD_{out\text{-}of\text{-}plane}$). Points are colored by the value of the maximum $\Delta$IB and only those with the second highest $\Delta IB < 0.67$ are shown.

## 4.5 Conclusions

We have developed a robust and efficient screening process of metal–organic frameworks (MOFs) for aligned heteroepitaxial growth on metal-hydroxide substrates, which we have validated by comparison with experimental measurements.[55] This process applies a geometrical lattice-matching algorithm[209,210] and extends previous methodologies for determining chemical compatibility at an interface[213] to filter and categorize MOF structures. We have proposed a novel metric for the favorability of interfacial binding, $\Delta$IB, which is simple and efficient to compute and accounts for both the energetic gain of bond formation and the energetic cost of broken and dangling bonds. Importantly, our screening process can be generalized to cases other than those studied in this work. The MOF in the experimental data set made from the TDC linker was found to be an outlier in the validation of the screening process, which we suggest could

be due to experimental difficulties associated with the solubility of that linker. This finding highlights the need to consider the chemistry of the MOF building blocks of any top candidates generated by computational screening before further experimental evaluation.

Our screening process does not consider the kinetics or thermodynamics of MOF growth from the substrate or the possibility of competing binding modes of the ligands used to form the MOF (for example, the nitrogen donors of bipyridine ligands may coordinate to copper species and compete with carboxylate groups). Furthermore, we have limited our analysis to a substrate in which the metal ion is in a single oxidation state and for which a single interfacial binding geometry is expected to dominate. Substrate interfaces at which multiple coordination geometries are possible at a given site would be less likely to yield aligned heteroepitaxial growth due to the less stringent constraints on binding, and therefore are potentially less promising for aligned heteroepitaxial growth. Neglecting to consider multiple binding geometries in the screening process in cases in which they are possible could potentially lead to incorrect predictions of aligned growth or of no growth. The screening algorithm in this work can readily be extended to consider different types of interfacial binding atoms and/or binding geometries in these circumstances. In addition, more computationally intensive simulations that quantitatively model different interfacial interactions could be used after initial high-throughput screening in the final selection of MOF candidates.[116,206,220] Nonetheless, the criteria we have implemented allow for an efficient screening process that considers the necessary conditions for aligned heteroepitaxial growth of MOF crystallites on the macroscale.

We used the screening process to compile a list of top candidates for aligned heteroepitaxial growth from the CORE MOF,[94] hMOF,[101] and TOBACCO[99,214] databases, which equated to screening approximately 20 000 structures in a few days on a desktop workstation. The screening results showed that MOFs that have at least one lattice plane with rectangular symmetry have the greatest propensity for aligned heteroepitaxial growth on $Cu(OH)_2$, a condition that is commonly satisfied only for certain MOF topologies (the **pcu** and **nbo-b** topologies were found to make up the majority of top candidates). This finding indicates a substrate-directing effect, whereby the symmetry of the substrate surface determines the preferential symmetry of MOF

binding sites at the interface and, hence, the symmetry of the MOF binding plane. Knowledge of substrate-dependent design principles like this allows for topologically targeted synthetic strategies, which offers an efficient route to viable experimental targets.[97,98,239] We identified seven (five of which are unique) experimentally realized MOF structures (REFCODES: ZAZBUZ, ZECKID, CEHPIP, NEJRUR, NEJSAY, NEJSEC, and UNABUH) that are likely to form aligned heteroepitaxial structures on $Cu(OH)_2$.

Furthermore, we used the hypothetical MOF database (hMOF), which comprises mostly **pcu** MOF structures, to show that the dimensions of the substrate lattice, which determine the registry and alignment of the MOF, do not constrain the available range of pore diameters and the anisotropy of the pore networks. A broad range of possible pore architectures means that the design of MOF thin films can be tailored toward target applications. The screening algorithm we have described offers a fast and robust way to find MOF candidates for further experimental analysis in the growing field of MOF thin films and device development. In addition, the categorization of vast MOF databases afforded the determination of design principles and structure–property relationships.

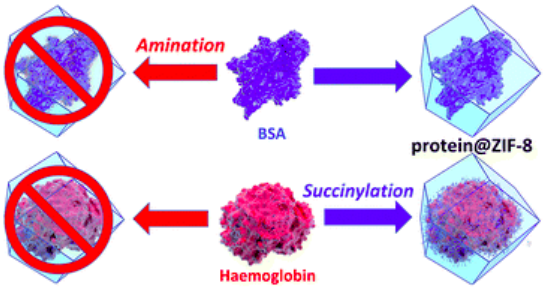
All code used in this work is freely available at `https://bitbucket.org/andrewtarzia/epitmof/src/master/`.

## Acknowledgements

A. T. was supported by an Australian Government RTP Scholarship and a PhD top-up scholarship from CSIRO Division of Materials Science and Engineering. C. J. D. acknowledges the JSPS for a short-term visiting professorship. P.F. acknowledges TU Graz for the Lead Project (LP-03) and the European Union's Horizon 2020 Programme (FP/2014-2020)/ERC Grant Agreement n.771834 -POPCRYSTAL.

# CHAPTER 5

# Prediction of MOF-based biomimetic mineralization




This chapter collates research that has appeared in the following publications. This chapter reports on original research I conducted during the period of my Higher Degree by Research candidature and is not subject to any obligations or contractual agreements with a third party that would constrain its inclusion in this thesis. I am the primary author of this chapter. I hereby certify the statement of contribution is accurate.

**Title:** Protein Surface Functionalisation as a General Strategy for Facilitating Biomimetic Mineralisation of ZIF-8

*Published*

Maddigan, N; Tarzia, A; Huang, D. M.; Sumby, C. J.; Bell, S. G; Falcaro, P.; Doonan, C. J. Protein Surface Functionalisation as a General Strategy for Facilitating Biomimetic Mineralisation of ZIF-8. *Chem. Sci.* 4217-4223 (2018).

**Principal author (candidate):** Andrew Tarzia

**Contribution:** 30%

Application of computational and theoretical analysis. Collation of results. Construction of figures corresponding to computational results. Writing and revision of computational sections of manuscript.

Signature: Date: 11/1/19

## Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i the candidate's stated contribution to the publication is accurate (as detailed above);

ii permission is granted for the candidate in include the publication in the thesis; and

iii the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

**Co-Author:** Natasha K. Maddigan

**Contribution:**

Performed experiments. Constructed figures. Wrote and drafted manuscript. Acted as primary author.

Signature: Date: 2/1/19

**Co-Author:** David M. Huang

**Contribution:**

Supervised the computational methods and theoretical calculations and assisted in revision of manuscript.

Signature: Date: 21/01/2019

**Co-Author:** Christopher J. Sumby

**Contribution:**

Supervised experimental methods, assisted in interpretation of results, assisted in writing and revision of manuscript.

Signature: Date: 21/1/19

**Co-Author:** Stephen G. Bell

**Contribution:**

Supervised experimental methods, assisted in interpretation of results, assisted in writing and revision of manuscript.

Signature: Date: 21/1/19

**Co-Author:** Paolo Falcaro

**Contribution:**

Conceived project, assisted in interpretation of experimental results, assisted in writing and revision of manuscript. Acted as corresponding author.

Signature: Date: 21/01/2019

**Co-Author:** Christian J. Doonan

**Contribution:**

Conceived project, assisted in interpretation of experimental results, assisted in writing and revision of manuscript. Acted as corresponding author.

Signature: Date: 21/1/2019

**Title:** Carbohydrates@MOFs

*Accepted*

Astria, E.; Thonhofer, M.; Ricco, R.; Liang, W.; Chemelli, A.; Tarzia, A.; Alt, K.; Hagemeyer, C. E.; Rattenberger, J.; Schroettner, H; Wrodnigg, T.; Amenitsch, H.; Huang, D. M.; Doonan, C. J.; Falcaro, P. Carbohydrates@MOFs. *Mater. Horiz.* **2019**, *accepted.* DOI: 10.1039/C8MH01611A

**Principal author (candidate):** Andrew Tarzia

**Contribution:** 5%

Application of computational and theoretical analysis. Collation of computational results. Construction of figures corresponding to computational results. Writing and revision of computational sections of manuscript.

Signature: Date: 11/1/19

## Co-Author Contributions

By signing the Statement of Authorship, each author certifies that:

i the candidate's stated contribution to the publication is accurate (as detailed above);

ii permission is granted for the candidate in include the publication in the thesis; and

iii the sum of all co-author contributions is equal to 100% less the candidate's stated contribution.

**Co-Author:** Weibin Liang

**Contribution:**

Carried out experimental analysis of MOF biocomposites.

Signature: Date: 22/01/2019

**Co-Author:** David M. Huang

**Contribution:**

Supervised computational and theoretical analysis done by the candidate. Assisted in manuscript preparation and drafting.

Signature: Date: 21/01/2019

**Co-Author:** Christian J. Doonan

**Contribution:**

Conceived project, assisted in interpretation of experimental results, assisted in writing and revision of manuscript. Acted as corresponding author. As the primary supervisor of the candidate and a corresponding author of this paper, I declare that this statement of authorship was signed on behalf of all remaining co-authors.

Signature: Date: 21/1/2019

**Co-Author:** Paolo Falcaro

**Contribution:**

Conceived project, assisted in interpretation of experimental results, supervised writing and revision of manuscript. Acted as corresponding author.

Signature: Date: 21/01/2019

## 5.1 Abstract

Biomacromolecules, such as proteins and carbohydrates, can be encapsulated inside MOF shells via a process called biomimetic mineralization. Herein, we present calculations that corroborate experimental results that suggest a vital role of the electrostatic properties of biomacromolecules on the biomimetic mineralization process and use calculations of these electrostatic properties to predict the propensity of biomacromolecules to induce biomineralization. We use theory and computation to model the interactions of proteins with their surrounding electrolyte to verify, for the first time, the role of negatively charged protein surfaces in the accumulation of nearby zinc ions, which is expected to enhance ZIF growth. Therefore, we find that the surface electrostatic potential of a protein is strongly correlated with its propensity for seeding ZIF-8 growth. The calculated pI of a protein was also found to predict whether a protein would seed ZIF-8 growth accurately. Furthermore, the effect of chemical functionalization of a protein, shown experimentally to allow for the control of the biomimetic mineralization process, on its propensity to seed ZIF growth can be predicted using the sequence pI. The efficiency and accuracy of the sequence pI as a predictor for biomimetic mineralization allows for the high-throughput screening of proteins for encapsulation inside MOFs. Importantly, these results were corroborated by an experimental and computational study of functionalized carbohydrate systems, which suggests that the prediction of biomimetic mineralization based on electrostatic interactions can be generally applied to a wide range of biomacromolecules.

## 5.2 Introduction

Metal–organic frameworks (MOFs) are a class of crystalline, porous materials synthesized via a modular approach from metal-based nodes and organic linkers.[33] Recently, MOFs have been used as encapsulating agents for biomacromolecules.[64–66,77] Specifically, the growth of zeolitic imidazolate framework-8 (ZIF-8), formed from $Zn^{2+}$ ions and 2-methyl imidazole (2mIM) linkers,[68,69] can be triggered by the presence of biomacromolecules in a facile process termed biomimetic mineralization,[67] which has been applied to the encapsulation of proteins, DNA, viruses and cells.[67,73–76] The

archetypal MOF for biomimetic mineralization, ZIF-8, is thermally stable and can be synthesized in biologically compatible conditions.[70–72] Encapsulation yields biomacromolecules with enhanced stability and protection from the surrounding environment, facilitates size-selective transport of enzyme substrates,[67,77,241] and allows for controlled release of cargo.[67,76] Furthermore, MOFs have been shown to act as carriers for therapeutic agents.[61,242,243] In summary, encapsulation of biomacromolecules inside MOFs offers many advantages for biotechnological and biomedical applications. However, to apply biomimetic mineralization generally, an understanding of how the encapsulation process occurs is needed.

The first report of biomimetic mineralization posited, based on experimental analysis, that the accumulation of MOF building blocks near bovine serum albumin (BSA) lead to enhanced kinetics for ZIF formation, but this analysis was limited to only one protein.[67] To further understanding of the biomimetic mineralization process, our experimental collaborators investigated a series of proteins[3] and carbohydrates (CHs)[4] for their propensity to seed ZIF-8 growth. In both cases, the experiments suggested that ZIF-8 formation requires the presence of a negatively charged seeding entity. Furthermore, modification of the charge on a biomacromolecule through chemical functionalization was found to be a general procedure for controlling the biomimetic mineralization process. However, experimental screening of biological entities for successful biomimetic mineralization is slow and expensive. Therefore, it is necessary to develop generalized methodologies and predictive techniques to expand the application of MOF-based biocomposites.

By solving the Poisson–Boltzmann (PB) equation, we modelled the interaction of proteins and carbohydrates with their surrounding electrolyte and confirmed that ZIF formation correlates with substantial enhancements of the zinc ion concentration near negatively charged biomacromolecules. The computational and experimental findings show that the calculated surface electrostatic potential and isoelectronic point (pI) of a protein can adequately predict biomimetic mineralization, which allows for fast screening before experimental testing. Finally, the computational results obtained from the distinct protein and carbohydrate models support the same general mechanism for biomimetic mineralization, which paves the way for further advances in the development of MOF-encapsulated biomacromolecules.

Herein, I present my contribution to two collaborative projects, for which I designed computational approaches to rationalize and build upon experimental findings. Below I include a brief description of the experiments that are crucial to the computational results; for a detailed summary, see Refs 3 and 4. In both cases, we carried out a systematic study of the growth of ZIF-8 in the presence of biomacromolecules, for conditions that do not lead to ZIF formation without the presence of a macromolecule.

## 5.3 Experimental results

### 5.3.1 Controlling biomimetic mineralization via protein surface functionalization

To understand the main features of a proteins surface chemistry that control the biomimetic mineralization process, the encapsulation of a range of proteins by ZIF-8 was studied. Table 5.1 shows the set of structurally distinct proteins that were screened for their ability to seed ZIF-8 growth under identical conditions ($0.5\,\mathrm{mg\,ml^{-1}}$ of protein dissolved in a solution of $1:4:278$ molar ratio of $Zn^{2+}$ : 2mIM : $H_2O$). All proteins were homogeneously dispersed and ZIF-8 formation does not occur without a seeding entity under these conditions. Importantly, this data shows that the biomimetic mineralization process is protein dependent. Furthermore, Table 5.1 suggests that proteins with low pIs are necessary to seed ZIF formation. A low pI is related to a higher proportion of acidic residues (aspartic acid side-chain $pK_a = 4.05$, and glutamic acid side-chain $pK_a = 4.45$),[244] which are negatively charged under these conditions (pH $\approx$ 11). Proteins with high pIs have a higher proportion of basic residues (lysine side-chain $pK_a = 10.0$, and arginine side-chain $pK_a = 12.0$),[244] which are positively charged under these conditions, leading to a repulsion of $Zn^{2+}$ ions and diminished ZIF-8 growth.

Table 5.1: Reported pI (pH at which the protein is uncharged), experimental zeta potential ($\zeta$) in a 2mIM solution at pH 11, and binary ZIF-8 growth result for each protein tested in this work. The yes/no descriptor for ZIF growth indicates the formation of a biocomposite with sodalite topology (determined by PXRD). Uncertainties are twice the standard error in the mean.

| protein | pI | $\zeta$ [mV] | ZIF-8 | modification | $\zeta$ [mV] | ZIF-8 |
|---|---|---|---|---|---|---|
| pepsin | 2.9[245] | $-30.9 \pm 1.4$ | yes | amination | $-7.9 \pm 0.6$ | no |
| BSA | 5.3[246] | $-36.4 \pm 1.4$ | yes | amination | $-5.8 \pm 0.2$ | no |
| lipase | 4–8[a][247] | $-31.7 \pm 0.3$ | yes | | | |
| catalase | 5.4[b][248] | $-30.4 \pm 0.6$ | yes | | | |
| HRP | 3.0-9.0[c][249] | $-36.4 \pm 1.0$ | yes | | | |
| haemoglobin | 8.1($\alpha$), 7.0($\beta$)[245] | $-21.0 \pm 2.4$ | no | succinylation | $-37.0 \pm 2.7$ | yes |
| | | | | acetylation | $-35.9 \pm 2.6$ | yes[d] |
| myoglobin | 7.6[245] | $-14.7 \pm 2.0$ | no | succinylation | $-36.6 \pm 0.2$ | yes |
| | | | | acetylation | $-36.1 \pm 3.6$ | yes[d] |
| trypsin | 10.7[245] | $-9.0 \pm 1.05$ | no | | | |
| lysozyme | 11, 11.3[245] | $6.6 \pm 0.2$ | no | | | |

[a] Broad experimental isoelectric region
[b] Computational value
[c] Seven isozymes
[d] Not phase pure

To systematically study the role of protein charge on ZIF formation, amino acid modifications were applied to a subset of proteins in Table 5.1. Amino acid modifications are used to enhance binding affinity during immobilization by altering electrostatic interactions.[250] Figure 5.1 shows the reaction schemes of the three surface modifications carried out to alter the charge of a protein. See Reference 3 for full experimental details. Succinylation and acetylation convert basic side-chains of lysine residues into acidic or non-ionizable groups, respectively, while amination converts acidic side-chains into basic groups. It was found that succinylation or acetylation of haemoglobin and myoglobin (corresponding to an increase of the negative charge of the proteins) enhanced ZIF growth in all cases, while amination of pepsin and BSA (corresponding to a decrease of the negative charge of the proteins) inhibited ZIF growth (Figure 5.2). In summary, there results suggested that amino acid modifications, which occur mostly on the surface, could be used to control the biomimetic mineralization process by modulating the surface charge of a protein.

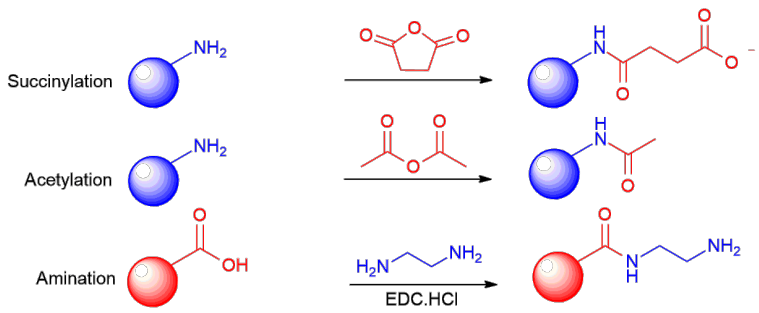



Figure 5.1: Surface modification reactions. Succinylation and acetylation reactions lower the pI of a protein by modification of exposed amine groups. Amination reactions cap carboxyl groups with a free amine, thus increasing the pI. (Source: Maddigan et. al., *Chem. Sci.* **2018**, *9*, 4217-4223.[3])

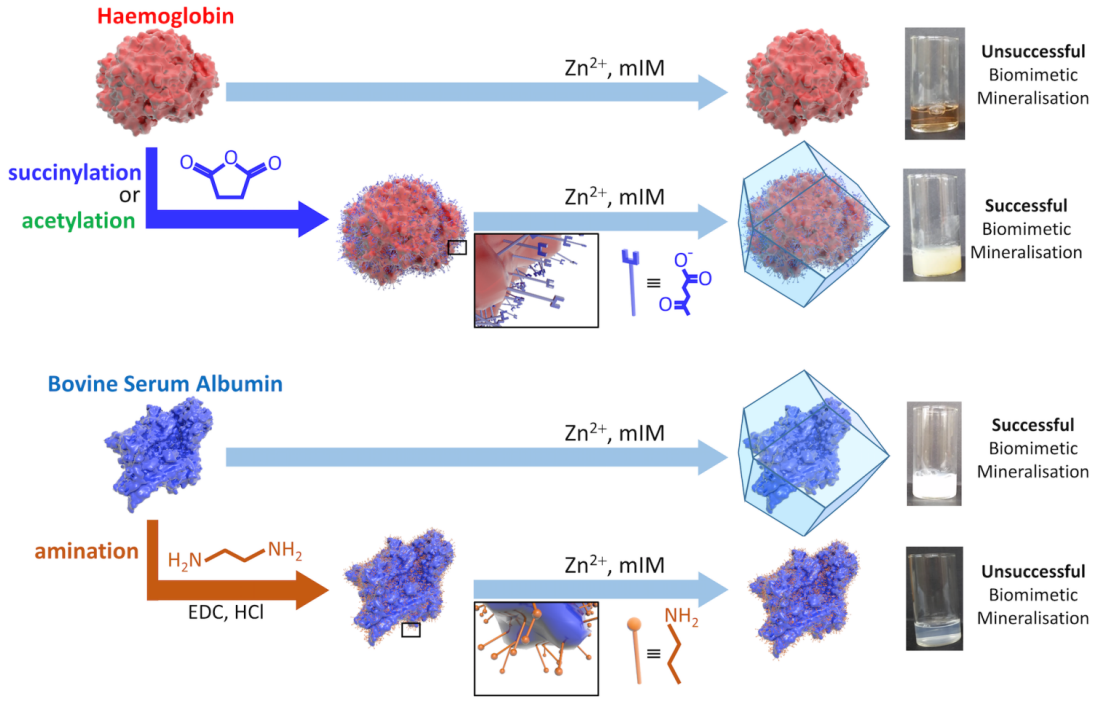



Figure 5.2: Schematic representations of the outcomes of biomimetic mineralization for two proteins, namely haemoglobin and BSA. Haemoglobin does not undergo biomimetic mineralization under standard conditions but can be chemically modified by acetylation or succinylation (shown) to increase the surface negative charge and facilitate ZIF-8 formation and encapsulation. BSA successfully undergoes biomimetic mineralization but amination introduces surface amine groups that are protonated under the conditions used for ZIF-8 formation and thereby prevent mineralization. (Source: Maddigan et. al., *Chem. Sci.* **2018**, *9*, 4217-4223.[3])

To quantify the role of the surface charge, the zeta potential ($\zeta$) of each protein (as well as the modified proteins) was determined in a 160 mM solution of 2mIM (pH $\approx$ 11) (shown in Table 5.1) and Milli-Q water (pH $\approx$ 7). Zeta potential measurements were

recorded in aqueous solutions using a Malvern Zetasizer nano and the Smoluchowski approximation.[3] The zeta potential is a measure of the electrostatic potential near the surface of a macromolecule in solution.[251] Table 5.1 shows that a zeta potential below c.a. −30 mV leads to ZIF-8 formation and that surface modifications can decrease (increase) a protein's zeta potential to below (above) this threshold to turn "on" ("off") biomimetic mineralization. These results show that amino acid modifications are an effective and general way to control biomimetic mineralization and that the zeta potential and pI are good predictors for ZIF growth. Experimental determination of the zeta potential and pI of a protein is slow and expensive. However, both of these properties can be calculated using theory, allowing for accurate and fast prediction of biomimetic mineralization.

### 5.3.2 Encapsulation of functionalized carbohydrates

In recent years, the importance of carbohydrates (CHs) to many biological functions has become more clear, and CH-based therapies have gained significant interest.[252–254] Inspired by the work described in Section 5.3.1, the ability of a range of carbohydrate molecules to induce biomimetic mineralization of ZIF-8 was investigated. By testing a series of mono- and polysaccharides, it was found that only carbohydrate chains with negatively charged functionalization lead to the reliable formation of ZIF biocomposites.[4] Importantly, the molecular weight of dextran polysaccharides was not found to impact the biomimetic mineralization process. Using the same experimental conditions as in Section 5.3.1, ZIF growth was tested in the presence of amino- (AM) and carboxymethyl-functionalized (CM) dextrans with similar molecular weight (10 kDa and 10−20 kDa, respectively). Only the solution containing CM-dextran showed rapid precipitation (Figure 5.3), which was confirmed by FT-IR to be a composite of CM-dextran and ZIF.[4] Again, these findings show that the presence of negatively charged macromolecules induces biomimetic mineralization.

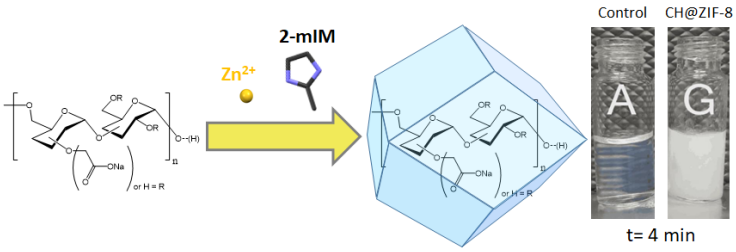



Figure 5.3: Schematic of the biomimetic mineralization of CM-dextran@Zn(2mIM)$_2$ according to the infrared investigation; photograph of a control sample (2mIM, Zn(OAc)$_2$ and water) and the same solution in presence of CM-dextran. (Source: Astria et. al., *Mater. Horiz.* **2019**, *accepted*.[4])

## 5.4 Computational methods

### 5.4.1 Calculation of the average charge of an ionizable residue

We calculated the average charge ($q_{\pm,i}$) of an ionizable residue of type $i$ in a biomacromolecule using the Henderson–Hasselbach equation,

$$q_{\pm,i} = \pm \frac{10^{\mp \mathrm{pH} \pm \mathrm{p}K_{\mathrm{a},i}}}{10^{\mp \mathrm{pH} \pm \mathrm{p}K_{\mathrm{a},i}} + 1}, \tag{5.1}$$

where $\mathrm{p}K_{\mathrm{a},i}$ is the $\mathrm{p}K_{\mathrm{a}}$ of the residue and the sign of $q_{\pm,i}$ is positive for positively charged residues (e.g. basic amino acids) and negative for negatively charged residues (e.g. acidic amino acids).

### 5.4.2 Calculation of the pI from protein sequence

For each protein, we extracted the sequence of natural amino acids from a FASTA file. We calculated the total protein charge as $Q = \sum_i N_i q_{\pm,i}$, where $q_{\pm,i}$ and $N_i$ are the average charge (Section 5.4.1) and number of ionizable residues of type $i$, respectively. The pH was varied until the total protein charge was within $0 \pm 0.0001e$ to determine the sequence pI. We used the Biopython module[244,255,256] to calculate the pI of an amino acid sequence, where the side-chain $\mathrm{p}K_{\mathrm{a},i}$ of all residue types were kept constant and defined in Ref. 244. Note that this method also considers the ionization of the N and C-termini of the polypeptide chain using Equation 5.1 and $\mathrm{p}K_{\mathrm{a}}$ defined in Ref. 244. The pI from this method is referred to as the 'sequence model' pI in this chapter

and we compare it to the pI calculated from a 3D model in Section 5.5.1.

### 5.4.3 Calculation of the average hydropathic index from protein sequence

The hydropathic index is a measure of an amino acid sequence's hydropathicity. Negative hydropathicities indicate an overall hydrophilic protein, whereas positive values indicate an overall hydrophobic protein. We calculated the hydropathic index for a protein sequence using the Biopython module[255,257] and the Kyte and Doolittle scale of residue hydropathicity,[258] which quantifies the hydropathicity of each residue. We reported the average hydropathy index for the entire sequence as a single value.

### 5.4.4 Surface modification of proteins

We calculated the pI of surface-modified proteins using Biopython[255] and assumed 100% efficiency of modification reactions on all target residues. Figure 5.1 shows the reaction schemes of the surface modification reactions used in this work. For the amination reaction, the p$K_a$ and charge of the functionalized aspartic acid or glutamic acid residues was taken to be that of lysine. For the acetylation and succinylation reactions, any lysine residues were either ignored in the calculation of the protein charge (acetylation) or their p$K_a$ and charge was taken to be that of glutamic acid (succinylation).

### 5.4.5 Calculation of electrostatic potential around a protein

We obtained all protein crystal structures used in this work from the Protein Data Bank[259] (PDB accession codes given in Table 5.2). Where available, we obtained a protein structure associated with the same organism as the experimental source. Each PDB file contains one or more polypeptide chain. We used only the first polypeptide chain in the PDB file for BSA because this protein is expected to exist as a monomer in solution. We used all chains in the PDB files in all other cases. We only modelled the polypeptide chain, i.e. we removed heteroatoms (ligands), bound ions or water molecules that were in the protein structures.

PROPKA 3.0[269,270] was used to estimate the p$K_a$ of each ionizable residue in each protein structure using a highly efficient, empirical method. PROPKA calculates the

Table 5.2: Protein Data Bank (PDB)[259] accession codes for all protein crystal structures.

| protein | PDB accession code |
|---|---|
| pepsin | 4pep[260] |
| bovine serum albumin (BSA) | 4f5s[261] |
| lipase, *Candida antarctica* lipase B (CALB) | 1tca[262] |
| catalase | 3re8[263] |
| peroxidase from horseradish (HRP) | 1w4w[264] |
| myoglobin | 2frf[265] |
| haemoglobin | 2dn2[266] |
| trypsin | 1s81[267] |
| lysozyme | 2vb1[268] |

shift in the p$K_a$ of an ionizable group in the 3D environment of a protein compared to that same group in water by calculating the difference between the free energies of deprotonation of the ionizable group in both environments. To do so in an efficient way, the 3D protein structure surrounding a given charge centre is treated as a small environmental perturbation on the water reference state of that charge centre. The total perturbation for a given ionizable group is taken as the sum of three terms: a Coulombic interaction term due to the charge-charge interactions between it and other charged groups in the protein, an energetic penalty for desolvation of the residue and an interaction term that accounts for non-Coulombic interactions with other atoms in the protein. The desolvation contribution of a charge centre is approximated by the volume of neighbouring atoms and their distance from the charge centre. All other interactions with the remainder of the protein (other than the Coulombic interactions) are approximated by two terms: a hydrogen bond term and a repulsive interaction term. Each of the above terms includes empirically fitted parameters, which were parameterized by comparing to experimental p$K_a$ values.[269] We have confirmed that similar results are obtained for the calculated p$K_a$'s using the more sophisticated DelPhi p$K_a$[271] to assign atom charges and protonation states (results not shown). DelPhi p$K_a$ uses a variable dielectric coefficient within the protein and the free energy difference between the protonated and deprotonated state of each ionizable residue within the 3D structure (using a PB-based approach to calculate the free energy difference) to obtain the p$K_a$ for each residue. The calculated p$K_a$ of each ionizable residue, given by PROPKA, was used to define the charge state of a 3D model (discussed below) of each protein and also to calculate the '3D model' pI using the Henderson-Hasselbach

equation.

Before analysing each crystal structure, we used the PDB2PQR software[272,273] to add missing heavy atoms, to make sure there were no overlapping atoms in the structure, to protonate the structure based on the p$K_a$'s calculated by PROPKA and the given pH and to assign atomic charges and radii from the AMBER[274] force field to each atom. We used PDB2PQR because it is efficient and commonly applied to prepare protein structures for electrostatic calculations.[149,275] Note that default parameters were used in general.[272,273] PDB2PQR protonates residues if their p$K_a$ is greater than the given pH, which is a crude method for setting the charge on a residue and will deviate from the total protein charge determined using the Henderson-Hasselbach equation (Equation 5.1), as in Section 5.4.2. Furthermore, we note that some protonation states derived from PROPKA are not supported by the AMBER force field, i.e. charge and radii parameters are not provided. In these cases the residue was treated as its pH 7 protonation state.

Using the SURFPOT module[150] within the DelPhi software[147] we solved the linearized PB equation,[148]

$$\nabla \cdot [\varepsilon(r)\nabla\psi(r)] - \varepsilon_0\varepsilon_r\kappa(r)^2\psi(r) = -\rho(r), \tag{5.2}$$

to calculate the electrostatic potential, $\psi(r)$, at position $r$. In Equation 5.2, $\rho(r)$ is the (fixed) charge density of the solute (protein), $\varepsilon(r)$ is the spatially varying dielectric permittivity, which is different in the protein and in the solution, $\varepsilon_0$ is the vacuum permittivity, $\varepsilon_r$ is the relative permittivity of water (80), and $\kappa(r)$ is the Debye screening parameter, given by

$$\kappa = \left(\frac{\varepsilon_r\varepsilon_0 k_B T}{2e^2 I}\right)^{-1/2} \tag{5.3}$$

for values of $r$ outside of the protein and taken to be zero for $r$ inside of the protein. $\kappa$ outside of the protein is the inverse of the Debye screening length $\lambda_D$, i.e. $\lambda_D = \kappa^{-1}$. In Equation 5.3, $e$ is the elementary charge, $k_B$ is the Boltzmann constant, $T$ is temperature, and $I$ is the ionic strength of the electrolyte solution, given by $I = \frac{1}{2}\sum_{i=\pm} c_{i,bulk} z_i^2$. Under the standard ZIF-8 synthesis conditions used throughout this work, the bulk zinc ion ($Zn^{2+}$) concentration ($c_{+,bulk}$) was 0.04 M and the concentration of counter ions ($c_{-,bulk}$), with a valency -1, was 0.08 M. Therefore, $\lambda_D = 8.8$ Å for all

of our calculations.

The zeta potential for each protein was estimated to be the average electrostatic potential on a surface at 4 Å from the van der Waals (vdW) surface of the protein. The zeta potential of a particle undergoing electrophoresis is defined by the electrostatic potential at the shear plane, which is not readily determined for heterogeneous and rough surfaces such as proteins. We expect the chosen surface at which the zeta potential was calculated to be a reasonable approximation for the shear plane and is similar to that used previously in the literature to estimate the zeta potential of proteins.[150] We used an interior protein dielectric coefficient of 4 and we confirmed that the average surface potential was not sensitive to this parameter within the range of 4–20 (results not shown), which agrees well with literature.[150] We note that the dielectric coefficient changes sharply at the interface between the protein and the solvent. We used a grid spacing of 0.5 Å, a probe radius (to define the protein surface) of 1.4 Å, which is equivalent to the radius of a water molecule, dipolar boundary conditions on the edge of the box, and a box size such that the longest dimension of the solute was 60% of the box size.

### 5.4.6 Calculation of electrostatic potential around a carbohydrate molecule

We modelled a dextran chain as an ion-permeable sphere with a radius given by the radius of gyration ($R_g$) of a freely-jointed chain of length $N$ segments, where $N$ was the degree of polymerization (DP), (Figure 5.4),[276] since single-molecule AFM studies of functionalized and native dextran chains have shown from their elasticity they they can be described approximately as freely-jointed chains with Kuhn lengths equal to the length of the glucose monomer (4.4 Å).[277,278] Although branching is common for dextran chains, the branching density of the experimental samples was unknown, so for simplicity we assumed that the dextran chains in our model were unbranched. Nevertheless, we do not expect the conclusions from the model to change significantly if branching were considered. We applied the ion-permeable sphere model (Figure 5.4) because most of the pervaded volume will be accessible to ions in solution (the average separation of monomers in the pervaded volume $\approx$ 5.7 Å, whereas the diameter of a $Zn^{2+}$ ion is 1.48 Å and the diameter of the acetate counter-ion is $\approx$ 3.72 Å).[279] Table 5.3 shows the parameters used.

Table 5.3: Parameters of the ion-permeable sphere model of dextrans.

| parameter | value |
|---|---|
| carbohydrate concentration | 0.72 mg mL$^{-1}$ |
| degree of polymerisation (DP) | 60 |
| Kuhn length ($l_k$) | 4.4 Å[277] |
| polymer molecular weight ($M$) | ≈10 kDa |
| radius of gyration ($R_g$) | 13.91 Å |
| bulk zinc concentration ($c_{+,bulk}$) | 0.04 M |
| p$K_a$ of CM | 4.0 |
| p$K_a$ of AM | 10.64[280] |
| pH | 11 |

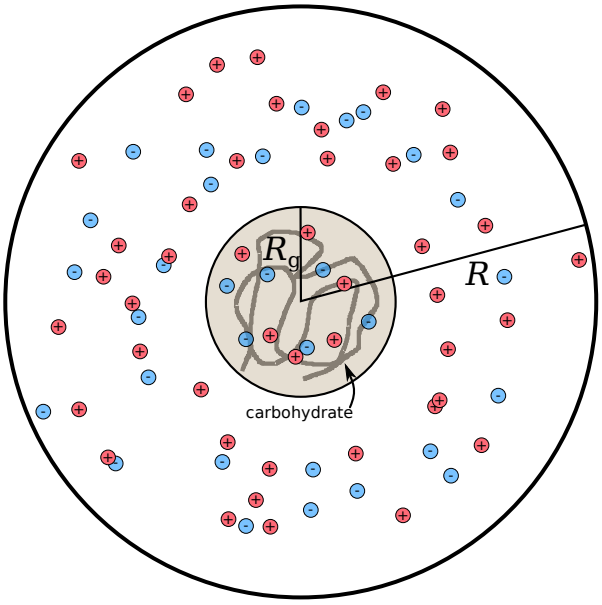



Figure 5.4: Schematic of the ion-permeable spherical model of a carbohydrate chain with a radius of gyration ($R_g$) in an electrolyte solution. The electrostatic potential was taken to be zero at a large distance, $R$, from the center of the ion-permeable sphere.

The average number of functionalized sites per glucose mononer is given by the degree of substitution (DS), $\alpha$, which can take any value from 0 to 3. DS was approximately 0.125 in the experiments, which corresponds to 1 functionalized group every eight monomers. We calculated the total charge ($Q_\pm$) on a dextran chain as $Q_\pm = \alpha N q_\pm$, where $q_\pm \equiv q_{\pm,i}$ is given by Equation 5.1 (only one residue type per dextran chain was considered) and the sign of $q_\pm$ is determined by the sign of the charge on the functionality (-1 for carboxymethyl (CM) functionalization and +1 for amine (AM) functionalization). A range of values have been reported for p$K_a$ of the CM functionality, but in all cases it is much smaller than the pH (≈11) used in the exper-

iments. Thus, each CM functionality is expected to have a full negative charge under the experimental conditions. Furthermore, there are multiple amine species associated with the AM functionality and it is expected that each amine will have a different p$K_a$. For simplicity, we have assumed a single p$K_a$ value, representative of a tertiary amine,[280] which will be partially positively charged at the pH used in the experiments. Figure 5.5 shows the volume charge density ($\rho_s = \frac{3Q_\pm}{4\pi R_g^3}$) of the carbohydrate chain as a function of DS for both functionalities with the specified parameters (Table 5.3).

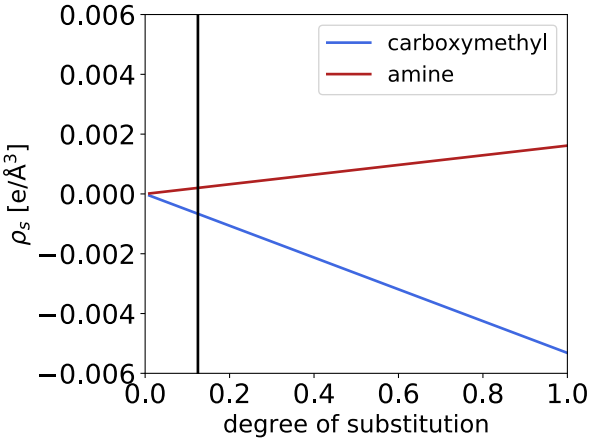



Figure 5.5: Volume charge density ($\rho_s$) of the ion-permeable spheres as a function of degree of substitution (DS) for carboxymethyl (CM) and amine (AM) functionalities. The black line represents the approximate DS used in the experiments.

Unlike the model described in Section 5.4.5, the ion-permeable sphere has radial symmetry; therefore, $r$ is simply the radial distance from the center of the ion-permeable sphere. Note that the electrolyte concentrations and Debye length are the same as in Section 5.4.5. We used the boundary value problem solver in the SciPy Python library[281] to solve the nonlinear PB equation around an ion-permeable sphere,[282]

$$\varepsilon_0 \varepsilon_r \nabla^2 \psi + e \left[ \sum_{i=\pm} c_{i,\text{bulk}} z_i \exp\left(\frac{-z_i e \psi}{k_B T}\right) \right] = -\rho(r), \tag{5.4}$$

where $\psi$ is the electrostatic potential at the radial coordinate $r$, and $\rho(r)$ is the charge density due to the carbohydrate chain, given by

$$\rho(r) = \begin{cases} \rho_s, & \text{if } 0 \le r \le R_g \\ 0, & \text{if } r > R_g \end{cases}. \tag{5.5}$$

All other quantities have been defined previously. Note that the entire system is assumed to be the same dielectric medium (i.e. $\varepsilon_r = 80$). The following boundary conditions were applied to Equation 5.4:

1.
$$\lim_{r \to 0} \frac{d\psi}{dr} = 0 \tag{5.6}$$

2.
$$\psi(r = R) \to 0 \tag{5.7}$$

$R$ was set to be a large enough value (100 Å) such that its specific value did not affect the calculated potential or ion concentrations near the ion-permeable sphere. We also applied the analytical solution to the linearized PB equation for an ion-permeable sphere reported by Ohshima and co-workers.[283] We note that the linearized PB equation is only applicable when $|\psi| \ll \frac{k_B T}{z_+ e} \approx 12\text{mV}$. Therefore, numerical solutions to the nonlinear PB equation were required in general.

For numerical stability, we replaced the step-function form of $\rho(r)$ in Equation 5.5 by the smooth function

$$\rho(r) = -\frac{\rho_s}{2}\left[\tanh\left(\frac{r - R_g}{2w}\right) - 1\right], \tag{5.8}$$

where $w$ defines the width of the transition of $\rho(r)$ from $\rho_s$ to 0. $w$ was set to $0.02\lambda_D$ and the final enhancement results were found to be robust to changes in $w$ in the range $0.01\lambda_D \leq w \leq 0.5\lambda_D$.

All code used for the ion-permeable sphere model is freely available at `https://bitbucket.org/andrewtarzia/sugar_source`.

### 5.4.7 Calculation of ion concentration and enhancement

In both the protein and carbohydrate models, the concentration of ions of type $i$ at position $r$ was calculated from the electrostatic potential ($\psi(r)$) using the Boltzmann equation,

$$c_i(r) = c_{i,\text{bulk}} \exp\left[\frac{-z_i e \psi(r)}{k_B T}\right]. \tag{5.9}$$

The zinc ion enhancement at coordinate $r$ was calculated from the zinc ion concentration ($c_+(r)$) as the ratio $\frac{c_+(r)}{c_{+,\text{bulk}}}$.

## 5.5 Results and discussion

### 5.5.1 Modelling the electrostatic properties of proteins

We applied the computational tools described above to verify the crucial role of electrostatic interactions in biomimetic mineralization of ZIF-8 and to provide an efficient and accurate means of predicting ZIF formation. We calculated an approximate zinc ion enhancement from the experimental zeta potential using the Boltzmann equation (Equation 5.9), assuming that the zeta potential was equal to the surface electrostatic potential. Figure 5.6a shows that a zinc ion concentration enhancement $> 10$ (or a zinc ion concentration $> 0.4\,\text{M}$) is required for ZIF formation. We calculated the average hydropathic index of a sequence for each protein in Table 5.1 (see Section 5.4.3 for details). Figure 5.6b shows that the hydropathy of a protein does not predict ZIF formation, suggesting that hydrophobic interactions do not play a crucial role.

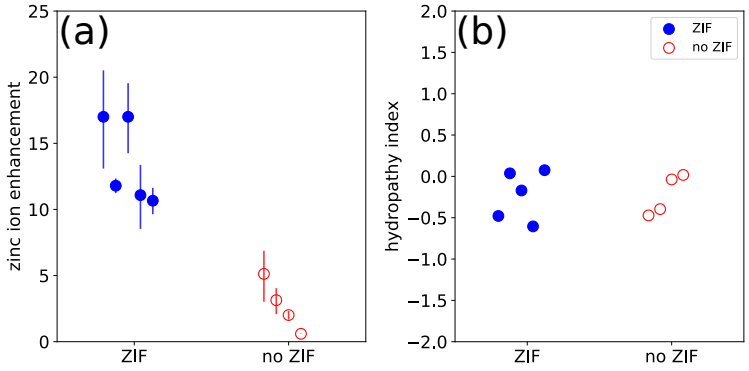



Figure 5.6: Categorical scatterplots of (a) the zinc ion enhancement calculated from the experimental zeta potentials at pH 11 and (b) the average hydropathicity of the peptide sequences for all unmodified proteins. Closed circles and the "ZIF" label denote proteins that form ZIF-8 and open circles and the "no ZIF" label denote proteins that do not form ZIF-8. Error bars in (a) were calculated from the standard errors reported in Table 5.1.

Table 5.1 shows that the pI of a protein is also a reasonable predictor for successful biomimetic mineralization. The pI can be calculated using a 3D model and the PROPKA software (described in Section 5.4.5)[269,270] or using a sequence model and the BioPython library (described in Section 5.4.2).[244,255,256] Figure 5.7a shows agreement

between the pI calculated using the 3D model and the sequence model and Figure 5.7b shows reasonable agreement between the pI calculated using the sequence model and the reported pI. Figure 5.7c shows that the pI calculated using the sequence model can adequately categorize the propensity of proteins to induce biomimetic mineralization for all proteins in Table 5.1: a pI of around 7 separates proteins that successfully undergo biomimetic mineralization from those that do not. For high-throughput screening of proteins for biomimetic mineralization, a sequence-based model is ideal because it avoids the need for a 3D structural model, which can be difficult to obtain and allows for straightforward predictions of the effect of amino acid modifications on the pI of a protein, as shown in Figure 5.8. We calculated the impact of amino acid modifications on the pI of a sequence by assuming 100% conversion of all target residues. We expect this approach to overestimate the calculated change in pI because it does not consider whether an amino acid is buried or not and assumes 100% reaction efficiency (see Section 5.4.4 for details). Nonetheless, Figure 5.8 shows that the calculated pI accurately predicts the effect of the amino acid modification on the ability of a protein to seed ZIF-8 growth. Furthermore, agreement with experimental pI values was obtained (e.g. aminated BSA has a pI $> 9.5$).[284]

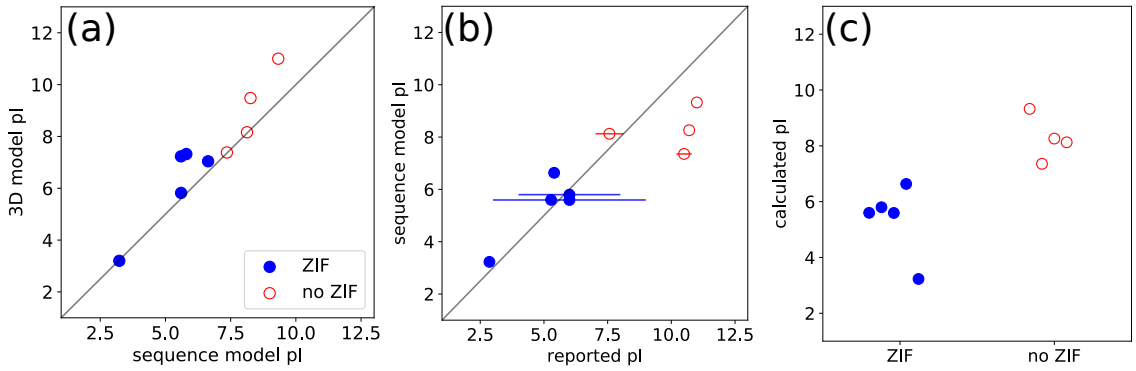



Figure 5.7: Parity plots comparing (a) the calculated pI values from the 3D model and the sequence model and (b) the calculated pI from the sequence model and the reported pI values for all unmodified proteins (the $y = x$ line is shown). Error bars in (b) represent ranges of pI values reported in Table 5.1. (c) Categorical scatter plot of the calculated pI from the sequence model of all proteins. ("ZIF" denotes ZIF formation and "no ZIF" denotes no ZIF formation.)

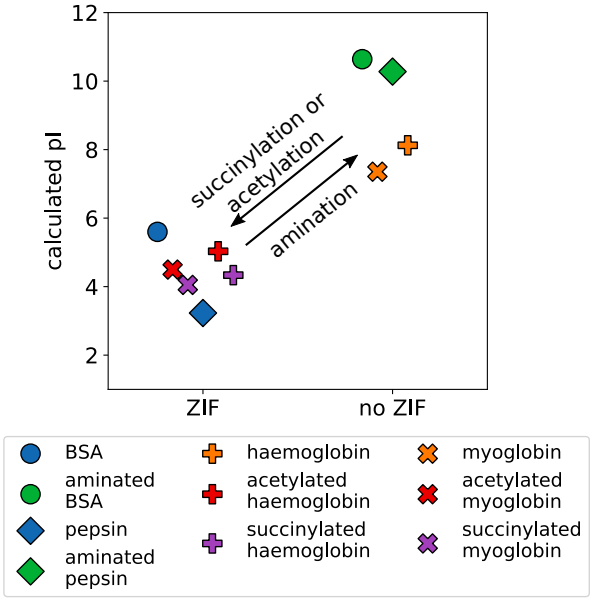



Figure 5.8: Categorical scatter plot of the calculated pI for BSA, pepsin, haemoglobin and myoglobin, with and without the surface modifications used in the experiments (Figure 5.2). The same protein is represented by the same marker. Arrows indicate the general outcome of each modification reaction. ("ZIF" denotes ZIF formation and "no ZIF" denotes no ZIF formation.)

We computed the electrostatic potential by solving the linearized PB equation around each protein in Table 5.1 to investigate its role in determining the ability of a protein to seed ZIF-8 growth (see Section 5.4.5 for details). The calculated average surface potentials from the 3D model of each protein show reasonable agreement with the experimental zeta potentials (Figure 5.9a–b). The average surface potentials for very highly charged proteins, such as BSA, catalase and pepsin, are overestimated compared with experimental zeta potentials, likely because of the use of the linearized PB equation, which breaks down in regimes of high zeta potential ($|\zeta| > \frac{k_\mathrm{B}T}{z_+e} \approx 12\mathrm{mV}$). Note that we did not use the nonlinear PB equation because it is much more computationally demanding than the linearized equation for complex 3D systems. The underestimation of the magnitude of the average surface potential compared with experimental zeta potentials at pH 11 for lipase and HRP is likely a result of differences between the experimental and calculated structures. Both proteins are expected to be glycosylated,[285,286] which is known to affect zeta potential measurements,[287] whereas the calculations used nonglycosylated structures. Additionally, HRP could be a mixture of different isoenzymes, which are enzymes that differ in peptide sequence but catalyze the same reaction, with vastly different electrostatic properties.[288] We note

that both proteins have reported pIs that span a broad range of values (Table 5.1), indicating a broad range of electrostatic properties for different samples.

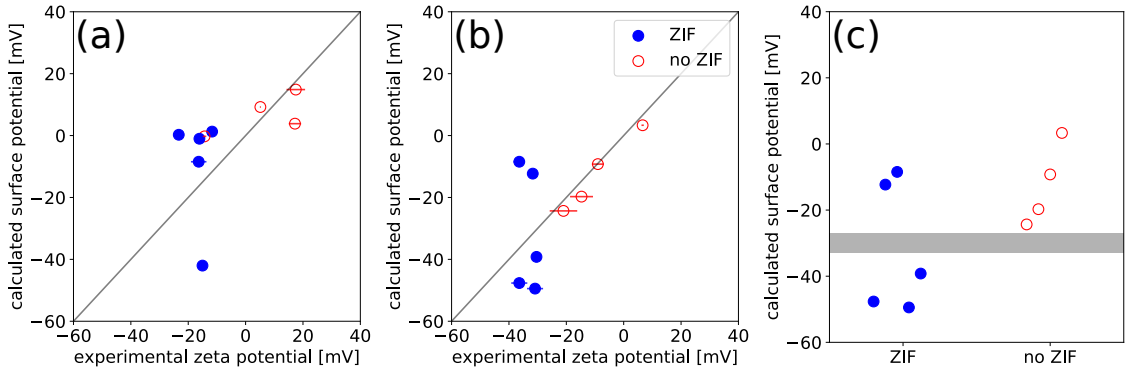



Figure 5.9: Parity plots comparing the calculated surface potential from our 3D model and experimental zeta potentials at (a) pH 7 and (b) pH 11 for all proteins (the $y = x$ line is shown). (c) Categorical scatter plot of the calculated surface potential from our 3D model at pH 11 for all proteins. Closed circles and the "ZIF" label denote proteins that form ZIF-8 and open circles and the "no ZIF" label denote proteins that do not form ZIF-8. The shaded region in (c) highlights the approximate boundary of the calculated surface potential that separates proteins that do and do not seed ZIF-8 growth.

In all of our calculations, we have approximated each protein by a static 3D model obtained from X-ray crystallography. We note that a static structure is unlikely to be representative of a protein structure in solution at pH 11. At high pHs, the presence of high-charge regions would lead to electrostatic repulsion and a degree of unfolding, which our model does not take into account. We also note that the interior of a protein has a highly variable dielectric coefficient and assuming a constant dielectric coefficient, as we have, can give rise to errors in the potential near the surface of proteins.[289] Furthermore, by taking the average surface potential to be equal to the experimental zeta potential for a heterogeneous protein surface we have assumed that the electric double layer surrounding the protein is thin compared with the size of the protein and that the linearized PB equation applies, which may not always be the case for the systems studied (discussed above).[142] The semiquantitative agreement with experiment in most cases in Figure 5.9a–b is encouraging, considering the approximations in the calculations. Finally, we emphasize that the efficiency of the linearized PB equation makes it more amenable to high-throughput computational screening than solving the full nonlinear equation for complex 3D models.

Figure 5.9c shows a categorical scatter plot of the calculated average surface poten-

tials at pH 11 for all proteins. These results support the experimental findings and show a reasonable ability to predict a protein's propensity to seed ZIF-8 formation. We show results at pH 11 because the initial solution (before adding zinc ions) in the biomimetic mineralization experiments is at approximately pH 11, but we note that upon addition of zinc ions, the pH quickly decreases to around 9, likely due to ZIF nucleation.[71] In addition, we found good agreement between the calculated surface potential at pH 11 and the experimental zeta potentials, which were measured at approximately pH 11 (Figure 5.9b).

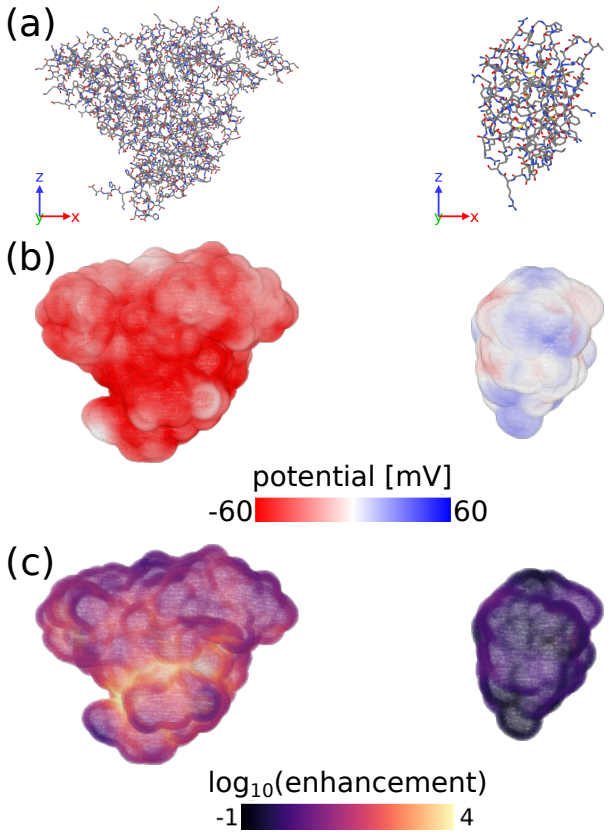



Figure 5.10: (a) Stick representations of protein crystal structures of (left) BSA and (right) lysozyme. Hydrogens are omitted for clarity. (b) Calculated surface potential and (c) log 10 of the zinc ion enhancement at the surface of both proteins. Structure and surface figures were made using OVITO.[187]

The calculated surface potential also provides comprehensive 3D information about the electrostatic interactions of the protein with the surrounding electrolyte solution. Figure 5.10 shows 3D maps of the electrostatic potential and zinc ion enhancement at pH 11 for BSA and lysozyme. These 3D maps clearly highlight the significant differences in the interactions with the surrounding electrolyte between proteins that do seed ZIF growth and those that do not.

### 5.5.2 Screening enzymes for biomimetic mineralization

Based on Figure 5.7c, we suggest that the pI of a protein calculated from its sequence can be used to accurately predict whether that protein will induce biomimetic mineralization of ZIF-8 or not. To highlight the broad applicability of this screening technique as well as biomimetic mineralization, we calculated the pI of all enzyme sequences with known Enzyme Commission numbers in the BRENDA database ($\sim 10^5$ sequences).[290,291] Figure 5.11 shows that the vast majority of enzyme sequences have pI values below the threshold for biomimetic mineralization. Therefore, there are many enzyme candidates that could be used. Furthermore, the effect of amino acid modifications on the pI could be predicted (Section 5.4.4) to determine if biomimetic mineralization is viable for proteins with pI values above 7.

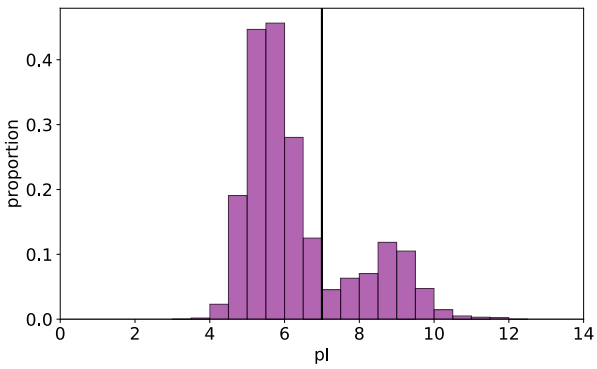



Figure 5.11: Normalized distribution of the pI for all enzyme sequences with known Enzyme Commission numbers from the BRENDA database.[290,291] The black vertical line highlights the threshold pI ($\approx$ 7) below which biomimetic mineralization is expected to occur, as determined from Figure 5.7c.

### 5.5.3 Modelling the electrostatic properties of carbohydrates

The experimental results in Section 5.3.2 support the hypothesis that electrostatic interactions drive the biomimetic mineralization process for carbohydrate molecules. To further understand these findings, we applied a computational model of the electrostatic potential and ion concentrations near dextran chains as a function of their charge. It is possible to modify the charge of a functionalized dextran chain by changing the degree of substitution (DS) of neutral hydroxyl groups with ionizable functional groups along the chain. In the experiments above, DS $\approx$ 0.125, which corresponds to about one ionizable group per eight glucose monomers. Unlike for proteins (Section 5.3.1),

the 3D structure of dextran chains is not well defined, but experimental measurements of their elasticity indicates that they can be approximated as freely jointed chains (see Section 5.4.6 for details).[277,278] To describe the electrostatics, we have treated dextran chains as ion-permeable spheres because most of the pervaded volume occupied by the dextran molecule is accessible to the ions in solution.[276] Therefore, to investigate the role of dextran functionalization on ZIF formation, we calculated the electrostatic potential and ion concentrations surrounding an ion-permeable sphere using numerical[282] and analytical[283] solutions to the PB equation, where positive and negative charges were added to a neutral dextran backbone as a function of DS (see Section 5.4.6 for details). Figure 5.12 shows the electrostatic potential surrounding AM and CM functionalized dextran chains for DS = 0.01, 0.04, 0.06, and 0.1 calculated using the analytical solution to the linearized PB equation and numerical solutions to the nonlinear PB equation. The numerical solution deviates significantly from the analytical solution for values of DS $\geq$ 0.04 for the CM functionalized dextran where $|\psi| > 12$mV, hence the linearized PB equation is not expected to apply (see Section 5.4.6 for details).

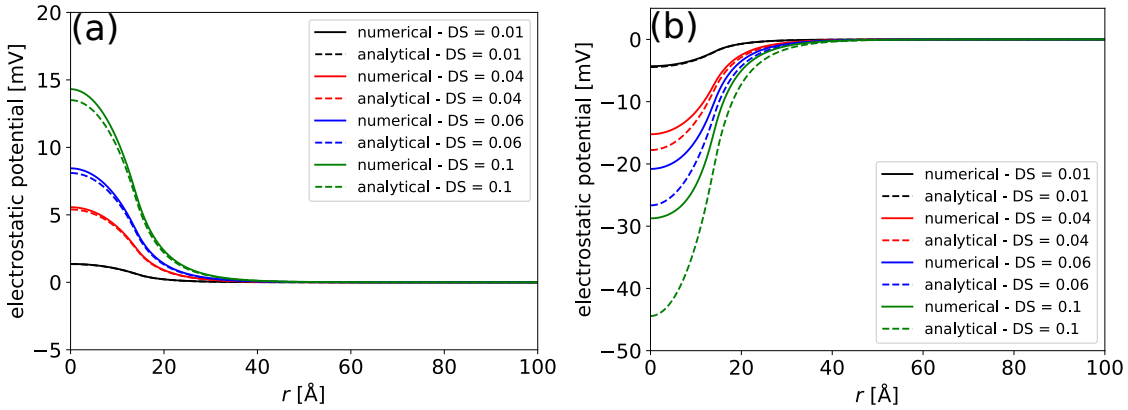



Figure 5.12: Comparison of the calculated electrostatic potential as a function of the radial coordinate $r$ for (a) AM and (b) CM functionalized dextrans for different values of DS using the analytical[283] (dashed lines) and numerical[282] (solid lines) solutions to the PB equation.

Figure 5.13 shows the electrostatic potential and zinc ion enhancement at the center of the ion-permeable sphere as a function of DS. The positively charged AM-dextran depletes zinc ions near the dextran chain, while the negatively charged CM-dextran chain enhances the zinc ion concentration. Importantly, at the DS used in the experiments (the black vertical line), the zinc ion enhancement is above the approximate threshold identified for the biomimetic mineralization of proteins in Section 5.3.1, where a zinc ion

enhancement ≈ 10 (calculated from experimental zeta potentials) was required for ZIF formation.[3] Figure 5.13 shows the electrostatic potential and zinc ion enhancement at the center of the ion-permeable sphere ($r = 0$) because the approximate size of the MOF precursors ($Zn^{2+}$ diameter = 1.48 Å and 2mIM diameter ≈ 5.2 Å) are smaller than the average separation of dextran monomers in the pervaded volume (≈5.7 Å). Therefore, it is possible that ZIF formation would occur anywhere within the pervaded volume of the carbohydrate chain. Thus, it can reasonably be expected that ZIF formation will be governed by the electrostatic potential and zinc ion enhancement at the center of the ion-permeable sphere where the zinc ion concentration is greatest, but we note that ZIF growth could be seeded at any point for $0 \leq r \leq R_g$. These findings, alongside those in Section 5.3.1, confirm the importance of a biomacromolecule's charge on the biomimetic mineralization process. They also show that the same threshold for the electrostatic potential governs successful encapsulation of distinct biomacromolecules.

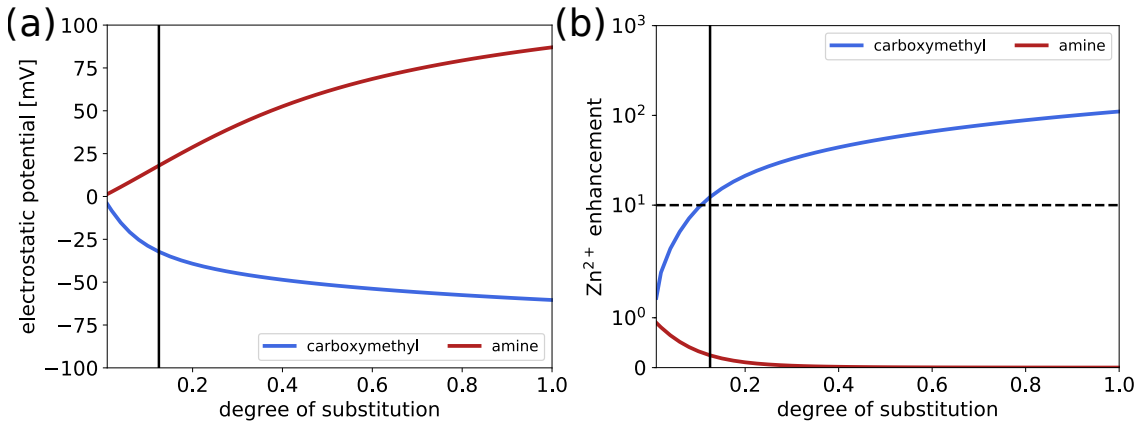



Figure 5.13: Calculated (a) electrostatic potential and (b) zinc ion enhancement at the center of the carbohydrate versus degree of carboxymethyl or amino functionalization.

## 5.6 Conclusions

In conclusion, by assessing the biomimetic mineralization of ZIF-8 by biomacromolecules under experimental conditions that usually do not lead to ZIF-8 growth, we have confirmed that the electrostatic properties of a biomacromolecule are the dominant factor in inducing ZIF-8 growth in water.[3,4] We showed that surface modification of proteins allows for modulation of their surface charge and control over the biomimetic mineralization process. Therefore, biomimetic mineralization may generally be applied using

this strategy. Experimental collaborators found that the zeta potential, which is related to a protein's surface charge, and the pI of a protein are good predictors for whether a protein will seed ZIF-8 growth.

To investigate the impact of the charge on a biomacromolecule on biomimetic mineralization at the molecular level, we modelled the surrounding electrostatic potential and ion distribution using the PB equation. For proteins, the average surface electrostatic potential, calculated based on a static 3D structure, was found to qualitatively agree with experimental zeta potentials at pH 7 and pH 11 and was able to predict ZIF-8 growth. By modelling the interactions between biomacromolecules and their surrounding electrolyte solution, we verified that the enhancement of the zinc ion concentration near negatively charged biomacromolecules drives biomimetic mineralization. Furthermore, we showed that the predictions made by a simple sequence-based model of the protein pI and a more physical 3D structure-based model of the surface electrostatic potential are similar and that both of these calculations agree with the experimental findings. Calculation of the pI of a protein is fast and general, allowing for the screening of proteins or enzymes for biomimetic mineralization before applying any experimental effort. By calculating the pI of $10^5$ enzyme sequences, we highlight the broad applicability of biomimetic mineralization because the majority of enzyme sequences have pI values below the threshold of 7.

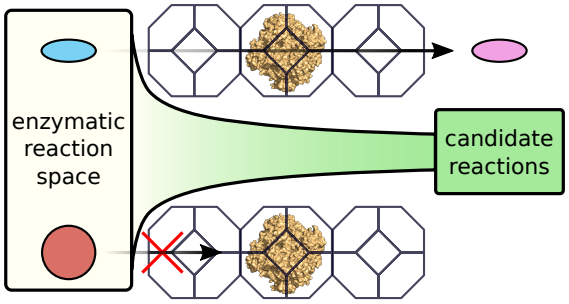
Inspired by these findings, we sought out to determine if they are transferable to other biomacromolecules. Experimental collaborators carried out a systematic study (varying synthesis conditions and carbohydrate (CH) properties) of CH-seeded biomimetic mineralization and found that only CHs functionalized with negatively charged groups reproducibly lead to ZIF-8 formation and CH encapsulation. We modelled a CH chain as an ion-permeable sphere and calculated the surrounding electrostatic potential and electrolyte concentration as a function of the degree of substitution of the CH chain and found that zinc ion enhancement increases significantly for small degrees of negatively charged functionalization of the CH backbone. By our calculations, the experimental degree of substitution on the CH chain produces a similar zinc ion enhancement as that determined from the experimental zeta potentials of proteins for which biomimetic mineralization occurs. These findings further solidify the role of electrostatic interactions and the enhancement of zinc ion concentration in seed-

ing crystallisation and suggest that biomimetic mineralization can be generalized to a broad range of biomacromolecules.

# CHAPTER 6

# Analysis of the enzymatic reaction space of enzyme@MOF biocomposites

This chapter contains unpublished work written in manuscript style.




## 6.1 Abstract

The encapsulation of enzymes inside metal–organic frameworks (MOFs) is a rapidly expanding field because of the potential to maintain enzyme activity in harsh industrial conditions. The ability to carry out an enzymatic reaction inside a MOF relies on the reactants and products being able to diffuse through the MOF pores. Therefore, to determine the viability of an enzymatic reaction inside a MOF, the molecular size of all components is calculated using a new, efficient approach that accurately matches measured kinetic diameters for a wide range of small molecules. Many of the enzymatic reactions tested in the literature in the MOF ZIF-8 were found to have molecules bigger than the proposed diffusion threshold for ZIF-8 (4–6.6 Å), suggesting that the substrates and products do not reach the encapsulated enzyme via the MOF pores but instead via defects in the MOF structure or that the reaction is occurring via surface-bound enzymes. Furthermore, the reported library of enzymatic reactions that have been carried out in ZIF-8 represents a small subset of all possible reactions. Hence, a screening process was developed to collect and classify all enzymatic reactions

available in an online database for viability in MOFs. A series of enzymatic reactions that allow for the systematic investigation of the impact of molecular size on activity in enzyme@MOF composites are suggested based on informatics descriptors of their molecular components.

## 6.2 Introduction

Metal–organic frameworks (MOFs) are a class of materials synthesized via a modular approach from metal-based nodes and organic linkers.[33] Recently, biomacromolecules have been successfully encapsulated within MOFs using a variety of approaches.[64,77] One such approach, termed biomimetic mineralization,[67,292] has been shown to encapsulate enzymes,[67] viruses[73,74] and cells[75,76] in a facile process that leads to enhanced stability of the biomacromolecules in harsh environments. To date, zeolitic imidazolate framework-8 (ZIF-8), composed of $Zn^{2+}$ ions and 2-methyl imidazole links,[68,69] is the most widely explored MOF for biomimetic mineralization because it is stable and can be synthesized in biologically compatible conditions.[70–72] Importantly, general approaches have been reported that can drive the encapsulation of a broad range of biomacromolecules.[3,293]

Enzymes are a class of protein whose function is to catalyze the biochemical transformation of a substrate to a product. Through encapsulation in MOF matrices (enzyme@MOF), enzymes are able to retain their activity in harsh conditions (e.g. elevated temperatures or proteolytic media) due to the protection afforded by the MOF matrix.[67,241,292,294–296] Improved durability is a crucial step toward the commercial application of enzymes as industrial catalysts.[250] The ZIF-8 matrix also affords size-selective transport of substrates to the active site of an enzyme via its pore network.[67,292] However, an essential and under-explored aspect of the application of enzymes inside ZIFs and other MOFs is the extent to which the pore network limits the mass transfer of substrates and products to and from the encapsulated enzyme.

A tremendous amount of research effort has gone into understanding the molecular sieving behaviour of ZIF-8 and other ZIFs because of their potential use in membrane and separation applications.[297–299] ZIF-8 allows for the diffusion of molecules significantly larger than its crystallographic pore aperture (3.4 Å) through rotations of the

organic linker and deformation of the pore window.[300–306] Gas-phase experiments have shown that molecules as big as 1,2,4-trimethylbenzene (kinetic diameter of 7.6 Å) were able to diffuse into ZIF-8, but the diffusivity of molecules this large was found to be slow.[307] A series of gas-phase experiments and simulations found that a molecule's diffusivity through ZIF-8 can be related to its kinetic diameter and decreases significantly for kinetic diameters around 4.0–4.5 Å.[300,307,308] This property of ZIF-8 gives rise to the kinetic separation of molecules of different size, e.g. propylene and propane.[309] Liquid-phase adsorption studies of ZIF-8 at room temperature (performed over 24h) have shown adsorption of *n*-hexane and benzene, but rejection of mesitylene,[310] which puts an approximate threshold for adsorption between 5.4–8.2 Å (based on the diameters of benzene[311] and mesitylene[312]). Furthermore, desalination experiments and simulations have shown complete rejection of ions by ZIF-8 membranes at room temperature, which narrows the diffusion threshold to approximately 6.6 Å based on the hydrated diameters of rejected ions.[298]

The experimental and simulation studies outlined above suggest that there exists an approximate size threshold (≈4.0–4.5 Å) where diffusion through ZIF-8 slows down by orders of magnitude.[300,307–309,313] Molecular size alone does not determine a hard limit for diffusion,[118] because diffusion depends significantly on many factors,[314] including (i) experimental conditions (such as temperature, pressure,[300] whether the adsorbate is in the gas or liquid phase[301] and adsorbate concentration[304,315]), (ii) the presence of defects,[316] (iii) crystal size,[317] (iv) pore aperture flexibility,[318,319] (v) molecular flexibility,[318–320] (vi) molecular shape[312,318–320] and, (vii) thermodynamic factors such as adsorbate-MOF interactions.[314,321,322] Nonetheless, the size of a molecule has been shown to approximate its ability to diffuse through ZIF-8 at a reasonable rate.[300,307,308] Importantly, a recent study of the diffusion of neutral and charged molecules in MOFs found a linear relationship between diffusivity and the relative size of the pores and adsorbate.[322] Hence, we have applied molecular size as the criterion to determine if an enzymatic reaction can occur inside crystalline ZIF-8 without significant mass transfer limitations.

During our analysis of the literature, we found that the set of enzymatic reactions that have been carried out inside ZIF-8 is a small subset of all possible enzymatic reactions. Therefore, we developed a computational screening process for viable enzymatic

reactions to assist the rapid expansion of the use of enzyme@MOF composites. The use of computational approaches to efficiently explore chemical space is crucial to multiple scientific fields, including materials science[17,87] and drug design,[323] because they offer a low-cost strategy to advance experimental endeavors. Our screening process extracts known enzymatic reactions from online databases and determines their viability for application inside MOFs. In this work, we used the size of the molecular components of a reaction, which can be related to their ability to diffuse through MOFs, as the primary determinant of viability. Multiple online databases that contain many thousands of possible enzymatic reactions exist (such as the KEGG,[324–326] BRENDA,[290,291] SABIO-RK,[327–329] and ATLAS[330] databases). Therefore, the ability to efficiently screen these databases for candidate reactions that can be expected to occur inside ZIF-8 and other MOFs without arduous and expensive experimental testing is important. We used our screening process to collect and analyze ∼4000 unique reactions from the KEGG database. From the resultant data set we suggest a series of reactions that could be used to test the impact of molecular size on the enzymatic activity inside MOFs.

## 6.3 Calculation of molecular size

We implemented an efficient algorithm to calculate the molecular size that correlates well with known kinetic diameters. Table C.2 shows the molecules and their kinetic diameters used to parameterize our algorithm. We chose this data set as these molecules have reported kinetic diameters[307,308,311] and a subset have been tested for diffusion through ZIF-8.[301,307,308,310,313,315,331–333] The algorithm effectively calculates the cross-sectional size of the van der Waals (vdW) volume of a given molecule and is similar to approaches used to determine the molecular size of adsorbates.[312] Our method differs in the use of multiple 3D conformers to efficiently account for molecular flexibility, which is crucial for modelling diffusion through porous materials.[320] We used RDKit (a cheminformatics Python toolkit)[334] to convert the SMILES string of a molecule into $N$ ($N = 100$ in this work) 3D conformers using the ETKDG algorithm, followed by energy minimization using the universal force field (UFF) (Figure 6.1a,b).[119,335] The ETKDG algorithm stochastically initializes atom positions in 3D space based on the bonding of the molecule and then uses empirical bond, angle and torsional-angle

information as well as a small set of basic structural constraints to build reasonable 3D conformers. We calculated a grid representation of the vdW volume of each conformer using RDKit (shown in Figure 6.1c). The grid is defined by a box margin, which sets the length that the grid extends beyond the positions of the atoms, and a grid spacing, which sets the distance between points in the grid (in the following section we discuss the choice of parameters used in this algorithm).

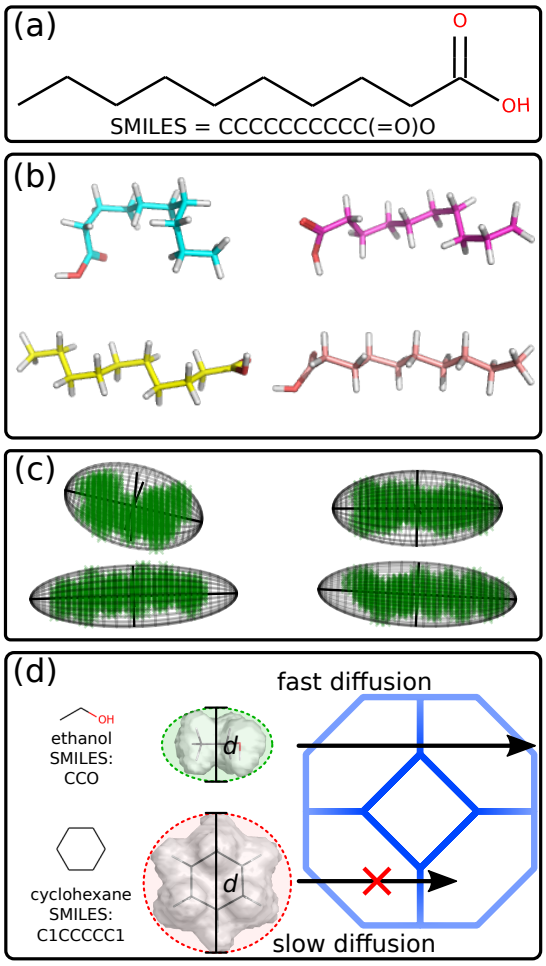



Figure 6.1: (a–c) Example calculation of the molecular size of *n*-decanoic acid from (a) a SMILES string (a 2D representation of *n*-decanoic acid is also shown). (b) Multiple 3D conformers of *n*-decanoic acid are generated (shown in distinct colors) and (c) the minimum volume ellipsoid (black grid) that encloses a grid representation of the vdW volume of each conformer (shown as green marks) is calculated. Atom positions are omitted for clarity in (c). (d) Schematic of the molecular size $d$ of ethanol and cyclohexane and its impact on diffusivity through ZIF-8 (represented by a sodalite cage). Example minimum volume ellipsoids that encompass the vdW volume of ethanol and cyclohexane (represented by gray clouds) are shown in green and red, respectively.[336,337]

We then calculated the minimum-volume enclosing ellipsoid (shown in black in Figure 6.1c) for the points in the vdW volume of each conformer using a minimization

algorithm based on the Khachiyan algorithm (we used a tolerance of 0.1 for the error in the solution with respect to the optimal value).[336,337] An ellipsoid is defined by its three principal diameters, where its second largest (or intermediate) diameter defines the smallest sized cross-section that is required to diffuse through the MOF pores (assuming cylindrical pores).[312] We assigned the size of a molecule from the minimum intermediate diameter ($d$) of all of its conformers. Our measure of molecular size accounts for molecular flexibility during diffusion[320] by assigning size based on the smallest possible value among all conformers, which is exemplified by the four different ellipsoids of *n*-decanoic acid in Figure 6.1c. Note that we extract the conformer with the smallest intermediate diameter, not the most stable conformer. We do not consider the charge state or solvation shell of a compound when calculating the molecular size, which could lead to an underestimation of $d$ for charged molecules. Finally, we applied a molecular weight limit of $500\,\mathrm{g\,mol^{-1}}$ because this method is not expected to apply for large, flexible molecules.[338]

The sensitivity of the calculated molecular size to the calculation parameters is studied in detail in Appendix C.2. Figure C.2 shows good agreement between the calculated molecular size $d$ and reported kinetic diameters for all molecules in Table C.2 using a box margin of 4 Å, a grid spacing of 0.5 Å, $N = 100$, and a vdW scale parameter of 0.8. Furthermore, Figure C.3 shows our methodology can accurately reproduce the relationship between self-diffusivity and molecular size. Nevertheless, discrepancies between the reported kinetic diameters and $d$ were found (∼1.0 Å) for flexible alkyl chains, such as *n*-heptane. The calculation of $d$ was designed to be efficient and does not consider many factors that may enhance or limit diffusion. For example, using $d$ alone correctly predicts that cyclohexane (6.2 Å) will diffuse easier than cyclooctene (6.9 Å),[339] but fails to predict the exclusion of *n*-decane (5.5 Å) and diffusion of aniline (6.3 Å)[340] or the exclusion of 2,2′-dimethylbutane (6.0 Å).[341] The failure of $d$ in these examples suggests that factors other than molecular size can determine diffusion through ZIF-8. Indeed, increased adsorbate-MOF dispersion interactions, compared to shorter alkyl chains, was found to impact *n*-decane adsorption inside ZIF-8.[342] This work does not consider substrate functionality and adsorbate-MOF interactions, which may have a significant impact on diffusion through the MOF pores. In general there is a trade-off between accuracy and computational efficiency, but the methodology

appears sufficient to approximate the kinetic diameters of small molecules accurately. Therefore, it was applied to approximate diffusion through ZIF-8 (Figure 6.1d).

## 6.4 Re-evaluation of previous enzyme@MOF measurements

Table 6.1 summarises the enzymatic reactions reported to have some degree of activity inside ZIF-8. We did not filter reactions based on the encapsulation method, but we did ignore reported reactions occurring at pH $< 6$ because ZIF-8 degrades under these conditions.[67] Table C.3 contains all substrates and products (that had to be able to diffuse through ZIF-8 for a reaction to occur) of all reactions in Table 6.1. Table 6.1 shows the Enzyme Commission (EC) numbers,[343] which classify an enzyme by the reaction it completes. EC numbers are represented by four numbers or levels, with the first level being the most general and the fourth level being the most specific. The EC numbers in Table 6.1 represent a small subset of possible EC numbers (15 of ∼7000), and hence, possible enzymatic reactions. Note that EC numbers have not been fully specified (to the fourth level) for many enzymatic reactions.[330,344]

Table 6.1: Reactions reported to be completed by enzymes encapsulated in ZIF-8 and their associated EC numbers. Similar reactions are grouped together.

| reaction | EC No. of enzyme |
|---|---|
| ethanol oxidation[296] | 1.1.1.1 |
| pyruvate reduction[296] | 1.1.1.27/1.1.1.28 |
| glucose oxidation[67,292,295,296,345–347] | 1.1.3.4, 1.1.5.2 |
| peroxide decomposition[67,77,292,294–296,346,348–350] | 1.11.1.5, 1.11.1.6, 1.11.1.7, 1.9.3.1 |
| lactate oxidation[346] | 1.13.12.4 |
| ester hydrolysis[77,350–354] | 3.1.1.3, 3.1.1.6 |
| lactose hydrolysis[296] | 3.2.1.23 |
| sucrose hydrolysis[346] | 3.2.1.26 |
| urea conversion[67,295,355] | 3.5.1.5 |
| production of 6-APA[348] | - |
| kinetic resolution[354] | - |
| catalase inhibition[294] | - |
| methylene blue reduction[296] | - |

As discussed above, the kinetic or molecular diameter of a molecule is related to its diffusivity through ZIF-8 and a significant decrease in diffusivity was found to occur for molecules with kinetic diameters in the range of 4–6.6 Å.[298,300,307,308,310] Therefore,

molecules with kinetic diameters greater than this threshold would be unlikely to diffuse through ZIF-8, severely limiting their viability for reaction inside ZIF-8. Based on these findings, we implemented an efficient descriptor of molecular size to predict the diffusion of small molecules through ZIF-8, and hence, the viability of a target reaction (see Section 6.3 for details). Figure 6.2 shows many molecules in Table C.3 are not expected to diffuse through ZIF-8 at a reasonable rate, based on their value of $d$. While the literature does not confirm complete exclusion of all molecules larger than 6.6 Å, it does suggest that mass transfer is limited for reactions that have components larger than this size. Nonetheless, there are alternate pathways to alleviate mass transfer issues, which may explain these findings. Firstly, the presence of enzyme on the surface of ZIF-8 may lead to activity, which would be indistinguishable from the activity of enzymes inside the MOF,[8,77] and would not be limited by mass transfer through the MOF pores. Experiments have confirmed the presence of enzymes inside and near the surface of ZIF-8.[77,293] We note that simply immobilizing enzymes on the surface of MOF particles, and surfaces in general, can lead to enhanced activity or stability.[356,357] Secondly, different preparation methods[72,347,354,358] may lead to differing degrees of ZIF crystallinity, which could lead to large interstitial diffusion pathways. For example, the decreased crystallinity of ZIF-8 formed via mechanical synthesis may explain the reported diffusion of *p*-nitrophenyl octanoate (6.7 Å); in this case, diffusion inside ZIF-8 was confirmed by infra-red spectroscopy.[354] Finally, recent reports have shown the degradation of ZIF-8 particles under certain reaction conditions,[359] which would result in large diffusion pathways through the MOF. These results suggest that careful characterization of the presence of surface-bound enzymes and ZIF crystallinity before and after measuring enzyme activity should be carried out to ensure that the reported activity originates from enzymes inside the MOF. Chen and co-workers recently reported enzymatic cascade reactions (with large molecular components) inside crystalline ZIF-8 (crystallinity was confirmed by SEM and X-ray diffraction), for which they thoroughly checked for surface-bound enzyme.[296] Their results suggest that an as-yet unknown factor may explain the enzymatic activity inside ZIF-8 for reactions with large molecular components.

Figure 6.2 shows that many enzymatic reactions that have been reported to occur inside ZIF-8 are likely to be limited by mass transfer and so the reported activity is

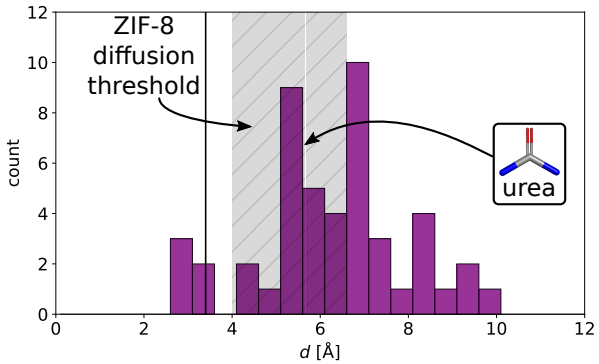



Figure 6.2: Molecular size $d$ of all molecular components of enzymatic reaction systems reported to occur within ZIF-8 (Table C.3). The vertical shaded region indicates the approximate range for the threshold for diffusion through ZIF-8 from the literature.[298,300,307,308,310] The black vertical line corresponds to the crystallographic pore diameter (3.4 Å) of ZIF-8. The 3D structure and molecular size of urea ($d \approx 5.2$ Å) is highlighted.

likely due to surface-bound enzymes or defects in the MOF structure. Enzyme@MOF reactions that require surface-bound enzymes or MOF defects to occur are unlikely to be useful in practical applications as surface-bound enzymes are likely to be denatured and MOF defects near encapsulated enzymes are unlikely to be straightforward to engineer in a reproducible fashion. These data suggest it will be important to systematically test the effect of molecular size on enymatic activity for enzymes that can be verified to be fully encapsulated in ZIF-8. To this end, we suggest two series of reactions from the reported reactions in Table 6.1. The first test examines the decomposition of peroxides (hydrogen peroxide, methyl ethyl ketone peroxide and tert-butyl hyperperoxide) of increasing size (Figure 6.3a).[77] The benefit of this series of reactions is that many enzymes show peroxidase activity and multiple assays exist for monitoring peroxide decomposition.[77] The second series of reactions examines *p*-nitrophenyl ester hydrolysis, which is commonly used as an assay for lipase activity.[77,350–354] Figure 6.3b shows the ester substrates and acid products that have been reported, which represent a systematic increase in molecular size, flexibility and hydrophobicity. Furthermore, lipases are generally stable to immobilization.[360] Importantly, the diffusion of *p*-nitrophenyl octanoate (C8) through mechanically synthesised ZIF-8 was confirmed by infra-red spectroscopy.[354] In both cases, a broad range of enzymes can be used, and the molecular properties can be systematically varied.

The values of $d$ in Figure 6.3 span values for which mass transfer is predicted to limit or not to limit the reactions. Therefore, the presence or absence of enzymatic

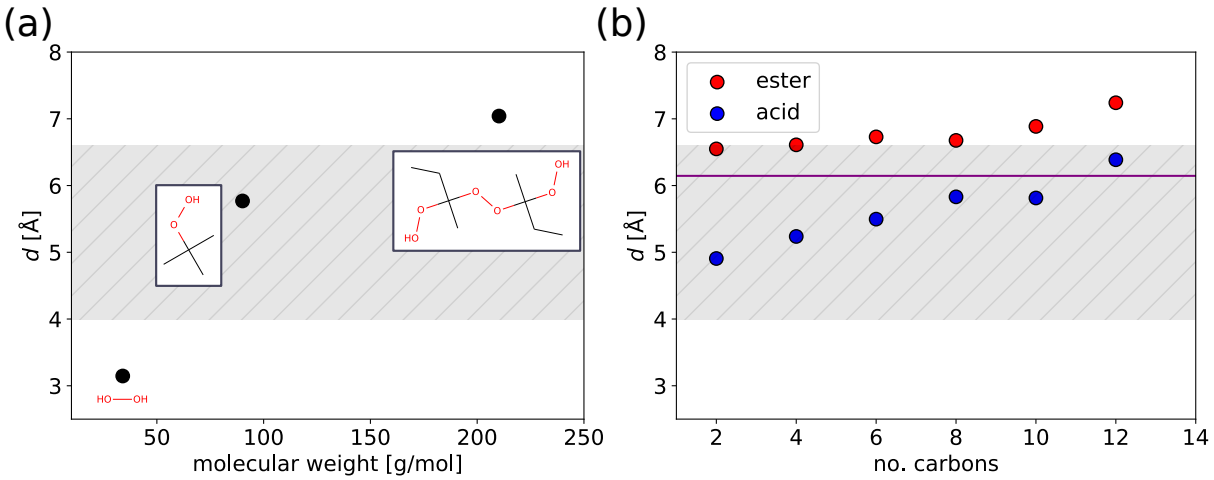



Figure 6.3: Molecular size $d$ calculated for (a) hydrogen peroxide, methyl ethyl ketone peroxide and tert-butyl hyperperoxide as a function of their molecular weights (chemical structures are also shown) and (b) the ester substrates and acid products of a series ester hydrolysis reactions as a function of the number of carbons in the acid chain. The purple line in (b) indicates $d$ of *p*-nitrophenol, which is another product of all of the ester hydrolysis reactions. The gray region indicates the approximate range for the threshold for diffusion of molecules through ZIF-8 from the literature.[298,300,307,308,310]

activity in ZIF-8 for these reaction series provides a sensitive test of molecular-size limitations on enzyme@MOF reactions. Based on the molecular-size threshold for diffusion in ZIF-8, the reaction series in Figure 6.3b may not show significant activity because all of the reactants (the esters) are near or above the threshold. We note that small increments in the pore size of ZIF-8 can be achieved using solvent-assisted linker exchange (SALE), which can increase the pore size of ZIF-8 by ~1 Å.[361] By using SALE, for example, it may be possible to monitor the impact of ZIF pore size on enzymatic activity for this series of reactions. The relationship between the molecular size of the substrates and products of a reaction, the MOF pore size and enzymatic activity remains unclear, but an understanding of this relationship is crucial for the design and control of enzyme@MOF composites in the future. In the following sections we describe a screening process that will aid in the selection of viable enzyme@MOF reactions, for which we have assumed the size of the molecular components to be the critical factor determining reaction viability. From our screening process we suggest further reactions that could be used to understand the relationship between molecular size and enzymatic activity in MOFs.

## 6.5 Identification of new enzyme@MOF reactions

Progress toward a better understanding and improved control over the encapsulation of enzymes in MOFs is being undertaken[293,362] and, in general, conditions that lead to successful encapsulation can be found[64,73,74,295] through strategies such as amino acid modification.[3] Nonetheless, Table 6.1 shows a limited scope of enzymatic reactions that have been tested inside ZIF-8. The cost of purified enzymes and the time required to prepare and screen new enzymes is significant. Hence, accurate predictions of which reactions may be possible inside a given MOF before doing costly experiments are valuable. To this end, we have designed an efficient screening process to explore the known enzymatic reaction space (Figure 6.4). The screening process uses molecular information from databases of biochemical transformations to determine the viability of a given reaction and suggest the best candidates for further experiments. The KEGG database[324–326] of ∼11 000 curated enzymatic reactions was used to compile a list of possible reactions with known molecular components.

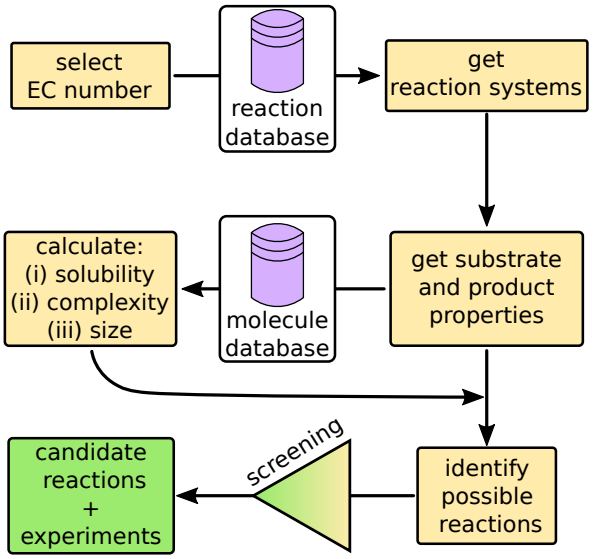



Figure 6.4: Flowchart of the process used to screen enzymatic reactions in MOFs. A series of reaction systems are collected for all EC numbers from a database of enzymatic reactions (the KEGG database was used).[324–326] The molecular structures of the components of each reaction system are extracted and analyzed to provide a list of possible reactions, which are then screened for target properties.

Based on the importance of molecule size in determining diffusivity in ZIF-8, we used the size of the largest component ($d$) of each reaction as the first check for vi-

ability. Figure 6.5a shows the distribution of the maximum component size of all reaction systems collected from the KEGG database and the cumulative number of reactions as a function of the maximum $d$ of the components of each reaction. We excluded reactions with components with molecular weights $> 500\,\mathrm{g\,mol^{-1}}$, which explains the plateau in Figure 6.5a around 10 Å. We found 36 reactions in this data set with maximum component sizes $< 4.5$ Å, where mass transfer should not be limited in ZIF-8.[300,307,308] The number of reactions increases dramatically (up to ∼1000) near the approximate diffusion threshold of ZIF-8, indicating that many new reactions may be possible with current experimental methodologies. Nonetheless, Figure 6.5a shows that the small pore aperture of ZIF-8 is a limiting factor in the utility of encapsulated enzymes. Figures 6.5a and C.4 show that the distributions of the maximum component size are centered around 7–9 Å. Therefore, it is important to target ZIFs with pore-limiting diameters around that size. The pore size and chemistry of ZIF materials can be systematically controlled through stiffening the framework using electric fields,[363,364] the formation of mixed-linker ZIFs,[333,365,366] structuralization[367,368] and linker exchange (during and post-synthesis).[313,331,361,363,369–372] Furthermore, there are many experimentally known ZIF structures that have been reported with larger pore apertures (up to ≈13 Å).[68,69,373–375] The vertical dashed line in Figure 6.5a highlights the number of unique reactions (∼4000) accessible to experimentally realized ZIFs with larger pore apertures (the largest having a pore-limiting diameter of 13.1 Å).[373] Encapsulation of enzymes in other ZIFs has been reported, which suggests that the chemistry is transferable.[8,77,241,294] Figure 6.5a shows that even small increases in the pore size can lead to significant expansion of the reaction scope of enzyme@MOF composites. Methods such as SALE,[361] that allow for the systematic modification of pore sizes of ZIFs provide another means of studying the effect of molecular size without changing the enzymatic reaction being studied. Our results indicate that an increase in pore size by ∼1 Å, which can be achieved using SALE,[361] will allow access to hundreds of new reactions.

Alongside the maximum component size, we calculated a series of properties of each reaction system that are expected to be relevant for high-throughput screening of enzymatic reactions in MOFs. The aqueous solubility (logS) and hydrophobicity (logP) of each component is crucial for the experimental viability of a reaction in aque-

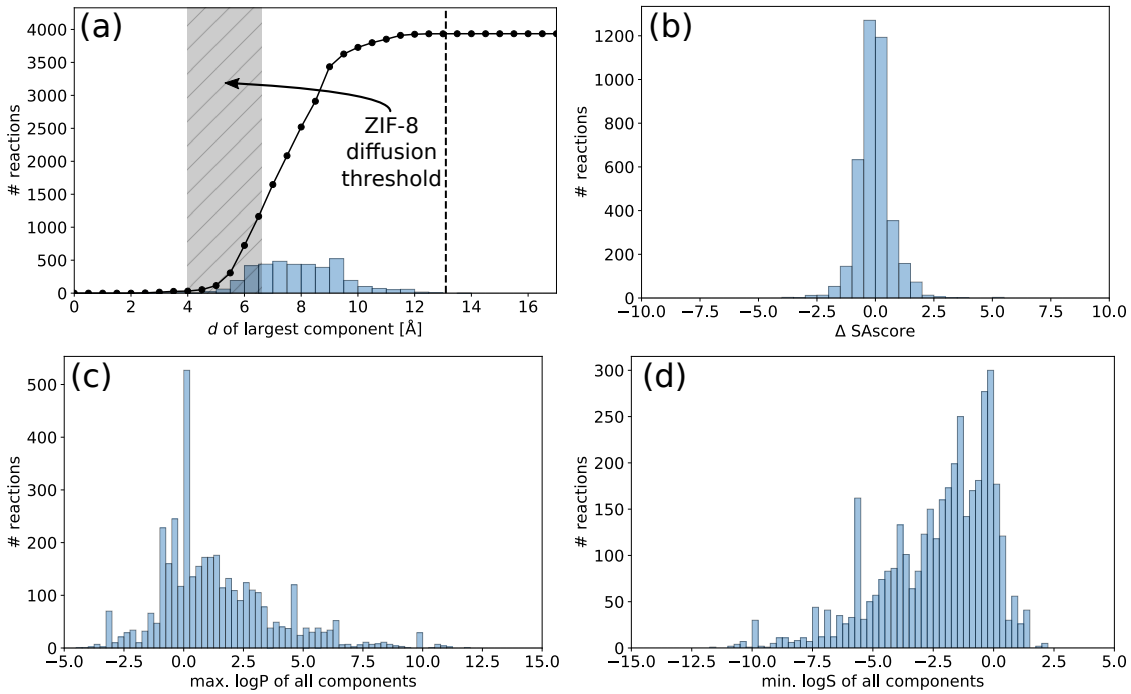



Figure 6.5: Distributions of the (a) molecular size $d$ of the largest component, (b) change in SAscore, (c) logP of the most hydrophobic component and (d) logS of the least water soluble component for all reaction systems collected. The black curve in (a) shows the cumulative number of reactions extracted from the KEGG database[324–326] as a function of the $d$ of the largest molecular component of each reaction. The hatched region in (a) indicates the approximate range for the threshold for diffusion through ZIF-8 from the literature.[298,300,307,308,310] The vertical dashed line in (a) represents the largest pore-limiting diameter of any experimentally realized ZIF structure.[373,374] Only unique reactions are shown in all distributions.

ous solution.[376] Experimental evidence suggests that the hydrophobicity of ZIF-8 does not inhibit the diffusion of water through its pores.[297,298] Therefore, we consider the aqueous solubility of the reaction components to be a more critical factor, because all reported reactions, so far, have been done in water. Nonetheless, increased hydrophobicity of the MOF pores relative to water could potentially limit the diffusion from water into the MOF or vice versa if a molecule is extremely hydrophilic or hydrophobic, respectively (ignoring the impact of hydration and surface tension). Alternatively, less hydrophobic ZIFs (i.e. MAF-7) could be used.[8,371] We used the AlogP98 method (implemented in RDKit and termed "logP" in this paper) to determine the hydrophobicity (defined by the octanol-water partition coefficient) of a molecule.[334,377,378] This method defines logP as a sum of contributions from all of its atoms, taking into account intramolecular interactions based on an atoms neighbours.[377] We predicted the aqueous solubility of a molecule (represented here by logS) from a molecules logP and

structure (e.g. molecular weight, number of rotatable bonds and the proportion of atoms in the molecule with aromatic bonds).[376] We applied an updated version[379] of the ESOL approach[376] implemented using RDKit.[334]

The value added by a reaction, based on the molecular complexity of the products compared to the reactants, is another useful descriptor for screening candidate reactions. Synthetic accessibility (given by a synthetic accessibility score, SAscore) is a cheminformatic estimate of how difficult it would be to synthesize a molecule.[380] SAscore takes into account molecular complexity and the "historical synthetic knowledge". Molecular complexity has been defined in multiple ways (see Appendix C.5 for details), but in the calculation of the SAscore, complexity is defined by the size of the molecule and the presence of complex fragments (such as macrocycles). Historical synthetic knowledge is taken into account by scoring each fragment present in a molecule, where a fragments score is determined by its prevalence in the PubChem database (fragments that appear with frequently in molecules in the PubChem database are assumed to be more synthetically accessible).[380] The SAscore is a number ranging from 1 to 10, where a value of 10 implies that a molecule is difficult to synthesize. A high value-added reaction would take an easy to make compound (lower SAscore) and produce a difficult to make compound (higher SAscore). Interestingly, Figure C.5 shows that biological transformations do not often induce large changes in synthetic accessibility. Importantly, we can calculate the SAscore, as well as the logS and logP of each molecule efficiently using RDKit.[334]

The impact of encapsulation in MOFs on enzyme structure is still an open question,[8,362] but the immobilization of proteins on surfaces and on/in MOFs has been shown to lead to conformational changes from their native structures, which may lead to a loss or gain in activity.[352,381] The encapsulation process may also impact enzyme activity by interrupting the binding of co-factors or resulting in MOF building blocks inhibiting the enzyme active site. Therefore, it is crucial to undertake an experimental analysis of the integrity of an enzyme before and after testing bioactivity. Many factors, including the ability of an enzyme to retain its native configuration, determine the activity of immobilized enzymes.[360,382] Therefore, predicting the bioactivity of an immobilized enzyme is not trivial and still not fully understood. Thus, we have not considered this factor in our screening process but this would be a crucial step to

screening candidate enzymatic reactions.

Figure 6.5b–c show distributions of the change in SAscore, logP of the most hydrophobic component and logS of the least water soluble component of all reaction systems collected (we further analyze these data in Appendix C.5). Note that we did not use the change in SAscore in the following analysis because the KEGG database does not specify the directionality of its reactions (we manually determined the directionality of the final set of candidate reactions). We designed an example process that uses the above data to screen the ∼4000 candidate reactions (Figure 6.5) for a reaction that adds value and can be monitored in situ to probe enzymatic activity inside a MOF. Figure 6.6 shows the applied screening process. Firstly, we extracted reactions with maximum component sizes ($d$) in the range of 5–7 Å. This size regime is at the diffusion threshold for ZIF-8 and ∼1–2 Å above the size at which diffusivity was found to decrease by orders of magnitude (≈4–4.5 Å).[300,307,308] Therefore, the impact of molecular size on enzymatic activity can be tested. Furthermore, because the maximum component size is within ∼1–2 Å of the diffusion threshold, it is possible to apply SALE[361] to monitor the change in activity as a function of the change in pore size. 1233 reactions remained after this first step. Secondly, we extracted the reactions that included components with a nitrile group on at least one side of the reaction (determined by the RDKit fragment classifier).[334] Nitrile groups have been used as probes in protein systems because of their distinct infra-red signal (≈2200 $cm^{-1}$) in a region that is relatively uncluttered by protein backbone signals.[383] Therefore, the presence of a nitrile signal may allow for the reaction to be followed in situ by infra-red spectroscopy. From 33 reactions, we extracted 21 reactions for which the least soluble and most hydrophobic molecule was at least more soluble and less hydrophobic, respectively, than $n$-pentane (based on the calculated values of logS and logP of $n$-pentane, which is an example of a solvent that is immiscible in water). Finally, we manually determined the directionality and value added (defined by the cost of the substrates and products) of each reaction using the primary literature available in the KEGG database and the ZINC database of purchasable compounds.[81]

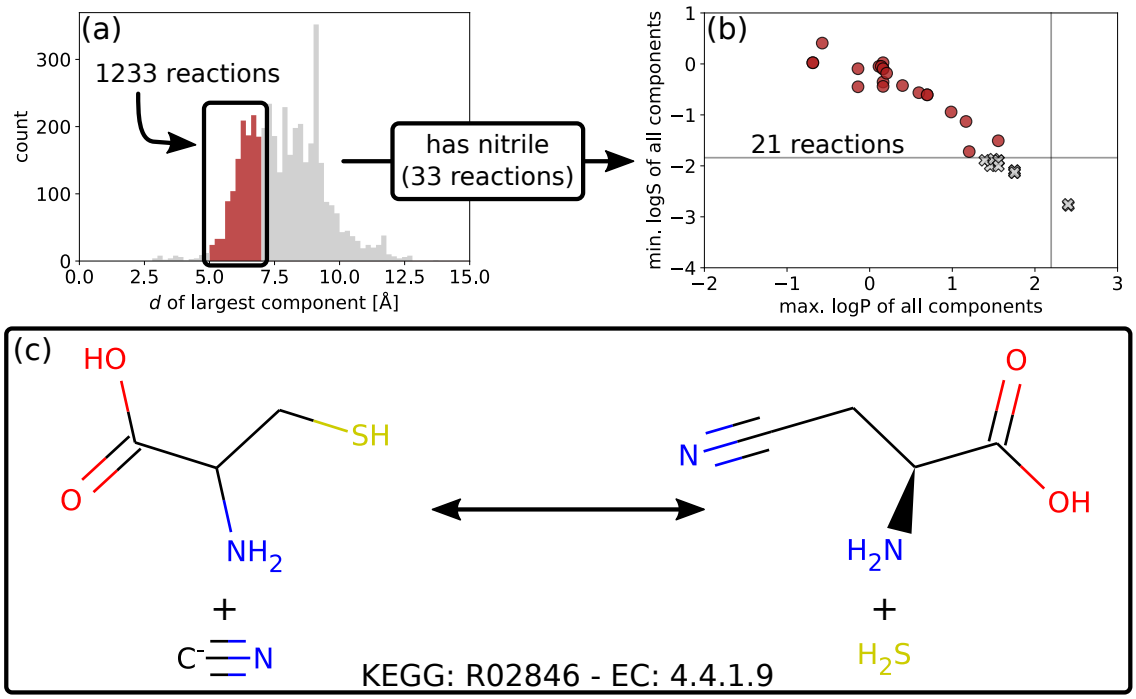



Figure 6.6: An example application of our screening process to identify novel enzyme@MOF reactions. (a) Reactions are filtered by the $d$ of their largest component (reactions in the red region of the histogram are extracted), the presence of a nitrile functionality on at least one side of the reaction and (b) the logP and logS of their most hydrophobic and least water-soluble component, respectively. Reactions (shown by the gray crosses) with maximum logP or minimum logS values above or below, respectively, that of *n*-pentane (shown by the vertical and horizontal gray lines, respectively) are excluded. (c) A candidate reaction, with its KEGG reaction identifier, EC number, and molecular components given.

Figure 6.6c shows the selected enzymatic transformation of cysteine (a natural amino acid, $\sim\$5\,g^{-1}$)[384] to 3-cyano-*L*-alanine ($\sim\$300\,g^{-1}$),[384] which is carried out by $\beta$-cyano-*L*-alanine synthase (with an EC number 4.4.1.9).[385,386] $\beta$-cyano-*L*-alanine synthase requires pyridoxal 5′-phosphate as a cofactor to carry out the reaction in Figure 6.6c.[385] We note that any cofactors, which can be relatively large molecules such as pyridoxal 5′-phosphate, should be considered when determining the viability of a reaction. However, many cofactors remain bound to the enzyme during a reaction. Therefore, through careful preparation, any required cofactor could be bound to the enzyme before encapsulation, which avoids the need for their diffusion through the MOF pores. We note that whether or not the encapsulation process interrupts the binding of cofactors would need to be confirmed by experiments. In this work, we have collated all molecular components of each reaction as present in the KEGG database,[324–326] assuming that any other required cofactors may be introduced before encapsulation.

While the change in SAscore for this reaction is relatively small (0.26), the fact that the substrate is an abundant amino acid may explain why there is such a price discrepancy. The infra-red signal of the nitrile group in the product ($\approx$2200 $cm^{-1}$) will differ enough from the signal of the cyanide ions in the reactants ($\approx$2050 $cm^{-1}$)[387] such that it will be possible to monitor reaction progress in situ.[383] Figure 6.6 is an example of an efficient screening process for new and interesting enzymatic reactions to test inside MOFs. In future, the cost of substrates and products could be considered automatically using the ZINC database to speed up this process further.[81] Furthermore, screening for other functionalities and molecular fingerprints using RDKit[334] is possible and larger data sets of enzymatic reactions (e.g. the ATLAS,[330] BRENDA[290,291] or SABIO-RK[327–329] databases) could be used to broaden the diversity of these results.

## 6.6 Conclusion

The encapsulation of enzymes in MOFs allows for enhanced stability in harsh environments, which could expand their use in industrial applications. ZIF-8 allows for the diffusion of molecules much larger than its crystallographic pore aperture.[307] An analysis of the literature puts a diffusion threshold for ZIF-8 around 4–6.6 Å,[298,300,307,308,310] with diffusion (in the gas phase) decreasing by orders of magnitude (compared to smaller molecules) for molecules with kinetic diameters around 4–4.5 Å.[307,308] Therefore, the kinetic diameter could be used to estimate diffusivity through ZIF-8.[300,307,308] Based on these results, we introduced an efficient measure of molecular size to determine the likelihood that a component can diffuse through ZIF-8 (Section 6.3). The method was found to agree with reported kinetic diameters (Figure C.2) using a stochastic description of the conformational space of each molecule. Therefore, we applied this simple descriptor of molecular size as a first approximation for the likelihood of diffusion through ZIF-8.

We have shown that many reactions reported to occur inside ZIF-8 (Table 6.1) are expected to be strongly limited by mass transfer through the pore network due to the size of the molecular components and therefore are unlikely to be due to completely encapsulated enzymes (Figure 6.2). We suggest that the presence of surface-bound enzymes or MOF defects may explain the reported activity. This finding highlights

the importance of undertaking a careful analysis of the integrity of enzyme@MOF composites. Furthermore, screening of enzymatic reactions by the size of molecular components alone will lead to false negatives under experimental conditions where surface-bound enzymes or MOF defects are prevalent. However, the lack of controllability and predictability in material properties under such conditions is undesirable for practical applications. We suggest series of reactions from the list of previously tested reactions that could be used to further investigate the impact of molecular size on enzyme activity in enzyme@MOF composites.

The available literature shows a narrow exploration of enzymatic reaction space in ZIF-8 (Table 6.1). Recent studies have shown that systematic selection of experimental conditions[293] and chemical modification of proteins and enzymes,[3] can be used to achieve encapsulation in general. Furthermore, encapsulation of enzymes in other ZIFs[241,294] and MOFs[295,388] suggests that achieving enzyme encapsulation inside MOFs is not a limiting factor. Therefore, the ability to intelligently explore reaction space while avoiding expensive and arduous experiments is necessary. To that end, we have reported an efficient methodology to screen for reactions that may be able to diffuse through ZIF-8 and other MOFs. We collected ∼4000 unique enzymatic reactions using the KEGG database[324–326] and, in concert with other established informatics techniques, we determined the viability of these reactions for application inside MOFs. We found that ∼1000 of the collected reactions may be viable inside ZIF-8. Furthermore, the distributions of molecular size of all collected reactions are centered around 7–9 Å, suggesting that MOFs with pore sizes in this range should be targeted. Finally, we designed a screening process to find candidate reactions that could be monitored in situ (by searching for specific functionalities) and that added significant value from substrates to products (based on cost per gram). In this process, we screened for reactions with maximum component sizes near the diffusion threshold to further investigate the impact of molecular size on enzymatic activity in ZIF-8. We found a candidate reaction that corresponds to the transformation of a cheap amino acid (cysteine) into a significantly more expensive nitrile-containing compound, where the presence of the nitrile functionality allows for in situ monitoring of reaction progress by infra-red spectroscopy (Figure 6.6). We have shown that this method can be used to assist experimental endeavours by efficiently exploring enzymatic reaction space and designing

new experiments that can help answer some open questions in the expanding field of enzyme encapsulation in MOFs.

Approximating diffusivity through ZIF-8 by molecular size alone does not consider multiple factors that may play an important role in the activity of enzymes in MOFs such as substrate- and product-MOF interactions and other experimental factors.[314] More computationally expensive calculations of molecular diffusion through ZIF-8 and other porous materials[118,138,300,304,306,316,389,390] as well as analysis of adsorbate-MOF interactions[338,391] could be used to consider the role of factors other than molecular size. The analysis presented herein does not explicitly consider the nature of the MOF beyond pore size, i.e. MOFs with different flexibilities or pore chemistries may behave differently to ZIF-8. Nevertheless, we expect screening reactions based on molecular size to be a useful first approximation. Finally, a crucial factor that has not been considered in our screening process is whether an enzyme will retain its activity after being encapsulated in a MOF matrix. Predicting whether enzymes retain their activity after immobilization is not trivial and this remains an open question in the field.

All code used in this work is freely available at `https://bitbucket.org/andrewtarzia/psp_source/src/master/`.

## Acknowledgements

A. T. was supported by an Australian Government RTP Scholarship and a PhD top-up scholarship from CSIRO Division of Materials Science and Engineering. A. T. would like to thank Dr. Jesse Teo, Natasha Maddigan, Dr. Weibin Liang, Dr Justin Spence and A/Prof Stephen Bell for discussions about this work.

# CHAPTER 7

# Conclusion

## 7.1 Summary and future directions

This thesis describes computational approaches to explore the chemical space of functional porous materials efficiently. We have used computational methods to build upon and provide molecular-level insight into experimental findings associated with a range of different assembly phenomena.

In Chapter 3, we developed a coarse-grained (CG) molecular dynamics (MD) simulation model to study the formation mechanism of a series of four porous aromatic frameworks (PAFs).[1] Experimental findings showed an unexpected trend in the porosity of these PAFs as a function of the structure of their monomer. Our results showed that relatively weak dispersion forces induce interpenetration in PAF materials, which lowers their porosity. Furthermore, we found that bulky reaction intermediates in some PAF reaction mechanisms effectively eliminated interpenetration, which may be a way to improve the porosity in future experiments. Finally, our analysis suggested that the kinetic growth mechanism of PAFs robustly leads to highly defective and highly porous networks. Our physical model of PAF formation facilitates a molecular-level understanding of amorphous materials, for which experimental characterization is difficult. We modelled the porosity of discrete PAF clusters using an approximate method that treats each cluster as periodic. Recently developed geometrical approaches for modelling the porosity of discrete porous organic cages, which have distinct internal and external accessible surface areas, could be used to improve the characterization of early-time PAF clusters formed using our model.[392,393] Our CG simulation model could be extended to other porous materials to study their formation. In particular, crystalline materials formed using tetrahedral monomers, structurally similar to those modelled in Chapter 3, show polymorphism as a function of their synthesis conditions,

which is expected to be a result of the interplay between kinetic and thermodynamic control.[6] Only small modifications to our existing model would be required to study the driving force for the different polymorphs in these systems.

In Chapter 4, we developed a high-throughput screening method to find known metal–organic frameworks (MOFs) that would be good candidates for aligned heteroepitaxial growth.[2] Importantly, we expanded the number of known MOFs likely to undergo aligned growth and uncovered some design principles. Specifically, we found evidence for a substrate-directing effect, by which the in-plane symmetry of the substrate selects for MOF interfaces with similar symmetries. The algorithm we developed is not limited to the specific case that was studied (copper-based MOFs grown from a $Cu(OH)_2$ substrate via carboxylate binding sites) and an extension of our work would be to study a broader set of substrates, MOFs and binding geometries. Such a study may help to further understand the proposed substrate-directing effects by exploring a more diverse set of structures, while still providing top candidates (both substrate and MOF candidates) for expansion of the field of MOF thin films. When considering different binding geometries, it is necessary to account for the relative binding energies of each geometry with the substrate. An extension of our current model could achieve this by including a relative binding energy term into a scoring function analogous to $\Delta$IB. The relative binding energy of different binding site configurations as a function of their 2D position with respect to the substrate could be parameterized using density functional theory (DFT) calculations.[116] Therefore, from the relative positions in 2D of each atom in the MOF and substrate binding planes relative binding energies could be calculated and included in the scoring function based on the DFT parameterization. Furthermore, our 2D description of the interface is simple and efficient, but we note that a 3D description of the interface may be necessary to consider more complicated MOF-substrate binding interactions. Recent studies that systematically sliced zeolites to form reasonable interfaces may offer guidance for the process of systematically slicing MOF structures.[225,394] Finally, the formation mechanism underpinning the growth of aligned MOFs from a sacrificial substrate remains unclear.[55] As explored in Chapters 3 and 5, molecular simulation and theory are an excellent strategy for studying such processes, and it would be valuable to understand how the kinetics and thermodynamics of MOF formation control their macroscale alignment.

In Chapter 5, we used theory and computation to verify experimental evidence that the electrostatic properties of a biomacromolecule are the dominant factor in inducing zeolitic imidazolate framework-8 (ZIF-8) growth in water.[3,4] We used the Poisson–Boltzmann (PB) equation to verify that increased negative charge of a biomacromolecule leads to an increased concentration of zinc ions in the surrounding solution, which is expected to enhance zeolitic imidazolate framework (ZIF) formation kinetics. These findings were transferable to carbohydrate systems,[4] which suggests that biomimetic mineralization can generally be applied using chemical functionalization to a broad range of biomacromolecules. Importantly, we can calculate the zeta potential and pI of a protein efficiently, which we show to predict ZIF-8 formation in the presence of proteins accurately. These findings show the potential for computational screening to advance experimental endeavors in this burgeoning field of MOF chemistry by screening biomacromolecules for their ability to seed MOF growth before carrying out experiments. We emphasized both the efficiency of these calculations and the generality of biomimetic mineralization by calculating the pI of $\sim 10^5$ enzyme sequences, which shows that many enzymes have pI values low enough to seed ZIF-8 growth. We studied only a small set of all of the protein systems that exist in nature, but our improved understanding of how to control the biomimetic mineralization process can guide future experiments. However, a molecular-level understanding of how the enhanced zinc-ion concentration induces biomimetic mineralization is still lacking. A first step to understanding ZIF self-assembly in the presence of biomacromolecules would be to develop a kinetic model that considers the concentration (calculated using the methods reported in Chapter 5) and the dynamics of ZIF precursors. The kinetic model could be parameterized using previously reported models[72,395] and experimental investigations of ZIF formation.[71,72,395–398]

A small set of possible enzymatic reactions have been carried out inside ZIF-8 to-date. Therefore, we developed a screening process for viable enzymatic reactions that can be studied inside enzyme@MOF composites. We defined the first check for viability to be whether the substrates and products can diffuse through the MOF shell to and from the target enzyme. Thus, we implemented an efficient descriptor for molecular size to approximate a molecules ability to diffuse through ZIF-8. Our descriptor considered the flexibility of molecules and was found to match reported kinetic diameters.

Using this descriptor we have shown that the components of many of the reactions reported to occur inside ZIF-8 are unlikely to be able to diffuse through ZIF-8. We suggest that surface-bound enzymes or defects in the MOF structure may explain the reported enzyme activity. Finally, we screened ∼11 000 enzymatic reactions from the KEGG database[324–326] using our molecular size descriptor and other cheminformatic descriptors. Expanding our screening process to other databases of enzymatic reactions, such as the BRENDA database,[290,291] for a complete survey of the enzymatic reaction space is crucial. Furthermore, the SABIO-RK database[327–329] specifies (when available) the enzyme peptide sequence that completes each the reaction. Therefore it may be possible to screen based on enzyme properties also. Unfortunately, it is not trivial to predict whether an enzyme retains its activity after immobilization.[360,382] Therefore, we did not include an analysis of the enzymes in our screening process, but we expect this to be a crucial factor in the application of enzyme@MOF composites. It may be possible to use bioinformatic tools to efficiently score proteins (based on predicted properties such as thermostability) to approximately predict whether an enzyme is stable to encapsulation, but at this stage, very little experimental data exists to test the predictions of any proposed model of the stability of MOF-encapsulated enzymes. Ultimately, we determined that the small pore aperture of ZIF-8 excludes a majority of enzymatic reactions in the KEGG database. However, the role of molecular size on the activity of enzymes in ZIF-8 remains unclear. Furthermore, it remains unclear whether reported activity from enzyme@MOF composites is occurring inside the MOF. Therefore, we suggested a series of reactions that allow for the systematic analysis of the impact of molecular size on enzymatic activity inside ZIF-8. However, we suggest that the development of experimental methodologies that unambiguously confirm that the observed activity is from enzymes inside MOFs is crucial to the advancement of this field.

Finally, we have highlighted in this thesis the benefits of computational analysis and screening processes in advancing materials science, especially in relatively new fields. However, the usefulness of the screening methods described in Chapters 4 and 6 rests, in part, in their ability to accurately select out the best candidates for future experiments. Hence, for all reported screening processes, further experimental analysis of the top candidates is required to confirm the accuracy of the applied simple descriptors.

# APPENDIX A

# Supporting information: Molecular insight into assembly mechanisms of porous aromatic frameworks

## A.1 Proposed reaction mechanisms for PAF synthesis

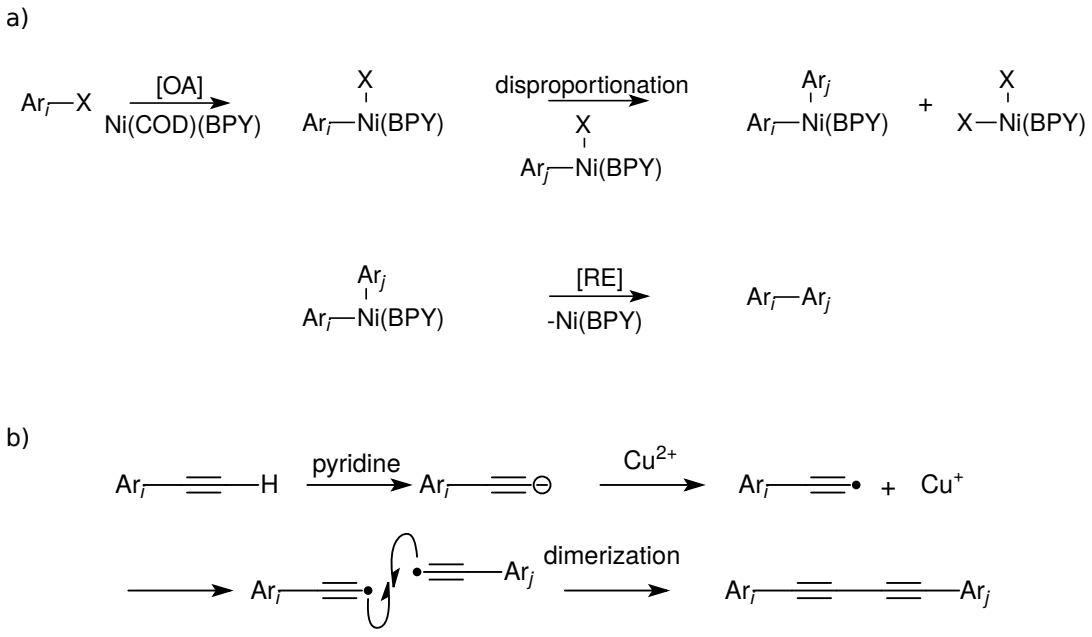



Figure A.1: The proposed mechanisms of the homo-coupling of $Ar_i$ and $Ar_j$ via (a) the Yamamoto reaction and (b) the Eglinton reaction. BPY is 2,2′-bipyridine, COD is 1,5-cyclooctadiene.

Two reactions are used in experiments to synthesize the four porous aromatic frameworks (PAFs) studied in this work: the Yamamoto homo-coupling[176] and the Eglinton homo-coupling.[177] The proposed mechanism of the two reactions are in Figure A.1.

## A.2 Coarse-grained mapping of PAFs

Table A.1: Principal diameters of the ellipsoidal coarse-grained sites obtained from the Connolly volumes for all moieties of all monomers.

| moiety | $a_i$ [Å] | $b_i$ [Å] | $c_i$ [Å] |
|---|---|---|---|
| carbon | 3.5 | 3.5 | 3.5 |
| phenyl | 7.3 | 7.3 | 3.4 |
| alkyne | 4.7 | 3.5 | 3.5 |
| adamantane | 7.3 | 7.0 | 7.4 |

As illustrated in Figure A.2 the coarse-graining procedure replaced different functional moieties of the PAF monomers with ellipsoidal particles. The size and shape of each ellipsoid was specified by the principal diameters, ($a_i$, $b_i$ and $c_i$, determined from Connolly volumes, which were obtained using the "Atom Volumes and Surfaces" Tool within Materials Studio)[178] given in Table A.1. The shape of the bond-forming ellipsoid was modified in the binding direction to yield the correct bond length upon binding. Each PAF monomer was treated as a rigid body, which means structural fluctuations were quenched.

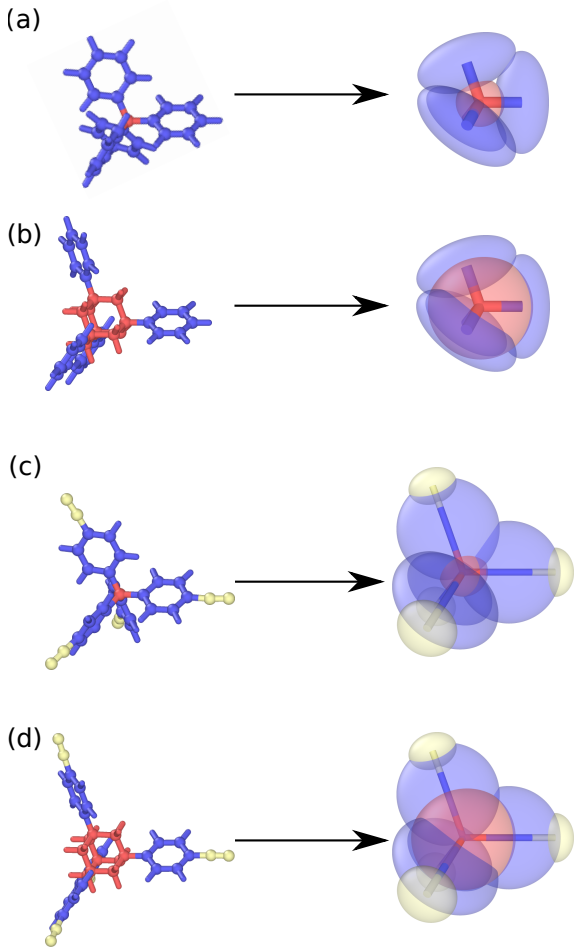



Figure A.2: Mapping of atomistic monomers of (a) PAF-1, (b) PPN-3, (c) PPN-1, and (d) PPN-2 to coarse-grained monomers. Colors of atoms in all-atom models match the ellipsoid that replaces them.

## A.3 Simulation methodology and parameters

The molecular dynamics (MD) simulation program LAMMPS[184] was used to perform GPU-accelerated Langevin dynamics of rigid bodies representing PAF monomers composed of ellipsoidal particles. The ellipsoidal particles interacted via the Gay-Berne (GB) potential.[180,181] The GB potential is an anisotropic and shifted Lennard-Jones (LJ) 12-6 interaction, which can be written as a product of 3 terms,

$$U = U_{\mathrm{r}} \cdot \eta \cdot \chi, \tag{A.1}$$

parameterized by the ellipsoid shapes and relative interaction energies. A shape matrix is specified for ellipsoid $i$ as $\mathbf{S}_i = \mathrm{diag}(a_i, b_i, c_i)/\sigma$, where $\sigma$ is the distance scale used

in simulations. Relative well depths $\varepsilon_{ia}$, $\varepsilon_{ib}$ and $\varepsilon_{ic}$ for particles interacting along the corresponding principal axes (end-to-end, side-to-side, and face-to-face interactions) give the matrix $\mathbf{E}_i = \mathrm{diag}(\varepsilon_{ia}, \varepsilon_{ib}, \varepsilon_{ic})$ for ellipsoid $i$. The orientation of ellipsoid $i$ is determined by the rotation matrix $\mathbf{A}_i$, which represents the transformation from the lab frame to the body frame.

In Equation A.1, $U_\mathrm{r}$ represents the shifted LJ interaction, which is determined by the distance of closest approach $h_{ij}$, the distance scale $\sigma$, energy scale $\varepsilon$ and shift factor $\gamma$ (set to 1 in all cases),

$$U_\mathrm{r} = 4\varepsilon(\rho_{ij}^{12} - \rho_{ij}^{6}), \tag{A.2}$$

$$\rho_{ij} = \frac{\sigma}{h_{ij} + \gamma\sigma}, \tag{A.3}$$

$$h_{ij} = r_{ij} - \left[\frac{1}{2}\widehat{\boldsymbol{r}}_{ij}^{\mathrm{T}}\mathbf{G}_{ij}^{-1}\widehat{\boldsymbol{r}}_{ij}\right]^{1/2}, \tag{A.4}$$

where $\boldsymbol{r}_{ij} = \boldsymbol{r}_i - \boldsymbol{r}_j$, with $\boldsymbol{r}_i$ and $\boldsymbol{r}_j$ the positions of particles $i$ and $j$, respectively, $r_{ij} = |\boldsymbol{r}_{ij}|$, $\widehat{\boldsymbol{r}}_{ij} = \boldsymbol{r}_{ij}/r_{ij}$, and $\mathbf{G}_{ij} = \mathbf{A}_i^{\mathrm{T}}\mathbf{S}_i^2\mathbf{A}_i + \mathbf{A}_j^{\mathrm{T}}\mathbf{S}_j^2\mathbf{A}_j$. In addition to the distance of closest approach, the interaction anisotropy is characterized by the distance-independent terms $\eta$ and $\chi$ that control the interaction strength based on the particle shapes and relative well depths, respectively, with

$$\eta = \left[\frac{2s_i s_j}{\det(\mathbf{G}_{ij})}\right]^{\upsilon/2}, \tag{A.5}$$

$$s_i = [a_i b_i + c_i c_i][a_i b_i]^{1/2} \tag{A.6}$$

and

$$\chi = \left[2\widehat{\boldsymbol{r}}_{ij}^{\mathrm{T}}\mathbf{B}_{ij}^{-1}\widehat{\boldsymbol{r}}_{ij}\right]^{\mu}, \tag{A.7}$$

$$\mathbf{B}_{ij} = \mathbf{A}_i^{\mathrm{T}}\mathbf{E}_i^2\mathbf{A}_i + \mathbf{A}_j^{\mathrm{T}}\mathbf{E}_j^2\mathbf{A}_j. \tag{A.8}$$

The parameters $\upsilon$ and $\mu$ in Equations A.5 and A.7 are exponents that can be tuned to adjust the angular dependence of the potential; both were set to 1 in all simulations, for which the homo-coupling interaction becomes negligible when the ellipsoids deviate by more than about $\pm 15°$ from collinear at the simulation temperature. The angular dependence of the potential was made strict so that bound monomers approximately modelled the flexibility of a carbon-carbon bond. We used a modular approach to

distinguish three types of interactions: primary (1°), secondary (2°) and steric. The potential functions used to represent primary interactions were chosen to model irreversible carbon-carbon bond formation by using a deep well, as shown in Table A.2 and Figure 3.2. Note that the small differences between energy values for the primary interactions of PPN-3 compared with the other three PAFs will have a negligible effect on the simulations because the very strong interactions lead to irreversible binding in all cases (a value of $100k_{\mathrm{B}}T$ was used for all PAFs, resulting in a different value for PPN-3 due to the different temperature, $T$, used in the experimental synthesis). Secondary interactions (such as $\pi \cdots \pi$, alkyne $\cdots$ alkyne and alkyne $\cdots \pi$ interactions) were modelled using potentials with relatively shallow wells, as shown in Table A.2 and Figure 3.2, that were matched approximately to literature values for these types of interactions and which give reversible binding.[182,183] To approximate the steric interactions due to the reaction intermediates in the synthesis of PAF-1 and PPN-3, we introduced an isotropic potential between phenyl and monomer-core sites with a Gaussian well,

$$U_{\mathrm{steric}}\left(r_{ij}\right) = -\alpha \exp\left[-\beta(r_{ij} - \delta)^2\right], \tag{A.9}$$

where $\alpha$ determines the well depth, $\beta$ determines the well width and $\delta$ determines the position of the well (set to 12.2 Å = 9.3 Å (approximately the size of the catalyst as measured by its Connolly volume) + 2.9 Å (phenyl arm length)). A well depth of approximately 30 kJ mol$^{-1}$ for both PAF-1 and PPN-3, which matches the energy of nickel–ligand binding,[175] was used as nickel acts as the catalyst in the Yamamoto mechanism.

Table A.2: Maximum potential well depth for all interaction types.

| sites | interaction | energy [kJ mol$^{-1}$] |
|---|---|---|
| homo-coupling[a] | end-to-end | 294[b] / 319[c] |
| phenyl–phenyl | edge-to-edge | 8.82 |
| phenyl–phenyl | face-to-face | 5.88 |
| phenyl–alkyne | edge-to-edge | 8.82 |
| phenyl–alkyne | face-to-face | 5.88 |
| alkyne–alkyne | edge-to-edge | 2.94 |
| alkyne–alkyne | face-to-face | 2.94 |

[a] Phenyl–phenyl for PAF-1 and PPN-3; Alkyne–alkyne for PPN-1 and PPN-2.
[b] PAF-1, PPN-1 and PPN-2.
[c] PPN-3.

PACKMOL was used to initialize a system of randomly placed, non-overlapping monomers at the desired monomer density ($\rho_{\text{mon}}$ in Table A.3).[186] To push apart overlapping atoms, the system was evolved with soft non-bonded potentials before carrying out simulations with interactions specified by the GB potential with an integration time step (given in Table A.4 for each system) defined as one tenth of the Langevin dynamics relaxation time, which is inversely proportional to the predicted diffusion coefficient. For ease of comparison to the simulations of PAF-1 (performed at 80 °C), the temperature of the reactions for PPN-1 and PPN-2 was taken to be 80 °C, which is approximately the reaction temperature used in the Eglinton reactions that were carried out under reflux conditions in the experiments using a 1:4 mixture of methanol:pyridine (the boiling points of methanol and pyridine are 65.7 °C and 115.3 °C, respectively).[159]

Table A.3: Experimental and simulation parameters for the four PAF systems studied. $\eta$ is the solvent viscosity.

| PAF | solvent | $T_{\text{exp}}$ [°C] | $\rho_{\text{mon}}$ [mol L$^{-1}$] | $\eta$ [mPa s] |
|---|---|---|---|---|
| PAF-1 | N,N-dimethylformamide | 80 | 0.013 | 0.47[399] |
| PPN-1 | 4:1 pyridine:methanol | reflux | 0.022 | 0.60[400] |
| PPN-2 | 4:1 pyridine:methanol | reflux | 0.022 | 0.60[400] |
| PPN-3 | 1:2 N,N-dimethylformamide:toluene | 110 | 0.013 | 0.67[401] |

Langevin dynamics was used to simulate the presence of solvent molecules implicitly. The functional form of the Langevin equation applied to rigid bodies[123,185] is

$$F(t) = -\nabla U(x) - \xi_{\text{t}} m \boldsymbol{v}(t) + f_{\text{R}}(t), \tag{A.10}$$

$$\tau(t) = -\nabla U(x) - \xi_{\text{r}} I \boldsymbol{\omega}(t) + \tau_{\text{R}}(t), \tag{A.11}$$

where $F(t)$ and $\tau(t)$ are the total force and torque, respectively, on a body, $\nabla U(x)$ is the conservative force due to the interaction potentials acting on the body, $\xi_{\text{t}}$ and $\xi_{\text{r}}$ are the translational and rotational friction coefficients, respectively, $m$ is the monomer mass, $I$ is the moment of inertia, $\boldsymbol{v}(t)$ and $\boldsymbol{\omega}(t)$ are the translational and rotational velocities of the bodies at time $t$, respectively, and $f_{\text{R}}(t)$ and $\tau_{\text{R}}(t)$ are the random force and random torque, respectively, applied to the bodies. As no simple analytical expression exists to describe the diffusion of ellipsoids in general, we have assumed that the monomers diffuse like spherically symmetric Brownian particles, for which the

Stokes-Einstein equation gives the translational diffusion coefficient as

$$D_{\mathrm{t}} = \frac{k_{\mathrm{B}}T}{6\pi\eta R}, \tag{A.12}$$

where $\eta$ is the solvent viscosity,[399–401] $R$ is the radius of the particle, $k_{\mathrm{B}}$ is the Boltzmann constant and $T$ is the temperature. The translational friction coefficient, $\xi_{\mathrm{t}}$, can be determined from

$$\xi_{\mathrm{t}} = \frac{k_{\mathrm{B}}T}{D_{\mathrm{t}}m}. \tag{A.13}$$

For spherical Brownian particles, the rotational coefficients, $\xi_{\mathrm{r}}$ and $\tau_{\mathrm{R}}(t)$, are obtained by scaling $\xi_{\mathrm{t}}$ and $f_{\mathrm{R}}(t)$ by $\frac{10}{3}$ and $\sqrt{\frac{10}{3}}$, respectively. The random force and torque, have zero mean and a white-noise spectrum and are related to the respective friction coefficients via the fluctuation-dissipation theorem.

The translational and rotational diffusion coefficients of each monomer were estimated using HydroPRO,[402–404] which computes the hydrodynamic properties of rigid molecules using the properties of the solvent and a bead model of the molecules. The value of $D_{\mathrm{t}}$ obtained was then used in Equation A.13 to determine the value of $\xi_{\mathrm{t}}$ used in the simulations, from which $f_{\mathrm{R}}(t)$, $\xi_{\mathrm{r}}$ and $\tau_{\mathrm{R}}(t)$ were determined, as described above.

Translational diffusion coefficients were calculated from simulations of 400 non-interacting PAF monomers by fitting the mean squared displacement (MSD) of the monomer center-of-mass to a straight line (Figure A.3), i.e.

$$\langle \mid \boldsymbol{r}(t+t') - \boldsymbol{r}(t') \mid^2 \rangle \sim 6D_{\mathrm{t}}t. \tag{A.14}$$

The rotational diffusion coefficients were calculated from the same set of simulations by fitting the natural logarithm of the decay of the orientation autocorrelation function to a straight line (Figure A.4), i.e.

$$\ln\langle \hat{\boldsymbol{u}}_i(t') \cdot \hat{\boldsymbol{u}}_i(t+t') \rangle \sim 2D_{\mathrm{r}}t. \tag{A.15}$$

$\hat{\boldsymbol{u}}_i(t)$ is the unit vector pointing along the $i^{\mathrm{th}}$ arm of the tetrahedral monomer at time $t$. The translational ($D_{\mathrm{t}}$) and rotational ($D_{\mathrm{r}}$) diffusion coefficient measured from the simulations were found to agree well with those computed by HydroPRO (see Table A.4), suggesting that the assumption that the symmetric tetrahedral monomers

diffuse like spherical particles is reasonable.

Table A.4: Langevin dynamics simulation parameters. $\xi_t$ is the friction coefficient applied to the Langevin dynamics. $D_{t,HP}$ and $D_{r,HP}$ are the translational and rotational diffusion coefficients calculated by HydroPRO.

| PAF | translational $[10^{-9}\,m^2\,s^{-1}]$ | | rotational $[10^{-9}\,s^{-1}]$ | | | |
|---|---|---|---|---|---|---|
| | $D_{t,HP}$ | $D_t$ | $D_{r,HP}$ | $D_r$ | $\xi_t$ $[10^{12}\,s^{-1}]$ | time step [fs] |
| PAF-1 | 1.138 | 1.160 | 3.537 | 4.282 | 8.13 | 12.5 |
| PPN-1 | 0.640 | 0.634 | 0.939 | 1.286 | 11.11 | 8.8 |
| PPN-2 | 0.590 | 0.600 | 0.694 | 1.082 | 9.35 | 10.8 |
| PPN-3 | 0.755 | 0.765 | 1.707 | 1.987 | 9.62 | 10.5 |

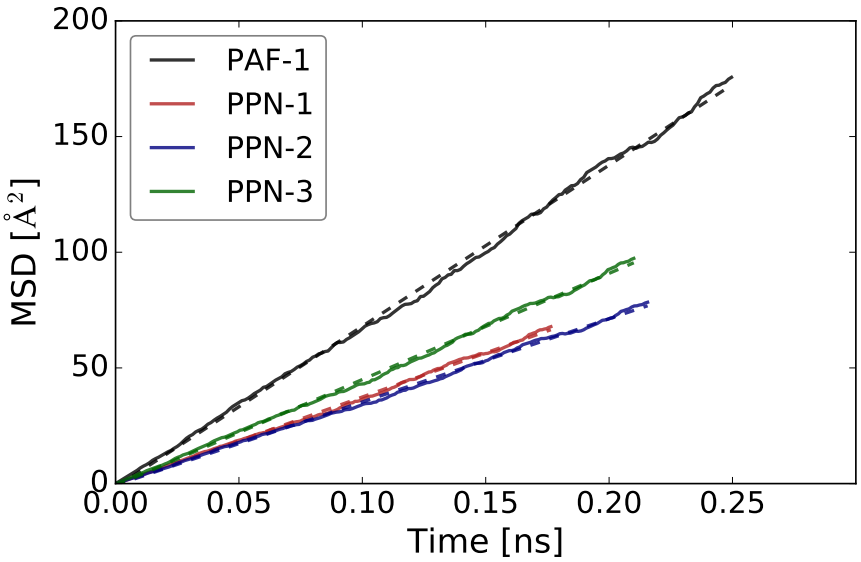



Figure A.3: The mean square displacement (MSD) for each PAF (solid lines) and linear fits to these curves (dashed lines). The slope is proportional to the translational diffusion coefficient.

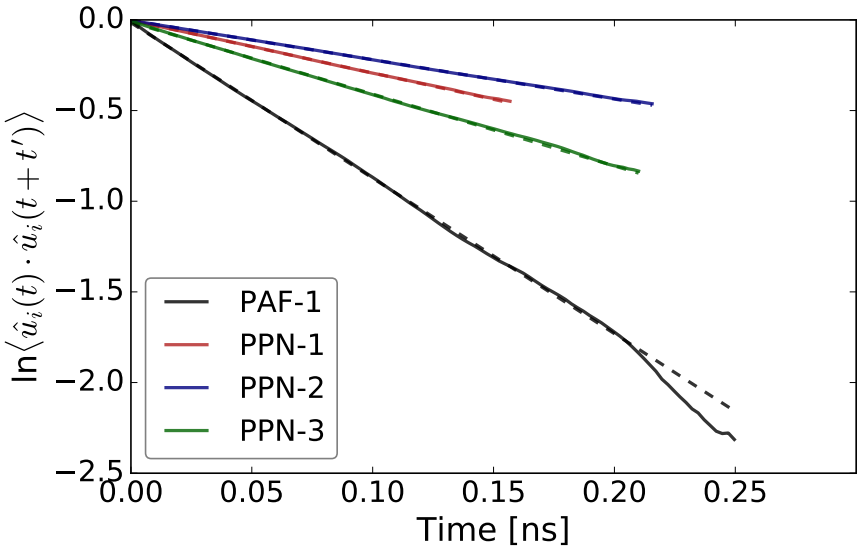



Figure A.4: Natural logarithm of the decay of the orientational auto-correlation function (solid lines) and linear fits to these curves (dashed lines). The slope is proportional to the rotation diffusion coefficient.

## A.4 Compression simulation methodology

To facilitate assembly and to capture the growth of larger clusters than observed in the constant-volume simulations, the simulation box was isotropically compressed at a constant rate. The pressure was monitored as the simulation box was compressed and plotted as a rolling average as a function of time, as shown in Figure A.5. The simulation was stopped when the pressure started to rise precipitously (the cut-off point was set at an average pressure of $5 \times 10^5$ atm), shown by the vertical lines in Figure A.5, indicating that the particles were starting to overlap. The initial and final box sizes are reported in Table A.5. Simulations were otherwise identical to the constant-volume simulations described above. Compression was carried out at two different rates to verify that there was no effect of compression rate on the results – these are denoted 'fast' and 'slow', where the slow compression rate was half the fast compression rate. Table A.5 shows the times at which each simulation reached its cut-off pressure at each rate. In addition, similar simulations were carried out in the presence of primary interactions only at the 'fast' rate (see Table A.6 below).

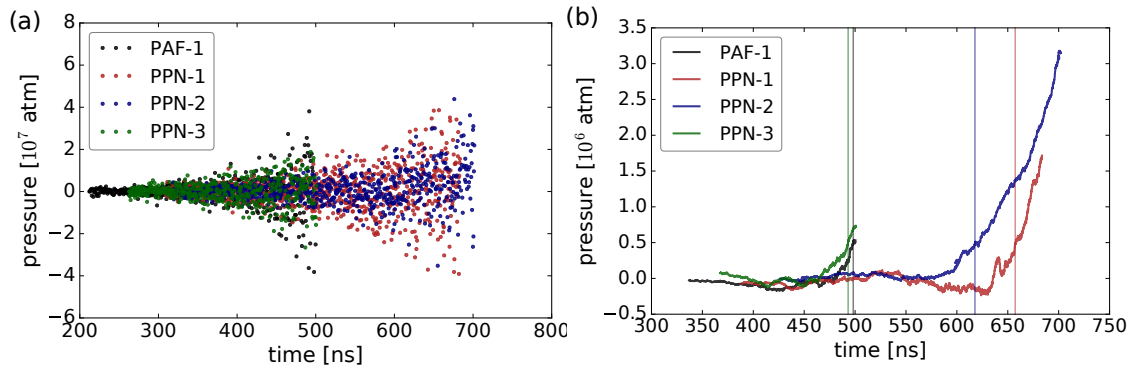



Figure A.5: (a) Instantaneous pressure as each PAF system is compressed at the fast rate. (b) Rolling average of the pressure from (a) with vertical lines highlighting the time point at which the cut-off pressure was met and analysis was undertaken.

Table A.5: Simulation parameters for assembly with compression from experimental concentration with box dimension $l_\mathrm{i}$ to the final analysis state with box dimension $l_\mathrm{f}$ at both compression rates and the time $t_\mathrm{f}$ at which the cut-off pressure was reached. 1° and 2° denote primary and secondary interactions, respectively.

| system | interactions | $l_\mathrm{i}$ [Å] | $l_\mathrm{f}$ - fast [Å] | $t_\mathrm{f}$ - fast [ns] | $l_\mathrm{f}$ - slow [Å] | $t_\mathrm{f}$ - slow [ns] |
|---|---|---|---|---|---|---|
| PAF-1 | 1°+2°+steric | 1006 | 276.5 | 473 | 251.4 | 982 |
| PPN-3 | 1°+2°+steric | 1006 | 341.7 | 478 | 361.5 | 927 |
| PPN-1 | 1°+2° | 842 | 286.7 | 637 | 297.0 | 1241 |
| PPN-2 | 1°+2° | 842 | 337.2 | 598 | 327.5 | 1202 |

Table A.6: Simulation parameters for systems with primary interactions only for assembly with compression from experimental concentration with box dimension $l_\mathrm{i}$ to the final analysis state with box dimension $l_\mathrm{f}$ and the time $t_\mathrm{f}$ at which the cut-off pressure was reached.

| system | interactions | $l_\mathrm{i}$ [Å] | $l_\mathrm{f}$ [Å] | $t_\mathrm{f}$ [ns] |
|---|---|---|---|---|
| PAF-1 | 1° | 1006 | 351.6 | 426 |
| PPN-3 | 1° | 1006 | 349.1 | 473 |
| PPN-1 | 1° | 842 | 388.2 | 509 |
| PPN-2 | 1° | 842 | 442.2 | 483 |

## A.5 Cluster selection and reverse mapping methods

Monomers were assigned to clusters at selected time steps along the simulation trajectory of each PAF using the Density-based spatial clustering of applications with noise (DBSCAN) clustering algorithm,[188] which identifies contiguous regions in which the local particle density at any point stays above a specified threshold. Shown in

Table A.7, the threshold parameters were chosen such that diffuse PAF chains were excluded from clusters. A distance parameter (taken as the monomer–monomer binding distance for each PAF, $r_{\mathrm{bind}}$) was also specified to define the radius around each particle within which the density of neighbouring particles was measured. Selected clusters were reverse-mapped to their atomistic representations by a one-to-one mapping, because monomers were assumed to be rigid, from the CG representation. Following reverse mapping, the energies of the atomistic clusters were relaxed using the Open Babel software[405] and the UFF forcefield.[119,120]

Table A.7: Minimum number of neighbours threshold used in the DBSCAN algorithm for all PAFs for all interaction types.

| material | 1° | 1° + 2° | 1° + 2° + steric |
|---|---|---|---|
| PAF-1 | 12 | 18 | 12 |
| PPN-3 | 12 | 18 | 12 |
| PPN-1 | 24 | 30 | - |
| PPN-2 | 24 | 30 | - |

## A.6 Typical clusters from constant-volume simulations

Typical clusters in their coarse-grained representation are shown in Figures A.6–A.9 for all four PAFs for all types of interactions studied. These clusters were extracted from constant-volume simulations at the point in time when the free monomer concentration fell to 1 % of its original value, which occurred after 100–200 ns. Therefore these clusters are not as big as those formed by the end of the simulations, which were carried out for ∼1–2 µs in all cases, but do highlight the different morphologies obtained and the effect of different interactions.

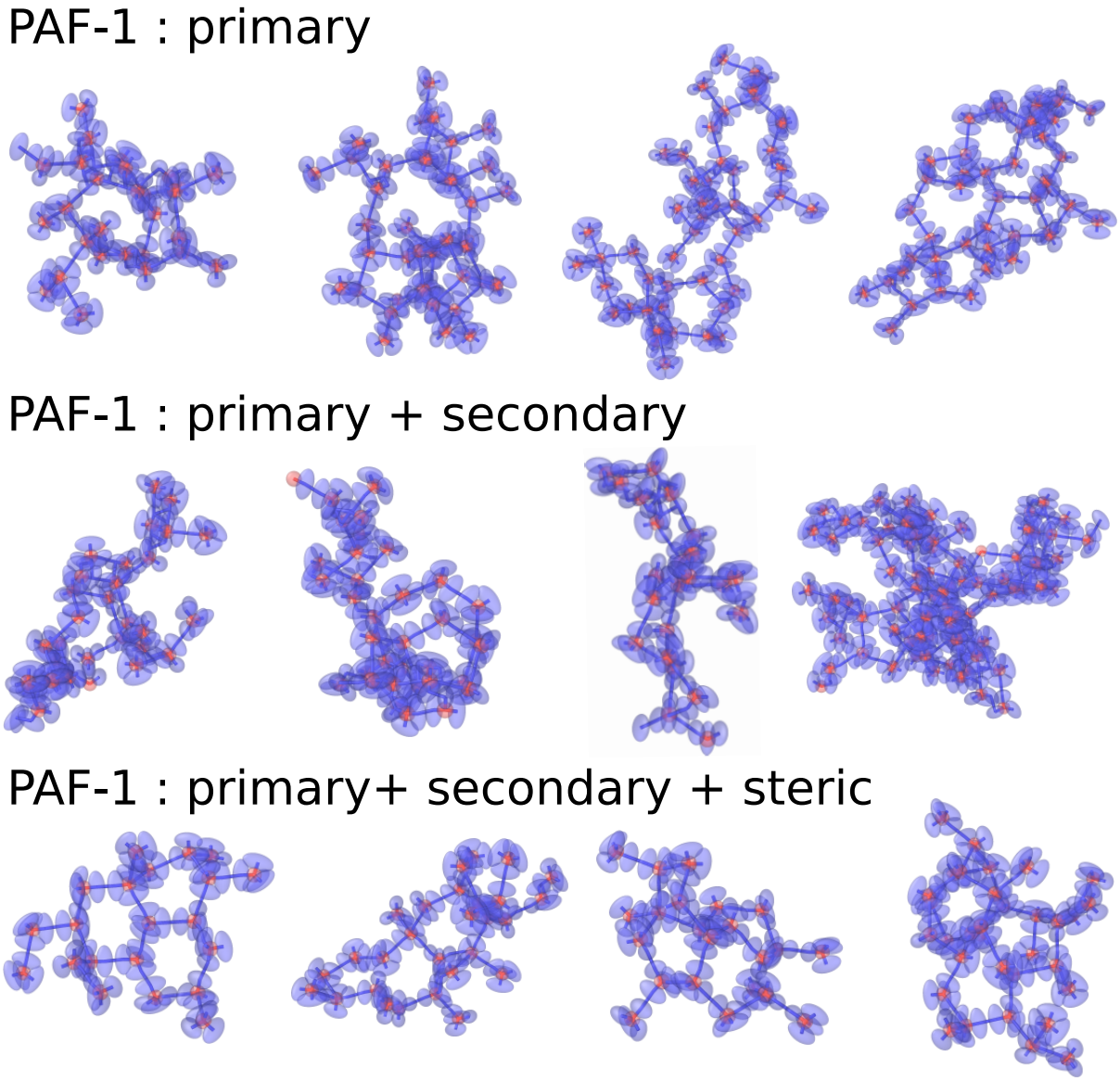



Figure A.6: Typical clusters extracted from simulations of PAF-1 for all interaction types.

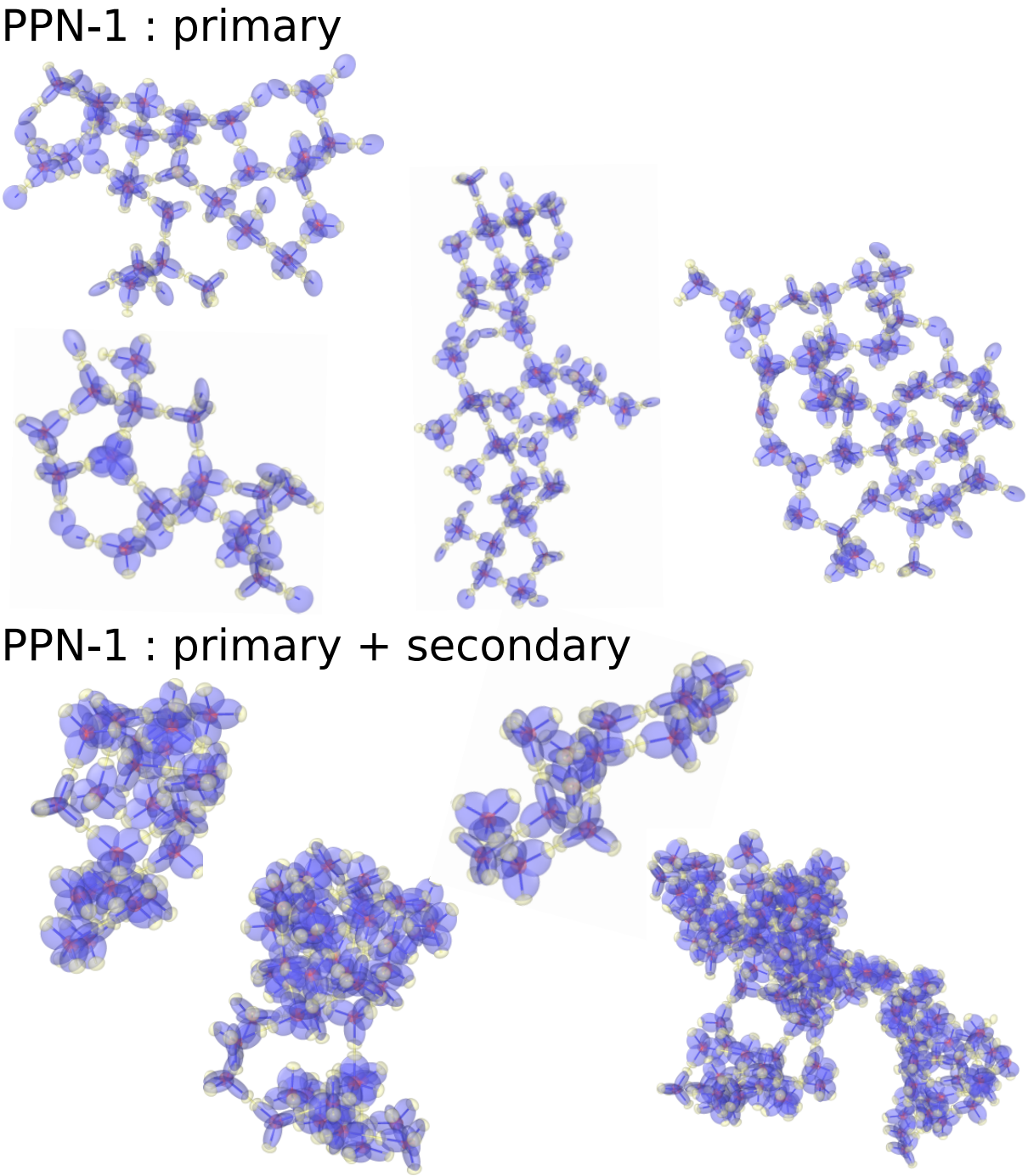



Figure A.7: Typical clusters extracted from simulations of PPN-1 for all interaction types.

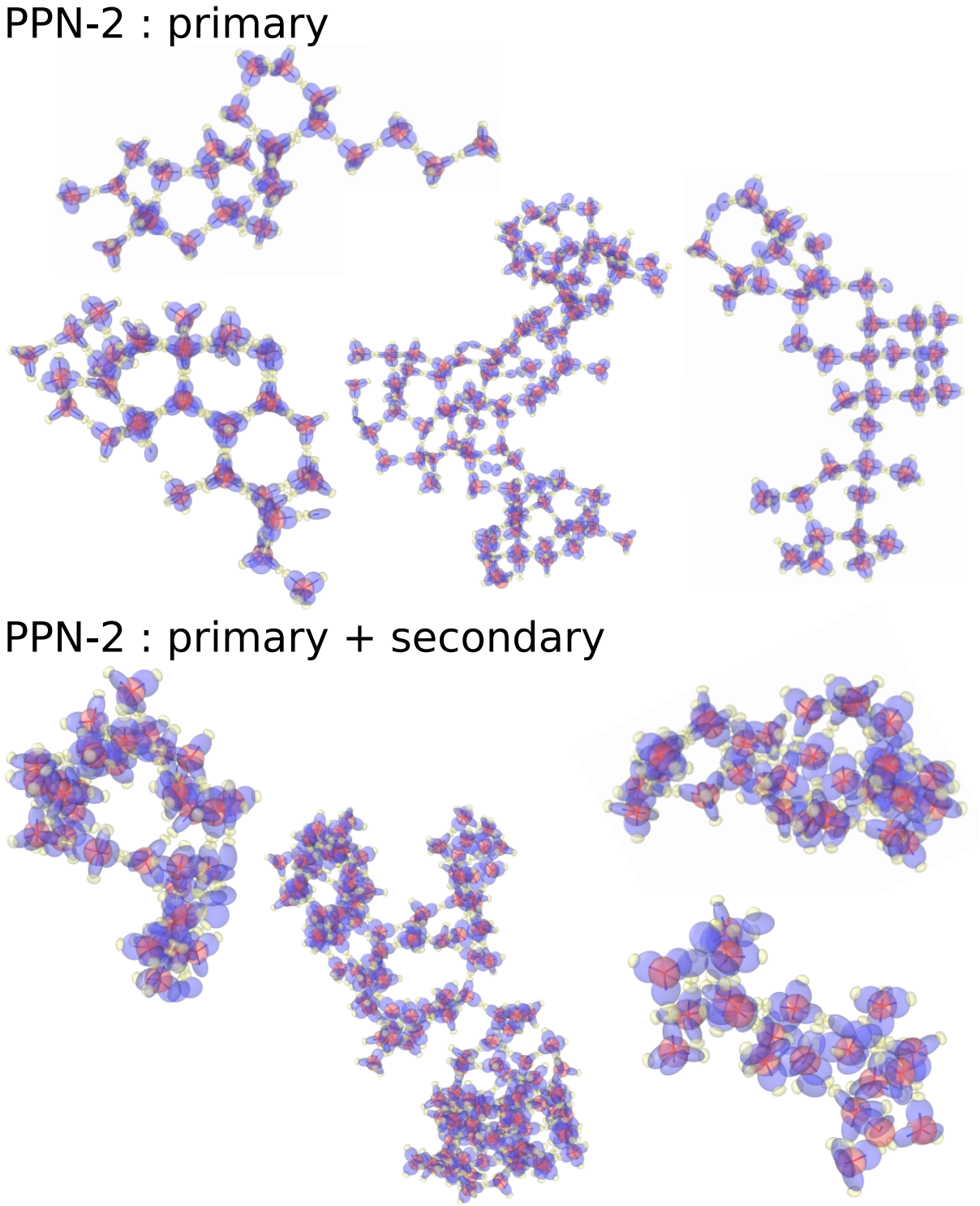



Figure A.8: Typical clusters extracted from simulations of PPN-2 for all interaction types.

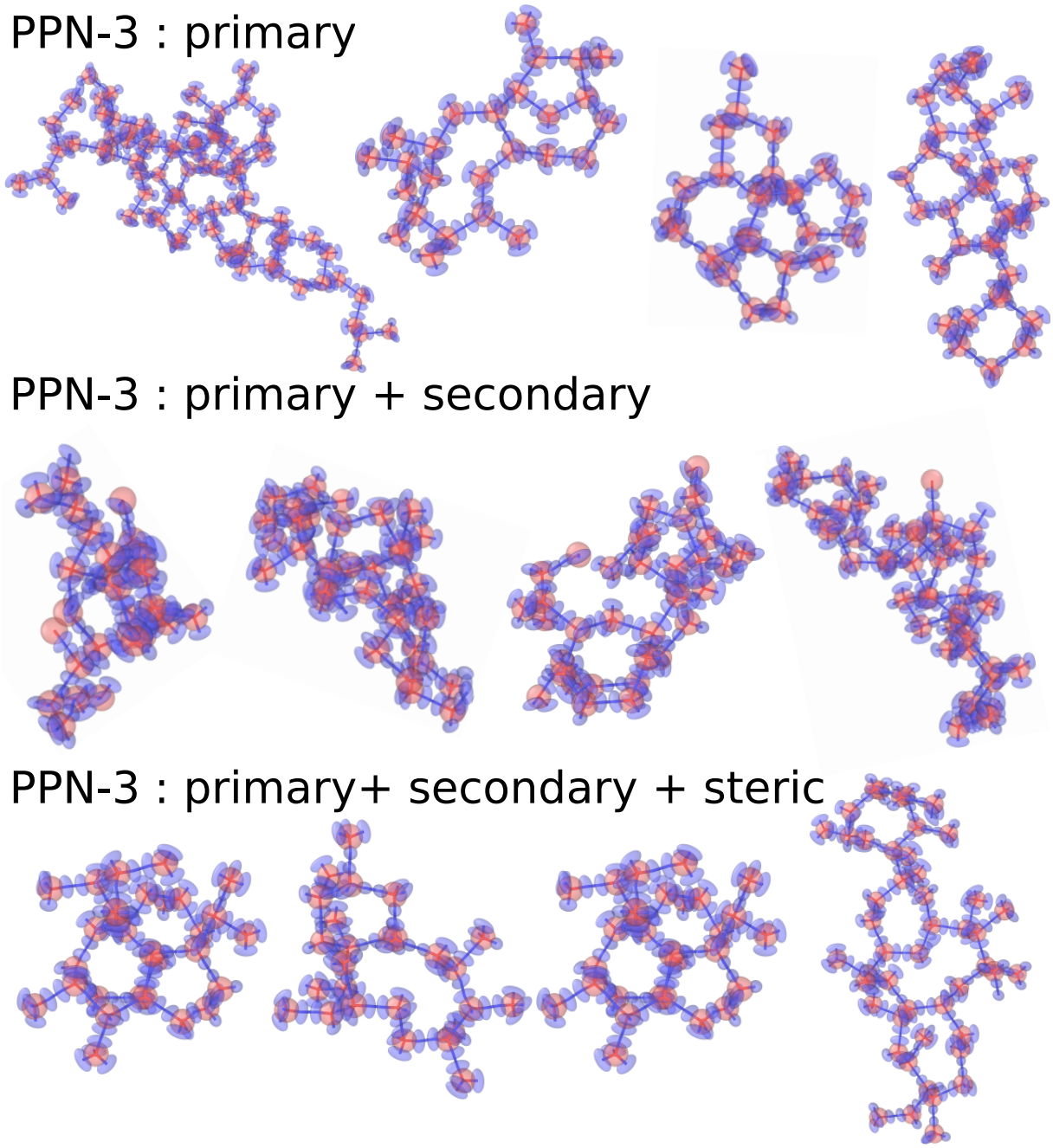



Figure A.9: Typical clusters extracted from simulations of PPN-3 for all interaction types.

## A.7 Porosity analysis of clusters in uncompressed systems

All constant-volume simulations yielded many clusters of each PAF, for which structural properties such as density, pore surface area and pore volumes could be determined. As described earlier, clusters were extracted at certain time steps from simulation trajectories and reverse mapped to the atomistic representation. All of the above properties are affected by the external surface of the finite-sized clusters, so a method

was developed to minimize this effect in order to approximate the properties of the bulk material:

1. A cluster was selected and reverse mapped using the previously described method. The cluster was rotated to align its principal axes with a reference Cartesian frame. The dimensions of the cuboid of minimum dimensions centered at the center-of-mass of the cluster and aligned with the reference Cartesian axes that enclosed the whole cluster were determined (step 1 in Figure A.10).

2. Properties (e.g. density, pore surface area, pore volume) were measured for the fraction of the cluster enclosed by a smaller box with dimensions proportional to those of the cuboid determined in step 1 for progressively increasing box sizes (step 2–3 in Figure A.10). For the porosity analysis, it was assumed that periodic boundary conditions applied to the atoms enclosed by the box.

3. The measured density from the analysis in step 2 generally had a maximum as a function of increasing fraction of the cluster considered (Figure A.11) – very small fractions were not representative of the bulk material whereas large fractions enclosed a significant amount of external void space. So the fraction that yields the maximum density is expected to be most representative of the bulk PAF. The measured value of any other property (e.g. pore surface area, as illustrated in Figure A.11) that best approximated that of the bulk PAF was thus assumed to be the value at this density maximum.

4. The reported values of each property was taken to be the mass-weighted average of the approximate bulk value measured for each cluster.

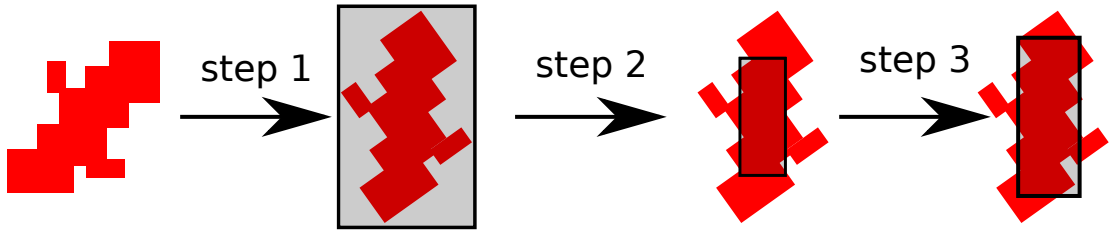



Figure A.10: Schematic of the algorithm used to measure cluster properties. The red polyhedra represents a cluster of PAF selected from a LAMMPS trajectory. The cluster is first rotated to align the principal axes with a reference Cartesian frame. The next two steps illustrates the incrementing of the size of the box centered at the centre-of-mass of the cluster in which properties are measured to produce a plot such as the one seen in Figure A.11.

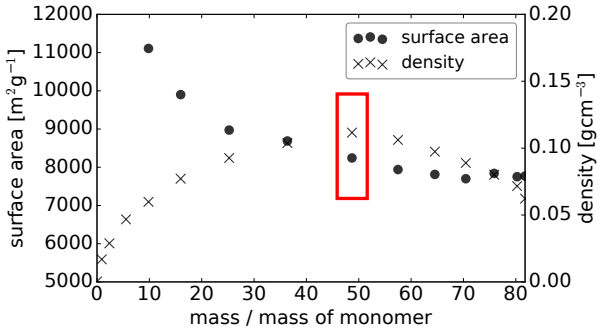



Figure A.11: Example of incremental cluster property measurements applied to a PAF-1 cluster, showing the change in density (crosses) and gravimetric surface area (circles) with increasing fraction of the cluster considered. The red box highlights the cluster fraction of maximum density, for which the surface area is taken to be the best approximation for that of the bulk material.

## A.8 Random-sampling method for analysis of compressed systems

The compression simulations produced one connected cluster that exceeded the computer memory requirements for analysis using RASPA2 and Zeo++, so we analyzed smaller random samples of the simulated system using the following procedure:

1. At the desired time step the full simulation box was reverse-mapped to the atomistic representation without structural relaxation.

2. Random points within the simulation box were then selected and for each point all atoms within a box centered around that point with dimensions equal to some

percentage of the full simulation box dimensions were extracted (the effect of the sampled box size was analyzed below).

3. Properties (e.g. density, pore surface area, pore volume, pore size distribution and X-ray diffraction pattern) were measured for each random sample. The reported value of each property was taken to be the mass-weighted average of the measured value from each sample.

Figures A.12 and A.13 show that using two different sampling box sizes, comprising 15% and 30% of the total simulation cell, respectively, in this analysis yields essentially the same results for the pore surface area, indicating that the results for the smaller box size are representative of the simulation cell. Using the smaller box size generally affords slightly higher density samples because the algorithm described above tends to avoid large void spaces. For the data in Figures A.12 and A.13, a similar incrementing algorithm to the one used above for clusters was applied and shows approximately linear relationship between the total pore surface area and mass or volume, which indicates that the gravimetric and volumetric surface area has converged to a limiting value for quite small samples of the framework.

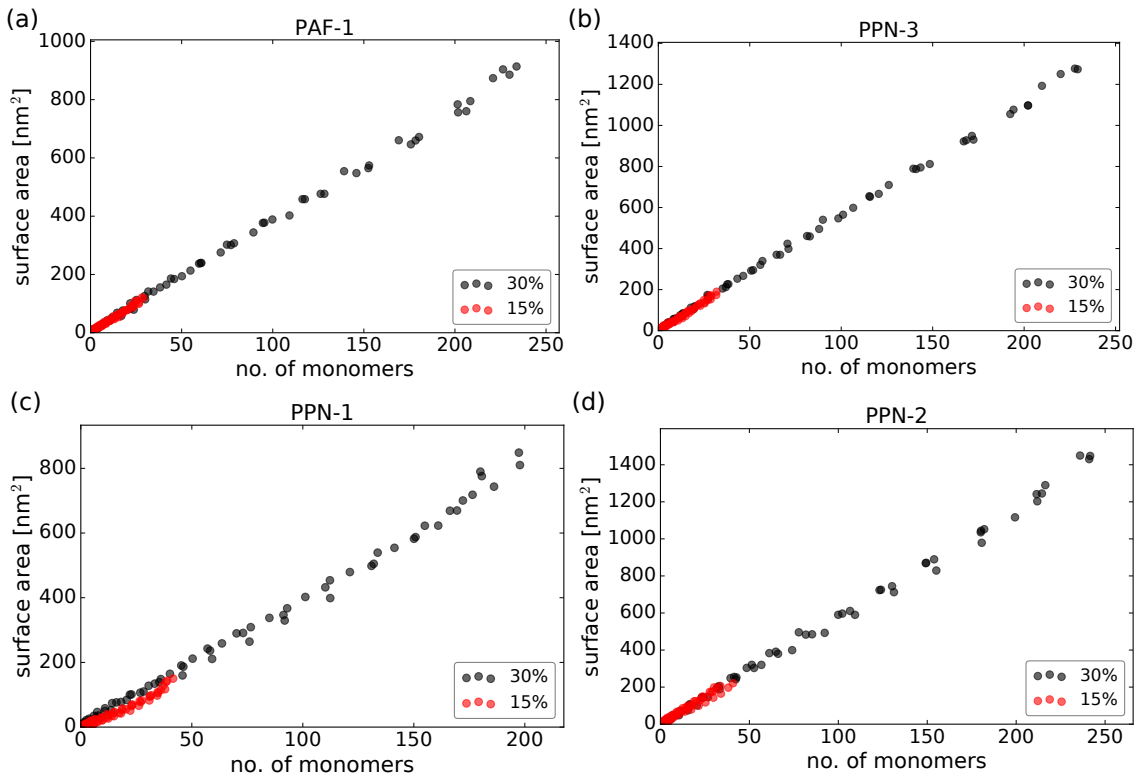



Figure A.12: Absolute surface area as a function of number of monomers for each volume increment of each random sample for all four PAFs with both sampling box sizes.

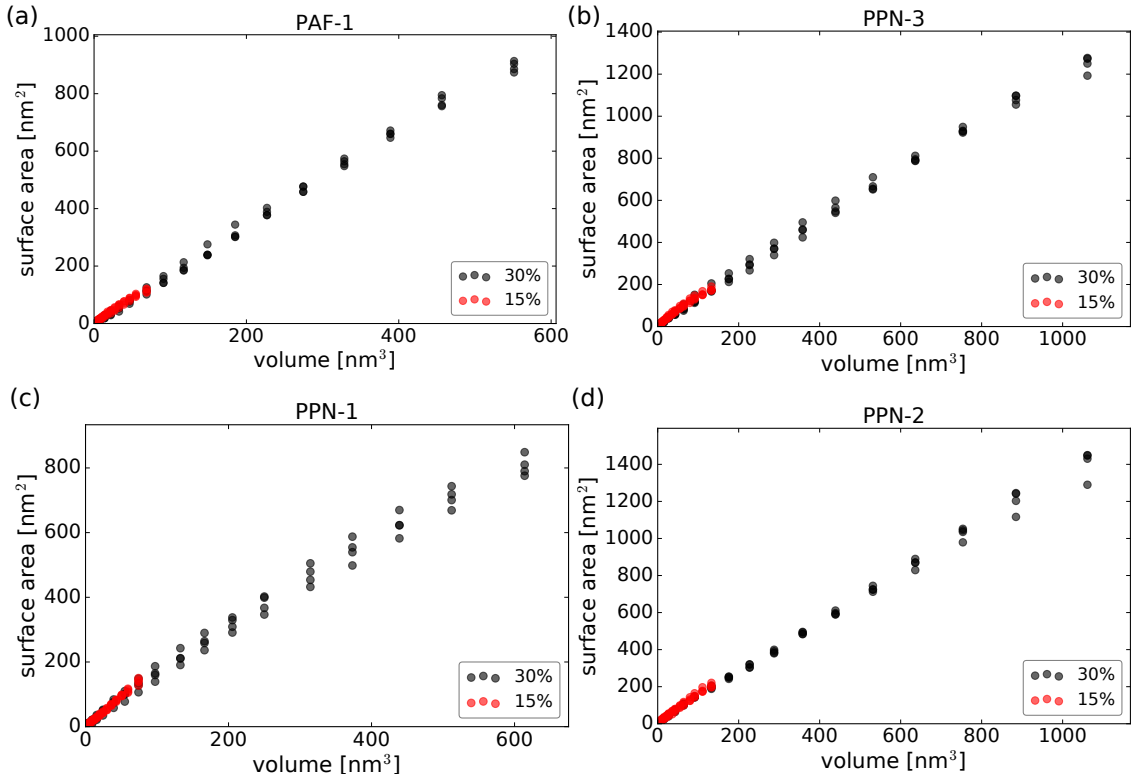



Figure A.13: Absolute surface area as a function of box volume for each volume increment of each random sample for all four PAFs with both sampling box sizes.

## A.9 Energies of misbound PAF monomer pairs

From the core-core RDFs in Figure 3.4 we defined misbound monomer pairs as any pair within the simulation configuration with a core-to-core distance less than $r_{\text{bind}} - 2\sigma_{\text{bind}}$, where $r_{\text{bind}}$ is the position of the most prominent peak in the core–core RDF, which corresponds to the core–core distance between two irreversibly bound monomers, and $\sigma_{\text{bind}}$ is the peak width. This definition gives an approximate means of quantifying misbinding. For all misbound monomer pairs present in the final configuration of simulations run with primary and secondary interactions, the contribution to the interaction energy between pairs of monomers from primary interactions and from secondary interactions were separately calculated; a normalized distribution of the pair energy contributions is shown in Figure A.14. For PAF-1 and PPN-3 the distributions for both types of interactions generally overlap, while for PPN-1 and PPN-2 there is a clear distinction between the two types of interactions, with the distribution of secondary interaction energies skewed to significantly more negative values.

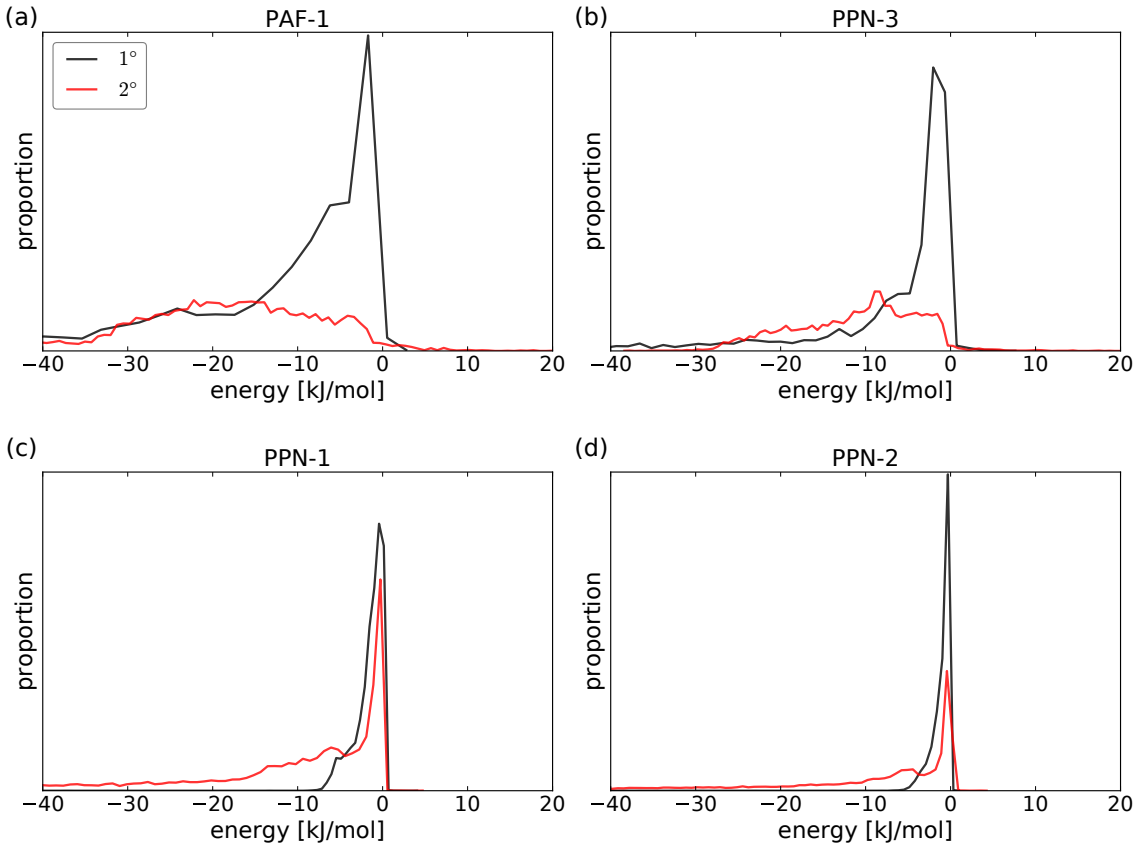



Figure A.14: The distribution of pair-wise energies of misbound monomer pairs due to primary and secondary interactions for (a) PAF-1, (b) PPN-3, (c) PPN-1, and (d) PPN-2.

## A.10 Kinetics of PAF growth

The kinetics of PAF growth in the simulations was quantified for all four PAFs by measuring the number of free monomers and average cluster size as a function of time, as shown in Figures A.15 and A.16, respectively. Figure A.15 shows similar rates of free monomer loss in the presence of primary and primary + secondary interactions, with the rate when secondary interactions are added being slightly faster. Introduction of the steric potential slows the rate of free monomer loss, as expected due to the presence of a barrier to primary and secondary binding. The average cluster size versus time in Figure A.16 shows a similar decrease in growth rate when steric interactions are present. In all four cases the growth of the average cluster size with secondary interactions is smaller than without secondary interactions, which is an effect of the formation of more compact clusters due to the secondary interactions, as more compact clusters are less likely to collide.

We rationalize the approximately linear behaviour of the average cluster size with time at early times (Figure 3.6) in terms of the Smoluchowski coagulation model,[193,194] which is a general model for describing irreversible coagulation. The Smoluchowski coagulation model assumes that assembly occurs by bimolecular collisions of clusters and is irreversible, the system is well-mixed and can be fully specified by the distribution of the number of monomers in each cluster.[195] These assumptions are expected to hold reasonably well in our simulations. A linear relationship between the average cluster size and time, as observed to a good approximation in all of our simulations at early times, indicates that the bimolecular coagulation rate constant $K_{ij}$ between a cluster containing $i$ units and a cluster containing $j$ units is constant, i.e. $K_{ij} = K$. Furthermore, the loss of free monomers with time follows the exponential decay expected for a constant coagulation rate constant $K$.[195] Assuming the coagulation rate constant can be written as

$$K = \alpha k_{\mathrm{s}}, \tag{A.16}$$

where $\alpha$ is the sticking probability upon collision and $k_{\mathrm{s}}$ is the bimolecular collision rate constant. Assuming that the clusters can be approximated as hard spheres of radius $R_i$ and $R_j$, respectively, which undergo Brownian motion with diffusion coefficients $D_i$ and $D_j$, respectively, the collision rate constant is

$$k_{\mathrm{s}} = 4\pi \left(R_i + R_j\right)\left(D_i + D_j\right). \tag{A.17}$$

To confirm the size-independent behaviour at early times, $k_{\mathrm{s}}$ was calculated for a random selection of clusters extracted at 100 ns from constant-volume simulations for all PAFs for all interaction types. The radius of each cluster in our simulations was taken to be its radius of gyration. Each cluster was placed on its own in a very large simulation box, in which its motion was simulated for ∼1 ns using an identical Langevin dynamics algorithm and parameters to those applied in the assembly simulations. The translational diffusion coefficient was calculated from the MSD of the cluster center of mass using the same method described in Section A.3. Diffusion coefficients as a function of number of monomers in a cluster are shown in Figure A.17. The radius of

gyration of each cluster (shown in Figure A.18) was calculated as

$$R_{\mathrm{g}} = \sqrt{\lambda_1^2 + \lambda_2^2 + \lambda_3^2}, \tag{A.18}$$

where $\lambda_i$ is the $i^{\mathrm{th}}$ eigenvalue of the gyration tensor $\boldsymbol{Q}$,

$$Q_{mn} = \frac{1}{N} \sum_{i=1}^{N} r_m^{(i)} r_n^{(i)}, \tag{A.19}$$

where $N$ is the number of monomers in a cluster and $r_m^{(i)}$ and $r_n^{(i)}$ are the $m^{\mathrm{th}}$ and $n^{\mathrm{th}}$ cartesian coordinates of the position vector $\boldsymbol{r}^{(i)}$ of the $i^{\mathrm{th}}$ monomer, respectively. $k_{\mathrm{s}}$ was calculated for all pairs of clusters using Equation A.17 and is plotted as a function of the total number of monomers in the pair of clusters, $N_i + N_j$, in Figure A.19, which shows a large spread of points with a relatively weak relationship with cluster size, indicating approximate size-independent behaviour. The bimolecular coagulation rate constant $K$, was obtained from the slope of the data in Figure 3.6 using the equation for the average cluster size $\langle \mathrm{N(t)} \rangle$ versus time from the Smoluchowski coagulation model with constant $K$,

$$\langle \mathrm{N(t)} \rangle = \langle \mathrm{N(0)} \rangle + \frac{1}{2} K C_1 t, \tag{A.20}$$

where $C_1$ is the concentration of monomers in solution. The value of $K$, mean value of $k_{\mathrm{s}}$ in Figure A.19, and the sticking probability $\alpha$ calculated from Equation A.16 are reported in Table 3.1 for each PAF and for each interaction type.

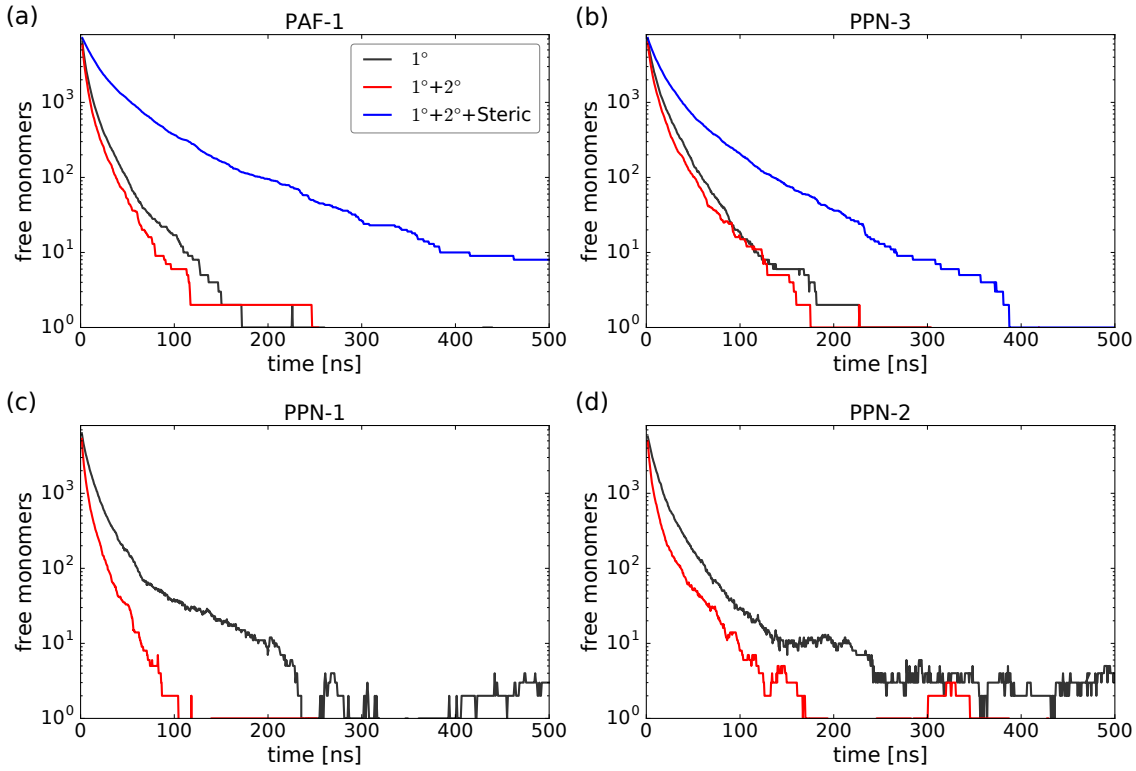



Figure A.15: Number of free monomers versus time for (a) PAF-1, (b) PPN-3, (c) PPN-1, and (d) PPN-2 for all types of interactions.

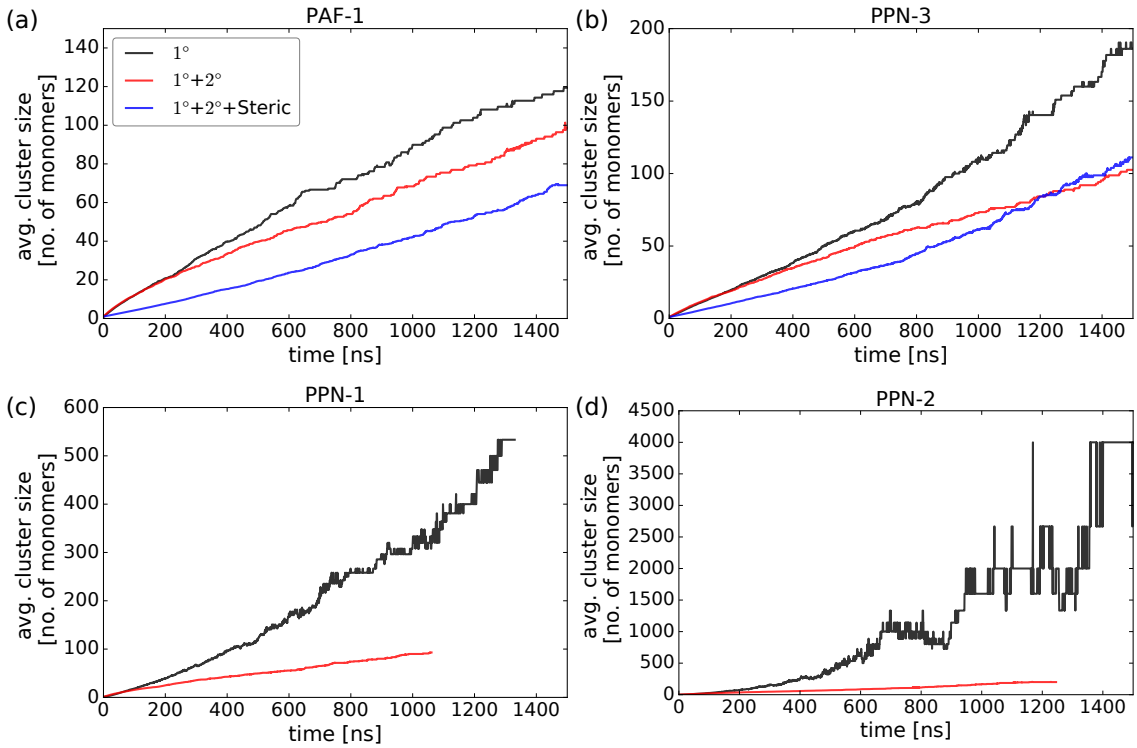



Figure A.16: Average cluster size for (a) PAF-1,(b) PPN-3, (c) PPN-1, and (d) PPN-2 for all types of interactions.

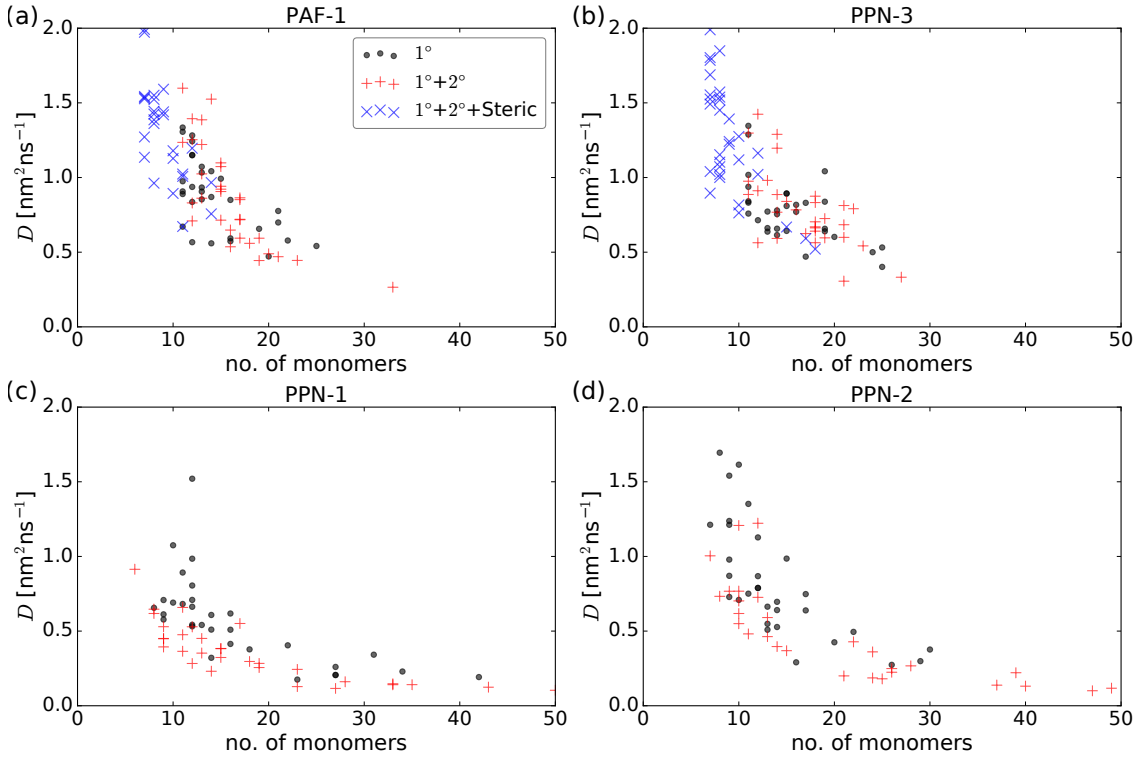



Figure A.17: Translational diffusion coefficient of clusters selected after 100 ns of framework growth for (a) PAF-1,(b) PPN-3, (c) PPN-1, and (d) PPN-2 for all types of interactions.

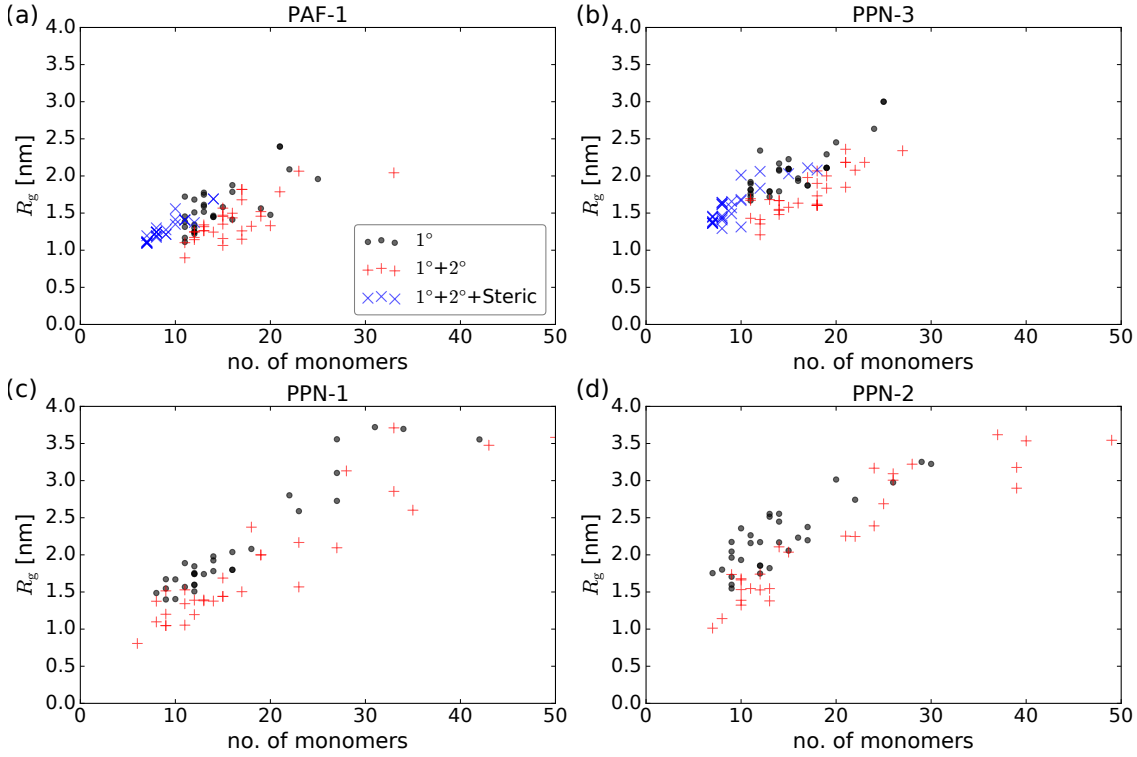



Figure A.18: Radius of gyration of clusters selected after 100 ns of framework growth for (a) PAF-1,(b) PPN-3, (c) PPN-1, and (d) PPN-2 for all types of interactions.

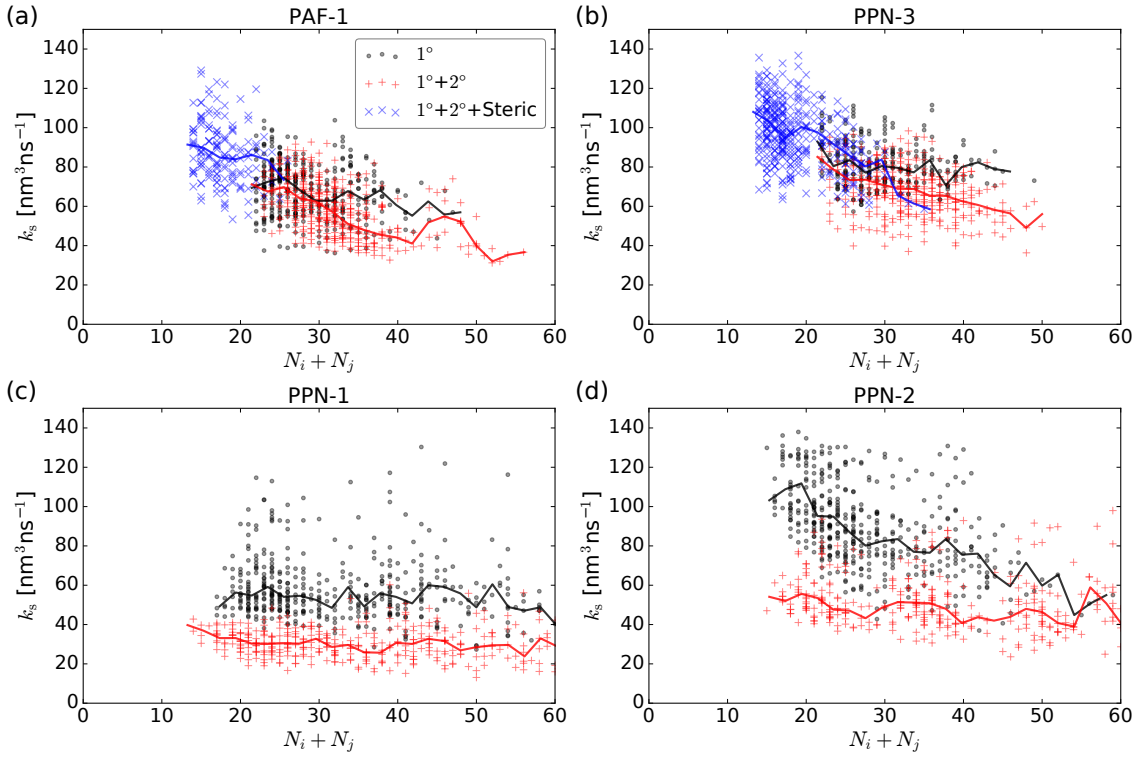



Figure A.19: Bimolecular collis ion rate constant, $k_s$, of clusters selected after 100 ns of framework growth for (a) PAF-1,(b) PPN-3, (c) PPN-1, and (d) PPN-2 for all types of interactions. Lines represent the mean of $k_s$ as a function of $N_i + N_j$.

## A.11 Relative shape anisotropy of PAF clusters

The relative shape anisotropy ($\kappa^2$) of each cluster extracted from the final configuration in the constant-volume simulations of all four PAFs is shown in Figure A.20 as a function of the radius of gyration ($R_g$), which is a measure of cluster size. $\kappa^2$ was calculated from the eigenvalues ($\lambda_1$, $\lambda_2$ and $\lambda_3$) of the gyration tensor as

$$\kappa^2 = 1 - 3(\lambda_1\lambda_2 + \lambda_1\lambda_3 + \lambda_2\lambda_3)/R_g^4. \tag{A.21}$$

$\kappa^2$ equals zero for a sphere and one for a linear rod. Figure A.20 shows a wide spread of values for all PAFs for all types of interactions, which suggests a limited effect on shape anisotropy due to cluster size, monomer type or interaction type.

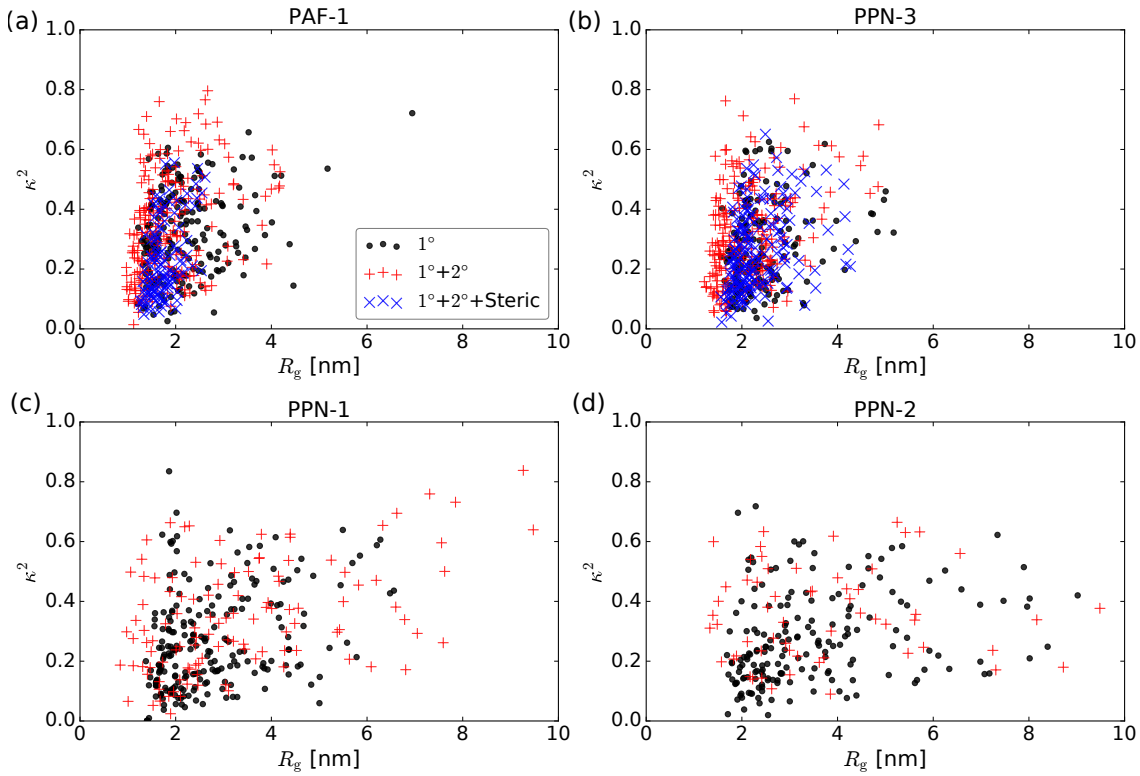



Figure A.20: Relative shape anisotropy $\kappa^2$ as a function of the radius of gyration $R_g$ of each cluster identified for (a) PAF-1,(b) PPN-3, (c) PPN-1, and (d) PPN-2 for all types of interactions.

## A.12 Porosity in constant-volume simulations

Table A.8: Nitrogen-accessible surface areas and pore volumes obtained from constant-volume simulations reported in the main text. Primary interactions are denoted by 1°, secondary interactions by 2°. Errors are one standard deviation. Simulated values for a perfectly ordered diamondoid (dia) network are also shown.

| material | expt. | dia. | 1° only | 1°+2° | 1°+2°+steric |
|---|---|---|---|---|---|
| | | | Volumetric Surface Area [$m^2\,cm^{-3}$] | | |
| PAF-1 | - | 1930 | 1390 ± 550 | 1720 ± 240 | 2250 ± 72 |
| PPN-3 | - | 1300 | 1350 ± 360 | 1640 ± 250 | 1400 ± 310 |
| PPN-1 | - | 730 | 890 ± 420 | 760 ± 420 | - |
| PPN-2 | - | 750 | 450 ± 390 | 700 ± 450 | - |
| | | | Pore Volume [$cm^3g^{-1}$] | | |
| PAF-1 | 0.89–1.44[a] | 1.45 | 4.63 ± 2.86 | 1.41 ± 0.77 | 1.79 ± 0.20 |
| PPN-3 | 1.70[159]/1.69[b]/2.67[c] | 3.38 | 5.06 ± 2.35 | 2.02 ± 0.68 | 4.53 ± 1.12 |
| PPN-1 | 0.45[159]/0.91[d] | 12.36 | 12.72 ± 6.26 | 8.79 ± 4.52 | - |
| PPN-2 | 1.26[159] | 9.48 | 30.52 ± 14.13 | 12.01 ± 5.42 | - |
| | | | Pore Volume Fraction [%] | | |
| PAF-1 | - | 47 | 65 ± 17 | 42 ± 14 | 52 ± 6 |
| PPN-3 | - | 66 | 70 ± 17 | 52 ± 12 | 67 ± 8 |
| PPN-1 | - | 86 | 82 ± 10 | 78 ± 12 | - |
| PPN-2 | - | 85 | 91 ± 9 | 84 ± 12 | - |

[a] Range of experimental data.[165]
[b] Synthesized at 80 °C.[161]
[c] Synthesized at room temperature in a different solvent (DMF/THF).[28]
[d] Synthesized under palladium catalyzed conditions.[161]

## A.13 Porosity in compression simulations

Table A.9: Nitrogen-accessible surface areas, densities and pore volumes obtained from compression simulations reported in the main text. Primary interactions are denoted by 1°. Errors are one standard deviation. "1° only" and "full fast" simulations were run at the fast rate with 1° interactions only and the full set of interactions, respectively, and "full slow" simulations were run at the slow rate with the full set of interactions (where the slow rate was half the fast rate). Simulated values for a perfectly ordered diamondoid (dia) network are also shown.

| material | expt. | dia. | 1° only | full fast | full slow |
|---|---|---|---|---|---|
| | | Gravimetric Surface Area [$m^2\,g^{-1}$] | | | |
| PAF-1 | 5600[27]/3639[29] | 5930 | 8460 ± 440 | 9100 ± 330 | 7150 ± 720 |
| PPN-3 | 2840[159]/3180[a]/4221[b] | 6623 | 8340 ± 350 | 8990 ± 460 | 9020 ± 390 |
| PPN-1 | 1249[159]/1470[c] | 10560 | 10920 ± 440 | 5460 ± 330 | 6200 ± 320 |
| PPN-2 | 1764[159] | 8340 | 10670 ± 370 | 7300 ± 500 | 7330 ± 680 |
| | | Volumetric Surface Area [$m^2\,cm^{-3}$] | | | |
| PAF-1 | - | 1930 | 1270 ± 80 | 1790 ± 90 | 2050 ± 80 |
| PPN-3 | - | 1300 | 1440 ± 50 | 1380 ± 80 | 1400 ± 80 |
| PPN-1 | - | 730 | 1160 ± 100 | 1830 ± 60 | 1800 ± 90 |
| PPN-2 | - | 750 | 1100 ± 70 | 1590 ± 140 | 1580 ± 120 |
| | | Density [$gcm^{-3}$] | | | |
| PAF-1 | - | 0.324 | 0.15 ± 0.01 | 0.20 ± 0.01 | 0.29 ± 0.03 |
| PPN-3 | - | 0.196 | 0.17 ± 0.01 | 0.15 ± 0.01 | 0.16 ± 0.01 |
| PPN-1 | - | 0.069 | 0.11 ± 0.01 | 0.34 ± 0.02 | 0.29 ± 0.01 |
| PPN-2 | - | 0.089 | 0.10 ± 0.01 | 0.22 ± 0.03 | 0.22 ± 0.03 |
| | | Pore Volume [$cm^3g^{-1}$] | | | |
| PAF-1 | 0.89–1.44[d] | 1.45 | 4.70 ± 0.42 | 2.97 ± 0.35 | 1.61 ± 0.22 |
| PPN-3 | 1.70[159]/1.69[a]/2.67[b] | 3.38 | 3.78 ± 0.28 | 4.49 ± 0.41 | 4.44 ± 0.51 |
| PPN-1 | 0.45[159]/0.91[c] | 12.36 | 7.26 ± 0.96 | 1.17 ± 0.17 | 1.59 ± 0.13 |
| PPN-2 | 1.26[159] | 9.48 | 7.40 ± 0.70 | 2.66 ± 0.52 | 2.69 ± 0.44 |
| | | Pore Volume Fraction [%] | | | |
| PAF-1 | - | 47 | 70 ± 2 | 58 ± 3 | 46 ± 3 |
| PPN-3 | - | 66 | 76 ± 2 | 68 ± 2 | 68 ± 3 |
| PPN-1 | - | 86 | 76 ± 3 | 39 ± 4 | 46 ± 2 |
| PPN-2 | - | 85 | 65 ± 2 | 56 ± 5 | 57 ± 4 |

[a] Synthesized at 80 °C.[161]
[b] Synthesized at room temperature in a different solvent (DMF/THF).[28]
[c] Synthesized under palladium catalyzed conditions.[161]
[d] Range of experimental data.[165]

## A.14 X-ray diffraction patterns

X-ray diffraction (XRD) patterns were calculated using RASPA2[131] for all clusters extracted from the constant-volume simulations without applying the incremental method described above and averaged. For the compression simulations, the XRD patterns were calculated for each random sample selected using the methods described earlier and averaged.

The XRD patterns of structures extracted from the constant-volume simulations and compressed simulations provide further evidence of the high degree of disorder in the simulated structures when compared with the XRD patterns of modelled diamondoid structures of each PAF (purple lines in Figure A.21 and Figure A.22). Similarly to previous simulations of PAF-1,[165] much better agreement with experiment was obtained with XRD patterns of amorphous structures than the crystalline structures. The agreement between XRD patterns from our model and powder X-ray diffraction (PXRD) patterns from experiments is poor at large angles, which suggests there is a higher degree of interpenetration (larger angles implies atoms at shorter distances) in the experimental systems.

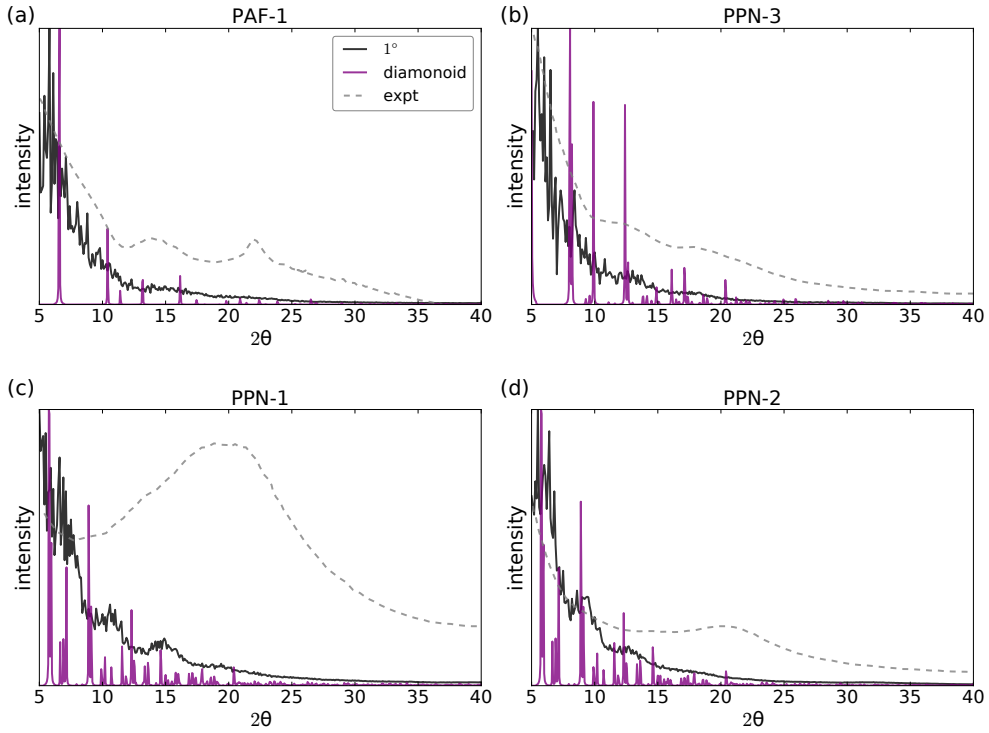



Figure A.21: Simulated X-ray diffraction patterns from constant-volume simulations, previously reported experiments[27,159] and modelled diamondoid frameworks of (a) PAF-1, (b) PPN-3, (c) PPN-1, and (d) PPN-2.

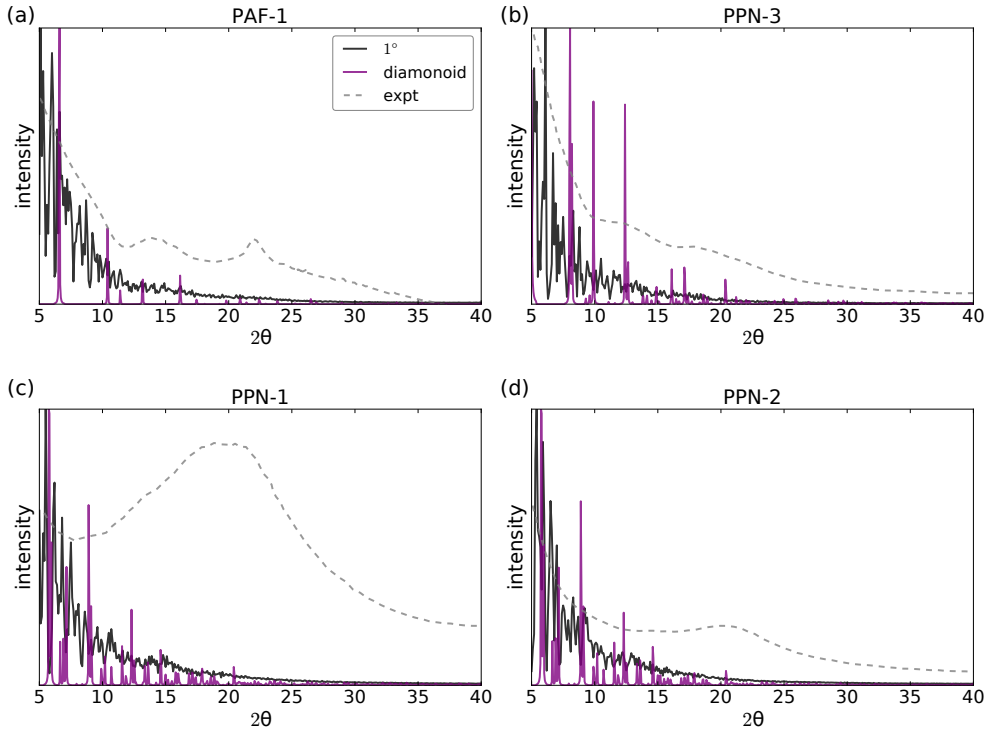



Figure A.22: Simulated X-ray diffraction patterns from compression simulations, previously reported experiments[27,159] and modelled diamondoid frameworks of (a) PAF-1, (b) PPN-3, (c) PPN-1, and (d) PPN-2.

## A.15 Unbound terminal sites in PAFs

In the experiments used to synthesize PAFs, an unbound terminal site (or dangling bond) is either the unreacted bromine group in PAF-1 or PPN-3 or the unreacted acetylene in PPN-1 or PPN-2. In all reported syntheses of these four materials, FTIR and NMR show no evidence of either unreacted terminal site at the end of the reaction. In our simulations, an unbound terminal site was defined as a primary binding site with no other primary site within some cut-off binding distance ($r_\mathrm{b}$), which are shown in Table A.10 and were set to encompass the possible distances of the primary binding interaction subject to thermal fluctuations. Figure A.23 shows that the number of unbound terminal sites does not depend strongly on the type of interaction for PAF-1 and PPN-3. For PPN-1 and PPN-2, the introduction of secondary interactions increased the number of misbound monomers, which increased the proportion of unbound monomers in PPN-1 and PPN-2. The same was not true for PAF-1 and PPN-3 as there were less misbound monomers for those PAFs. Similarly, Figure A.24 shows that compression of the system had a small effect. Drops in the proportion of unbound terminals occurred when clusters came together, but due to the high degree of disorder and the irreversibility of the primary interactions, many unbound terminal sites still existed. Note the rate of compression had very little effect.

Table A.10: Cut-off distances ($r_\mathrm{b}$) used to determine if two sites were bound.

| material | $r_\mathrm{b}$ [Å] |
|---|---|
| PAF-1 | 5.42 |
| PPN-1 | 3.42 |
| PPN-2 | 3.42 |
| PPN-3 | 5.42 |

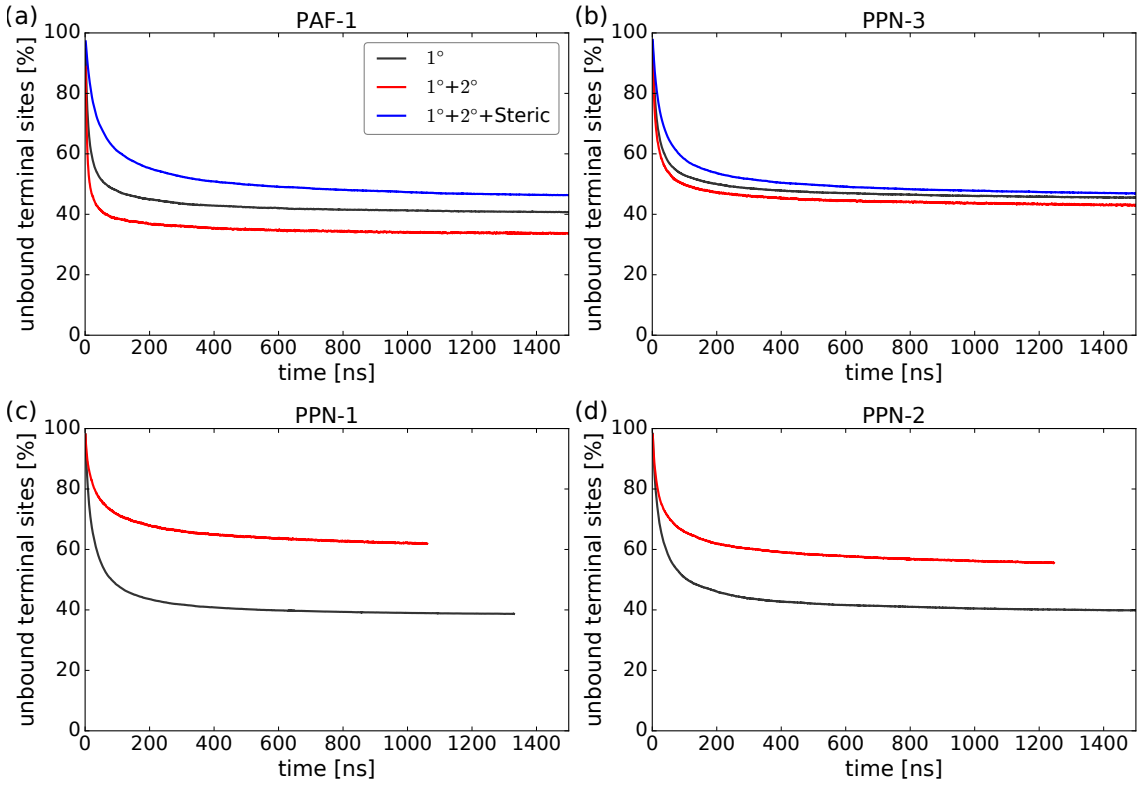



Figure A.23: Percent of unbound terminal sites in all four PAFs, for each type of interaction in constant-volume simulations.

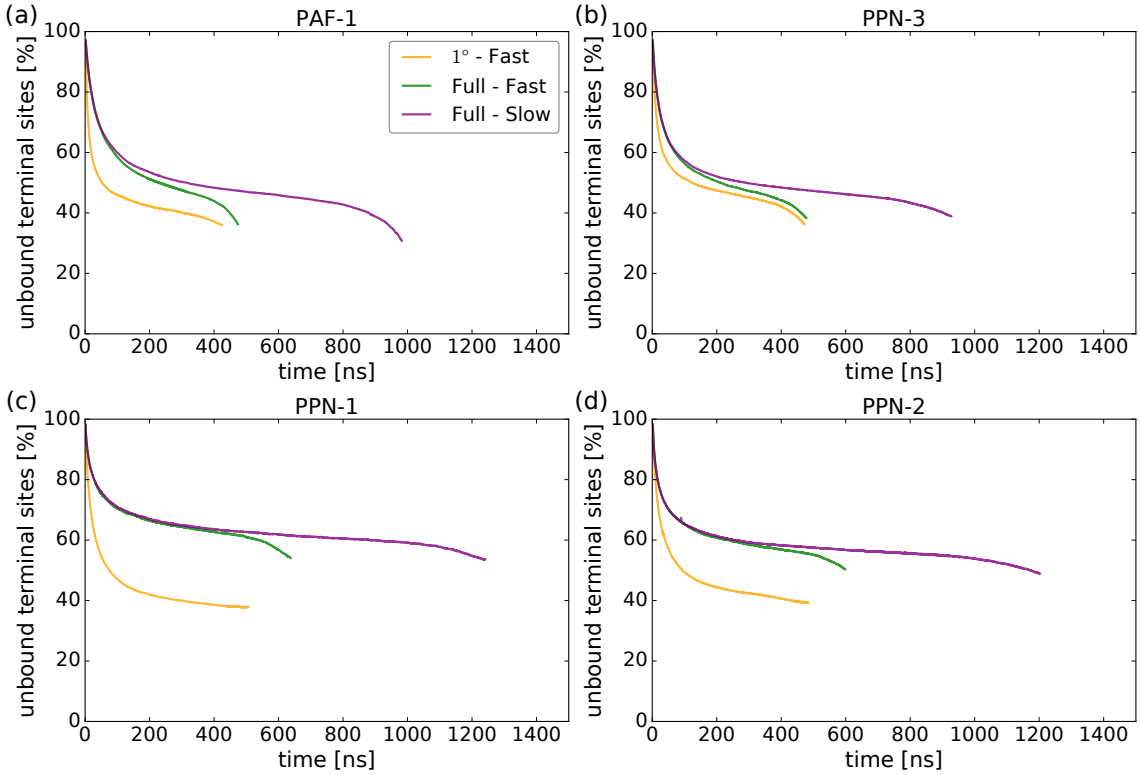



Figure A.24: Percent of unbound terminal sites in all four PAFs, in the presence of primary interactions only or the full set of interactions in compression simulations.

# APPENDIX B

# Supporting information: High-throughput screening of metal–organic frameworks for macroscale heteroepitaxial alignment

## B.1 Identifying structures with appropriate chemistry

The first step of the screening process tests whether a MOF crystal structure contains the desired substrate binding site anywhere within its crystal structure. Figure B.1 shows the binding site used when screening MOFs for aligned growth from $Cu(OH)_2$ (a metal-bound carboxylate group). First, for simplicity we determine whether the crystal structure contains copper and no other transition or alkali metals. Second, we search a $2 \times 2 \times 2$ supercell of the MOF crystal structure for the desired binding site using the following substeps (see Figure B.1 for naming convention):

i. Find an oxygen (O1) within a specified cut-off distance of a copper atom (M1) and a carbon atom (C).

ii. Check that C has another oxygen (O2) within a specified cut-off distance.

iii. Check that O2 has a copper atom (M2) within a specified cut-off distance.

If all substeps are passed at least once for a given structure, the structure moves on to the next step of the screening process. Two atoms are classified as bonded if they are within a distance given by the covalent radii of the two atoms + a skin distance of 0.3 Å.[215] We have confirmed that this step of the screening process is robust in general, but the large skin distance means that the bonding criterion is lenient. The definition above requires that both oxygens be bound to copper atoms, but not that those copper

atoms be different. Therefore the search is not limited only to the copper paddle wheel structure.

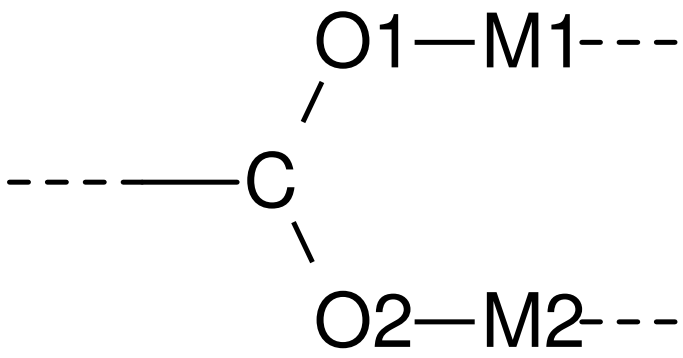



Figure B.1: Schematic of the desired binding site, where M is copper. Labels correspond to those described in Section B.1.

## B.2 Parameterization of the lattice-matching algorithm

Using the lattice-matching algorithm[209,210] implemented in the pymatgen software,[216] supercells are tested up to some maximum multiple of the unit cell areas of the MOF ($A_{\mathrm{MOF}}$) and substrate ($A_{\mathrm{sub}}$). The area ratio matching criterion determines the multiples $m$ and $n$ of the MOF and substrate unit cells, respectively, that are tested using

$$\left| \frac{A_{\mathrm{MOF}}}{A_{\mathrm{sub}}} - \frac{n}{m} \right| < \alpha_{\mathrm{AR}}, \tag{B.1}$$

where $\alpha_{\mathrm{AR}}$ is the maximum area ratio tolerance, which was set to 15%. For all $m$ and $n$ that pass this test, an array of possible supercells are produced by the lattice-matching algorithm. The mismatch tolerances in the length of and the angle between the supercell vectors were set to 10% and 2%, respectively. Figure B.2 shows that the maximum $\Delta$IB for a small set of MOFs is robust to increasing values of all tolerances. Decreases in the length mismatch tolerance or $\alpha_{\mathrm{AR}}$ from the chosen value result in substantial changes in the maximum $\Delta$IB as supercells with high atomic overlap are excluded in the lattice-matching algorithm.

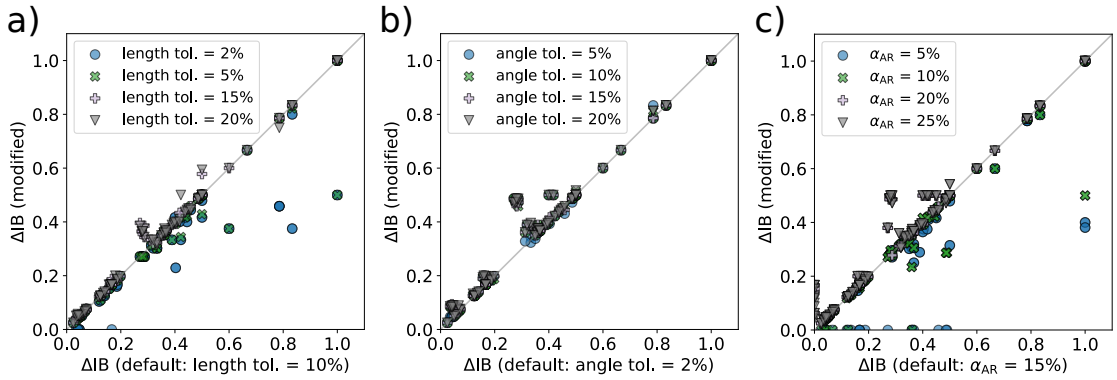



Figure B.2: Parity plots of the maximum ΔIB for each binding plane for all Miller planes for MOF crystal structures in a small data set (including structures from the experimental validation data set and CORE MOF database) comparing the effects of varying the (a) length tolerances, (b) angle tolerances, and (c) max area ratio tolerances ($\alpha_{\mathrm{AR}}$) in the lattice-matching algorithm. Modified values are given in the figure legends, default values used for screening in the main paper are given in the $x$-axis titles. Modified parameters may change whether a particular Miller plane has any matching lattices for which a ΔIB can be calculated. Cases with no matching lattices were assigned maximum ΔIB values of zero.

### B.2.1 Upper bound on unit-cell expansions

The smaller unit cell of the (010) slab of $Cu(OH)_2$ compared with most MOFs (due to their low densities) means that in most cases the number of unit-cell multiples required to build a matching supercell at the MOF–substrate interface will be much higher for the substrate than for the MOF. An upper bound on the match area is required because matching lattices can always be found for large enough supercells without necessarily yielding favorable binding between atoms at the interface. Limiting the maximum unit-cell multiple also makes the algorithm more efficient because it reduces the number of lattices that require testing. We chose the maximum area of the MOF and substrate supercells to be equal to the smaller of 9× the MOF unit-cell area or 90× the substrate unit-cell area, since this allowed all MOFs in the experimental validation data set to pass the lattice-matching step (required for validation of all steps of the screening process) and provided robust results for the full screening process, while remaining computationally efficient. Figure B.3 shows that a large increase in the maximum number of MOF and substrate unit-cell multiples does not significantly change the maximum ΔIB for a test set of MOF structures. The calculations with the larger maximum area do not offer any substantial changes in the maximum ΔIB for interfaces

already tested with the default parameters (although the maximum ASO (not shown) for some of the structures does change) and are much more computationally expensive. Note that increasing the maximum area limit does allow for new interfaces to be tested, but none of those interfaces were found to have high maximum ΔIB values (Figure B.3).

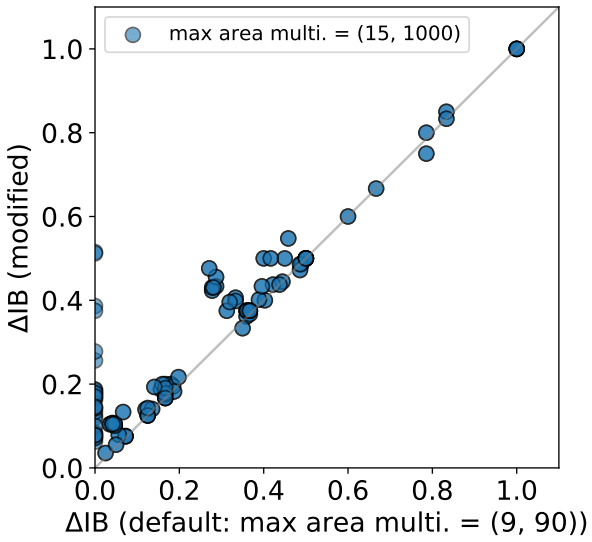



Figure B.3: Parity plot of the maximum ΔIB for each binding plane for all Miller planes for MOF crystal structures in a small data set (which includes structures from the experimental validation data set and CORE MOF database) when the MOF maximum area is capped at the smaller of 9× the MOF unit cell area or 90× the substrate unit cell area compared with the smaller of 15× the MOF unit cell area or 1000× the substrate unit cell area. Modified parameters may change whether a particular Miller plane has any matching lattices for which a ΔIB can be calculated. Cases with no matching lattices were assigned maximum ΔIB values of zero.

## B.3 Building binding planes

Each interface is made up of two binding planes, which are 2D structures containing only the binding atoms on each side of the interface. In this work, the binding atoms were the specified metal (Cu for copper hydroxide) for the substrate, and the carboxylate oxygens for the MOF. The extraction of binding planes from the 3D crystal structures differs for the MOF and substrate due to the more complex topology of the MOF structure. The algorithms used to extract the binding planes for the substrate and MOF are outlined in the following sections.

### B.3.1 Determining bonded components

A neighbor list approach (built into the Atomic Simulation Environment (ASE) code)[215] is used to identify bonded components within a crystal structure. Two atoms are considered bonded if the distance between them is less then the sum of their covalent radii plus the skin distance (0.3 Å). Each atom is assigned to a molecule based on whether it is part of a connected network of neighbors (bonded atoms). The atoms comprising each organic linker are determined by identifying the separate connected networks that are formed after all metals are removed from the structure. In some cases, the large skin distance used gives rise to unphysical bonds between the carbons in a carboxylate functionality and the associated copper atom (i.e. a bond between C and M1 in Figure B.1), which results in overcounting of the number of bonds broken at an interface (see Section B.4). For this reason, any copper–carbon or hydrogen-containing broken bonds are not counted during the calculation of the ΔIB. For simplicity we have assumed all copper–carbon bonds correspond to the aforementioned unphysical case. Indeed, none of the MOFs in the screened databases that passed the lattice-matching step contain a copper–carbon bond.

### B.3.2 Substrate interfaces

The following steps are used to produce binding planes of the substrates:

i. Read in the substrate crystal structure.

ii. Build a specific 3D slab of the substrate for the target Miller plane (this could be generalized to any Miller plane) using the SlabGenerator function in pymatgen.[216,219] Slabs are produced such that the $z$ axis is normal to the surface. A slab has cell parameters (specifically $a$, $b$, and $\gamma$) that define the unit cell lattice vectors used in the lattice-matching algorithm. Conversion of slabs to primitive structures was not attempted.

iii. Remove all atoms that are not substrate binding atoms (i.e. all non-copper atoms when testing $Cu(OH)_2$).

iv. Slice the slab at the binding atom with the highest $z$ coordinate and collect all binding atoms within a buffer zone of ±0.8 Å to produce a 2D binding plane.

### B.3.3 MOF interfaces

The following steps are used to produce binding planes for all MOFs:

i. Read in the MOF crystal structure. For consistency, all crystal structures are converted to primitive symmetry by pymatgen.[216]

ii. Build 3D slabs of the MOF structure for all structurally distinct Miller planes with indices $-1 \leq h, k, l \leq 1$ using the SlabGenerator function in pymatgen.[216,219] Slabs are produced such that the $z$ axis is normal to the surface. A slab has cell parameters (specifically $a$, $b$, and $\gamma$) that define the unit cell lattice vectors used in the lattice-matching algorithm. Conversion of slabs to primitive structures was not attempted.

iii. For each slab, duplicate the structure in the $z$ direction and search for a specific binding functionality. A search for metal-bound carboxylate functionalities like that in step 1 of the overall screening process is carried out, but specifically for binding groups pointing in the $z$ direction towards the interface (see Figure B.4). This is determined by the angle between the normal to the plane containing the carbon and two oxygen atoms in a carboxylate group ($\boldsymbol{n}_{\mathrm{OCO}}$) and the $z$ axis: any functionalities with angles less than 30° are excluded. Finally, we make sure that the $z$ coordinates of the two oxygens in the carboxylate group are within approximately one oxygen covalent radius (0.6 Å) of one another ($d_{\mathrm{OO}}$ in Figure B.4). If all of these criteria are satisfied, then the crystal structure will be sliced at the $z$ coordinate of this binding functionality (in practice the structure is sliced at the maximum of the $z$ coordinates of the two oxygen binding atoms). Before slicing the crystal structure, merge all carboxylate functionalities that also pass all the above criteria and are within a range of $z$ values (the "slicing buffer", set to 1.2 Å, which is approximately twice the oxygen covalent radius) of the first carboxylate group into one collection of binding atoms and one slicing position. Changing the parameters used to determine binding interfaces would result in changing the number and nature of the binding interfaces that are tested by our screening process. Nevertheless, physically reasonable variations of the parameters around the chosen values do not affect the results of the screening process for a test data set

that includes the experimental validation data set, as shown in Figure B.5.

iv. For each slicing position (which may be a collection of more than one carboxylate functionality), if organic bonds are broken during the slicing, enforce a cleaning procedure of the binding interface that introduces missing linker defects, which are produced when the binding plane slices through part of an organic linker. If any linker containing a binding site oxygen atom is sliced in this process, the binding plane is discarded.

v. If there are no nonbinding atoms in the cleaned 3D binding interface with a $z$ coordinate less than the $z$ coordinate of the slicing position (i.e. there are no overlapping atoms at the interface), extract the 2D binding plane from the cleaned 3D binding interface by extracting all binding atoms (carboxylate oxygens in our case). Any Miller plane may have possible binding planes at multiple slicing positions. Without the use of expensive energy calculations, it is not possible to determine which binding plane configuration is most likely to form an interface, and so all unique binding plane configurations for each Miller plane are tested. We exclude any binding planes that are duplicates, based on the number of binding atoms and their $x$ and $y$ coordinates. Only oxygens that match the carboxylate functionality criteria are extracted as part of the binding plane and each carboxylate oxygen atom is only considered in one binding plane per MOF Miller plane.

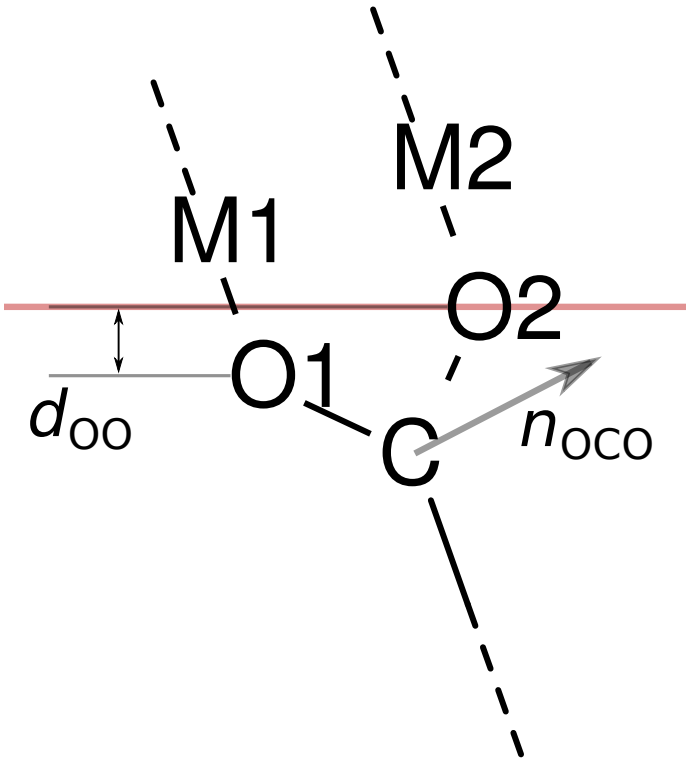



Figure B.4: Schematic of the definition of a binding site used for determining the binding interface, where M1 and M2 are copper atoms (note that M1 and M2 can be the same atom), $d_{\mathrm{OO}}$ is the difference between the $z$ coordinates of the two carboxylate oxygens, $\boldsymbol{n}_{\mathrm{OCO}}$ is the normal to the the plane containing the two oxygens and carbon in the carboxylate group, and the red line represents the plane where the MOF crystal structure is sliced.

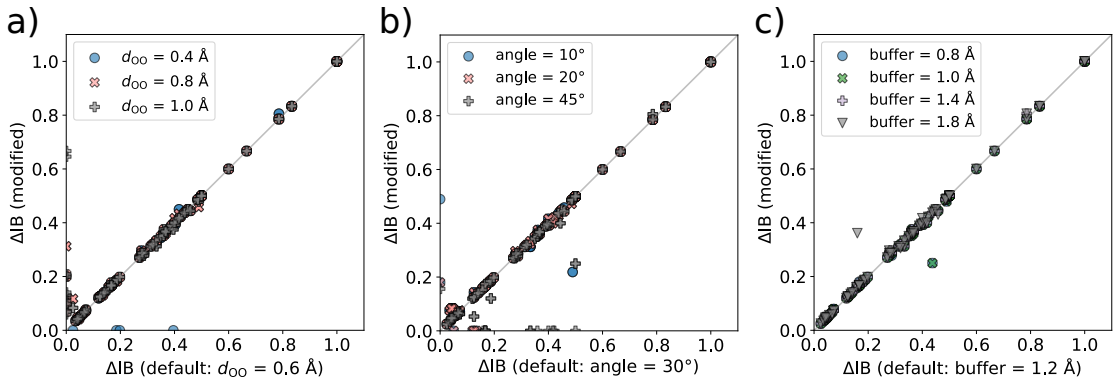



Figure B.5: Parity plots of the maximum $\Delta\mathrm{IB}$ for each binding plane for all Miller planes for MOF structures in a small data set (which includes structures from the experimental validation data set and CORE MOF databases) comparing the effects of the default parameters in the binding plane building algorithm and modified (a) $d_{\mathrm{OO}}$, (b) angle allowed between $\boldsymbol{n}_{\mathrm{OCO}}$ and the $z$ axis, and (c) slicing buffer. Modified values are given in the figure legends and default values used for screening in the main paper are given in the $x$-axis titles. Modified parameters may change whether a particular Miller plane has any matching lattices for which a $\Delta\mathrm{IB}$ can be calculated. Cases with no matching lattices were given maximum $\Delta\mathrm{IB}$ values of zero.

### B.3.4 Building supercells

Using the reduction scheme and lattice-matching algorithm,[209,210] a list of supercells of both the MOF and substrate unit cells with areas matching within some tolerance are calculated. For each pair of supercells of the MOF and substrate deemed to be matching, the unit cell atom positions are mapped to their respective supercells, using the MPInterfaces software,[221] which uses the pymatgen Python library.[216] Once both sides of the interface have been transformed to their supercells, the MOF supercell is rotated about the $z$ axis such that its shortest vector is parallel to the shortest vector of the substrate supercell. The lattice-matching algorithm can produce MOF supercells that are mirror images of each other, but except when the angle between the primitive supercell vectors is close to 90°, one of these supercells corresponds to the MOF and substrate being on the same side of the interface, which is unphysical and so this supercell is discarded. Once both sides of the interface have been transformed to their supercells, the ASO is calculated from the coordinates of the atoms in the $xy$ plane. For each MOF supercell, the supercell lattice vectors $u$ and $v$ are multiplied by integers $I_u$ and $I_v$ such that they span at least three times each of the MOF unit cell vectors, $a$ and $b$. The substrate supercell vectors are then multiplied by $I_u$ and $I_v$. This expansion of the supercell is used to ensure that the ASO accounts for the lattice mismatch between the supercells. Without applying this expansion, examples of $1 \times 1$ supercells were found with binding planes with large lattice mismatches that produced high ASOs. In all of these cases, the high ASOs are decreased after expansion as the lattice mismatch causes atoms in neighboring unit cells to no longer overlap with the substrate atoms. Note that the supercell without expansion is tested during the lattice-matching step, and upon passing, is expanded for the ASO calculation.

## B.4 Monte Carlo algorithm for ASO maximization

A Metropolis Monte Carlo (MC) algorithm is applied to find the maximum possible ASO for each interface using rigid translations of the MOF binding atoms.[213] The MC algorithm allows for translations of the MOF atom positions along the MOF supercell lattice vectors by a random number uniformly distributed in the range of ±1 multi-

plied by a maximum displacement that is varied as the MC simulation proceeds so as to maintain an acceptance rate of 20% for MC moves. Because of the lattice mismatch, periodic boundary conditions are not applied in the MC algorithm to avoid spurious boundary effects. To stop the MOF supercell translations from sampling positions too far away from the center of the substrate supercell unnecessarily, if the cumulative displacement of the MOF atoms after step $i$ is greater than the size of the MOF supercell in the $x$ or $y$ dimension, the translation in step $i+1$ is forced away from the boundary of the supercell by requiring the MC moves to reduce the cumulative displacement along the specified dimension. The ASO is calculated after each translation and a move is accepted if the new ASO at step $i$ ($\mathrm{ASO}_i$) is greater than the previous ASO ($\mathrm{ASO}_{i-1}$) or if

$$\exp\left(-\beta\left(-\mathrm{ASO}_i + \mathrm{ASO}_{i-1}\right)\right) > R, \tag{B.2}$$

where $\beta$ is a parameter that acts like an inverse effective temperature and $R$ is a random number selected from a uniform distribution between 0 and 1. Each matched pair of supercells undergoes $M_\mathrm{T}$ trials of $M_\mathrm{S}$ MC steps. The initial positions of the MOF atoms relative to the substrate are randomized in each trial. The maximum ASO obtained from all $M_\mathrm{T}$ trials is taken to be the ASO for that pair of supercells.

Figure B.6 shows that the MC algorithm is robust to changes to the parameters (Table B.1), but a range of configurations with similar ASOs are found for interfaces with a low maximum ASO, resulting in significant uncertainty in the value of the maximum ASO for these low-ASO interfaces. The time taken to calculate the maximum ASO scales linearly with the number of trials or steps. Thus, we have used values for the number of trials and steps that are as small as possible, while still obtaining robust results for the maximum $\Delta$IB for each MOF crystal structure (see Figure B.6a and b). Figure B.6c shows minimal changes in the maximum $\Delta$IB for all crystal structures for wide range of $\beta$ values.

Table B.1: Description and value for each parameter in the Metropolis MC algorithm used to calculate the maximum ASO.

| parameter | value | description |
|---|---|---|
| $M_\mathrm{T}$ | 40 | no. of MC simulations (trials) per interface |
| $M_\mathrm{S}$ | 400 | no. of steps in each individual MC simulations |
| $\Delta$ | 0.5 Å | initial maximum displacement along each supercell lattice vector |
| $\beta$ | 50 | MC acceptance parameter |

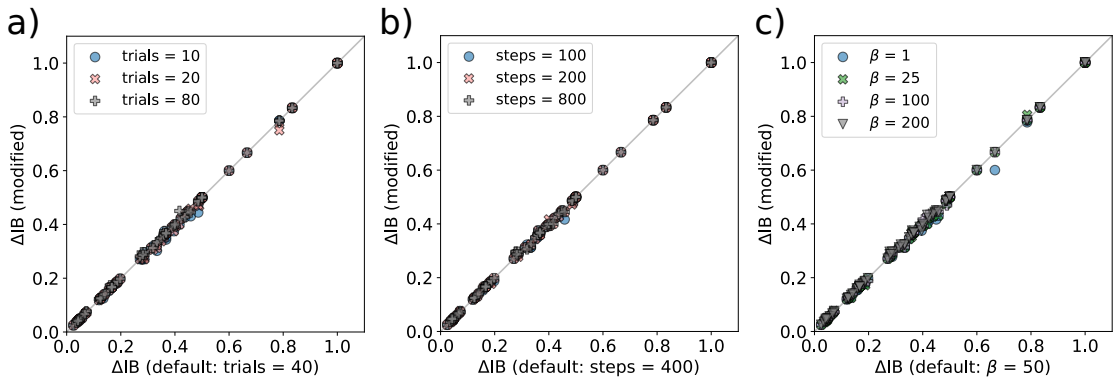



Figure B.6: Parity plots of the maximum ΔIB for each binding plane for all Miller planes for MOF crystal structures in a small data set (which includes structures from the experimental validation data set and CORE MOF database) comparing the default parameters (Table B.1) and modified (a) number of trials, (b) number of steps, and (c) $\beta$. Modified values are given in the figure legends and default values are given in the $x$-axis titles.

## B.5 Calculating binding at the interface

For each matching pair of supercells calculated from the lattice-matching algorithm, the ASO is calculated from the number of bonds that can be formed between binding sites of the MOF and binding sites of the substrate. Atoms on either side of the interface are defined to be coincident if the distance between them, in the $xy$ plane, is less than or equal to the arithmetic mean of the ionic radii of the MOF and substrate-binding atoms.[222,223] Unlike the previous application of the ASO, we only consider specific binding atoms at the interface and not all atoms.[213] In this work we have only allowed one bond per atom on either side of the interface, but this could be generalized to allow multiple bonds per atom. Our algorithm calculates the ASO for each supercell for each binding plane for each Miller plane for a given MOF and substrate pair. From all ASOs, a ΔIB is calculated post processing. In general, we report a single ΔIB and supercell for each MOF crystal structure, which is the maximum possible ΔIB for all Miller planes of that structure. The following steps are implemented to calculate the ASO and ΔIB:

i For each Miller plane of a given MOF crystal structure, build the binding planes (described in Section B.3).

ii For each binding plane, obtain an array of possible MOF–substrate supercell pairs (using the lattice-matching algorithm).[209,210]

iii For each supercell pair, carry out MC simulations (described in Section B.4) to obtain the maximum ASO for that supercell pair.

iv For each supercell pair, calculate ΔIB using the configuration from the previous step that yields the maximum ASO for that supercell pair.

v Take the maximum ΔIB of the MOF–substrate pair to be the highest ΔIB for all supercell pairs, binding planes, and MOF Miller planes.

### B.5.1 Identifying equivalent MOF–substrate supercell pairs

Two MOF–substrate supercell pairs are effectively equivalent, regardless of their size, if the orientation of the MOF surface in the two MOF supercells with respect to the substrate surface is the same within some tolerance. We identify equivalent MOF–substrate supercell pairs as those in which the in-plane positions (relative to the substrate) of equivalent MOF binding sites in the unit cells of the two MOF supercells is no greater than 0.5 Å. This corresponds to an in-plane rotation of the MOF unit cells with respect to one another of no more than $\approx 10°$. (In practice, the minimum pairwise distance between atoms in the two MOF unit cells is calculated and the two supercell pairs are deemed to be equivalent if the largest minimum pairwise distance is less than 0.5 Å.)

## B.6 Experimental validation data set

### B.6.1 Substrate

Copper(II) hydroxide ($Cu(OH)_2$) has been shown to form highly aligned macroscale structures, such as nanobelts or nanotubes, on silica substrates.[55] Aligned $Cu(OH)_2$ has also been shown to act as a sacrifical epitaxial substrate for MOF growth, and when epitaxial registry between the MOF lattice and substrate lattice exists, the MOF lattice has been shown to grow aligned with the substrate lattice. $Cu(OH)_2$ has *Cmc21* symmetry with the (010) Miller plane exposed at the surface.[55,217] Note that experimental results of aligned heteroepitaxial growth were only available for the $Cu(OH)_2$ substrate at the time of submission, but the previously described experimental method[55] and our

screening process are not limited to this substrate.

### B.6.2 MOFs

Table B.2: Common name, organic linkers, dimensionality and binary yes/no result of alignment experiment of all MOFs in the experimental data set used to validate the screening algorithm. Structures of the organic linkers are shown in Figure B.7.

| name | linkers | dimensionality | aligned? |
|---|---|---|---|
| CuBDC | 1,4-BDC | 2D | Yes[55] |
| CuBPDC | 4,4'-BPDC | 2D | Yes[55] |
| CuNDC | 2,6-NDC | 2D | Yes[55] |
| CuBDC-DABCO | 1,4-BDC + DABCO | 3D | Yes[55] |
| CuBDC-BPY | 1,4-BDC + 4,4'-BPY | 3D | Yes[a] |
| CuTDC | 1,4-TDC | 2D | No[55] |
| CuBTC | 1,3,6-BTC | 3D | No[55] |

[a] See Figure B.8

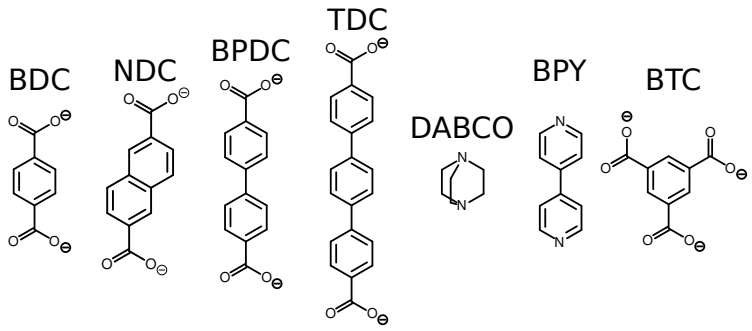



Figure B.7: Organic linkers in the experimental MOF data set given by their common names.

Table B.2 and Figure B.7 give the MOFs and organic linkers for which there are experimental results on aligned heteroepitaxial growth on $Cu(OH)_2$. We used one or more crystal structures for each of these MOFs to parameterize and validate the screening process. For the MOFs (CuBDC, CuBPDC, CuNDC, CuTDC), we used Materials Studio[178] to build crystal structures with experimental values for the lattice parameters (Table B.3) and the UFF4MOF[120,121] force field and the General Utility Lattice Program (GULP)[406,407] to energy minimize the atom positions, while the cell parameters were constrained. All 2D layered materials were built with an eclipsed (P4) layering structure.[206,408] For CuBDC-BPY, we used the CIF NEJRUR from the CORE MOF

database,[94] which is derived from an experimental crystal structure.[409] For CuBDC-DABCO-a, we used an experimental crystal structure for an isostructural MOF with the fluorinated BDC ligand,[410] which we also used as a starting point to build a non-fluorinated crystal structure, CuBDC-DABCO-b, with experimental lattice parameters (Table B.3) using Materials Studio, GULP and UFF4MOF. For CuBTC (also known as HKUST-1), we extracted crystal structures from the CORE MOF database (structures in the CORE MOF database were extracted from the Cambridge Structural Database (CSD)),[94] the hypothetical MOF database,[101] and one structure from the Crystallography Open Database (COD).[411,412] Table B.3 shows the lattice parameters for all crystal structures used for all MOFs in the experimental data set.

Table B.3: Names, lattice parameters and citations (and CSD REFCODES if available) of all crystal structures used in the experimental parameterization database.

| name | CIF | $a$ | $b$ | $c$ | $\alpha$ | $\beta$ | $\gamma$ |
|---|---|---|---|---|---|---|---|
| CuBDC-a | [413] | 10.61 | 5.80 | 10.61 | 90.0 | 90.0 | 90.0 |
| CuBDC-b | [408] | 11.19 | 5.80 | 11.19 | 90.0 | 90.0 | 90.0 |
| CuNDC | [408] | 13.35 | 5.90 | 13.35 | 90.0 | 90.0 | 90.0 |
| CuBPDC | [408] | 15.49 | 5.90 | 15.49 | 90.0 | 90.0 | 90.0 |
| CuBDC-DABCO-a | ACATAA[a] [410] | 10.87 | 10.87 | 9.67 | 90.0 | 90.0 | 90.0 |
| CuBDC-DABCO-b | [414] | 10.60 | 10.60 | 9.50 | 90.0 | 90.0 | 90.0 |
| CuBDC-BPY | NEJRUR[94,409] | 10.81 | 10.85 | 14.02 | 87.5 | 89.0 | 86.0 |
| CuTDC | [408] | 19.84 | 5.90 | 19.84 | 90.0 | 90.0 | 90.0 |
| CuBTC-a | [101] | 18.63 | 18.63 | 18.63 | 120 | 90 | 60 |
| CuBTC-b | 2300380[b] [415] | 26.30 | 26.30 | 26.30 | 90.0 | 90.0 | 90.0 |
| CuBTC-c | XAMDUM[94,416] | 18.62 | 18.62 | 18.62 | 60.0 | 60.0 | 60.0 |
| CuBTC-d | XAMDUM01[94,416] | 18.60 | 18.60 | 18.60 | 60.0 | 60.0 | 60.0 |
| CuBTC-e | XAMDUM02[94,416] | 18.59 | 18.59 | 18.59 | 60.0 | 60.0 | 60.0 |
| CuBTC-f | XAMDUM03[94,416] | 18.57 | 18.57 | 18.57 | 60.0 | 60.0 | 60.0 |
| CuBTC-g | XAMDUM04[94,416] | 18.35 | 18.35 | 18.35 | 60.0 | 60.0 | 60.0 |
| CuBTC-h | XAMDUM05[94,416] | 18.51 | 18.51 | 18.51 | 60.0 | 60.0 | 60.0 |
| CuBTC-i | XAMDUM06[94,416] | 18.60 | 18.60 | 18.60 | 60.0 | 60.0 | 60.0 |
| CuBTC-j | XAMDUM07[94,416] | 18.66 | 18.66 | 18.66 | 60.0 | 60.0 | 60.0 |

[a] Crystal structure of isostructural MOF with fluorinated BDC ligand.
[b] Non-P1 symmetry = Fm-3m; reference ID for Crystallography Open Database[411,412]

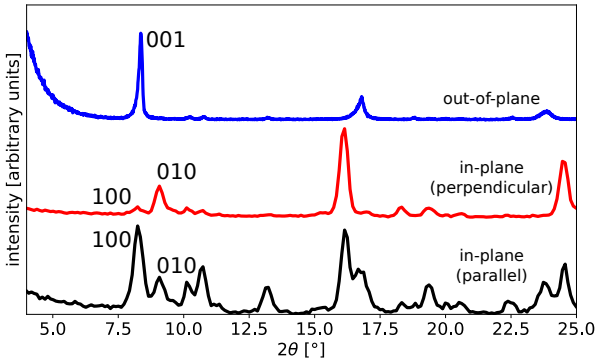



Figure B.8: X-ray diffraction patterns of CuBDC-BPY MOFs grown epitaxially on an aligned $Cu(OH)_2$ substrate, measured out-of-plane (blue line), in-plane perpendicular (red line), and in-plane parallel (black line) with respect to the longitudinal direction of the aligned $Cu(OH)_2$ nanobelts in the substrate, using methods described previously.[55] All patterns are normalized to their maximum values for the $2\theta$ range shown. Significant differences in the relative intensities of the peaks from $2\theta = 7.5$–$10^\circ$ between the three diffraction patterns support the macroscale in-plane and out-of-plane alignment of CuBDC-BPY. (Note that several peaks (in the 10–15° range) in the in-plane parallel plot could not be indexed and are likely due to an impurity phase, possibly a derivative phase of the interpenetrated framework structure. These peaks are very small in the out-of-plane measurement, indicating that the proportion of this impurity phase is less than $\sim 5\,\%$.)

## B.7 Effect of interpenetration

During the cleaning process used to produce the CORE MOF database, interpenetrating nets were retained by identifying the number of atoms, $N$, in the largest of all bonded components in a crystal structure, and retaining any components with at least $\frac{N}{2}$ atoms.[94] We used this definition to determine if a crystal structure was interpenetrated and, if so, we extracted each component with at least $\frac{N}{2}$ atoms and wrote that component to a new structure file, which we tested independently in step 3 of the screening process. Note that we did not check if the two separated nets were independently stable or reasonable structures. For CuBDC-BPY we used a structure from the CORE MOF database with the REFCODE: NEJRUR,[94,409] which, by the above definition, was interpenetrated with two identical nets. Figure B.9 shows the binding interfaces for the (100) Miller plane of the original crystal structure and the two separate nets. The maximum ASO of the full and separated structures were the same, but the $\Delta$IB of the full structure suffers from slicing the other net (see Figure B.10). Hence, for the screening process we removed the interpenetrated structures and only tested

each net separately. Note that interpenetration in CuBDC-BPY has been shown to be suppressed when it forms as a surface-anchored MOF using layer-by-layer liquid-phase epitaxy, but that was suggested to be due partially to the washing process.[417] Hence, experimental results do not clearly suggest that interpenetration is less favored, although Figure B.10 does suggest that at the interface, a non-interpenetrated structure will be more stable. Throughout this work, when necessary, we report the porosity of the interpenetrated version of any structure.

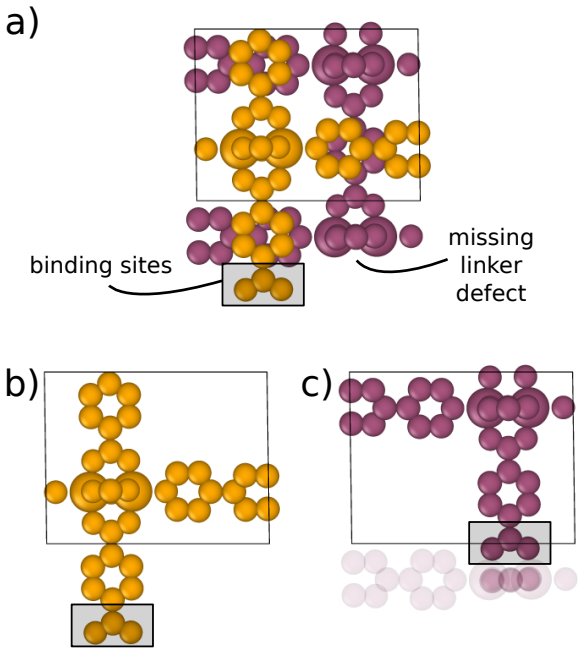



Figure B.9: (100) unit cell of the binding interface of the (a) interpenetrated and (b,c) separated nets of the NEJRUR crystal structure. Hydrogens are omitted for clarity and separate nets are shown in distinct colors. The carboxylate binding functionalities and missing linker defects in the interpenetrated structure are highlighted. Atoms that were removed to form the binding plane are semitransparent.

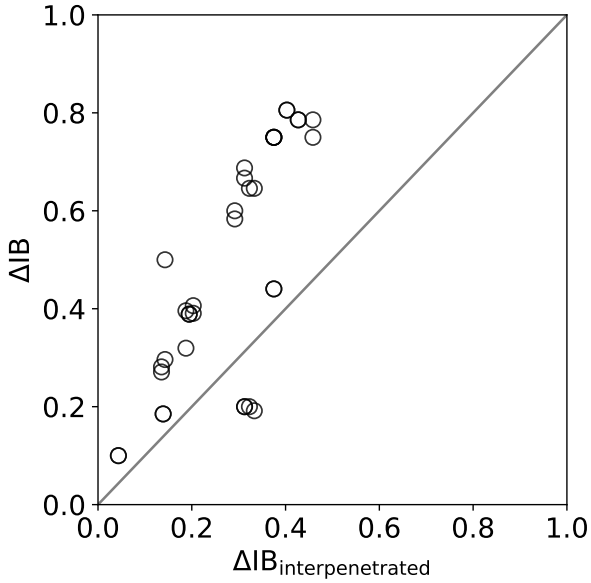



Figure B.10: Parity plot of the maximum $\Delta$IB obtained for all separated nets of interpenetrated MOFs in the CORE MOF database versus the maximum $\Delta$IB obtained for the interpenetrated MOF.

## B.8 Symmetry of binding plane of MOF candidates

Figures B.11 and B.12 show normalized distributions of the angle between the unit-cell vectors for the Miller plane with the maximum $\Delta$IB for all crystal structures in all three MOF databases that were screened. Separate histograms are plotted for MOFs that are predicted to undergo aligned heteroepitaxial growth on $Cu(OH)_2$ and for MOFs that are not. These results indicate that the top MOF candidates for aligned growth predominantly have binding interfaces with approximate rectangular symmetry (i.e. angles between in-plane unit-cell vectors close to 90°). Furthermore, a greater proportion of these top candidates have binding interfaces with rectangular symmetry compared with MOFs that are not expected to exhibit aligned growth, indicating that aligned heteroepitaxial growth on $Cu(OH)_2$ selects for MOF interfaces with rectangular symmetry.

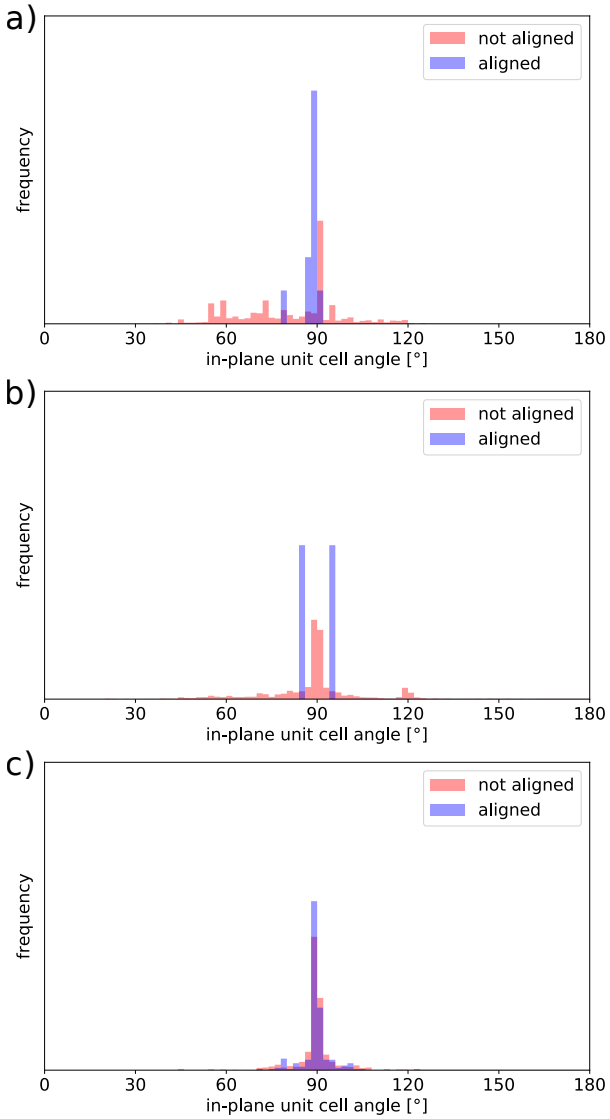



Figure B.11: Distribution of the in-plane unit cell lattice angle for all binding planes in the (a) CORE MOF, (b) TOBACCO, and (c) hMOF database. Binding planes predicted to undergo aligned heteroepitaxial growth on $Cu(OH)_2$ ($\Delta IB \geq 0.67$ and $\Delta IB_2 < 0.67$) are counted in the blue distributions, while all other binding planes are counted in the red distributions.

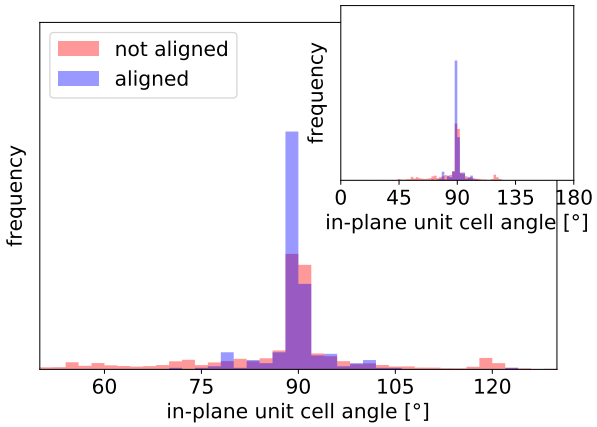



Figure B.12: Distribution of the in-plane unit-cell lattice angle of binding planes in the CORE MOF, TOBACCO, and hMOF databases. Binding planes predicted to undergo aligned heteroepitaxial growth on $Cu(OH)_2$ ($\Delta IB \geq 0.67$ and $\Delta IB_2 < 0.67$) are counted in the blue distribution, while all other binding planes are counted in the red distribution. The inset shows the full distribution of angles.

## B.9 Pore characteristics of hMOF candidates

Figure B.13 shows the maximum $\Delta$IB for all crystal structures in the hMOF database as a function of their pore-limiting diameter (PLD) and in-plane PLD, calculated using the "resex" command in Zeo++.[134] The in-plane PLD was taken as the maximum PLD along the two crystallographic dimensions that were parallel to the Miller plane bound to the substrate. Figure B.14 shows the normalized distribution of the PLDs and in-plane PLDs for MOFs with $\Delta$IB above and below the threshold for aligned heteroepitaxial growth. Both Figures B.13 and B.14 highlight a broad distribution of pore characteristics available for MOFs that are predicted to grow aligned on $Cu(OH)_2$.

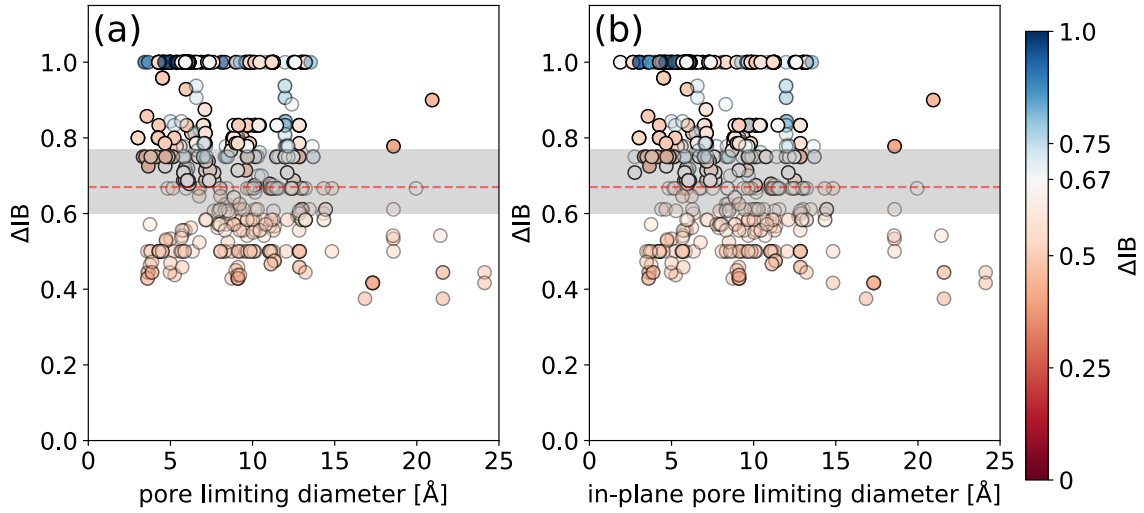



Figure B.13: Maximum $\Delta\mathrm{IB}$ versus (a) pore-limiting diameter and (b) in-plane pore-limiting diameter for all structures in the hMOF database whose maximum $\Delta\mathrm{IB}$ is associated with a Miller plane where $|h| + |k| + |l| = 1$. Points are colored by the value of the second highest $\Delta\mathrm{IB}$ ($\Delta\mathrm{IB}_2$). The shaded region indicates the approximate $\Delta\mathrm{IB}$ threshold for heteroepitaxial in-plane alignment determined in Section 4.4.1. The red line indicates the value of $\Delta\mathrm{IB}$ used as the threshold for top candidates.

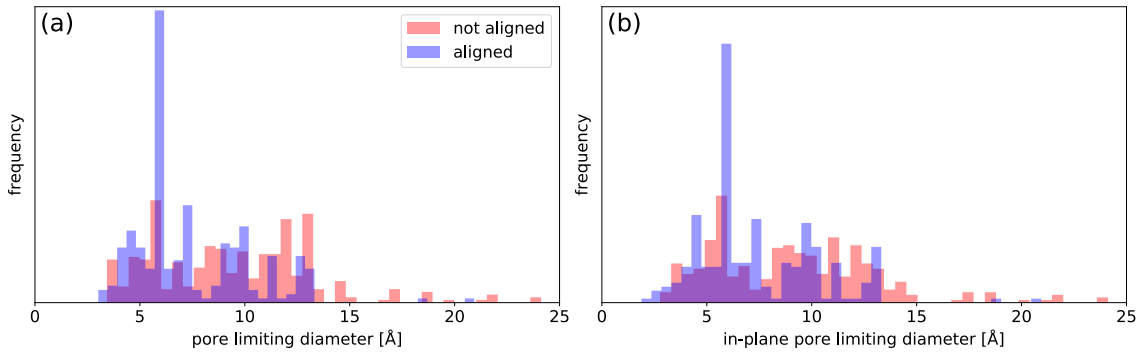



Figure B.14: Distribution of the (a) pore-limiting diameter and (b) in-plane pore-limiting diameter for all structures in the hMOF database whose maximum $\Delta\mathrm{IB}$ is associated with a Miller plane where $|h| + |k| + |l| = 1$.

# APPENDIX C

# Supporting information: Analysis of the enzymatic reaction space of enzyme@MOF biocomposites

## C.1 Enzyme Commission number

Table C.1: First level (most general) Enzyme Commission (EC) number classifications.

| first-level EC number | name |
|---|---|
| 1. | oxidoreductases |
| 2. | transferases |
| 3. | hydrolases |
| 4. | lyases |
| 5. | isomerases |
| 6. | ligases |

## C.2 Determining molecular accessibility

In this work we have implemented an efficient method for calculating the size of small molecules based on their van der Waals (vdW) volume (Section 6.3). In the following section we examine the sensitivity of the calculated molecular size on the parameters used in our method. Figure C.1 shows the effect of the grid spacing and the number of conformers ($N$) on the value of $d$ for a small set of molecules (we used a box margin of 4 Å in all calculations). We found that over the range of parameters tested, the change in $d$ is less than 0.6 Å and 0.3 Å for the grid spacing and number of conformers, respectively. Decreasing the grid spacing used or increasing $N$ increases computation time and memory requirements. Therefore, the choice of parameters is a balance between

accuracy and efficiency. In this work, we used values of 0.5 Å for the grid spacing and $N = 100$. The algorithm that builds the 3D conformers provides a stochastic sampling of conformational space (focused on chemically reasonable structures) for each molecule.[335] The number of conformers needed to sample the conformational space of a molecule approximately depends on the number of degrees of freedom in the molecule, which can be related to the number of heavy atoms or number of rotatable bonds.[335] Therefore, it is possible to select $N$ as a function of the number of rotatable bonds in a given molecule to improve the efficiency of our methodology.

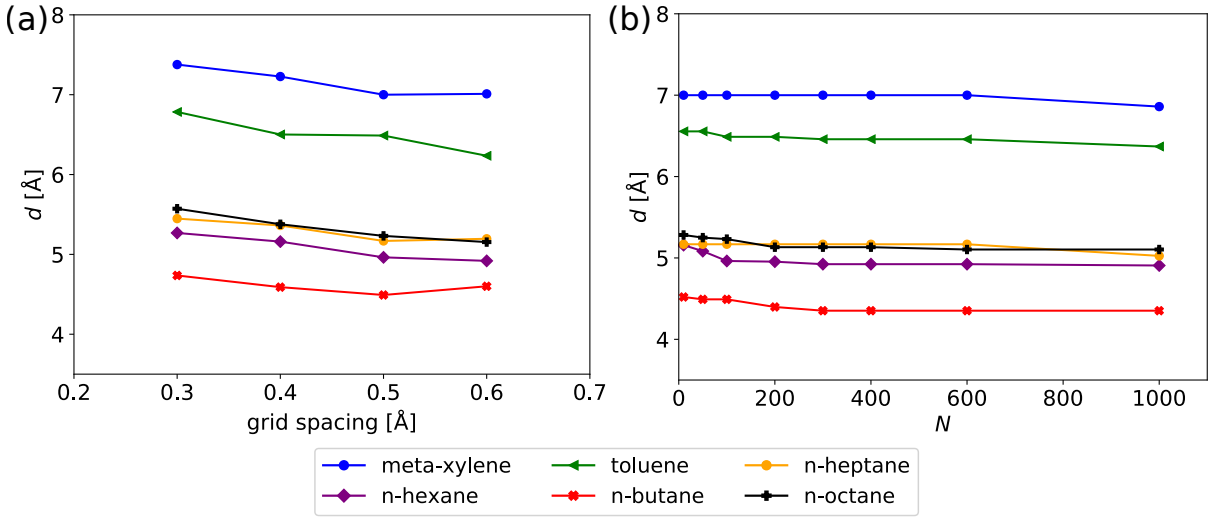



Figure C.1: Molecular size $d$ as a function of the (a) grid spacing and (b) number of conformers ($N$). For each plot, the unchanged parameters were set to: a grid spacing of 0.5 Å, $N = 100$, vdW scale parameter of 0.8 and box margin of 4.0 Å, which are the values used in this work.

Table C.2: Kinetic diameters of molecules used to parameterize our methodology for calculating the molecular size. All kinetic diameters were taken from Ref. 311 unless otherwise cited. Where applicable, the smaller value of a range was used in Figure C.2.

| name | kinetic diameter [Å] | name | kinetic diameter [Å] |
|---|---|---|---|
| He | 2.551 | dimethyl ether | 4.307 |
| Ne | 2.82 | ethane | 4.443 |
| Ar | 3.542 | ethene | 4.163 |
| Kr | 3.655 | ethanol | 4.530 |
| Xe | 4.047 | *n*-propane | 4.3–5.118 |
| $H_2$ | 2.827–2.89 | cyclopropane | 4.23–4.807 |
| $Cl_2$ | 4.217 | propene | 4.678 |
| $Br_2$ | 4.296 | acetone | 4.600 |
| $CO_2$ | 3.3 | *n*-butane | 4.687 |
| $O_2$ | 3.467 | 1-butene | 4.5 |
| $N_2$ | 3.64–3.80 | *i*-butane | 5.278 |
| $H_2O$ | 2.641 | 2,2-dimethylbutane | 6.2 |
| NO | 3.492 | cis-2-butene | 4.23 |
| CO | 3.69 | 1,3-butadiene | 5.2 |
| $N_2O$ | 3.828 | *n*-pentane | 4.5 |
| HCl | 3.339 | *i*-pentane | 5 |
| HBr | 3.353 | *neo*-pentane | 6.2–6.464 |
| $CS_2$ | 4.483 | 2-methyl pentane | 5.5 |
| COS | 4.130 | 2,2,4-trimethylpentane | 6.2 |
| $SO_2$ | 4.112 | 3-methylpentane | 5.5 |
| $H_2S$ | 3.623 | *n*-hexane | 4.3 |
| $NH_3$ | 2.900 | *n*-heptane | 4.3 |
| $NF_3$ | 3.62 | *n*-octane | 4.3 |
| $CCl_2F_2$ | 5.0 | cyclohexane | 6–6.182 |
| $CH_3Cl$ | 4.182 | benzene | 5.349–5.85 |
| $CH_2Cl_2$ | 4.898 | ethyl-benzene | 5.8 |
| $CHCl_3$ | 5.389 | para-xylene | 5.8 |
| $CCl_4$ | 5.947 | meta-xylene | 6.8 |
| $CF_4$ | 4.662 | ortho-xylene | 6.8 |
| $C_2F_6$ | 5.1 | *i*-butene | 4.8[308] |
| *n*-$C_6F_{14}$ | 7 | 1-butanol | 4.5[307] |
| methane | 3.758 | 2,3-dimethylbutane | 5.6[307] |
| methanol | 3.626 | 1,2,4-trimethylbenzene | 7.6[307] |
| acetylene | 3.3 | mesitylene | 8.2[a][312] |
| toluene | 5.25 | | |

[a] Calculated critical diameter

Figure C.2 compares the calculated molecular size $d$ and the reported kinetic diameters for all molecules in Table C.2. Deviations from the line $y = x$ in Figure C.2 are up to $\sim$1 Å, which is expected for larger or more flexible molecules or molecules with irregular shapes.[338] For example, $d$ of $n$-heptane and $n$-octane differ significantly from their reported kinetic diameters (by $\approx$0.9 Å in both cases). We used a constant random seed (1000) for the generation of 3D conformers, which allows for reproducibility over subsequent runs. Nonetheless, we can efficiently approximate the kinetic or molecular diameter of small molecules using this methodology. Note that the kinetic diameter of a molecule is a temperature dependent parameter, which our algorithm does not consider. Figure C.3 shows the relationship between simulated self-diffusivities of small molecules through ZIF-8[300] and the molecular size $d$ calculated using our method. The simulated values were shown to match well to experimental results.[300] As expected, self-diffusivity becomes significantly smaller for $d$ above approximately $\approx$4.0–4.5 Å.

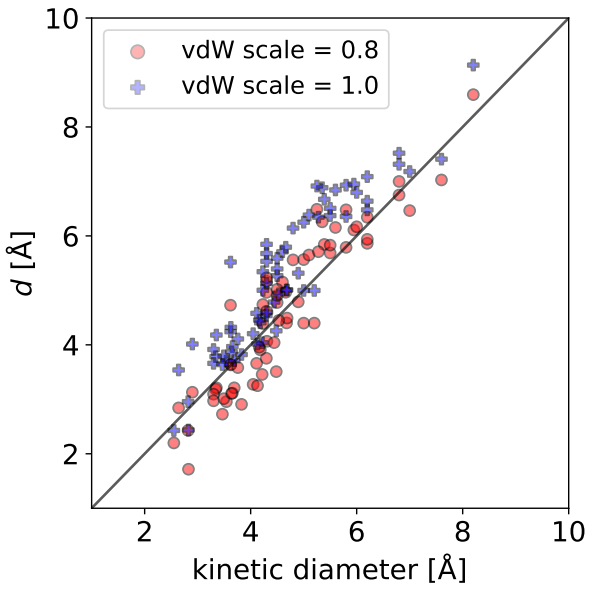



Figure C.2: Comparison of the calculated molecular size $d$ and reported kinetic diameters for all molecules in Table C.2 for two values of the RDKit vdW scale parameter.

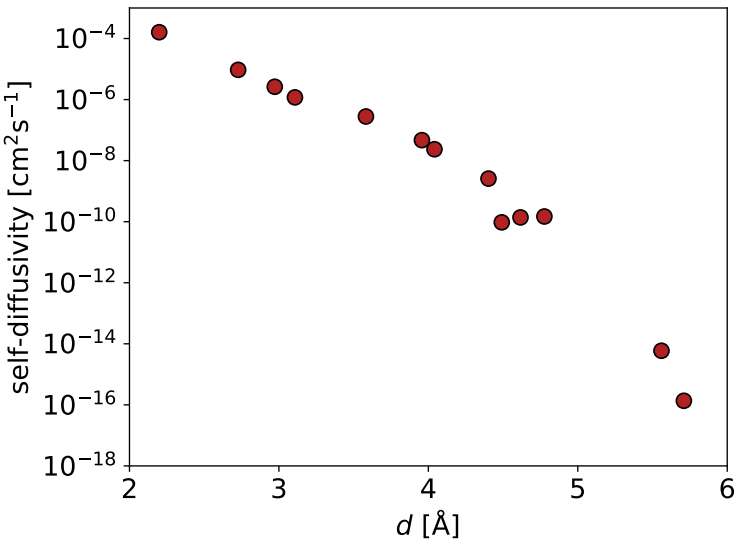



Figure C.3: Simulated self-diffusivities of small molecules in ZIF-8 extracted from Ref. 300 versus the calculated molecular size $d$ from this work.

## C.3 Components of reported enzyme@ZIF-8 reactions

Table C.3: Common names of molecules that have been components of enzymatic reactions carried out by enzymes encapsulated in ZIF-8.

| | | |
|---|---|---|
| water | hydrogen peroxide | oxygen |
| urea | carbon dioxide | ammonia |
| pyrogallol | purporogallin | 3-amino-1,2,4-triazole |
| ABTS | amplex red | resorufin |
| methyl ethyl ketone peroxide | tert-butyl hydroperoxide | methosulfate |
| 5-methylphenazin-5-ium | 2,6-dichloroindophenol | benzoquinone |
| *D*-glucose (ring) | *D*-glucono lactone | *D*-galactose |
| methylene blue | pyruvate | *L*-lactate |
| sucrose | *D*-fructose | beta-lactose |
| phosphate | acetic acid | butyric acid |
| hexanoic acid | octanoic acid | decanoic acid |
| dodecanoic acid | *p*-nitrophenol | vinyl acetate |
| octyl acetate | *p*-nitrophenyl acetate | *p*-nitrophenyl phosphate |
| *p*-nitrophenyl butyrate | *p*-nitrophenyl hexanoate | *p*-nitrophenyl octanoate |
| *p*-nitrophenyl decanoate | *p*-nitrophenyl dodecanoate | 2-octanol |
| penicillin-G | ethanol | acetaldehyde |

## C.4 Collection of enzyme reaction systems

The KEGG database includes ≈11 000 biochemical reactions, which are manually curated.[324–326] We iteratively collected the components and properties of all reactions for all EC numbers in the KEGG database, where EC numbers specify the reaction carried out by an enzyme. Because of the difficulties (mostly to do with nomenclature) associated with the curation of biochemical databases,[418] we avoided any cases for which molecular structures or properties were not unambiguously defined. For simplicity, we avoided name-based searching for chemical compounds except when secondary descriptors (such as database identifiers) were available to confirm the search result. We manually checked a subset of the collected reactions and molecules for agreement with the online database to confirm the reliability of the web interface. All reactions reported in the KEGG database as of the 5th of September, 2018, and their molecular components were collected using the steps below. In all cases, the structures extracted from a database were not modified or cleaned in any way and represent the structure contained online. The following steps are used to extract reaction systems from the KEGG database:

1. Collect reaction system information using the KEGG web service.[324–326]
2. Collect the molecular components of the reaction system.
    (a) Collect the KEGG identifier and role in reaction of each component. KEGG does not specify roles as substrate or products because all reactions are reported as reversible. Therefore, reactions were separated into the left-hand side and right-hand side.
    (b) Collect the molecular structure as a MOL file from the KEGG website using the KEGG identifier and KEGG web service. If no structural information was obtained, or the structure was generic, for any component of the reaction, then the reaction was skipped. The charge state of the molecule was not modified from what is provided on the KEGG website.
    (c) Convert KEGG MOL file into RDKit molecule and SMILES string.
    (d) Skip reactions that have any components that are generic (given by "*" in the SMILES string) or polymeric.

(e) Calculate the hydrophobicity (AlogP98, referred to as logP in this work),[377] synthetic accessibility score (SAscore)[380] and water solubility (logS)[376,379] using RDKit.

(f) Extract XLogP3-AA[378] and complexity[419,420] from the PubChem database.[80,421]

(g) Calculate molecular size using the method described in Section 6.3. If a molecule downloaded from the KEGG database cannot be converted to an RDKit molecule object, then the reaction was skipped.

## C.5 Analysis of enzyme reaction systems

As part of our screening process (Section C.4), we analyzed each reaction and its components using cheminformatics tools. For completeness, we do not report a certain property of a reaction if the property of interest could not be calculated for all components. All of the following methodologies are approximate and efficient. Hence, each candidate reaction should be checked more thoroughly before experimental testing. However, Figure 6.6 shows that these simple descriptors can effectively narrow the regions of interest in the chemical space associated with enzymatic reactions. We used the molecular size ($d$, see Section 6.3) to estimate the likelihood of diffusion of molecular components through ZIF-8. Figure C.4 shows the size of the maximum component of all reactions, separated by the first level EC numbers (Table C.1). The distributions of the size of the largest component are broad and centered around 7–9 Å in most cases, except for ligases (EC: 6) and transferases (EC: 2). For ligases, there are very few collected reactions and for transferases the sharp peak around 9 Å likely corresponds to co-enzyme molecules that act as donors in these classes of reactions.

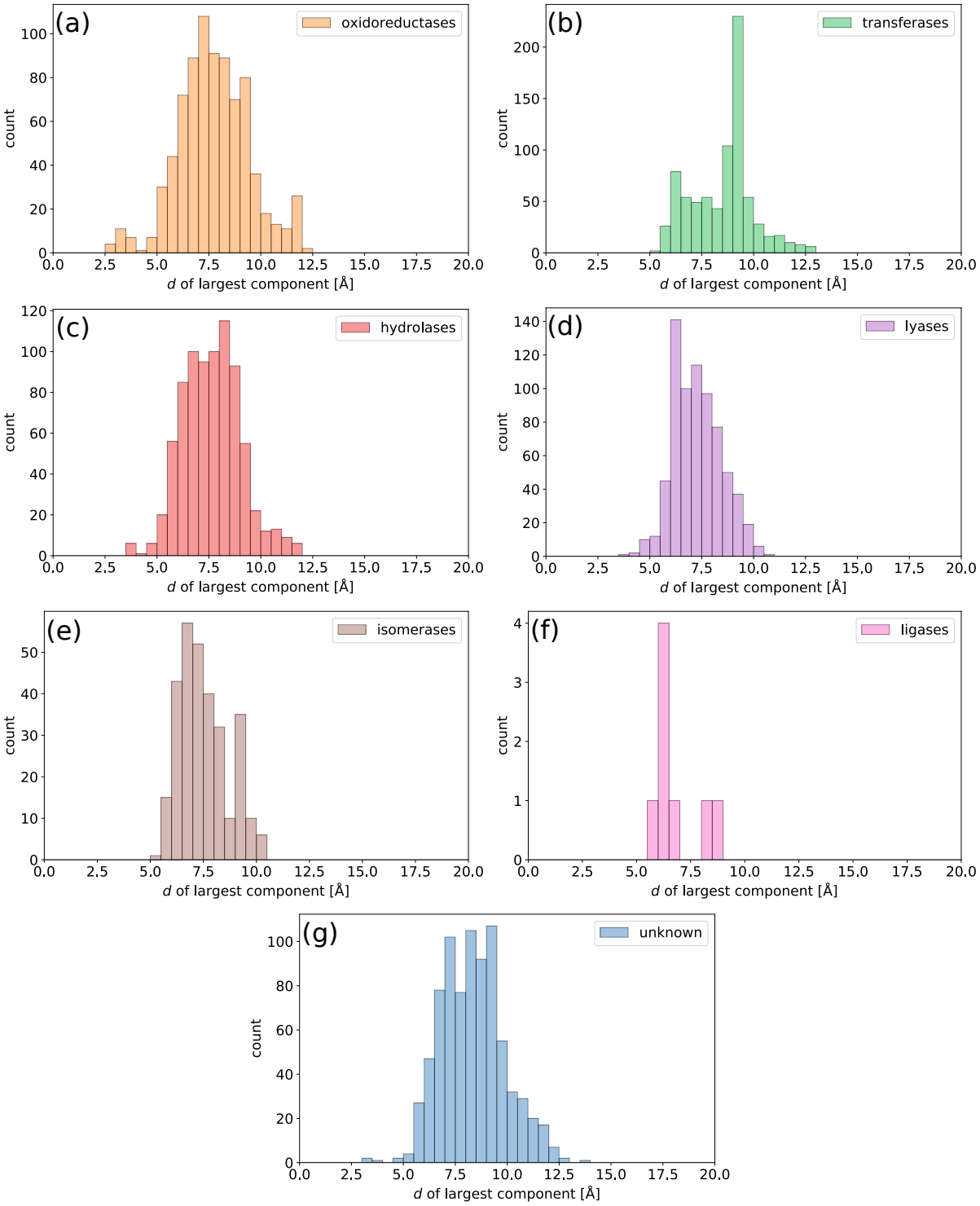



Figure C.4: Histogram of molecular size $d$ of the largest component of each reaction. Reactions are separated into their first level EC categories (Table C.1).

Figure C.5 shows the change in SAscore (as SAscore(right-hand side) − SAscore(left-hand side)) for all reactions collected from the KEGG database (separated by the first-level EC numbers), where the SAscore of each side of the reaction was that of the largest molecule (defined by the number of heavy atoms) on that side. We note that the SAscore can be over estimated for small molecules ($\lesssim 5$ heavy atoms) because their fragments are uncommon in the PubChem database (Figure C.6a). We also extracted the complexity[419,420] reported on the PubChem website,[80,421] which determines complexity by the bond network and symmetry of a molecule. Figure C.6b shows that SAscore and PubChem complexity are well correlated even though the measures of complexity are defined differently.

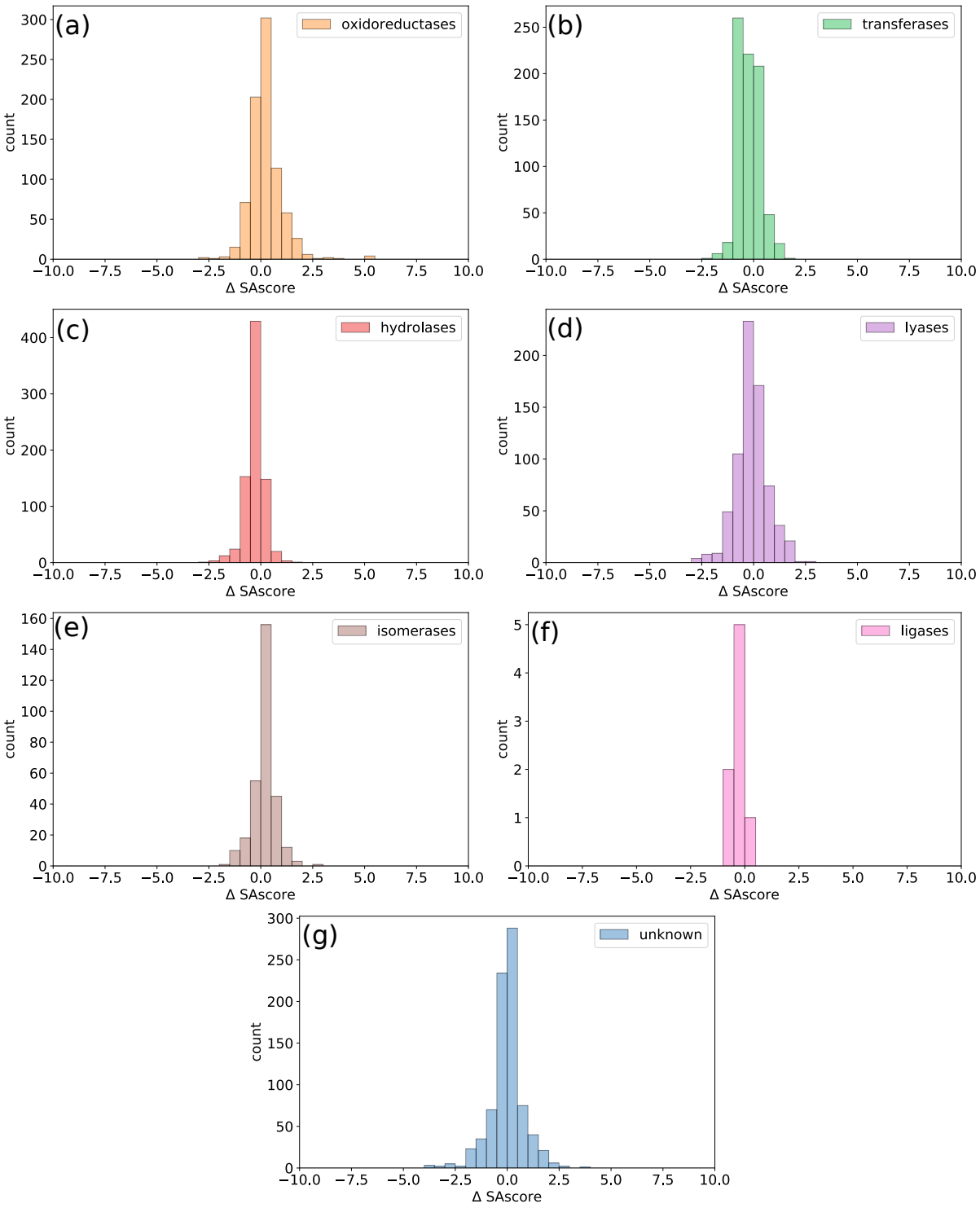



Figure C.5: Histograms of the change in SAscore of all reactions separated into their first level EC categories (Table C.1).

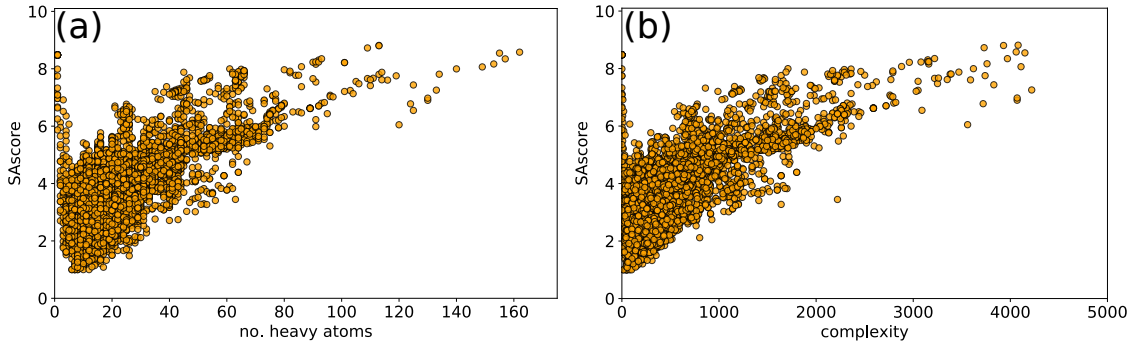



Figure C.6: SAscore as a function of (a) the number of heavy atoms and (b) the complexity (collected from PubChem)[80,421] for all molecules collected for all reaction systems.

There are many ways to calculate the hydrophobicity of a molecule.[422] In this work, we used the octanol–water partition coefficient as measured by the AlogP98 method (implemented in RDKit and termed "logP" in this paper).[334,377,378] Figure C.7 shows the logP of the most hydrophobic component of all reactions collected in this work. Some logP values are higher than that of *n*-pentane (logP $\approx$ 2.2), indicating that compounds in the KEGG database can be quite hydrophobic. Figure C.8 compares logP calculated using RDKit and XlogP3-AA[378] extracted from the PubChem database.[80,421] These methods are two of the most efficient and accurate methods for calculating logP.[422] We used the logP of a single charge state when dealing with an ionizable molecule. In particular, we used the charge state extracted from the KEGG database in all cases. Figure C.9 shows the logS of the least water-soluble component of all reactions collected. Similarly to the logP values, compounds less water-soluble than pentane (logS $\approx -1.8$) are present in the KEGG database.

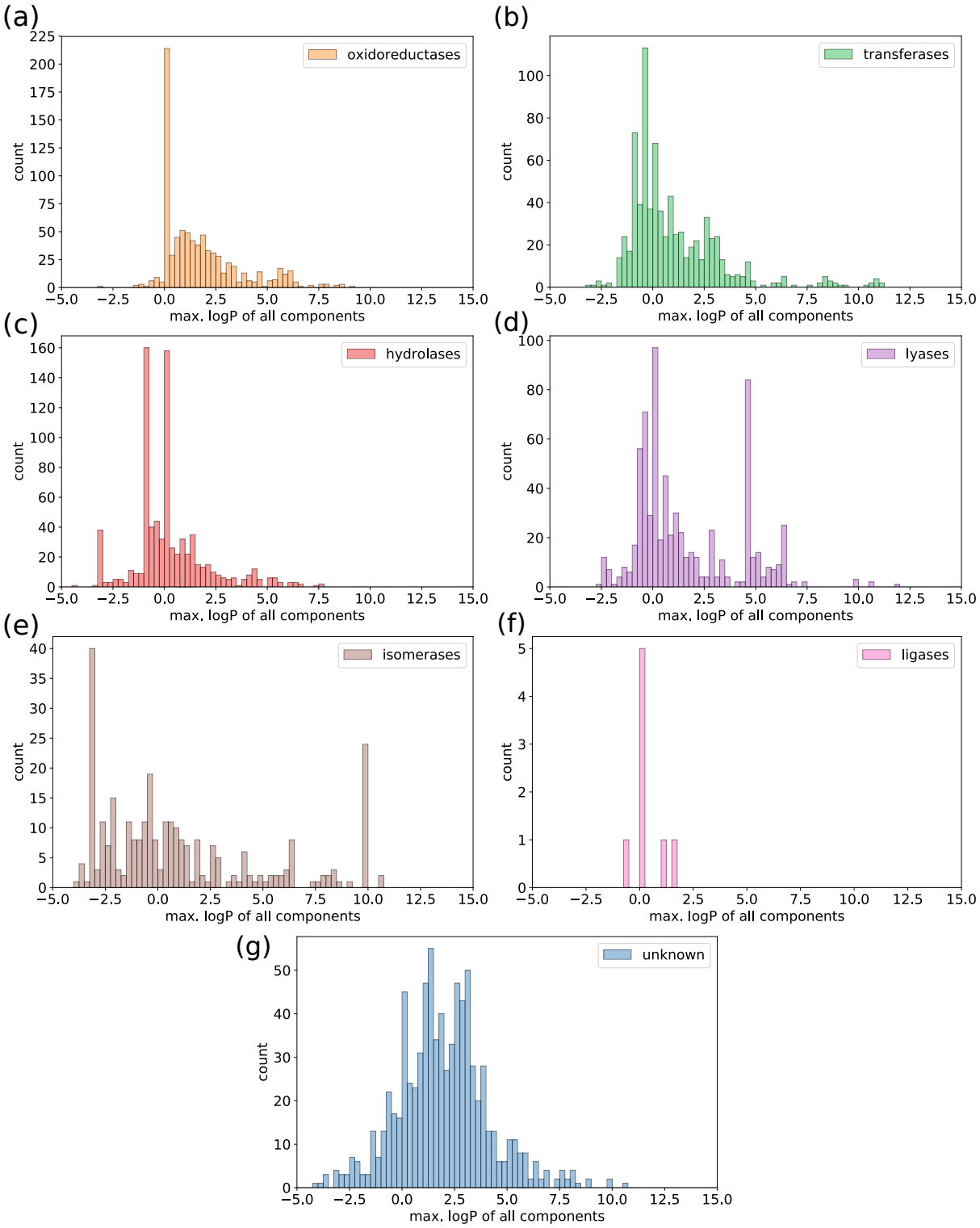



Figure C.7: Histograms of logP of the most hydrophobic component of each reaction separated into their first level EC categories (Table C.1).

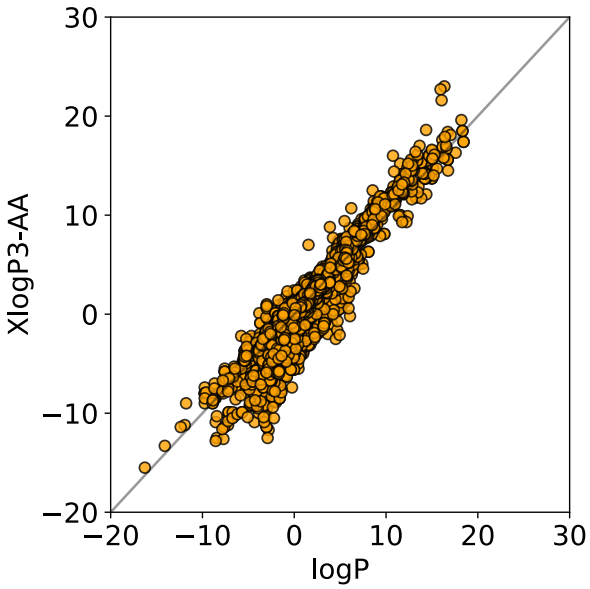



Figure C.8: LogP (AlogP-98)[377] versus XlogP3-AA (from PubChem)[80,378,421] for all molecules in all reaction systems collected.

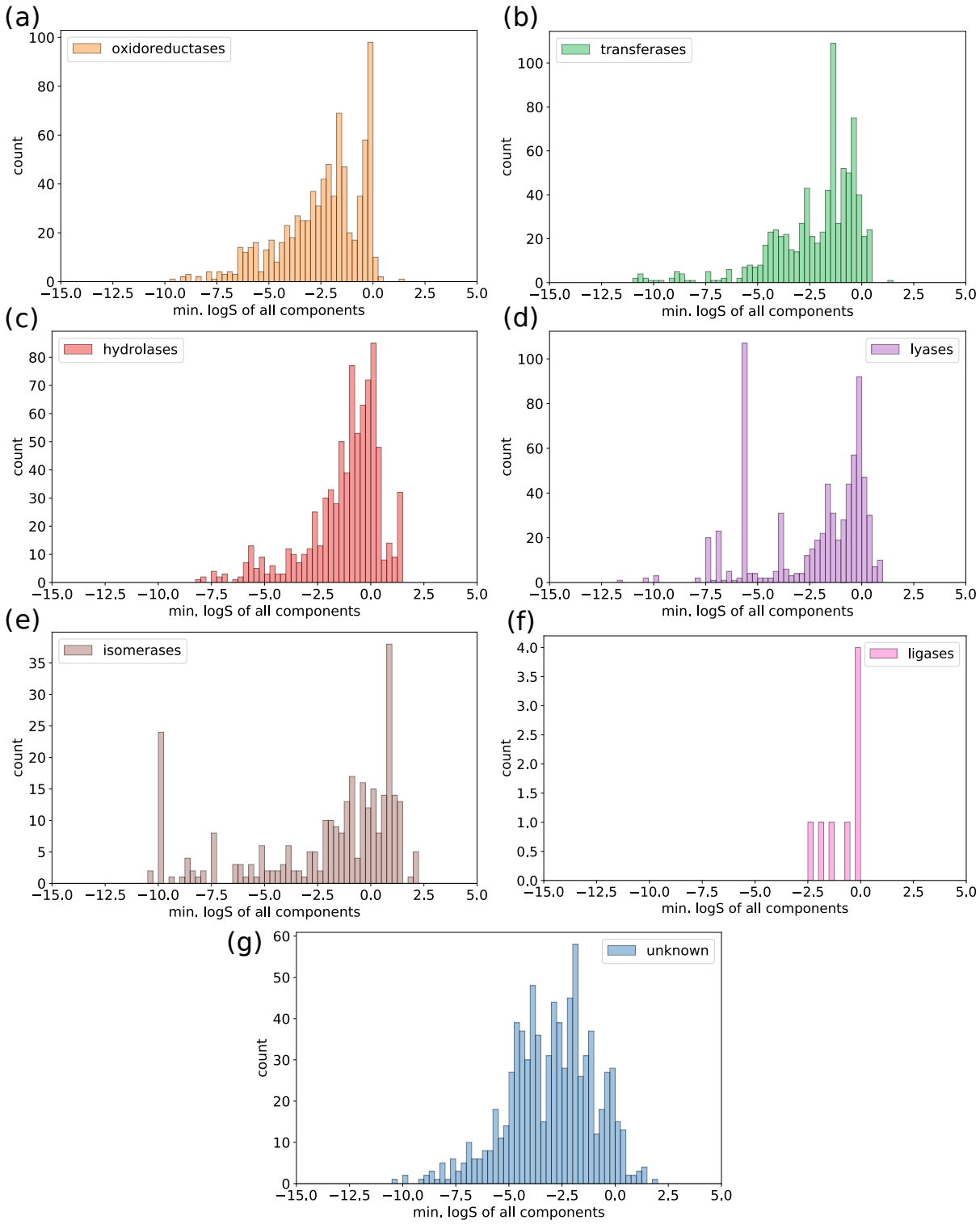



Figure C.9: Histograms of logS of the least water-soluble component of each reaction separated into their first level EC categories (Table C.1).

# References


[1] Tarzia, A.; Thornton, A. W.; Doonan, C. J.; Huang, D. M. Molecular Insight into Assembly Mechanisms of Porous Aromatic Frameworks. *J. Phys. Chem. C* **2017**, *121*, 16381–16392.

[2] Tarzia, A.; Takahashi, M.; Falcaro, P.; Thornton, A. W.; Doonan, C. J.; Huang, D. M. High-Throughput Screening of Metal–Organic Frameworks for Macroscale Heteroepitaxial Alignment. *ACS Appl. Mater. Interfaces* **2018**, *10*, 40938–40950.

[3] Maddigan, N. K.; Tarzia, A.; Huang, D. M.; Sumby, C. J.; Bell, S. G.; Falcaro, P.; Doonan, C. J. Protein Surface Functionalisation as a General Strategy for Facilitating Biomimetic Mineralisation of ZIF-8. *Chem. Sci.* **2018**, *9*, 4217–4223.

[4] Astria, E.; Thonhofer, M.; Ricco, R.; Liang, W.; Chemelli, A.; Tarzia, A.; Alt, K.; Hagemeyer, C.; Rattenberger, J.; Schroettner, H.; Wrodnigg, T.; Amenitsch, H.; Huang, D. M.; Doonan, C. J.; Falcaro, P. Carbohydrates@MOFs. *Mater. Horiz.* **2019**, Online early access. DOI: 10.1039/C8MH01611A.

[5] Huxley, M.; Coghlan, C. J.; Burgun, A.; Tarzia, A.; Sumida, K.; Sumby, C. J.; Doonan, C. J. Site-Specific Metal and Ligand Substitutions in a Microporous $Mn^{2+}$-Based Metal–Organic Framework. *Dalton Trans.* **2016**, *45*, 4431–4438.

[6] Morshedi, M.; Thomas, M.; Tarzia, A.; J. Doonan, C.; G. White, N. Supramolecular Anion Recognition in Water: Synthesis of Hydrogen-Bonded Supramolecular Frameworks. *Chem. Sci.* **2017**, *8*, 3019–3025.

[7] Sun, B.; M. Southam, H.; A. Butler, J.; K. Poole, R.; Burgun, A.; Tarzia, A.; Keene, F. R.; Collins, G. J. Synthesis, Isomerisation and Biological Properties of Mononuclear Ruthenium Complexes Containing the Bis[4(4′-Methyl-2,2′-Bipyridyl)]-1,7-Heptane Ligand. *Dalton Trans.* **2018**, *47*, 2422–2434.

[8] Liang, W.; Xu, H.; Carraro, F.; Maddigan, N. K.; Li, Q.; Bell, S. G.; Huang, D. M.; Tarzia, A.; Solomon, M.; Vaccari, L.; Amenitsch, H.; Sumby, C. J.; Falcaro, P.; Doonan, C. J. Enhanced Activity of Enzymes Encapsulated in Hydrophilic Metal–Organic Frameworks. *J. Am. Chem. Soc.* **2019**, *141*, 2348–2355.

[9] Davis, M. E. Ordered Porous Materials for Emerging Applications. *Nature* **2002**, *417*, 813–821.

[10] Rouquerol, J.; Avnir, D.; Fairbridge, C. W.; Everett, D. H.; Haynes, J. M.; Pernicone, N.; Ramsay, J. D. F.; Sing, K. S. W.; Unger, K. K. Recommendations for the Characterization of Porous Solids (Technical Report). *Pure Appl. Chem.* **1994**, *66*, 1739–1758.

[11] Eddaoudi, M.; Moler, D. B.; Li, H.; Chen, B.; Reineke, T. M.; O'Keeffe, M.; Yaghi, O. M. Modular Chemistry: Secondary Building Units as a Basis for the Design of Highly Porous and Robust Metal-Organic Carboxylate Frameworks. *Acc. Chem. Res.* **2001**, *34*, 319–330.

[12] Rowsell, J. L. C.; Yaghi, O. M. Metal–Organic Frameworks: A New Class of Porous Materials. *Microporous Mesoporous Mater.* **2004**, *73*, 3–14.

[13] Dawson, R.; Cooper, A. I.; Adams, D. J. Nanoporous Organic Polymer Networks. *Prog. Polym. Sci.* **2012**, *37*, 530–563.

[14] Chaoui, N.; Trunk, M.; Dawson, R.; Schmidt, J.; Thomas, A. Trends and Challenges for Microporous Polymers. *Chem. Soc. Rev.* **2017**, *46*, 3302–3321.

[15] Xu, Y.; Jin, S.; Xu, H.; Nagai, A.; Jiang, D. Conjugated Microporous Polymers: Design, Synthesis and Application. *Chem. Soc. Rev.* **2013**, *42*, 8012–8031.

[16] Yaghi, O. M.; O'Keeffe, M.; Ockwig, N. W.; Chae, H. K.; Eddaoudi, M.; Kim, J. Reticular Synthesis and the Design of New Materials. *Nature* **2003**, *423*, 705–714.

[17] Thornton, A. W.; Simon, C. M.; Kim, J.; Kwon, O.; Deeg, K. S.; Konstas, K.; Pas, S. J.; Hill, M. R.; Winkler, D. A.; Haranczyk, M.; Smit, B. Materials Genome in Action: Identifying the Performance Limits of Physical Hydrogen Storage. *Chem. Mater.* **2017**, *29*, 2844–2854.

[18] Jiang, J.-X.; Su, F.; Trewin, A.; Wood, C.; Campbell, N.; Niu, H.; Dickinson, C.; Ganin, A.; Rosseinsky, M.; Khimyak, Y.; Cooper, A. Conjugated Microporous Poly(Aryleneethynylene) Networks. *Angew. Chem. Int. Ed.* **2007**, *46*, 8574–8578.

[19] Cooper, A. I. Conjugated Microporous Polymers. *Advanced Materials* **2009**, *21*, 1291–1295.

[20] Martin, C. F.; Stockel, E.; Clowes, R.; Adams, D. J.; Cooper, A. I.; Pis, J. J.; Rubiera, F.; Pevida, C. Hypercrosslinked Organic Polymer Networks as Potential Adsorbents for Pre-Combustion CO2 Capture. *J. Mater. Chem.* **2011**, *21*, 5475–5483.

[21] Tan, L.; Tan, B. Hypercrosslinked Porous Polymer Materials: Design, Synthesis, and Applications. *Chem. Soc. Rev.* **2017**, *46*, 3322–3356.

[22] McKeown, N. B.; Budd, P. M. Polymers of Intrinsic Microporosity (PIMs): Organic Materials for Membrane Separations, Heterogeneous Catalysis and Hydrogen Storage. *Chem. Soc. Rev.* **2006**, *35*, 675–683.

[23] Ramimoghadam, D.; Gray, E. M.; Webb, C. J. Review of Polymers of Intrinsic Microporosity for Hydrogen Storage Applications. *Int. J. Hydrog. Energy* **2016**, *41*, 16944–16965.

[24] Ben, T.; Qiu, S. Porous Aromatic Frameworks: Synthesis, Structure and Functions. *CrystEngComm* **2013**, *15*, 17–26.

[25] Pei, C.; Ben, T.; Qiu, S. Great Prospects for PAF-1 and Its Derivatives. *Mater. Horiz.* **2015**, *2*, 11–21.

[26] Trewin, A.; Cooper, A. I. Porous Organic Polymers: Distinction from Disorder? *Angew. Chem. Int. Ed.* **2010**, *49*, 1533–1535.

[27] Ben, T.; Ren, H.; Ma, S.; Cao, D.; Lan, J.; Jing, X.; Wang, W.; Xu, J.; Deng, F.; Simmons, J. M.; Qiu, S.; Zhu, G. Targeted Synthesis of a Porous Aromatic Framework with High Stability and Exceptionally High Surface Area. *Angew. Chem. Int. Ed.* **2009**, *121*, 9621–9624.

[28] Yuan, D.; Lu, W.; Zhao, D.; Zhou, H.-C. Highly Stable Porous Polymer Networks with Exceptionally High Gas-Uptake Capacities. *Adv. Mater.* **2011**, *23*, 3723–3725.

[29] Konstas, K.; Taylor, J. W.; Thornton, A. W.; Doherty, C. M.; Lim, W. X.; Bastow, T. J.; Kennedy, D. F.; Wood, C. D.; Cox, B. J.; Hill, J. M.; Hill, A. J.; Hill, M. R. Lithiated Porous Aromatic Frameworks with Exceptional Gas Storage Capacity. *Angew. Chem. Int. Ed.* **2012**, *51*, 6639–6642.

[30] Lau, C. H.; Konstas, K.; Thornton, A. W.; Liu, A. C. Y.; Mudie, S.; Kennedy, D. F.; Howard, S. C.; Hill, A. J.; Hill, M. R. Gas-Separation Membranes Loaded with Porous Aromatic Frameworks That Improve with Age. *Angew. Chem. Int. Ed.* **2015**, *54*, 2669–2673.

[31] Lau, C. H.; Konstas, K.; Doherty, C. M.; Kanehashi, S.; Ozcelik, B.; Kentish, S. E.; Hill, A. J.; Hill, M. R. Tailoring Physical Aging in Super Glassy Polymers with Functionalized Porous Aromatic Frameworks for CO2 Capture. *Chem. Mater.* **2015**, *27*, 4756–4762.

[32] Li, H.; Eddaoudi, M.; O'Keeffe, M.; Yaghi, O. M. Design and Synthesis of an Exceptionally Stable and Highly Porous Metal–Organic Framework. *Nature* **1999**, *402*, 276–279.

[33] Furukawa, H.; Cordova, K. E.; O'Keeffe, M.; Yaghi, O. M. The Chemistry and Applications of Metal–Organic Frameworks. *Science* **2013**, *341*, 1230444.

[34] Moghadam, P. Z.; Li, A.; Wiggin, S. B.; Tao, A.; Maloney, A. G. P.; Wood, P. A.; Ward, S. C.; Fairen-Jimenez, D. Development of a Cambridge Structural Database Subset: A Collection of Metal–Organic Frameworks for Past, Present, and Future. *Chem. Mater.* **2017**, *29*, 2618–2625.

[35] Deng, H.; Doonan, C. J.; Furukawa, H.; Ferreira, R. B.; Towne, J.; Knobler, C. B.; Wang, B.; Yaghi, O. M. Multiple Functional Groups of Varying Ratios in Metal–Organic Frameworks. *Science* **2010**, *327*, 846–850.

[36] Eddaoudi, M.; Kim, J.; Rosi, N.; Vodak, D.; Wachter, J.; O'Keeffe, M.; Yaghi, O. M. Systematic Design of Pore Size and Functionality in Isoreticular MOFs and Their Application in Methane Storage. *Science* **2002**, *295*, 469–472.

[37] Islamoglu, T.; Goswami, S.; Li, Z.; Howarth, A. J.; Farha, O. K.; Hupp, J. T. Postsynthetic Tuning of Metal–Organic Frameworks for Targeted Applications. *Acc. Chem. Res.* **2017**, *50*, 805–813.

[38] Evans, J. D.; Sumby, C. J.; Doonan, C. J. Post-Synthetic Metalation of Metal–Organic Frameworks. *Chem. Soc. Rev.* **2014**, *43*, 5933–5951.

[39] Cohen, S. M. Postsynthetic Methods for the Functionalization of Metal–Organic Frameworks. *Chem. Rev.* **2012**, *112*, 970–1000.

[40] Suh, M. P.; Park, H. J.; Prasad, T. K.; Lim, D.-W. Hydrogen Storage in Metal–Organic Frameworks. *Chem. Rev.* **2012**, *112*, 782–835.

[41] Sumida, K.; Rogow, D. L.; Mason, J. A.; McDonald, T. M.; Bloch, E. D.; Herm, Z. R.; Bae, T.-H.; Long, J. R.

Carbon Dioxide Capture in Metal–Organic Frameworks. *Chem. Rev.* **2012**, *112*, 724–781.

[42] Kreno, L. E.; Leong, K.; Farha, O. K.; Allendorf, M.; Van Duyne, R. P.; Hupp, J. T. Metal–Organic Framework Materials as Chemical Sensors. *Chem. Rev.* **2012**, *112*, 1105–1125.

[43] Doonan, C. J.; Sumby, C. J. Metal–Organic Framework Catalysis. *CrystEngComm* **2017**, *19*, 4044–4048.

[44] Ricco, R.; Pfeiffer, C.; Sumida, K.; Sumby, C. J.; Falcaro, P.; Furukawa, S.; Champness, N. R.; Doonan, C. J. Emerging Applications of Metal–Organic Frameworks. *CrystEngComm* **2016**, *18*, 6532–6542.

[45] Dechnik, J.; Gascon, J.; Doonan, C. J.; Janiak, C.; Sumby, C. J. Mixed-Matrix Membranes. *Angew. Chem. Int. Ed.* **2017**, *56*, 9292–9310.

[46] Stassen, I.; Burtch, N.; Talin, A.; Falcaro, P.; Allendorf, M.; Ameloot, R. An Updated Roadmap for the Integration of Metal–Organic Frameworks with Electronic Devices and Chemical Sensors. *Chem. Soc. Rev.* **2017**, *46*, 3185–3241.

[47] Shekhah, O.; Liu, J.; Fischer, R. A.; Woll, C. MOF Thin Films: Existing and Future Applications. *Chem. Soc. Rev.* **2011**, *40*, 1081–1106.

[48] Haraguchi, T.; Otsubo, K.; Kitagawa, H. Emergence of Surface- and Interface-Induced Structures and Properties in Metal–Organic Framework Thin Films. *Eur. J. Inorg. Chem.* **2018**, *2018*, 1697–1706.

[49] Furukawa, S.; Reboul, J.; Diring, S.; Sumida, K.; Kitagawa, S. Structuring of Metal–Organic Frameworks at the Mesoscopic/Macroscopic Scale. *Chem. Soc. Rev.* **2014**, *43*, 5700–5734.

[50] Moitra, N.; Fukumoto, S.; Reboul, J.; Sumida, K.; Zhu, Y.; Nakanishi, K.; Furukawa, S.; Kitagawa, S.; Kanamori, K. Mechanically Stable, Hierarchically Porous $Cu_3(Btc)_2$ (HKUST-1) Monoliths via Direct Conversion of Copper(Ii) Hydroxide-Based Monoliths. *Chem. Commun.* **2015**, *51*, 3511–3514.

[51] Sumida, K.; Moitra, N.; Reboul, J.; Fukumoto, S.; Nakanishi, K.; Kanamori, K.; Furukawa, S.; Kitagawa, S. Mesoscopic Superstructures of Flexible Porous Coordination Polymers Synthesized via Coordination Replication. *Chem. Sci.* **2015**, *6*, 5938–5946.

[52] Zhuang, J.-L.; Terfort, A.; Wöll, C. Formation of Oriented and Patterned Films of Metal–Organic Frameworks by Liquid Phase Epitaxy: A Review. *Coord. Chem. Rev.* **2016**, *307*, 391–424.

[53] Stassen, I.; Campagnol, N.; Fransaer, J.; Vereecken, P.; De Vos, D.; Ameloot, R. Solvent-Free Synthesis of Supported ZIF-8 Films and Patterns through Transformation of Deposited Zinc Oxide Precursors. *CrystEngComm* **2013**, *15*, 9308–9311.

[54] Falcaro, P.; Ricco, R.; Doherty, C. M.; Liang, K.; Hill, A. J.; Styles, M. J. MOF Positioning Technology and Device Fabrication. *Chem. Soc. Rev.* **2014**, *43*, 5513–5560.

[55] Falcaro, P.; Okada, K.; Hara, T.; Ikigaki, K.; Tokudome, Y.; Thornton, A. W.; Hill, A. J.; Williams, T.; Doonan, C.; Takahashi, M. Centimetre-Scale Micropore Alignment in Oriented Polycrystalline Metal–Organic Framework Films via Heteroepitaxial Growth. *Nat. Mater.* **2017**, *16*, 342–348.

[56] Heinke, L.; Gliemann, H.; Tremouilhac, P.; Wöll, C. *The Chemistry of Metal–Organic Frameworks*; John Wiley & Sons, Ltd, 2016; pp 523–550, DOI: 10.1002/9783527693078.ch17.

[57] Liu, B.; Shekhah, O.; Arslan, H. K.; Liu, J.; Wöll, C.; Fischer, R. A. Enantiopure Metal–Organic Framework Thin Films: Oriented SURMOF Growth and Enantioselective Adsorption. *Angew. Chem. Int. Ed.* **2012**, *51*, 807–810.

[58] Biemmi, E.; Scherb, C.; Bein, T. Oriented Growth of the Metal Organic Framework $Cu_3(BTC)_2(H_2O)_3{\cdot}xH_2O$ Tunable with Functionalized Self-Assembled Monolayers. *J. Am. Chem. Soc.* **2007**, *129*, 8054–8055.

[59] Liu, J.; Shekhah, O.; Stammer, X.; Arslan, H. K.; Liu, B.; Schüpbach, B.; Terfort, A.; Wöll, C. Deposition of Metal–Organic Frameworks by Liquid-Phase Epitaxy: The Influence of Substrate Functional Group Density on Film Orientation. *Materials* **2012**, *5*, 1581–1592.

[60] Delen, G.; Ristanović, Z.; Mandemaker, L. D. B.; Weckhuysen, B. M. Mechanistic Insights into Growth of Surface-Mounted Metal–Organic Framework Films Resolved by Infrared (Nano-) Spectroscopy. *Chem. Eur. J.* **2018**, *24*, 187–195.

[61] Horcajada, P.; Gref, R.; Baati, T.; Allan, P. K.; Maurin, G.; Couvreur, P.; Férey, G.; Morris, R. E.; Serre, C. Metal–Organic Frameworks in Biomedicine. *Chem. Rev.* **2012**, *112*, 1232–1268.

[62] Chen, W.; Wu, C. Synthesis, Functionalization, and Applications of Metal–Organic Frameworks in Biomedicine. *Dalton Trans.* **2018**, *47*, 2114–2133.

[63] Simagina, A. A.; Polynski, M. V.; Vinogradov, A. V.; Pidko, E. A. Towards Rational Design of Metal–Organic

Framework-Based Drug Delivery Systems. *Russ. Chem. Rev.* **2018**, *87*, 831–858.

[64] Drout, R. J.; Robison, L.; Farha, O. K. Catalytic Applications of Enzymes Encapsulated in Metal–Organic Frameworks. *Coord. Chem. Rev.* **2019**, *381*, 151–160.

[65] Doonan, C.; Riccò, R.; Liang, K.; Bradshaw, D.; Falcaro, P. Metal–Organic Frameworks at the Biointerface: Synthetic Strategies and Applications. *Acc. Chem. Res.* **2017**, *50*, 1423–1432.

[66] Riccò, R.; Liang, W.; Li, S.; Gassensmith, J. J.; Caruso, F.; Doonan, C.; Falcaro, P. Metal–Organic Frameworks for Cell and Virus Biology: A Perspective. *ACS Nano* **2018**, *12*, 13–23.

[67] Liang, K.; Ricco, R.; Doherty, C. M.; Styles, M. J.; Bell, S.; Kirby, N.; Mudie, S.; Haylock, D.; Hill, A. J.; Doonan, C. J.; Falcaro, P. Biomimetic Mineralization of Metal–Organic Frameworks as Protective Coatings for Biomacromolecules. *Nat. Commun.* **2015**, *6*, 7240.

[68] Park, K. S.; Ni, Z.; Côté, A. P.; Choi, J. Y.; Huang, R.; Uribe-Romo, F. J.; Chae, H. K.; O'Keeffe, M.; Yaghi, O. M. Exceptional Chemical and Thermal Stability of Zeolitic Imidazolate Frameworks. *Proc. Natl. Acad. Sci. U.S.A.* **2006**, *103*, 10186–10191.

[69] Huang, X.-C.; Lin, Y.-Y.; Zhang, J.-P.; Chen, X.-M. Ligand-Directed Strategy for Zeolite-Type Metal–Organic Frameworks: Zinc(II) Imidazolates with Unusual Zeolitic Topologies. *Angew. Chem. Int. Ed.* **2006**, *45*, 1557–1559.

[70] Pan, Y.; Liu, Y.; Zeng, G.; Zhao, L.; Lai, Z. Rapid Synthesis of Zeolitic Imidazolate Framework-8 (ZIF-8) Nanocrystals in an Aqueous System. *Chem. Commun.* **2011**, *47*, 2071–2073.

[71] Jian, M.; Liu, B.; Liu, R.; Qu, J.; Wang, H.; Zhang, X. Water-Based Synthesis of Zeolitic Imidazolate Framework-8 with High Morphology Level at Room Temperature. *RSC Adv.* **2015**, *5*, 48433–48441.

[72] Kida, K.; Okita, M.; Fujita, K.; Tanaka, S.; Miyake, Y. Formation of High Crystalline ZIF-8 in an Aqueous Solution. *CrystEngComm* **2013**, *15*, 1794–1801.

[73] Li, S.; Dharmarwardana, M.; Welch, R. P.; Ren, Y.; Thompson, C. M.; Smaldone, R. A.; Gassensmith, J. J. Template-Directed Synthesis of Porous and Protective Core–Shell Bionanoparticles. *Angew. Chem. Int. Ed.* **2016**, *55*, 10691–10696.

[74] Li, S.; Dharmarwardana, M.; Welch, R. P.; Benjamin, C. E.; Shamir, A. M.; Nielsen, S. O.; Gassensmith, J. J. Investigation of Controlled Growth of Metal–Organic Frameworks on Anisotropic Virus Particles. *ACS Appl. Mater. Interfaces* **2018**, *10*, 18161–18169.

[75] Liang, K.; Richardson, J. J.; Doonan, C. J.; Mulet, X.; Ju, Y.; Cui, J.; Caruso, F.; Falcaro, P. An Enzyme-Coated Metal–Organic Framework Shell for Synthetically Adaptive Cell Survival. *Angew. Chem. Int. Ed.* **2017**, *56*, 8510–8515.

[76] Liang, K.; Richardson, J. J.; Cui, J.; Caruso, F.; Doonan, C. J.; Falcaro, P. Metal–Organic Framework Coatings as Cytoprotective Exoskeletons for Living Cells. *Adv. Mater.* **2016**, *28*, 7910–7914.

[77] Lyu, F.; Zhang, Y.; Zare, R. N.; Ge, J.; Liu, Z. One-Pot Synthesis of Protein-Embedded Metal–Organic Frameworks with Enhanced Biological Activities. *Nano Lett.* **2014**, *14*, 5761–5765.

[78] Colón, Y. J.; Snurr, R. Q. High-Throughput Computational Screening of Metal–Organic Frameworks. *Chem. Soc. Rev.* **2014**, *43*, 5735–5749.

[79] Sliwoski, G.; Kothiwale, S.; Meiler, J.; Lowe, E. W. Computational Methods in Drug Discovery. *Pharmacol. Rev.* **2014**, *66*, 334–395.

[80] Kim, S.; Thiessen, P. A.; Bolton, E. E.; Chen, J.; Fu, G.; Gindulyte, A.; Han, L.; He, J.; He, S.; Shoemaker, B. A.; Wang, J.; Yu, B.; Zhang, J.; Bryant, S. H. PubChem Substance and Compound Databases. *Nucleic Acids Res.* **2016**, *44*, D1202–D1213.

[81] Sterling, T.; Irwin, J. J. ZINC 15 – Ligand Discovery for Everyone. *J. Chem. Inf. Model.* **2015**, *55*, 2324–2337.

[82] Gaulton, A. et al. The ChEMBL Database in 2017. *Nucleic Acids Res.* **2017**, *45*, D945–D954.

[83] Lipinski, C. A.; Lombardo, F.; Dominy, B. W.; Feeney, P. J. Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings. *Adv. Drug Deliv. Rev.* **1997**, *23*, 3–25.

[84] Coudert, F.-X.; Fuchs, A. H. Computational Characterization and Prediction of Metal–Organic Framework Properties. *Coord. Chem. Rev.* **2016**, *307*, 211–236.

[85] Boyd, P. G.; Lee, Y.; Smit, B. Computational Development of the Nanoporous Materials Genome. *Nat. Rev. Mater.* **2017**, *2*, No. 17037.

[86] Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. Commentary: The Materials Project: A Materials Genome Approach to Accelerating Materials Innovation. *APL Mater.* **2013**, *1*, 011002.

[87] Simon, C. M.; Kim, J.; Gómez-Gualdrón, D. A.; Camp, J. S.; Chung, Y. G.; Martin, R. L.; Mercado, R.; Deem, M. W.; Gunter, D.; Haranczyk, M.; Sholl, D. S.; Snurr, R. Q.; Smit, B. The Materials Genome in Action: Identifying the Performance Limits for Methane Storage. *Energy Environ. Sci.* **2015**, *8*, 1190–1199.

[88] Bao, Y.; Martin, R. L.; Haranczyk, M.; Deem, M. W. In Silico Prediction of MOFs with High Deliverable Capacity or Internal Surface Area. *Phys. Chem. Chem. Phys.* **2015**, *17*, 11962–11973.

[89] Thornton, A. W.; Winkler, D. A.; Liu, M. S.; Haranczyk, M.; Kennedy, D. F. Towards Computational Design of Zeolite Catalysts for CO2 Reduction. *RSC Adv.* **2015**, *5*, 44361–44370.

[90] Simon, C. M.; Mercado, R.; Schnell, S. K.; Smit, B.; Haranczyk, M. What Are the Best Materials To Separate a Xenon/Krypton Mixture? *Chem. Mater.* **2015**, *27*, 4459–4475.

[91] Lee, K.; Howe, J. D.; Lin, L.-C.; Smit, B.; Neaton, J. B. Small-Molecule Adsorption in Open-Site Metal–Organic Frameworks: A Systematic Density Functional Theory Study for Rational Design. *Chem. Mater.* **2015**, *27*, 668–678.

[92] Colón, Y. J.; Fairen-Jimenez, D.; Wilmer, C. E.; Snurr, R. Q. High-Throughput Screening of Porous Crystalline Materials for Hydrogen Storage Capacity near Room Temperature. *J. Phys. Chem. C* **2014**, *118*, 5383–5389.

[93] Moghadam, P. Z.; Islamoglu, T.; Goswami, S.; Exley, J.; Fantham, M.; Kaminski, C. F.; Snurr, R. Q.; Farha, O. K.; Fairen-Jimenez, D. Computer-Aided Discovery of a Metal–Organic Framework with Superior Oxygen Uptake. *Nat. Commun.* **2018**, *9*, 1378.

[94] Chung, Y. G.; Camp, J.; Haranczyk, M.; Sikora, B. J.; Bury, W.; Krungleviciute, V.; Yildirim, T.; Farha, O. K.; Sholl, D. S.; Snurr, R. Q. Computation-Ready, Experimental Metal–Organic Frameworks: A Tool To Enable High-Throughput Screening of Nanoporous Crystals. *Chem. Mater.* **2014**, *26*, 6185–6192.

[95] Martin, R. L.; Simon, C. M.; Smit, B.; Haranczyk, M. In Silico Design of Porous Polymer Networks: High-Throughput Screening for Methane Storage Materials. *J. Am. Chem. Soc.* **2014**, *136*, 5006–5022.

[96] Martin, R. L.; Simon, C. M.; Medasani, B.; Britt, D. K.; Smit, B.; Haranczyk, M. In Silico Design of Three-Dimensional Porous Covalent Organic Frameworks via Known Synthesis Routes and Commercially Available Species. *J. Phys. Chem. C* **2014**, *118*, 23790–23802.

[97] Boyd, P. G.; Woo, T. K. A Generalized Method for Constructing Hypothetical Nanoporous Materials of Any Net Topology from Graph Theory. *CrystEngComm* **2016**, *18*, 3777–3792.

[98] Addicoat, M. A.; Coupry, D. E.; Heine, T. AuToGraFS: Automatic Topological Generator for Framework Structures. *J. Phys. Chem. A* **2014**, *118*, 9607–9614.

[99] Colón, Y. J.; Gómez-Gualdrón, D. A.; Snurr, R. Q. Topologically Guided, Automated Construction of Metal–Organic Frameworks and Their Evaluation for Energy-Related Applications. *Cryst. Growth Des.* **2017**, *17*, 5801–5810.

[100] Martin, R. L.; Haranczyk, M. Construction and Characterization of Structure Models of Crystalline Porous Polymers. *Cryst. Growth Des.* **2014**, *14*, 2431–2440.

[101] Wilmer, C. E.; Leaf, M.; Lee, C. Y.; Farha, O. K.; Hauser, B. G.; Hupp, J. T.; Snurr, R. Q. Large-Scale Screening of Hypothetical Metal–Organic Frameworks. *Nat. Chem.* **2012**, *4*, 83–89.

[102] Yoneya, M.; Tsuzuki, S.; Aoyagi, M. Simulation of Metal–Organic Framework Self-Assembly. *Phys. Chem. Chem. Phys.* **2015**, *17*, 8649–8652.

[103] Biswal, D.; Kusalik, P. G. Probing Molecular Mechanisms of Self-Assembly in Metal–Organic Frameworks. *ACS Nano* **2017**, *11*, 258–268.

[104] Nguyen, V.; Grünwald, M. Microscopic Origins of Poor Crystallinity in the Synthesis of Covalent Organic Framework COF-5. *J. Am. Chem. Soc.* **2018**, *140*, 3306–3311.

[105] Li, H.; Chavez, A. D.; Li, H.; Li, H.; Dichtel, W. R.; Bredas, J.-L. Nucleation and Growth of Covalent Organic Frameworks from Solution: The Example of COF-5. *J. Am. Chem. Soc.* **2017**, *139*, 16310–16318.

[106] Hendon, C. H.; Wittering, K. E.; Chen, T.-H.; Kaveevivitchai, W.; Popov, I.; Butler, K. T.; Wilson, C. C.; Cruickshank, D. L.; Miljanić, O. Š.; Walsh, A. Absorbate-Induced Piezochromism in a Porous Molecular Crystal. *Nano Lett.* **2015**, *15*, 2149–2154.

[107] Butler, K. T.; Hendon, C. H.; Walsh, A. Electronic Chemical Potentials of Porous Metal–Organic Frameworks.

*J. Am. Chem. Soc.* **2014**, *136*, 2703–2706.

[108] Butler, K. T.; Hendon, C. H.; Walsh, A. Electronic Structure Modulation of Metal–Organic Frameworks for Hybrid Devices. *ACS Appl. Mater. Interfaces* **2014**, *6*, 22044–22050.

[109] Ryder, M. R.; Donà, L.; Vitillo, J. G.; Civalleri, B. Understanding and Controlling the Dielectric Response of Metal–Organic Frameworks. *ChemPlusChem* **2018**, *83*, 308–316.

[110] Vanduyfhuys, L.; Vandenbrande, S.; Wieme, J.; Waroquier, M.; Verstraelen, T.; Van Speybroeck, V. Extension of the QuickFF Force Field Protocol for an Improved Accuracy of Structural, Vibrational, Mechanical and Thermal Properties of Metal–Organic Frameworks. *J. Comput. Chem.* **2018**, *39*, 999–1011.

[111] Tan, J.-C.; Civalleri, B.; Lin, C.-C.; Valenzano, L.; Galvelis, R.; Chen, P.-F.; Bennett, T. D.; Mellot-Draznieks, C.; Zicovich-Wilson, C. M.; Cheetham, A. K. Exceptionally Low Shear Modulus in a Prototypical Imidazole-Based Metal–Organic Framework. *Phys. Rev. Lett.* **2012**, *108*, 095502.

[112] Rogge, S. M. J.; Waroquier, M.; Van Speybroeck, V. Reliably Modeling the Mechanical Stability of Rigid and Flexible Metal–Organic Frameworks. *Acc. Chem. Res.* **2018**, *51*, 138–148.

[113] Evans, J. D.; Fraux, G.; Gaillac, R.; Kohen, D.; Trousselet, F.; Vanson, J.-M.; Coudert, F.-X. Computational Chemistry Methods for Nanoporous Materials. *Chem. Mater.* **2017**, *29*, 199–212.

[114] Ye, J.; Gagliardi, L.; Cramer, C. J.; Truhlar, D. G. Computational Screening of MOF-Supported Transition Metal Catalysts for Activity and Selectivity in Ethylene Dimerization. *J. Catal.* **2018**, *360*, 160–167.

[115] Butler, K. T.; Hendon, C. H.; Walsh, A. Designing Porous Electronic Thin-Film Devices: Band Offsets and Heteroepitaxy. *Faraday Discuss.* **2017**, *201*, 207–219.

[116] Bristow, J. K.; Butler, K. T.; Svane, K. L.; Gale, J. D.; Walsh, A. Chemical Bonding at the Metal–Organic Framework/Metal Oxide Interface: Simulated Epitaxial Growth of MOF-5 on Rutile $TiO_2$. *J. Mater. Chem. A* **2017**, *5*, 6226–6232.

[117] Shaw, D. E.; Maragakis, P.; Lindorff-Larsen, K.; Piana, S.; Dror, R. O.; Eastwood, M. P.; Bank, J. A.; Jumper, J. M.; Salmon, J. K.; Shan, Y.; Wriggers, W. Atomic-Level Characterization of the Structural Dynamics of Proteins. *Science* **2010**, *330*, 341–346.

[118] Verploegh, R. J.; Wu, Y.; Boulfelfel, S. E.; Sholl, D. S. Quantitative Predictions of Molecular Diffusion in Binary Mixed-Linker Zeolitic Imidazolate Frameworks Using Molecular Simulations. *J. Phys. Chem. C* **2018**, *122*, 5627–5638.

[119] Rappe, A. K.; Casewit, C. J.; Colwell, K. S.; Goddard, W. A.; Skiff, W. M. UFF, a Full Periodic Table Force Field for Molecular Mechanics and Molecular Dynamics Simulations. *J. Am. Chem. Soc.* **1992**, *114*, 10024–10035.

[120] Addicoat, M. A.; Vankova, N.; Akter, I. F.; Heine, T. Extension of the Universal Force Field to Metal–Organic Frameworks. *J. Chem. Theory Comput.* **2014**, *10*, 880–891.

[121] Coupry, D. E.; Addicoat, M. A.; Heine, T. Extension of the Universal Force Field for Metal–Organic Frameworks. *J. Chem. Theory Comput.* **2016**, *12*, 5215–5225.

[122] Frenkel, D.; Smit, B. *Understanding Molecular Simulation: From Algorithms to Applications*, 2nd ed.; Academic Press: San Diego, 2002.

[123] Sun, X.; Lin, T.; Gezelter, J. D. Langevin Dynamics for Rigid Bodies of Arbitrary Shape. *J. Chem. Phys.* **2008**, *128*, 234107.

[124] Kubo, R. The Fluctuation-Dissipation Theorem. *Rep. Prog. Phys.* **1966**, *29*, 255–284.

[125] Allen, M. P.; Tildesley, D. J. *Computer Simulation of Liquids*; Clarendon Press: Oxford, 1987.

[126] Morriss-Andrews, A.; Rottler, J.; Plotkin, S. S. A Systematically Coarse-Grained Model for DNA and Its Predictions for Persistence Length, Stacking, Twist, and Chirality. *J. Chem. Phys.* **2010**, *132*, No. 035105.

[127] Haxton, T. K.; Mannige, R. V.; Zuckermann, R. N.; Whitelam, S. Modeling Sequence-Specific Polymers Using Anisotropic Coarse-Grained Sites Allows Quantitative Comparison with Experiment. *J. Chem. Theory Comput.* **2015**, *11*, 303–315.

[128] Rouquerol, J.; Rouquerol, F.; Llewellyn, P.; Maurin, G.; Sing, K. S. W. *Adsorption by Powders and Porous Solids: Principles, Methodology and Applications*, 2nd ed.; Academic Press, 2013.

[129] Landers, J.; Gor, G. Y.; Neimark, A. V. Density Functional Theory Methods for Characterization of Porous Materials. *Colloids Surf. A* **2013**, *437*, 3–32.

[130] Park, J.; Howe, J. D.; Sholl, D. S. How Reproducible Are Isotherm Measurements in Metal–Organic Frameworks?

*Chem. Mater.* **2017**, *29*, 10487–10495.

[131] Dubbeldam, D.; Calero, S.; Ellis, D. E.; Snurr, R. Q. RASPA: Molecular Simulation Software for Adsorption and Diffusion in Flexible Nanoporous Materials. *Mol. Simul.* **2016**, *42*, 81–101.

[132] Walton, K. S.; Snurr, R. Q. Applicability of the BET Method for Determining Surface Areas of Microporous Metal-Organic Frameworks. *J. Am. Chem. Soc.* **2007**, *129*, 8552–8556.

[133] Düren, T.; Millange, F.; Férey, G.; Walton, K. S.; Snurr, R. Q. Calculating Geometric Surface Areas as a Characterization Tool for Metal-Organic Frameworks. *J. Phys. Chem. C* **2007**, *111*, 15350–15356.

[134] Willems, T. F.; Rycroft, C. H.; Kazi, M.; Meza, J. C.; Haranczyk, M. Algorithms and Tools for High-Throughput Geometry-Based Analysis of Crystalline Porous Materials. *Microporous Mesoporous Mater.* **2012**, *149*, 134–141.

[135] Martin, R. L.; Smit, B.; Haranczyk, M. Addressing Challenges of Identifying Geometrically Diverse Sets of Crystalline Porous Materials. *J. Chem. Inf. Model.* **2012**, *52*, 308–318.

[136] Lee, Y.; Barthel, S. D.; Dłotko, P.; Moosavi, S. M.; Hess, K.; Smit, B. Quantifying Similarity of Pore-Geometry in Nanoporous Materials. *Nat. Commun.* **2017**, *8*, 15396.

[137] Lee, Y.; Barthel, S. D.; Dłotko, P.; Moosavi, S. M.; Hess, K.; Smit, B. High-Throughput Screening Approach for Nanoporous Materials Genome Using Topological Data Analysis: Application to Zeolites. *J. Chem. Theory Comput.* **2018**, *14*, 4427–4437.

[138] Holden, D.; Jelfs, K. E.; Trewin, A.; Willock, D. J.; Haranczyk, M.; Cooper, A. I. Gas Diffusion in a Porous Organic Cage: Analysis of Dynamic Pore Connectivity Using Molecular Dynamics Simulations. *J. Phys. Chem. C* **2014**, *118*, 12734–12743.

[139] Jiang, S.; Jelfs, K. E.; Holden, D.; Hasell, T.; Chong, S. Y.; Haranczyk, M.; Trewin, A.; Cooper, A. I. Molecular Dynamics Simulations of Gas Selectivity in Amorphous Porous Molecular Solids. *J. Am. Chem. Soc.* **2013**, *135*, 17818–17830.

[140] Evans, J. D.; Huang, D. M.; Hill, M. R.; Sumby, C. J.; Sholl, D. S.; Thornton, A. W.; Doonan, C. J. Molecular Design of Amorphous Porous Organic Cages for Enhanced Gas Storage. *J. Phys. Chem. C* **2015**, *119*, 7746–7754.

[141] Honig, B.; Nicholls, A. Classical Electrostatics in Biology and Chemistry. *Science* **1995**, *268*, 1144–1149.

[142] Fogolari, F.; Zuccato, P.; Esposito, G.; Viglino, P. Biomolecular Electrostatics with the Linearized Poisson-Boltzmann Equation. *Biophys. J.* **1999**, *76*, 1–16.

[143] Zhang, Z.; Witham, S.; Alexov, E. On the Role of Electrostatics on Protein-Protein Interactions. *Phys. Biol.* **2011**, *8*, 035001.

[144] McCammon, J. A. Darwinian Biophysics: Electrostatics and Evolution in the Kinetics of Molecular Binding. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 7683–7684.

[145] Zhou, H.-X.; Pang, X. Electrostatic Interactions in Protein Structure, Folding, Binding, and Condensation. *Chem. Rev.* **2018**, *118*, 1691–1741.

[146] Andelman, D.; Introduction to Electrostatics in Soft and Biological Matter. In *Soft Condensed Matter Physics in Molecular and Cell Biology*, 1st; Poon, W.C.K.; Andelman, D.; Eds.; CRC Press: Boca Raton, 2006; 26.

[147] Li, L.; Li, C.; Sarkar, S.; Zhang, J.; Witham, S.; Zhang, Z.; Wang, L.; Smith, N.; Petukh, M.; Alexov, E. DelPhi: A Comprehensive Suite for DelPhi Software and Associated Resources. *BMC Biophys.* **2012**, *5*, 9.

[148] Nicholls, A.; Honig, B. A Rapid Finite Difference Algorithm, Utilizing Successive over-Relaxation to Solve the Poisson–Boltzmann Equation. *J. Comput. Chem.* **1991**, *12*, 435–445.

[149] Jurrus, E. et al. Improvements to the APBS Biomolecular Solvation Software Suite. *Protein Sci.* **2018**, *27*, 112–128.

[150] Chakravorty, A.; Jia, Z.; Li, L.; Alexov, E. A New DelPhi Feature for Modeling Electrostatic Potential around Proteins: Role of Bound Ions and Implications for Zeta-Potential. *Langmuir* **2017**, *33*, 2283–2295.

[151] Ahmed, A.; Xie, Z.; Konstas, K.; Babarao, R.; Todd, B. D.; Hill, M. R.; Thornton, A. W. Porous Aromatic Frameworks Impregnated with Fullerenes for Enhanced Methanol/Water Separation. *Langmuir* **2014**, *30*, 14621–14630.

[152] Shaibani, M.; Smith, S. J. D.; Banerjee, P. C.; Konstas, K.; Zafari, A.; Lobo, D. E.; Nazari, M.; Hollenkamp, A. F.; Hill, M. R.; Majumder, M. Framework-Mediated Synthesis of Highly Microporous Onion-like Carbon: Energy Enhancement in Supercapacitors without Compromising Power. *J. Mater. Chem. A* **2017**, *5*, 2519–2529.

[153] Goesten, M. G.; Szécsényi, A.; de Lange, M. F.; Bavykina, A. V.; Gupta, K. B. S. S.; Kapteijn, F.; Gascon, J.

Sulfonated Porous Aromatic Frameworks as Solid Acid Catalysts. *ChemCatChem* **2016**, *8*, 961–967.

[154] Li, Y.; Ben, T.; Zhang, B.; Fu, Y.; Qiu, S. Ultrahigh Gas Storage Both at Low and High Pressures in KOH-Activated Carbonized Porous Aromatic Frameworks. *Sci. Rep.* **2013**, *3*, 2420.

[155] Rangel-Rangel, E.; Verde-Sesto, E.; Rasero-Almansa, A. M.; Iglesias, M.; Sanchez, F. Porous Aromatic Frameworks (PAFs) as Efficient Supports for N-Heterocyclic Carbene Catalysts. *Catal. Sci. Technol.* **2016**, *6*, 6037–6045.

[156] Verde-Sesto, E.; Merino, E.; Rangel-Rangel, E.; Corma, A.; Iglesias, M.; Sánchez, F. Postfunctionalized Porous Polymeric Aromatic Frameworks with an Organocatalyst and a Transition Metal Catalyst for Tandem Condensation–Hydrogenation Reactions. *ACS Sustain. Chem. Eng.* **2016**, *4*, 1078–1084.

[157] Van Humbeck, J. F.; McDonald, T. M.; Jing, X.; Wiers, B. M.; Zhu, G.; Long, J. R. Ammonia Capture in Porous Organic Polymers Densely Functionalized with Brønsted Acid Groups. *J. Am. Chem. Soc.* **2014**, *136*, 2432–2440.

[158] Barin, G.; Peterson, G. W.; Crocellà, V.; Xu, J.; Colwell, K. A.; Nandy, A.; Reimer, J. A.; Bordiga, S.; Long, J. R. Highly Effective Ammonia Removal in a Series of Brønsted Acidic Porous Polymers: Investigation of Chemical and Structural Variations. *Chem. Sci.* **2017**, *8*, 4399–4409.

[159] Lu, W.; Yuan, D.; Zhao, D.; Schilling, C. I.; Plietzsch, O.; Muller, T.; Bräse, S.; Guenther, J.; Blümel, J.; Krishna, R.; Li, Z.; Zhou, H.-C. Porous Polymer Networks: Synthesis, Porosity, and Applications in Gas Storage/Separation. *Chem. Mater.* **2010**, *22*, 5964–5972.

[160] Martin, R. L.; Shahrak, M. N.; Swisher, J. A.; Simon, C. M.; Sculley, J. P.; Zhou, H.-C.; Smit, B.; Haranczyk, M. Modeling Methane Adsorption in Interpenetrating Porous Polymer Networks. *J. Phys. Chem. C* **2013**, *117*, 20037–20042.

[161] Holst, J. R.; Stöckel, E.; Adams, D. J.; Cooper, A. I. High Surface Area Networks from Tetrahedral Monomers: Metal-Catalyzed Coupling, Thermal Polymerization, and "Click" Chemistry. *Macromolecules* **2010**, *43*, 8531–8538.

[162] Bunz, U. H. F.; Seehafer, K.; Geyer, F. L.; Bender, M.; Braun, I.; Smarsly, E.; Freudenberg, J. Porous Polymers Based on Aryleneethynylene Building Blocks. *Macromol. Rapid Commun.* **2014**, *35*, 1466–1496.

[163] Reece, C.; Willock, D. J.; Trewin, A. Modelling Analysis of the Structure and Porosity of Covalent Triazine-Based Frameworks. *Phys. Chem. Chem. Phys.* **2014**, *17*, 817–823.

[164] Jiang, J.-X.; Trewin, A.; Su, F.; Wood, C. D.; Niu, H.; Jones, J. T. A.; Khimyak, Y. Z.; Cooper, A. I. Microporous Poly(Tri(4-Ethynylphenyl)Amine) Networks: Synthesis, Properties, and Atomistic Simulation. *Macromolecules* **2009**, *42*, 2658–2666.

[165] Thomas, J. M. H.; Trewin, A. Amorphous PAF-1: Guiding the Rational Design of Ultraporous Materials. *J. Phys. Chem. C* **2014**, *118*, 19712–19722.

[166] Fayon, P.; Trewin, A. Formation Mechanism of Ultra Porous Framework Materials. *Phys. Chem. Chem. Phys.* **2016**, *18*, 16840–16847.

[167] Grünwald, M.; Geissler, P. L. Patterns without Patches: Hierarchical Self-Assembly of Complex Structures from Simple Building Blocks. *ACS Nano* **2014**, *8*, 5891–5897.

[168] Haxton, T. K.; Whitelam, S. Design Rules for the Self-Assembly of a Protein Crystal. *Soft Matter* **2012**, *8*, 3558–3562.

[169] Spiga, E.; Alemani, D.; Degiacomi, M. T.; Cascella, M.; Peraro, M. D. Electrostatic-Consistent Coarse-Grained Potentials for Molecular Simulations of Proteins. *J. Chem. Theory Comput.* **2013**, *9*, 3515–3526.

[170] Whitelam, S. Hierarchical Assembly May Be a Way to Make Large Information-Rich Structures. *Soft Matter* **2015**, *11*, 8225–8235.

[171] Abbott, L. J.; Colina, C. M. Atomistic Structure Generation and Gas Adsorption Simulations of Microporous Polymer Networks. *Macromolecules* **2011**, *44*, 4511–4519.

[172] Abbott, L. J.; McKeown, N. B.; Colina, C. M. Design Principles for Microporous Organic Solids from Predictive Computational Screening. *J. Mater. Chem. A* **2013**, *1*, 11950–11960.

[173] Abbott, L. J.; Colina, C. M. Formation of Microporosity in Hyper-Cross-Linked Polymers. *Macromolecules* **2014**, *47*, 5409–5415.

[174] Abbott, L. J.; Hughes, J. E.; Colina, C. M. Virtual Synthesis of Thermally Cross-Linked Copolymers from a Novel Implementation of Polymatic. *J. Phys. Chem. B* **2014**, *118*, 1916–1924.

[175] Yamamoto, T.; Yamamoto, A.; Ikeda, S. Organo (Dipyridyl) Nickel Complexes. I. Stability and Activation of

the Alkyl-Nickel Bonds of Dialkyl (Dipyridyl) Nickel by Coordination with Various Substituted Olefins. *J. Am. Chem. Soc.* **1971**, *93*, 3350–3359.

[176] Yamamoto, T.; Wakabayashi, S.; Osakada, K. Mechanism of C-C Coupling Reactions of Aromatic Halides, Promoted by $Ni(COD)_2$ in the Presence of 2,2-Bipyridine and $PPh_3$, to Give Biaryls. *J. Organomet. Chem.* **1992**, *428*, 223–237.

[177] Eglinton, G.; Galbraith, A. R. 182. Macrocyclic Acetylenic Compounds. Part I. Cyclotetradeca-1:3-Diyne and Related Compounds. *J. Chem. Soc.* **1959**, 889–896.

[178] *Materials Studio 5.0*; San Diego: Accelrys Software Inc., 2009.

[179] Connolly, M. L. Analytical Molecular Surface Calculation. *J. Appl. Crystallogr.* **1983**, *16*, 548–558.

[180] Brown, W. M.; Petersen, M. K.; Plimpton, S. J.; Grest, G. S. Liquid Crystal Nanodroplets in Solution. *J. Chem. Phys.* **2009**, *130*, 044901.

[181] Ebrahimi, D.; Whittle, A. J.; Pellenq, R. J.-M. Mesoscale Properties of Clay Aggregates from Potential of Mean Force Representation of Interactions Between Nanoplatelets. *J. Chem. Phys.* **2014**, *140*, 154309.

[182] Hargittai, M.; Hargittai, I. *Advances in Molecular Structure Research*; JAI Press, 1998; Vol. 4.

[183] Jorgensen, W. L.; Severance, D. L. Aromatic-Aromatic Interactions: Free Energy Profiles for the Benzene Dimer in Water, Chloroform, and Liquid Benzene. *J. Am. Chem. Soc.* **1990**, *112*, 4768–4774.

[184] Brown, W. M.; Wang, P.; Plimpton, S. J.; Tharrington, A. N. Implementing Molecular Dynamics on Hybrid High Performance Computers - Short Range Forces. *Comput. Phys. Commun.* **2011**, *182*, 898–911.

[185] Miller, T. F.; Eleftheriou, M.; Pattnaik, P.; Ndirango, A.; Newns, D.; Martyna, G. J. Symplectic Quaternion Scheme for Biophysical Molecular Dynamics. *J. Chem. Phys.* **2002**, *116*, 8649–8659.

[186] Martìnez, L.; Andrade, R.; Birgin, E. G.; Martìnez, J. M. PACKMOL: A Package for Building Initial Configurations for Molecular Dynamics Simulations. *J. Comput. Chem.* **2009**, *30*, 2157–2164.

[187] Stukowski, A. Visualization and Analysis of Atomistic Simulation Data with OVITO–the Open Visualization Tool. *Model. Simul. Mater. Sci. Eng.* **2010**, *18*, 015012.

[188] Ester, M.; Kriegel, H.-P.; Sander, J.; Xu, X. In *A Density-Based Algorithm for Discovering Clusters in Large Spatial Databases with Noise*. Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, Portland, Oregon, Aug 2–4, 1996. pp 226–231, Simoudis, E., Han, J., Fayyad, U., Eds.; AAAI Press: Portland, OR, 1996.

[189] Keys, A. S.; Iacovella, C. R.; Glotzer, S. C. Characterizing Complex Particle Morphologies Through Shape Matching: Descriptors, Applications, and Algorithms. *J. Comput. Phys.* **2011**, *230*, 6438–6463.

[190] Xu, X.; Nieuwenhuyzen, M.; James, S. L. A Nanoporous Metal–Organic Framework Based on Bulky Phosphane Ligands. *Angew. Chem. Int. Ed.* **2002**, *41*, 764–767.

[191] Rankine, D.; Avellaneda, A.; Hill, M. R.; Doonan, C. J.; Sumby, C. J. Control of Framework Interpenetration for in Situ Modified Hydroxyl Functionalised IRMOFs. *Chem. Commun.* **2012**, *48*, 10328–10330.

[192] Ferguson, A.; Liu, L.; Tapperwijn, S. J.; Perl, D.; Coudert, F.-X.; Van Cleuvenbergen, S.; Verbiest, T.; van der Veen, M. A.; Telfer, S. G. Controlled Partial Interpenetration in Metal–Organic Frameworks. *Nat. Chem.* **2016**, *8*, 250–257.

[193] Von Smoluchowski, M. Drei Vortrage Uber Diffusion. Brownsche Bewegung Und Koagulation von Kolloidteilchen. *Z. Phys.* **1916**, *17*, 557–585.

[194] Leyvraz, F. Scaling Theory and Exactly Solved Models in the Kinetics of Irreversible Aggregation. *Phys. Rep.* **2003**, *383*, 95–212.

[195] Mansbach, R. A.; Ferguson, A. L. Coarse-Grained Molecular Simulation of the Hierarchical Self-Assembly of $\pi$-Conjugated Optoelectronic Peptides. *J. Phys. Chem. B* **2017**, *121*, 1684–1706.

[196] Cruz-Chu, E. R.; Aksimentiev, A.; Schulten, K. Water-Silica Force Field for Simulating Nanodevices. *J. Phys. Chem. B* **2006**, *110*, 21497–21508.

[197] Moroni, M.; Le Moigne, J.; Luzzati, S. Rigid Rod Conjugated Polymers for Nonlinear Optics: 1. Characterization and Linear Optical Properties of Poly(Aryleneethynylene) Derivatives. *Macromolecules* **1994**, *27*, 562–571.

[198] Gawande, M. B.; Goswami, A.; Felpin, F. A.-X.; Asefa, T.; Huang, X.; Silva, R.; Zou, X.; Zboril, R.; Varma, R. S. Cu and Cu-Based Nanoparticles: Synthesis and Applications in Catalysis. *Chem. Rev.* **2016**, *116*, 3722–3811.

[199] Allen, S. E.; Walvoord, R. R.; Padilla-Salinas, R.; Kozlowski, M. C. Aerobic Copper-Catalyzed Organic Reactions. *Chem. Rev.* **2013**, *113*, 6234–6458.

[200] Wheeler, S. E.; Houk, K. N. Substituent Effects in the Benzene Dimer Are Due to Direct Interactions of the Substituents with the Unsubstituted Benzene. *J. Am. Chem. Soc.* **2008**, *130*, 10854–10855.

[201] Sinnokrot, M. O.; Sherrill, C. D. Substituent Effects in $\pi$-$\pi$ Interactions: Sandwich and T-Shaped Configurations. *J. Am. Chem. Soc.* **2004**, *126*, 7690–7697.

[202] Liu, J.; Wöll, C. Surface-Supported Metal–Organic Framework Thin Films: Fabrication Methods, Applications, and Challenges. *Chem. Soc. Rev.* **2017**, *46*, 5730–5770.

[203] Oldenburg, M.; Turshatov, A.; Busko, D.; Jakoby, M.; Haldar, R.; Chen, K.; Emandi, G.; Senge, M. O.; Wöll, C.; Hodgkiss, J. M.; Richards, B. S.; Howard, I. A. Enhancing the Photoluminescence of Surface Anchored Metal–Organic Frameworks: Mixed Linkers and Efficient Acceptors. *Phys. Chem. Chem. Phys.* **2018**, *20*, 11564–11576.

[204] Reboul, J.; Furukawa, S.; Horike, N.; Tsotsalas, M.; Hirai, K.; Uehara, H.; Kondo, M.; Louvain, N.; Sakata, O.; Kitagawa, S. Mesoscopic Architectures of Porous Coordination Polymers Fabricated by Pseudomorphic Replication. *Nat. Mater.* **2012**, *11*, 717–723.

[205] Stavila, V.; Volponi, J.; Katzenmeyer, A. M.; Dixon, M. C.; Allendorf, M. D. Kinetics and Mechanism of Metal–Organic Framework Thin Film Growth: Systematic Investigation of HKUST-1 Deposition on QCM Electrodes. *Chem. Sci.* **2012**, *3*, 1531–1540.

[206] Wang, Z.; Liu, J.; Lukose, B.; Gu, Z.; Weidler, P. G.; Gliemann, H.; Heine, T.; Wöll, C. Nanoporous Designer Solids with Huge Lattice Constant Gradients: Multiheteroepitaxy of Metal–Organic Frameworks. *Nano Lett.* **2014**, *14*, 1526–1529.

[207] Shekhah, O.; Wang, H.; Zacher, D.; Fischer, R. A.; Wöll, C. Growth Mechanism of Metal–Organic Frameworks: Insights into the Nucleation by Employing a Step-by-Step Route. *Angew. Chem. Int. Ed.* **2009**, *48*, 5038–5041.

[208] Wang, Z.; Weidler, P. G.; Azucena, C.; Heinke, L.; Wöll, C. Negative, Anisotropic Thermal Expansion in Monolithic Thin Films of Crystalline Metal–Organic Frameworks. *Microporous Mesoporous Mater.* **2016**, *222*, 241–246.

[209] Zur, A.; McGill, T. C. Lattice Match: An Application to Heteroepitaxy. *J. Appl. Phys.* **1984**, *55*, 378–386.

[210] Zur, A.; McGill, T. C.; Nicolet, M.-A. Transition-metal Silicides Lattice-matched to Silicon. *J. Appl. Phys.* **1985**, *57*, 600–603.

[211] Ding, H.; Dwaraknath, S. S.; Garten, L.; Ndione, P.; Ginley, D.; Persson, K. A. Computational Approach for Epitaxial Polymorph Stabilization through Substrate Selection. *ACS Appl. Mater. Interfaces* **2016**, *8*, 13086–13093.

[212] Raclariu, A.-M.; Deshpande, S.; Bruggemann, J.; Zhuge, W.; Yu, T.; Ratsch, C.; Shankar, S. A Fast Method for Predicting the Formation of Crystal Interfaces and Heterocrystals. *Comput. Mater. Sci.* **2015**, *108, Part A*, 88–93.

[213] Butler, K. T.; Kumagai, Y.; Oba, F.; Walsh, A. Screening Procedure for Structurally and Electronically Matched Contact Layers for High-Performance Solar Cells: Hybrid Perovskites. *J. Mater. Chem. C* **2016**, *4*, 1149–1158.

[214] Gómez-Gualdrón, D. A.; Colón, Y. J.; Zhang, X.; Wang, T. C.; Chen, Y.-S.; Hupp, J. T.; Yildirim, T.; Farha, O. K.; Zhang, J.; Snurr, R. Q. Evaluating Topologically Diverse Metal–Organic Frameworks for Cryo-Adsorbed Hydrogen Storage. *Energy Environ. Sci.* **2016**, *9*, 3279–3289.

[215] Larsen, A. H. et al. The Atomic Simulation Environment—a Python Library for Working with Atoms. *J. Phys. Condens. Matter* **2017**, *29*, 273002.

[216] Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. Python Materials Genomics (Pymatgen): A Robust, Open-Source Python Library for Materials Analysis. *Comput. Mater. Sci.* **2013**, *68*, 314–319.

[217] Oswald, H. R.; Reller, A.; Schmalle, H. W.; Dubler, E. Structure of Copper(II) Hydroxide, $Cu(OH)_2$. *Acta Crystallogr. C* **1990**, *46*, 2279–2284.

[218] Chen, J.; Fu, Y.; Samad, L.; Dang, L.; Zhao, Y.; Shen, S.; Guo, L.; Jin, S. Vapor-Phase Epitaxial Growth of Aligned Nanowire Networks of Cesium Lead Halide Perovskites ($CsPbX_3$, X = Cl, Br, I). *Nano Lett.* **2017**, *17*, 460–466.

[219] Sun, W.; Ceder, G. Efficient Creation and Convergence of Surface Slabs. *Surf. Sci.* **2013**, *617*, 53–59.

[220] Domenico, J.; Foster, M. E.; Spoerke, E. D.; Allendorf, M. D.; Sohlberg, K. Effect of Solvent and Substrate

on the Surface Binding Mode of Carboxylate-Functionalized Aromatic Molecules. *J. Phys. Chem. C* **2018**, *122*, 10846–10856.

[221] Mathew, K.; Singh, A. K.; Gabriel, J. J.; Choudhary, K.; Sinnott, S. B.; Davydov, A. V.; Tavazza, F.; Hennig, R. G. MPInterfaces: A Materials Project Based Python Tool for High-Throughput Computational Screening of Interfacial Systems. *Comput. Mater. Sci.* **2016**, *122*, 183–190.

[222] Ahrens, L. H. The Use of Ionization Potentials Part 1. Ionic Radii of the Elements. *Geochim. Cosmochim. Acta* **1952**, *2*, 155–169.

[223] Shannon, R. D. Revised Effective Ionic Radii and Systematic Studies of Interatomic Distances in Halides and Chalcogenides. *Acta Crystallogr. A* **1976**, *32*, 751–767.

[224] Furukawa, S.; Hirai, K.; Nakagawa, K.; Takashima, Y.; Matsuda, R.; Tsuruoka, T.; Kondo, M.; Haruki, R.; Tanaka, D.; Sakamoto, H.; Shimomura, S.; Sakata, O.; Kitagawa, S. Heterogeneously Hybridized Porous Coordination Polymer Crystals: Fabrication of Heterometallic Core–Shell Single Crystals with an In-Plane Rotational Epitaxial Relationship. *Angew. Chem. Int. Ed.* **2009**, *48*, 1766–1770.

[225] Witman, M.; Ling, S.; Boyd, P.; Barthel, S.; Haranczyk, M.; Slater, B.; Smit, B. Cutting Materials in Half: A Graph Theory Approach for Generating Crystal Surfaces and Its Prediction of 2D Zeolites. *ACS Cent. Sci.* **2018**, *4*, 235–245.

[226] Bristow, J. K.; Svane, K. L.; Tiana, D.; Skelton, J. M.; Gale, J. D.; Walsh, A. Free Energy of Ligand Removal in the Metal–Organic Framework UiO-66. *J. Phys. Chem. C* **2016**, *120*, 9276–9281.

[227] Fang, Z.; Bueken, B.; De Vos, D. E.; Fischer, R. A. Defect-Engineered Metal–Organic Frameworks. *Angew. Chem. Int. Ed.* **2015**, *54*, 7234–7254.

[228] St. Petkov, P.; Vayssilov, G. N.; Liu, J.; Shekhah, O.; Wang, Y.; Wöll, C.; Heine, T. Defects in MOFs: A Thorough Characterization. *ChemPhysChem* **2012**, *13*, 2025–2029.

[229] Altintas, C.; Avci, G.; Daglar, H.; Nemati Vesali Azar, A.; Velioglu, S.; Erucar, I.; Keskin, S. Database for CO2 Separation Performances of MOFs Based on Computational Materials Screening. *ACS Appl. Mater. Interfaces* **2018**, *10*, 17257–17268.

[230] Qiao, Z.; Xu, Q.; Jiang, J. High-Throughput Computational Screening of Metal–Organic Framework Membranes for Upgrading of Natural Gas. *J. Membr. Sci.* **2018**, *551*, 47–54.

[231] Evans, J. D.; Huang, D. M.; Haranczyk, M.; Thornton, A. W.; Sumby, C. J.; Doonan, C. J. Computational Identification of Organic Porous Molecular Crystals. *CrystEngComm* **2016**, *18*, 4133–4141.

[232] Mao, Y.; Su, B.; Cao, W.; Li, J.; Ying, Y.; Ying, W.; Hou, Y.; Sun, L.; Peng, X. Specific Oriented Metal–Organic Framework Membranes and Their Facet-Tuned Separation Performance. *ACS Appl. Mater. Interfaces* **2014**, *6*, 15676–15685.

[233] Amirjalayer, S.; Tafipolsky, M.; Schmid, R. Surface Termination of the Metal–Organic Framework HKUST-1: A Theoretical Investigation. *J. Phys. Chem. Lett.* **2014**, *5*, 3206–3210.

[234] Blatov, V. A.; Shevchenko, A. P.; Proserpio, D. M. Applied Topological Analysis of Crystal Structures with the Program Package ToposPro. *Cryst. Growth Des.* **2014**, *14*, 3576–3586.

[235] Alexandrov, E. V.; Blatov, V. A.; Kochetkov, A. V.; Proserpio, D. M. Underlying Nets in Three-Periodic Coordination Polymers: Topology, Taxonomy and Prediction from a Computer-Aided Analysis of the Cambridge Structural Database. *CrystEngComm* **2011**, *13*, 3947–3958.

[236] O'Keeffe, M.; Peskov, M. A.; Ramsden, S. J.; Yaghi, O. M. The Reticular Chemistry Structure Resource (RCSR) Database of, and Symbols for, Crystal Nets. *Acc. Chem. Res.* **2008**, *41*, 1782–1789.

[237] Bonneau, C.; O'Keeffe, M.; Proserpio, D. M.; Blatov, V. A.; Batten, S. R.; Bourne, S. A.; Lah, M. S.; Eon, J.-G.; Hyde, S. T.; Wiggin, S. B.; Öhrström, L. Deconstruction of Crystalline Networks into Underlying Nets: Relevance for Terminology Guidelines and Crystallographic Databases. *Cryst. Growth Des.* **2018**, *18*, 3411–3418.

[238] Otsubo, K.; Haraguchi, T.; Sakata, O.; Fujiwara, A.; Kitagawa, H. Step-by-Step Fabrication of a Highly Oriented Crystalline Three-Dimensional Pillared-Layer-Type Metal–Organic Framework Thin Film Confirmed by Synchrotron X-Ray Diffraction. *J. Am. Chem. Soc.* **2012**, *134*, 9605–9608.

[239] Martin, R. L.; Haranczyk, M. Insights into Multi-Objective Design of Metal–Organic Frameworks. *Cryst. Growth Des.* **2013**, *13*, 4208–4212.

[240] Hirai, K.; Sumida, K.; Meilikhov, M.; Louvain, N.; Nakahama, M.; Uehara, H.; Kitagawa, S.; Furukawa, S. Impact of Crystal Orientation on the Adsorption Kinetics of a Porous Coordination Polymer–Quartz Crystal

Microbalance Hybrid Sensor. *J. Mater. Chem. C* **2014**, *2*, 3336–3344.

[241] Shieh, F.-K.; Wang, S.-C.; Yen, C.-I.; Wu, C.-C.; Dutta, S.; Chou, L.-Y.; Morabito, J. V.; Hu, P.; Hsu, M.-H.; Wu, K. C.-W.; Tsung, C.-K. Imparting Functionality to Biocatalysts via Embedding Enzymes into Nanoporous Materials by a de Novo Approach: Size-Selective Sheltering of Catalase in Metal–Organic Framework Microcrystals. *J. Am. Chem. Soc.* **2015**, *137*, 4276–4279.

[242] Ricco, R.; Konstas, K.; Styles, M. J.; Richardson, J. J.; Babarao, R.; Suzuki, K.; Scopece, P.; Falcaro, P. Lead(II) Uptake by Aluminium Based Magnetic Framework Composites (MFCs) in Water. *J. Mater. Chem. A* **2015**, *3*, 19822–19831.

[243] Lismont, M.; Dreesen, L.; Wuttke, S. Metal–Organic Framework Nanoparticles in Photodynamic Therapy: Current Status and Perspectives. *Adv. Funct. Mater.* **2017**, *27*, 1606314.

[244] Bjellqvist, B.; Basse, B.; Olsen, E.; Celis, J. E. Reference Points for Comparisons of Two-Dimensional Maps of Proteins from Different Human Cell Types Defined in a pH Scale Where Isoelectric Points Correlate with Polypeptide Compositions. *Electrophoresis* **1994**, *15*, 529–539.

[245] Patrickios, C. S.; Yamasaki, E. N. Polypeptide Amino Acid Composition and Isoelectric Point II. Comparison between Experiment and Theory. *Anal. Biochem.* **1995**, *231*, 82–91.

[246] Foster, J. F.; Kaplan, L. J. Isoelectric Focussing Behavior of Bovine Plasma Albumin, Mercaptalbumin, and $\beta$-Lactoglobulins A and B. *Biochemistry* **1971**, *10*, 630–636.

[247] Trodler, P.; Nieveler, J.; Rusnak, M.; Schmid, R. D.; Pleiss, J. Rational Design of a New One-Step Purification Strategy for Candida Antarctica Lipase B by Ion-Exchange Chromatography. *J. Chromatogr. A* **2008**, *1179*, 161–167.

[248] Samejima, T.; Kamata, M.; Shibata, K. Dissociation of Bovine Liver Catalase at Low pH. *J. Biochem.* **1962**, *51*, 181–187.

[249] Shannon, L. M.; Kay, E.; Lew, J. Y. Peroxidase Isozymes from Horseradish Roots I. Isolation and Physical Properties. *J. Biol. Chem.* **1966**, *241*, 2166–2172.

[250] Schmid, A.; Dordick, J. S.; Hauer, B.; Kiener, A.; Wubbolts, M.; Witholt, B. Industrial Biocatalysis Today and Tomorrow. *Nature* **2001**, *409*, 258–268.

[251] Hunter, R. J. *Zeta Potential in Colloid Science: Principles and Applications*; Academic Press, London, 1981.

[252] Freeze, H. H. Genetic Defects in the Human Glycome. *Nat. Rev. Genet.* **2006**, *7*, 537–551.

[253] Osborn, H. M. I.; Evans, P. G.; Gemmell, N.; Osborne, S. D. Carbohydrate-Based Therapeutics. *J. Pharm. Pharmacol.* **2004**, *56*, 691–702.

[254] Mitragotri, S.; Burke, P. A.; Langer, R. Overcoming the Challenges in Administering Biopharmaceuticals: Formulation and Delivery Strategies. *Nat. Rev. Drug Discov.* **2014**, *13*, 655–672.

[255] Cock, P. J. A.; Antao, T.; Chang, J. T.; Chapman, B. A.; Cox, C. J.; Dalke, A.; Friedberg, I.; Hamelryck, T.; Kauff, F.; Wilczynski, B.; de Hoon, M. J. L. Biopython: Freely Available Python Tools for Computational Molecular Biology and Bioinformatics. *Bioinformatics* **2009**, *25*, 1422–1423.

[256] Bjellqvist, B.; Hughes, G. J.; Pasquali, C.; Paquet, N.; Ravier, F.; Sanchez, J.-C.; Frutiger, S.; Hochstrasser, D. The Focusing Positions of Polypeptides in Immobilized pH Gradients Can Be Predicted from Their Amino Acid Sequences. *Electrophoresis* **1993**, *14*, 1023–1031.

[257] Hamelryck, T.; Manderick, B. PDB File Parser and Structure Class Implemented in Python. *Bioinformatics* **2003**, *19*, 2308–2310.

[258] Kyte, J.; Doolittle, R. F. A Simple Method for Displaying the Hydropathic Character of a Protein. *J. Mol. Biol.* **1982**, *157*, 105–132.

[259] Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242.

[260] Sielecki, A. R.; Fedorov, A. A.; Boodhoo, A.; Andreeva, N. S.; James, M. N. Molecular and Crystal Structures of Monoclinic Porcine Pepsin Refined at 1.8 Å Resolution. *J. Mol. Biol.* **1990**, *214*, 143–170.

[261] Bujacz, A. Structures of Bovine, Equine and Leporine Serum Albumin. *Acta Crystallogr. D* **2012**, *68*, 1278–1289.

[262] Uppenberg, J.; Hansen, M. T.; Patkar, S.; Jones, T. The Sequence, Crystal Structure Determination and Refinement of Two Crystal Forms of Lipase B from Candida Antarctica. *Structure* **1994**, *2*, 293–308.

[263] Purwar, N.; McGarry, J. M.; Kostera, J.; Pacheco, A. A.; Schmidt, M. Interaction of Nitric Oxide with Catalase:

Structural and Kinetic Analysis. *Biochemistry* **2011**, *50*, 4491–4503.

[264] Carlsson, G. H.; Nicholls, P.; Svistunenko, D.; Berglund, G. I.; Hajdu, J. Complexes of Horseradish Peroxidase with Formate, Acetate, and Carbon Monoxide. *Biochemistry* **2005**, *44*, 635–642.

[265] Copeland, D. M.; Soares, A. S.; West, A. H.; Richter-Addo, G. B. Crystal Structures of the Nitrite and Nitric Oxide Complexes of Horse Heart Myoglobin. *J. Inorg. Biochem.* **2006**, *100*, 1413–1425.

[266] Park, S.-Y.; Yokoyama, T.; Shibayama, N.; Shiro, Y.; Tame, J. R. 1.25 Å Resolution Crystal Structures of Human Haemoglobin in the Oxy, Deoxy and Carbonmonoxy Forms. *J. Mol. Biol.* **2006**, *360*, 690–701.

[267] Transue, T. R.; Krahn, J. M.; Gabel, S. A.; DeRose, E. F.; London, R. E. X-Ray and NMR Characterization of Covalent Complexes of Trypsin, Borate, and Alcohols. *Biochemistry* **2004**, *43*, 2829–2839.

[268] Wang, J.; Dauter, M.; Alkire, R.; Joachimiak, A.; Dauter, Z. Triclinic Lysozyme at 0.65 Å Resolution. *Acta Crystallogr. D* **2007**, *63*, 1254–1268.

[269] Søndergaard, C. R.; Olsson, M. H. M.; Rostkowski, M.; Jensen, J. H. Improved Treatment of Ligands and Coupling Effects in Empirical Calculation and Rationalization of pKa Values. *J. Chem. Theory Comput.* **2011**, *7*, 2284–2295.

[270] Olsson, M. H. M.; Søndergaard, C. R.; Rostkowski, M.; Jensen, J. H. PROPKA3: Consistent Treatment of Internal and Surface Residues in Empirical pKa Predictions. *J. Chem. Theory Comput.* **2011**, *7*, 525–537.

[271] Wang, L.; Li, L.; Alexov, E. pKa Predictions for Proteins, RNAs, and DNAs with the Gaussian Dielectric Function Using DelPhi pKa. *Proteins* **2015**, *83*, 2186–2197.

[272] Dolinsky, T. J.; Nielsen, J. E.; McCammon, J. A.; Baker, N. A. PDB2PQR: An Automated Pipeline for the Setup of Poisson–Boltzmann Electrostatics Calculations. *Nucleic Acids Res.* **2004**, *32*, W665–W667.

[273] Dolinsky, T. J.; Czodrowski, P.; Li, H.; Nielsen, J. E.; Jensen, J. H.; Klebe, G.; Baker, N. A. PDB2PQR: Expanding and Upgrading Automated Preparation of Biomolecular Structures for Molecular Simulations. *Nucleic Acids Res.* **2007**, *35*, W522–W525.

[274] Wang, J.; Cieplak, P.; Kollman, P. A. How Well Does a Restrained Electrostatic Potential (RESP) Model Perform in Calculating Conformational Energies of Organic and Biological Molecules? *J. Comput. Chem.* **2000**, *21*, 1049–1074.

[275] Unni, S.; Huang, Y.; Hanson, R.; Tobias, M.; Krishnan, S.; Li, W. W.; Nielsen, J. E.; Baker, N. A. Web Servers and Services for Electrostatics Calculations with APBS and PDB2PQR. *J. Comput. Chem.* **2011**, *32*, 1488–1491.

[276] Rubinstein, M.; Colby, R. H. *Polymer Physics*; Oxford University Press, Oxford, 2003.

[277] Marszalek, P. E.; Oberhauser, A. F.; Pang, Y.-P.; Fernandez, J. M. Polysaccharide Elasticity Governed by Chair–Boat Transitions of the Glucopyranose Ring. *Nature* **1998**, *396*, 661–664.

[278] Rief, M.; Oesterhelt, F.; Heymann, B.; Gaub, H. E. Single Molecule Force Spectroscopy on Polysaccharides by Atomic Force Microscopy. *Science* **1997**, *275*, 1295–1297.

[279] Simoes, M. C.; Hughes, K. J.; Ingham, D. B.; Ma, L.; Pourkashanian, M. Estimation of the Thermochemical Radii and Ionic Volumes of Complex Ions. *Inorg. Chem.* **2017**, *56*, 7566–7573.

[280] Rayer, A. V.; Sumon, K. Z.; Jaffari, L.; Henni, A. Dissociation Constants (pKa) of Tertiary and Cyclic Amines: Structural and Temperature Dependences. *J. Chem. Eng. Data* **2014**, *59*, 3805–3813.

[281] Oliphant, T. E. Python for Scientific Computing. *Comput. Sci. Eng.* **2007**, *9*, 10–20.

[282] Wall, F. T.; Berkowitz, J. Numerical Solution to the Poisson-Boltzmann Equation for Spherical Polyelectrolyte Molecules. *J. Chem. Phys.* **1957**, *26*, 114–122.

[283] Ohshima, H.; Kondo, T. Electrostatic Double-Layer Interaction between Two Charged Ion-Penetrable Spheres: An Exactly Solvable Model. *J. Colloid Interface Sci.* **1993**, *155*, 499–505.

[284] Muckerheide, A.; Apple, R. J.; Pesce, A. J.; Michael, J. G. Cationization of Protein Antigens. I. Alteration of Immunogenic Properties. *J. Immunol.* **1987**, *138*, 833–837.

[285] Spadiut, O.; Rossetti, L.; Dietzsch, C.; Herwig, C. Purification of a Recombinant Plant Peroxidase Produced in Pichia Pastoris by a Simple 2-Step Strategy. *Protein Expr. Purif.* **2012**, *86*, 89–97.

[286] Høegh, I.; Patkar, S.; Halkier, T.; Hansen, M. T. Two Lipases from Candida Antarctica: Cloning and Expression in Aspergillus Oryzae. *Can. J. Bot.* **1995**, *73*, 869–875.

[287] Kreuß, M.; Strixner, T.; Kulozik, U. The Effect of Glycosylation on the Interfacial Properties of Bovine Caseino-

macropeptide. *Food Hydrocoll.* **2009**, *23*, 1818–1826.

[288] Hoyle, M. C. High Resolution of Peroxidase-Indoleacetic Acid Oxidase Isoenzymes from Horseradish by Isoelectric Focusing. *Plant Physiol.* **1977**, *60*, 787–793.

[289] Li, L.; Li, C.; Zhang, Z.; Alexov, E. On the Dielectric "Constant" of Proteins: Smooth Dielectric Function for Macromolecular Modeling and Its Implementation in DelPhi. *J. Chem. Theory Comput.* **2013**, *9*, 2126–2136.

[290] Jeske, L.; Placzek, S.; Schomburg, I.; Chang, A.; Schomburg, D. BRENDA in 2019: A European ELIXIR Core Data Resource. *Nucleic Acids Res.* **2019**, *47*, D542–D549.

[291] Placzek, S.; Schomburg, I.; Chang, A.; Jeske, L.; Ulbrich, M.; Tillack, J.; Schomburg, D. BRENDA in 2017: New Perspectives and New Tools in BRENDA. *Nucleic Acids Res.* **2017**, *45*, D380–D388.

[292] Wu, X.; Ge, J.; Yang, C.; Hou, M.; Liu, Z. Facile Synthesis of Multiple Enzyme-Containing Metal–Organic Frameworks in a Biomolecule-Friendly Environment. *Chem. Commun.* **2015**, *51*, 13408–13411.

[293] Liang, W.; Ricco, R.; Maddigan, N. K.; Dickinson, R. P.; Xu, H.; Li, Q.; Sumby, C. J.; Bell, S. G.; Falcaro, P.; Doonan, C. J. Control of Structure Topology and Spatial Distribution of Biomacromolecules in Protein@ZIF-8 Biocomposites. *Chem. Mater.* **2018**, *30*, 1069–1077.

[294] Liao, F.-S.; Lo, W.-S.; Hsu, Y.-S.; Wu, C.-C.; Wang, S.-C.; Shieh, F.-K.; Morabito, J. V.; Chou, L.-Y.; Wu, K. C.-W.; Tsung, C.-K. Shielding against Unfolding by Embedding Enzymes in Metal–Organic Frameworks via a de Novo Approach. *J. Am. Chem. Soc.* **2017**, *139*, 6530–6533.

[295] Chen, G.; Huang, S.; Kou, X.; Wei, S.; Huang, S.; Jiang, S.; Shen, J.; Zhu, F.; Ouyang, G. A Convenient and Versatile Amino-Acid-Boosted Biomimetic Strategy for the Nondestructive Encapsulation of Biomacromolecules within Metal–Organic Frameworks. *Angew. Chem. Int. Ed.* **2019**, *58*, 1463–1467.

[296] Chen, W.-H.; Vázquez-González, M.; Zoabi, A.; Abu-Reziq, R.; Willner, I. Biocatalytic Cascades Driven by Enzymes Encapsulated in Metal–Organic Framework Nanoparticles. *Nat. Catal.* **2018**, *1*, 689–695.

[297] Gupta, K. M.; Zhang, K.; Jiang, J. Water Desalination through Zeolitic Imidazolate Framework Membranes: Significant Role of Functional Groups. *Langmuir* **2015**, *31*, 13230–13237.

[298] Zhu, Y.; Gupta, K. M.; Liu, Q.; Jiang, J.; Caro, J.; Huang, A. Synthesis and Seawater Desalination of Molecular Sieving Zeolitic Imidazolate Framework Membranes. *Desalination* **2016**, *385*, 75–82.

[299] Lai, Z. Development of ZIF-8 Membranes: Opportunities and Challenges for Commercial Applications. *Curr. Opin. Chem. Eng.* **2018**, *20*, 78–85.

[300] Verploegh, R. J.; Nair, S.; Sholl, D. S. Temperature and Loading-Dependent Diffusion of Light Hydrocarbons in ZIF-8 as Predicted Through Fully Flexible Molecular Simulations. *J. Am. Chem. Soc.* **2015**, *137*, 15760–15771.

[301] Peralta, D.; Chaplais, G.; Paillaud, J.-L.; Simon-Masseron, A.; Barthelet, K.; Pirngruber, G. D. The Separation of Xylene Isomers by ZIF-8: A Demonstration of the Extraordinary Flexibility of the ZIF-8 Framework. *Microporous Mesoporous Mater.* **2013**, *173*, 1–5.

[302] Moggach, S. A.; Bennett, T. D.; Cheetham, A. K. The Effect of Pressure on ZIF-8: Increasing Pore Size with Pressure and the Formation of a High-Pressure Phase at 1.47 GPa. *Angew. Chem. Int. Ed.* **2009**, *48*, 7087–7089.

[303] Fairen-Jimenez, D.; Moggach, S. A.; Wharmby, M. T.; Wright, P. A.; Parsons, S.; Düren, T. Opening the Gate: Framework Flexibility in ZIF-8 Explored by Experiments and Simulations. *J. Am. Chem. Soc.* **2011**, *133*, 8900–8902.

[304] Chokbunpiam, T.; Chanajaree, R.; Saengsawang, O.; Reimann, S.; Chmelik, C.; Fritzsche, S.; Caro, J.; Remsungnen, T.; Hannongbua, S. The Importance of Lattice Flexibility for the Migration of Ethane in ZIF-8: Molecular Dynamics Simulations. *Microporous Mesoporous Mater.* **2013**, *174*, 126–134.

[305] Chokbunpiam, T.; Fritzsche, S.; Caro, J.; Chmelik, C.; Janke, W.; Hannongbua, S. Importance of ZIF-90 Lattice Flexibility on Diffusion, Permeation, and Lattice Structure for an Adsorbed $H_2/CH_4$ Gas Mixture: A Re-Examination by Gibbs Ensemble Monte Carlo and Molecular Dynamics Simulations. *J. Phys. Chem. C* **2017**, *121*, 10455–10462.

[306] Coudert, F.-X. Molecular Mechanism of Swing Effect in Zeolitic Imidazolate Framework ZIF-8: Continuous Deformation upon Adsorption. *ChemPhysChem* **2017**, *18*, 2732–2738.

[307] Zhang, K.; Lively, R. P.; Zhang, C.; Chance, R. R.; Koros, W. J.; Sholl, D. S.; Nair, S. Exploring the Framework Hydrophobicity and Flexibility of ZIF-8: From Biofuel Recovery to Hydrocarbon Separations. *J. Phys. Chem. Lett.* **2013**, *4*, 3618–3622.

[308] Zhang, C.; Lively, R. P.; Zhang, K.; Johnson, J. R.; Karvan, O.; Koros, W. J. Unexpected Molecular Sieving

Properties of Zeolitic Imidazolate Framework-8. *J. Phys. Chem. Lett.* **2012**, *3*, 2130–2134.

[309] Hara, N.; Yoshimune, M.; Negishi, H.; Haraya, K.; Hara, S.; Yamaguchi, T. Diffusive Separation of Propylene/Propane with ZIF-8 Membranes. *J. Membr. Sci.* **2014**, *450*, 215–223.

[310] Diestel, L.; Bux, H.; Wachsmuth, D.; Caro, J. Pervaporation Studies of *n*-Hexane, Benzene, Mesitylene and Their Mixtures on Zeolitic Imidazolate Framework-8 Membranes. *Microporous Mesoporous Mater.* **2012**, *164*, 288–293.

[311] Li, J.-R.; Kuppler, R. J.; Zhou, H.-C. Selective Gas Adsorption and Separation in Metal–Organic Frameworks. *Chem. Soc. Rev.* **2009**, *38*, 1477–1504.

[312] Webster, C. E.; Drago, R. S.; Zerner, M. C. Molecular Dimensions for Adsorptives. *J. Am. Chem. Soc.* **1998**, *120*, 5509–5516.

[313] Eum, K.; Jayachandrababu, K. C.; Rashidi, F.; Zhang, K.; Leisen, J.; Graham, S.; Lively, R. P.; Chance, R. R.; Sholl, D. S.; Jones, C. W.; Nair, S. Highly Tunable Molecular Sieving and Adsorption Properties of Mixed-Linker Zeolitic Imidazolate Frameworks. *J. Am. Chem. Soc.* **2015**, *137*, 4191–4197.

[314] Gao, W.-Y.; Cardenal, A. D.; Wang, C.-H.; Powers, D. C. In Operando Analysis of Diffusion in Porous Metal–Organic Framework Catalysts. *Chem. Eur. J.* **2019**, *25*, 3465–3476.

[315] Cousin Saint Remi, J.; Rémy, T.; Van Hunskerken, V.; van de Perre, S.; Duerinck, T.; Maes, M.; De Vos, D.; Gobechiya, E.; Kirschhock, C. E. A.; Baron, G. V.; Denayer, J. F. M. Biobutanol Separation with the Metal–Organic Framework ZIF-8. *ChemSusChem* **2011**, *4*, 1074–1077.

[316] Han, C.; Verploegh, R. J.; Sholl, D. S. Assessing the Impact of Point Defects on Molecular Diffusion in ZIF-8 Using Molecular Simulations. *J. Phys. Chem. Lett.* **2018**, 4037–4044.

[317] Tanaka, S.; Fujita, K.; Miyake, Y.; Miyamoto, M.; Hasegawa, Y.; Makino, T.; Van der Perre, S.; Cousin Saint Remi, J.; Van Assche, T.; Baron, G. V.; Denayer, J. F. M. Adsorption and Diffusion Phenomena in Crystal Size Engineered ZIF-8 MOF. *J. Phys. Chem. C* **2015**, *119*, 28430–28439.

[318] Gounaris, C. E.; Wei, J.; Floudas, C. A.; Ranjan, R.; Tsapatsis, M. Rational Design of Shape Selective Separations and Catalysis: Lattice Relaxation and Effective Aperture Size. *AIChE J.* **2010**, *56*, 611–632.

[319] First, E. L.; Gounaris, C. E.; Floudas, C. A. Predictive Framework for Shape-Selective Separations in Three-Dimensional Zeolites and Metal–Organic Frameworks. *Langmuir* **2013**, *29*, 5599–5608.

[320] Haranczyk, M.; Sethian, J. A. Navigating Molecular Worms inside Chemical Labyrinths. *Proc. Natl. Acad. Sci. U.S.A.* **2009**, *106*, 21472–21477.

[321] Luebbers, M. T.; Wu, T.; Shen, L.; Masel, R. I. Effects of Molecular Sieving and Electrostatic Enhancement in the Adsorption of Organic Compounds on the Zeolitic Imidazolate Framework ZIF-8. *Langmuir* **2010**, *26*, 15625–15633.

[322] Jee, A.-Y.; Yanai, N.; Granick, S. Comparing Geometry and Chemistry When Confined Molecules Diffuse in Monodisperse Metal–Organic Framework Pores. *J. Phys. Chem. Lett.* **2018**, 6399–6403.

[323] Gómez-Bombarelli, R.; Wei, J. N.; Duvenaud, D.; Hernández-Lobato, J. M.; Sánchez-Lengeling, B.; Sheberla, D.; Aguilera-Iparraguirre, J.; Hirzel, T. D.; Adams, R. P.; Aspuru-Guzik, A. Automatic Chemical Design Using a Data-Driven Continuous Representation of Molecules. *ACS Cent. Sci.* **2018**, *4*, 268–276.

[324] Kanehisa, M.; Goto, S. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Res.* **2000**, *28*, 27–30.

[325] Kanehisa, M.; Sato, Y.; Kawashima, M.; Furumichi, M.; Tanabe, M. KEGG as a Reference Resource for Gene and Protein Annotation. *Nucleic Acids Res.* **2016**, *44*, D457–D462.

[326] Kanehisa, M.; Furumichi, M.; Tanabe, M.; Sato, Y.; Morishima, K. KEGG: New Perspectives on Genomes, Pathways, Diseases and Drugs. *Nucleic Acids Res.* **2017**, *45*, D353–D361.

[327] Rojas, I.; Golebiewski, M.; Kania, R.; Krebs, O.; Mir, S.; Weidemann, A.; Wittig, U. SABIO-RK: A Database for Biochemical Reactions and Their Kinetics. *BMC Syst. Biol.* **2007**, *1*, S6.

[328] Wittig, U.; Kania, R.; Golebiewski, M.; Rey, M.; Shi, L.; Jong, L.; Algaa, E.; Weidemann, A.; Sauer-Danzwith, H.; Mir, S.; Krebs, O.; Bittkowski, M.; Wetsch, E.; Rojas, I.; Müller, W. SABIO-RK—Database for Biochemical Reaction Kinetics. *Nucleic Acids Res.* **2012**, *40*, D790–D796.

[329] Wittig, U.; Rey, M.; Weidemann, A.; Kania, R.; Müller, W. SABIO-RK: An Updated Resource for Manually Curated Biochemical Reaction Kinetics. *Nucleic Acids Res.* **2018**, *46*, D656–D660.

[330] Hadadi, N.; Hafner, J.; Shajkofci, A.; Zisaki, A.; Hatzimanikatis, V. ATLAS of Biochemistry: A Repository of All Possible Biochemical Reactions for Synthetic Biology and Metabolic Engineering Studies. *ACS Synth. Biol.*

**2016**, *5*, 1155–1166.

[331] Zhang, C.; Koros, W. J. Tailoring the Transport Properties of Zeolitic Imidazolate Frameworks by Post-Synthetic Thermal Modification. *ACS Appl. Mater. Interfaces* **2015**, *7*, 23407–23411.

[332] Peralta, D.; Chaplais, G.; Simon-Masseron, A.; Barthelet, K.; Chizallet, C.; Quoineaud, A.-A.; Pirngruber, G. D. Comparison of the Behavior of Metal–Organic Frameworks and Zeolites for Hydrocarbon Separations. *J. Am. Chem. Soc.* **2012**, *134*, 8115–8126.

[333] Li, K.; Olson, D. H.; Seidel, J.; Emge, T. J.; Gong, H.; Zeng, H.; Li, J. Zeolitic Imidazolate Frameworks for Kinetic Separation of Propane and Propene. *J. Am. Chem. Soc.* **2009**, *131*, 10368–10369.

[334] Landrum, G. A. RDKit: Open-Source Cheminformatics. `http://www.rdkit.org/`, (accessed November 1, 2018).

[335] Riniker, S.; Landrum, G. A. Better Informed Distance Geometry: Using What We Know To Improve Conformation Generation. *J. Chem. Inf. Model.* **2015**, *55*, 2562–2574.

[336] Moshtagh, N. Minimum Volume Enclosing Ellipsoids. `http://citeseerx.ist.psu.edu/viewdoc/summary?doi=10.1.1.116.7691`, (accessed December 18, 2018).

[337] Imelfort, M. Ellipsoid: Example Code for Enclosing Points in an Elipsoid. `https://github.com/minillinim/ellipsoid`, (accessed June 17, 2018).

[338] Haldoupis, E.; Nair, S.; Sholl, D. S. Efficient Calculation of Diffusion Limitations in Metal Organic Framework Materials: A Tool for Identifying Materials for Kinetic Separations. *J. Am. Chem. Soc.* **2010**, *132*, 7528–7539.

[339] Kuo, C.-H.; Tang, Y.; Chou, L.-Y.; Sneed, B. T.; Brodsky, C. N.; Zhao, Z.; Tsung, C.-K. Yolk–Shell Nanocrystal@ZIF-8 Nanostructures for Gas-Phase Heterogeneous Catalysis with Selectivity Control. *J. Am. Chem. Soc.* **2012**, *134*, 14345–14348.

[340] Sim, H. Y. F.; Lee, H. K.; Han, X.; Koh, C. S. L.; Phan-Quang, G. C.; Lay, C. L.; Kao, Y.-C.; Phang, I. Y.; Yeow, E. K. L.; Ling, X. Y. Concentrating Immiscible Molecules at Solid@MOF Interfacial Nanocavities to Drive an Inert Gas-Liquid Reaction at Ambient Conditions. *Angew. Chem. Int. Ed.* **2018**, *57*, 17058–17062.

[341] Ferreira, A. F. P.; Mittelmeijer-Hazeleger, M. C.; Granato, M. A.; Martins, V. F. D.; Rodrigues, A. E.; Rothenberg, G. Sieving Di-Branched from Mono-Branched and Linear Alkanes Using ZIF-8: Experimental Proof and Theoretical Explanation. *Phys. Chem. Chem. Phys.* **2013**, *15*, 8795–8804.

[342] Chang, N.; Gu, Z.-Y.; Yan, X.-P. Zeolitic Imidazolate Framework-8 Nanocrystal Coated Capillary for Molecular Sieving of Branched Alkanes from Linear Alkanes along with High-Resolution Chromatographic Separation of Linear Alkanes. *J. Am. Chem. Soc.* **2010**, *132*, 13645–13647.

[343] IUPAC-IUBMB Joint Commission on Biochemical Nomenclature (JCBN) and Nomenclature Committee of IUBMB (NC-IUBMB). *Eur. J. Biochem.* **1999**, *264*, 607–609.

[344] Hadadi, N.; MohamadiPeyhani, H.; Miskovic, L.; Seijo, M.; Hatzimanikatis, V. Assigning Enzyme Sequences to Orphan and Novel Reactions Using Knowledge of Substrate Reactive Sites. *bioRxiv* **2018**, DOI: `https://doi.org/10.1101/210039`.

[345] Wu, X.; Yang, C.; Ge, J.; Liu, Z. Polydopamine Tethered Enzyme/Metal–Organic Framework Composites with High Stability and Reusability. *Nanoscale* **2015**, *7*, 18883–18886.

[346] Cheng, H.; Zhang, L.; He, J.; Guo, W.; Zhou, Z.; Zhang, X.; Nie, S.; Wei, H. Integrated Nanozymes with Nanoscale Proximity for in Vivo Neurochemical Monitoring in Living Brains. *Anal. Chem.* **2016**, *88*, 5489–5497.

[347] Zhang, Q.; Zhang, L.; Dai, H.; Li, Z.; Fu, Y.; Li, Y. Biomineralization-Mimetic Preparation of Robust Metal–Organic Frameworks Biocomposites Film with High Enzyme Load for Electrochemical Biosensing. *J. Electroanal. Chem.* **2018**, *823*, 40–46.

[348] Du, Y.; Gao, J.; Zhou, L.; Ma, L.; He, Y.; Huang, Z.; Jiang, Y. Enzyme Nanocapsules Armored by Metal–Organic Frameworks: A Novel Approach for Preparing Nanobiocatalyst. *Chem. Eng. J.* **2017**, *327*, 1192–1197.

[349] Tadepalli, S.; Yim, J.; Cao, S.; Wang, Z.; Naik, R. R.; Singamaneni, S. Metal–Organic Framework Encapsulation for the Preservation and Photothermal Enhancement of Enzyme Activity. *Small* **2018**, *14*, 1702382.

[350] Wu, X.; Yang, C.; Ge, J. Green Synthesis of Enzyme/Metal–Organic Framework Composites with High Stability in Protein Denaturing Solvents. *Bioresour. Bioprocess.* **2017**, *4*, 24–31.

[351] Wang, Y.; Ryu, B. H.; Yoo, W.; Lee, C. W.; Kim, K. K.; Lee, J. H.; Kim, T. D. Identification, Characterization, Immobilization, and Mutational Analysis of a Novel Acetylesterase with Industrial Potential (LaAcE) from Lactobacillus Acidophilus. *Biochim. Biophys. Acta, Gen. Subj.* **2018**, *1862*, 197–210.

[352] Nadar, S. S.; Rathod, V. K. Encapsulation of Lipase within Metal–Organic Framework (MOF) with Enhanced Activity Intensified under Ultrasound. *Enzyme Microb. Technol.* **2018**, *108*, 11–20.

[353] Pitzalis, F.; Carucci, C.; Naseri, M.; Fotouhi, L.; Magner, E.; Salis, A. Lipase Encapsulation onto ZIF-8: A Comparison between Biocatalysts Obtained at Low and High Zinc/2-Methylimidazole Molar Ratio in Aqueous Medium. *ChemCatChem* **2018**, *10*, 1578–1585.

[354] He, H.; Han, H.; Shi, H.; Tian, Y.; Sun, F.; Song, Y.; Li, Q.; Zhu, G. Construction of Thermophilic Lipase-Embedded Metal–Organic Frameworks via Biomimetic Mineralization: A Biocatalyst for Ester Hydrolysis and Kinetic Resolution. *ACS Appl. Mater. Interfaces* **2016**, *8*, 24517–24524.

[355] Liang, K.; J. Coghlan, C.; G. Bell, S.; Doonan, C.; Falcaro, P. Enzyme Encapsulation in Zeolitic Imidazolate Frameworks: A Comparison between Controlled Co-Precipitation and Biomimetic Mineralisation. *Chem. Commun.* **2016**, *52*, 473–476.

[356] Liu, Q.; Chapman, J.; Huang, A.; Williams, K. C.; Wagner, A.; Garapati, N.; Sierros, K. A.; Dinu, C. Z. User-Tailored Metal–Organic Frameworks as Supports for Carbonic Anhydrase. *ACS Appl. Mater. Interfaces* **2018**, *10*, 41326–41337.

[357] Zhang, C.; Wang, X.; Hou, M.; Li, X.; Wu, X.; Ge, J. Immobilization on Metal–Organic Framework Engenders High Sensitivity for Enzymatic Electrochemical Detection. *ACS Appl. Mater. Interfaces* **2017**, *9*, 13831–13836.

[358] Kaneti, Y. V.; Dutta, S.; Hossain, M. S. A.; Shiddiky, M. J. A.; Tung, K.-L.; Shieh, F.-K.; Tsung, C.-K.; Wu, K. C.-W.; Yamauchi, Y. Strategies for Improving the Functionality of Zeolitic Imidazolate Frameworks: Tailoring Nanoarchitectures for Functional Applications. *Adv. Mater.* **2017**, *29*, 1700213.

[359] Linder-Patton, O. M.; de Prinse, T. J.; Furukawa, S.; Bell, S. G.; Sumida, K.; Doonan, C. J.; Sumby, C. J. Influence of Nanoscale Structuralisation on the Catalytic Performance of ZIF-8: A Cautionary Surface Catalysis Study. *CrystEngComm* **2018**, *20*, 4926–4934.

[360] Singh, R. K.; Tiwari, M. K.; Singh, R.; Lee, J.-K. From Protein Engineering to Immobilization: Promising Strategies for the Upgrade of Industrial Enzymes. *Int. J. Mol. Sci.* **2013**, *14*, 1232–1277.

[361] Karagiaridi, O.; Lalonde, M. B.; Bury, W.; Sarjeant, A. A.; Farha, O. K.; Hupp, J. T. Opening ZIF-8: A Catalytically Active Zeolitic Imidazolate Framework of Sodalite Topology with Unsubstituted Linkers. *J. Am. Chem. Soc.* **2012**, *134*, 18790–18796.

[362] Pan, Y.; Li, H.; Farmakes, J.; Xiao, F.; Chen, B.; Ma, S.; Yang, Z. How Do Enzymes Orient When Trapped on Metal–Organic Framework (MOF) Surfaces? *J. Am. Chem. Soc.* **2018**, *140*, 16032–16036.

[363] Hou, Q.; Wu, Y.; Zhou, S.; Wei, Y.; Caro, J.; Wang, H. Ultra-Tuning of the Aperture Size in Stiffened ZIF-8_Cm Frameworks with Mixed-Linker Strategy for Enhanced $CO_2/CH_4$ Separation. *Angew. Chem. Int. Ed.* **2019**, *58*, 327–331.

[364] Knebel, A.; Geppert, B.; Volgmann, K.; Kolokolov, D. I.; Stepanov, A. G.; Twiefel, J.; Heitjans, P.; Volkmer, D.; Caro, J. Defibrillation of Soft Porous Metal–Organic Frameworks with Electric Fields. *Science* **2017**, *358*, 347–351.

[365] Chaplais, G.; Fraux, G.; Paillaud, J.-L.; Marichal, C.; Nouali, H.; Fuchs, A. H.; Coudert, F.-X.; Patarin, J. Impacts of the Imidazolate Linker Substitution ($CH_3$, Cl, or Br) on the Structural and Adsorptive Properties of ZIF-8. *J. Phys. Chem. C* **2018**, *122*, 26945–26955.

[366] Berens, S.; Chmelik, C.; Hillman, F.; Kaerger, J.; Jeong, H.-K.; Vasenkov, S. Ethane Diffusion in Mixed Linker Zeolitic Imidazolate Framework-7-8 by Pulsed Field Gradient NMR in Combination with Single Crystal IR Microscopy. *Phys. Chem. Chem. Phys.* **2018**, *20*, 23967–23975.

[367] Shen, K.; Zhang, L.; Chen, X.; Liu, L.; Zhang, D.; Han, Y.; Chen, J.; Long, J.; Luque, R.; Li, Y.; Chen, B. Ordered Macro-Microporous Metal–Organic Framework Single Crystals. *Science* **2018**, *359*, 206–210.

[368] Zhang, H.; Huo, J.; Yang, H.; Li, F.; Duan, C.; Xi, H. Green and Rapid Preparation of Hierarchically Porous Metal–Organic Zeolites and Simulation of Their Growth. *J. Mater. Chem. A* **2019**, *7*, 1022–1029.

[369] Thompson, J. A.; Blad, C. R.; Brunelli, N. A.; Lydon, M. E.; Lively, R. P.; Jones, C. W.; Nair, S. Hybrid Zeolitic Imidazolate Frameworks: Controlling Framework Porosity and Functionality by Mixed-Linker Synthesis. *Chem. Mater.* **2012**, *24*, 1930–1936.

[370] Thompson, J. A.; Brunelli, N. A.; Lively, R. P.; Johnson, J. R.; Jones, C. W.; Nair, S. Tunable CO2 Adsorbents by Mixed-Linker Synthesis and Postsynthetic Modification of Zeolitic Imidazolate Frameworks. *J. Phys. Chem. C* **2013**, *117*, 8198–8207.

[371] Zhang, J.-P.; Zhu, A.-X.; Lin, R.-B.; Qi, X.-L.; Chen, X.-M. Pore Surface Tailored SOD-Type Metal–Organic Zeolites. *Adv. Mater.* **2011**, *23*, 1268–1271.

[372] Morabito, J. V.; Chou, L.-Y.; Li, Z.; Manna, C. M.; Petroff, C. A.; Kyada, R. J.; Palomba, J. M.; Byers, J. A.; Tsung, C.-K. Molecular Encapsulation beyond the Aperture Size Limit through Dissociative Linker Exchange in Metal–Organic Framework Crystals. *J. Am. Chem. Soc.* **2014**, *136*, 12540–12543.

[373] Banerjee, R.; Phan, A.; Wang, B.; Knobler, C.; Furukawa, H.; O'Keeffe, M.; Yaghi, O. M. High-Throughput Synthesis of Zeolitic Imidazolate Frameworks and Application to $CO_2$ Capture. *Science* **2008**, *319*, 939–943.

[374] Phan, A.; Doonan, C. J.; Uribe-Romo, F. J.; Knobler, C. B.; O'Keeffe, M.; Yaghi, O. M. Synthesis, Structure, and Carbon Dioxide Capture Properties of Zeolitic Imidazolate Frameworks. *Acc. Chem. Res.* **2010**, *43*, 58–67.

[375] Tian, Y.-Q.; Zhao, Y.-M.; Chen, Z.-X.; Zhang, G.-N.; Weng, L.-H.; Zhao, D.-Y. Design and Generation of Extended Zeolitic Metal–Organic Frameworks (ZMOFs): Synthesis and Crystal Structures of Zinc(II) Imidazolate Polymers with Zeolitic Topologies. *Chem. Eur. J.* **2007**, *13*, 4146–4154.

[376] Delaney, J. S. ESOL: Estimating Aqueous Solubility Directly from Molecular Structure. *J. Chem. Inf. Comput. Sci.* **2004**, *44*, 1000–1005.

[377] Wildman, S. A.; Crippen, G. M. Prediction of Physicochemical Parameters by Atomic Contributions. *J. Chem. Inf. Comput. Sci.* **1999**, *39*, 868–873.

[378] Cheng, T.; Zhao, Y.; Li, X.; Lin, F.; Xu, Y.; Zhang, X.; Li, Y.; Wang, R.; Lai, L. Computation of Octanol-Water Partition Coefficients by Guiding an Additive Model with Knowledge. *J. Chem. Inf. Model.* **2007**, *47*, 2140–2148.

[379] Walters, P. An Implementation of Delaney's ESOL Method Using the RDKit. `https://github.com/PatWalters/solubility`, (accessed December 12, 2018).

[380] Ertl, P.; Schuffenhauer, A. Estimation of Synthetic Accessibility Score of Drug-like Molecules Based on Molecular Complexity and Fragment Contributions. *J. Cheminform.* **2009**, *1*, 8–18.

[381] Migliorini, E.; Weidenhaupt, M.; Picart, C. Practical Guide to Characterize Biomolecule Adsorption on Solid Surfaces (Review). *Biointerphases* **2018**, *13*, No. 06D303.

[382] Kienle, D. F.; Falatach, R. M.; Kaar, J. L.; Schwartz, D. K. Correlating Structural and Functional Heterogeneity of Immobilized Enzymes. *ACS Nano* **2018**, *12*, 8091–8103.

[383] Deb, P.; Haldar, T.; Kashid, S. M.; Banerjee, S.; Chakrabarty, S.; Bagchi, S. Correlating Nitrile IR Frequencies to Local Electrostatics Quantifies Noncovalent Interactions of Peptides and Proteins. *J. Phys. Chem. B* **2016**, *120*, 4034–4046.

[384] AstaTech. `http://astatechinc.com/`, (accessed December 29, 2018).

[385] Akopyan, T. N.; Braunstein, A. E.; Goryachenkova, E. V. Beta-Cyanoalanine Synthase: Purification and Characterization. *Proc. Natl. Acad. Sci. U.S.A.* **1975**, *72*, 1617–1621.

[386] Maruyama, A.; Saito, K.; Ishizawa, K. $\beta$-Cyanoalanine Synthase and Cysteine Synthase from Potato: Molecular Cloning, Biochemical Characterization, and Spatial and Hormonal Regulation. *Plant Mol. Biol.* **2001**, *46*, 749–760.

[387] Jobe, D. J.; Westaway, K. C. Fourier Transform Infrared Spectroscopy Studies of Cyanide Ion Solutions of Dimethylformamide and Aqueous Dimethylformamide. *Can. J. Chem.* **1993**, *71*, 1353–1361.

[388] Li, Z.; Xia, H.; Li, S.; Pang, J.; Zhu, W.; Jiang, Y. In Situ Hybridization of Enzymes and Their Metal–Organic Framework Analogues with Enhanced Activity and Stability by Biomimetic Mineralisation. *Nanoscale* **2017**, *9*, 15298–15302.

[389] Haldoupis, E.; Watanabe, T.; Nair, S.; Sholl, D. S. Quantifying Large Effects of Framework Flexibility on Diffusion in MOFs: $CH_4$ and $CO_2$ in ZIF-8. *ChemPhysChem* **2012**, *13*, 3449–3452.

[390] Holden, D.; Chong, S. Y.; Chen, L.; Jelfs, K. E.; Hasell, T.; Cooper, A. I. Understanding Static, Dynamic and Cooperative Porosity in Molecular Materials. *Chem. Sci.* **2016**, *7*, 4875–4879.

[391] Amrouche, H.; Creton, B.; Siperstein, F.; Nieto-Draghi, C. Prediction of Thermodynamic Properties of Adsorbed Gases in Zeolitic Imidazolate Frameworks. *RSC Adv.* **2012**, *2*, 6028–6035.

[392] Miklitz, M.; Jelfs, K. E. Pywindow: Automated Structural Analysis of Molecular Pores. *J. Chem. Inf. Model.* **2018**, *58*, 2387–2391.

[393] García, I. G.; Bernabei, M.; Haranczyk, M. Toward Automated Tools for Characterization of Molecular Porosity. *J. Chem. Theory Comput.* **2019**, *15*, 787–798.

[394] Knio, O.; Medford, A. J.; Nair, S.; Sholl, D. S. Database of Computation-Ready 2D Zeolitic Slabs. *Chem. Mater.* **2019**, *31*, 353–364.

[395] Saha, S.; Wiebcke, M.; Huber, K. Insight into Fast Nucleation and Growth of Zeolitic Imidazolate Framework-71 by In Situ Static Light Scattering at Variable Temperature and Kinetic Modeling. *Cryst. Growth Des.* **2018**, *18*, 4653–4661.

[396] Saliba, D.; Ammar, M.; Rammal, M.; Al-Ghoul, M.; Hmadeh, M. Crystal Growth of ZIF-8, ZIF-67, and Their Mixed-Metal Derivatives. *J. Am. Chem. Soc.* **2018**, *140*, 1812–1823.

[397] Cravillon, J.; Nayuk, R.; Springer, S.; Feldhoff, A.; Huber, K.; Wiebcke, M. Controlling Zeolitic Imidazolate Framework Nano- and Microcrystal Formation: Insight into Crystal Growth by Time-Resolved In Situ Static Light Scattering. *Chem. Mater.* **2011**, *23*, 2130–2141.

[398] Yeung, H. H.-M.; Sapnik, A. F.; Massingberd-Mundy, F.; Gaultois, M. W.; Wu, Y.; Fraser, D. A. X.; Henke, S.; Pallach, R.; Heidenreich, N.; Magdysyuk, O. V.; Vo, N. T.; Goodwin, A. L. Control of Metal–Organic Framework Crystallization by Metastable Intermediate Pre-Equilibrium Species. *Angew. Chem. Int. Ed.* **2019**, *58*, 566–571.

[399] Tsierkezos, N. G.; Filippou, A. C. Thermodynamic Investigation of N,N-Dimethylformamide/Toluene Binary Mixtures in the Temperature Range from 278.15 to 293.15 K. *J. Chem. Thermodyn.* **2006**, *38*, 952–961.

[400] Dikio, E. D.; Nelana, S. M.; Isabirye, D. A.; Ebenso, E. E. Density, Dynamic Viscosity and Derived Properties of Binary Mixtures of Methanol, Ethanol, n-Propanol, and n-Butanol with Pyridine at T=(293.15, 303.15, 313.15 and 323.15) K. *Int. J. Electrochem. Sci* **2012**, *7*, 11101–11122.

[401] Bernal-García, J. M.; Guzmán-López, A.; Cabrales-Torres, A.; Estrada-Baltazar, A.; Iglesias-Silva, G. A. Densities and Viscosities of (N,N-Dimethylformamide + Water) at Atmospheric Pressure from (283.15 to 353.15) K. *J. Chem. Eng. Data* **2008**, *53*, 1024–1027.

[402] Ortega, A.; Amorós, D.; Garćıa de la Torre, J. Prediction of Hydrodynamic and Other Solution Properties of Rigid Proteins from Atomic- and Residue-Level Models. *Biophys. J.* **2011**, *101*, 892–898.

[403] de la Torre, J. G.; Huertas, M. L.; Carrasco, B. Calculation of Hydrodynamic Properties of Globular Proteins from Their Atomic-Level Structure. *Biophys. J.* **2000**, *78*, 719–730.

[404] Amorós, D.; Ortega, A.; de la Torre, J. G. Prediction of Hydrodynamic and Other Solution Properties of Partially Disordered Proteins with a Simple, Coarse-Grained Model. *J. Chem. Theory Comput.* **2013**, *9*, 1678–1685.

[405] O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *J. Cheminform.* **2011**, *3*, 1–14.

[406] Gale, J. D. GULP: A Computer Program for the Symmetry-Adapted Simulation of Solids. *J. Chem. Soc. Faraday Trans.* **1997**, *93*, 629–637.

[407] Gale, J. D.; Rohl, A. L. The General Utility Lattice Program (GULP). *Mol. Simul.* **2003**, *29*, 291–341.

[408] Liu, J.; Lukose, B.; Shekhah, O.; Arslan, H. K.; Weidler, P.; Gliemann, H.; Bräse, S.; Grosjean, S.; Godt, A.; Feng, X.; Müllen, K.; Magdau, I.-B.; Heine, T.; Wöll, C. A Novel Series of Isoreticular Metal Organic Frameworks: Realizing Metastable Structures by Liquid Phase Epitaxy. *Sci. Rep.* **2012**, *2*, 921.

[409] Sakata, Y.; Furukawa, S.; Kondo, M.; Hirai, K.; Horike, N.; Takashima, Y.; Uehara, H.; Louvain, N.; Meilikhov, M.; Tsuruoka, T.; Isoda, S.; Kosaka, W.; Sakata, O.; Kitagawa, S. Shape-Memory Nanopores Induced in Coordination Frameworks by Crystal Downsizing. *Science* **2013**, *339*, 193–196.

[410] Kitaura, R.; Iwahori, F.; Matsuda, R.; Kitagawa, S.; Kubota, Y.; Takata, M.; Kobayashi, T. C. Rational Design and Crystal Structure Determination of a 3-D Metal-Organic Jungle-Gym-like Open Framework. *Inorg. Chem.* **2004**, *43*, 6522–6524.

[411] Gražulis, S.; Chateigner, D.; Downs, R. T.; Yokochi, A. F. T.; Quirós, M.; Lutterotti, L.; Manakova, E.; Butkus, J.; Moeck, P.; Le Bail, A. Crystallography Open Database – an Open-Access Collection of Crystal Structures. *J. Appl. Crystallogr.* **2009**, *42*, 726–729.

[412] Gražulis, S.; Daškevič, A.; Merkys, A.; Chateigner, D.; Lutterotti, L.; Quirós, M.; Serebryanaya, N. R.; Moeck, P.; Downs, R. T.; Le Bail, A. Crystallography Open Database (COD): An Open-Access Collection of Crystal Structures and Platform for World-Wide Collaboration. *Nucleic Acids Res.* **2012**, *40*, D420–D427.

[413] Arslan, H. K.; Shekhah, O.; Wieland, D. C.; Paulus, M.; Sternemann, C.; Schroer, M. A.; Tiemeyer, S.; Tolan, M.; Fischer, R. A.; Woll, C. Intercalation in Layered Metal–Organic Frameworks: Reversible Inclusion of an Extended Pi-System. *J. Am. Chem. Soc.* **2011**, *133*, 8158–61.

[414] Heck, R.; Shekhah, O.; Zybaylo, O.; Weidler, P. G.; Friedrich, F.; Maul, R.; Wenzel, W.; Wöll, C. Loading of Two Related Metal–Organic Frameworks (MOFs), $[Cu_2(bdc)_2(dabco)]$ and $[Cu_2(ndc)_2(dabco)]$, with Ferrocene. *Polymers* **2011**, *3*, 1565–1574.

[415] Yakovenko, A. A.; Reibenspies, J. H.; Bhuvanesh, N.; Zhou, H.-C. Generation and Applications of Structure Envelopes for Porous Metal–Organic Frameworks. *J. Appl. Crystallogr.* **2013**, *46*, 346–353.

[416] Graham, A. J.; Tan, J.-C.; Allan, D. R.; Moggach, S. A. The Effect of Pressure on Cu-Btc: Framework Compression vs. Guest Inclusion. *Chem. Commun.* **2012**, *48*, 1535–1537.

[417] Shekhah, O.; Wang, H.; Paradinas, M.; Ocal, C.; Schupbach, B.; Terfort, A.; Zacher, D.; Fischer, R. A.; Woll, C. Controlling Interpenetration in Metal–Organic Frameworks by Liquid-Phase Epitaxy. *Nat. Mater.* **2009**, *8*, 481–484.

[418] Wittig, U.; Rey, M.; Kania, R.; Bittkowski, M.; Shi, L.; Golebiewski, M.; Weidemann, A.; Müller, W.; Rojas, I. Challenges for an Enzymatic Reaction Kinetics Database. *FEBS J.* **2014**, *281*, 572–582.

[419] Bertz, S. H. The First General Index of Molecular Complexity. *J. Am. Chem. Soc.* **1981**, *103*, 3599–3601.

[420] Hendrickson, J. B.; Huang, P.; Toczko, A. G. Molecular Complexity: A Simplified Formula Adapted to Individual Atoms. *J. Chem. Inf. Comput. Sci.* **1987**, *27*, 63–67.

[421] Kim, S.; Thiessen, P. A.; Cheng, T.; Yu, B.; Bolton, E. E. An Update on PUG-REST: RESTful Interface for Programmatic Access to PubChem. *Nucleic Acids Res.* **2018**, *46*, W563–W570.

[422] Mannhold, R.; Poda, G. I.; Ostermann, C.; Tetko, I. V. Calculation of Molecular Lipophilicity: State-of-the-Art and Comparison of LogP Methods on More than 96,000 Compounds. *J. Pharm. Sci* **2009**, *98*, 861–893.