Karlsruhe Institute of Technology

FAKULTÄT FÜR CHEMIE & BIOWISSENSCHAFTEN

High-Throughput screening of the CoRE MOF database to predict cross-linking of Metal-Organic Frameworks

Zur Erlangung des akademischen Grades eines

DOKTORS DER NATURWISSENSCHAFTEN

(Dr. rer. nat.)

von der KIT-Fakultät für Chemie und Biowissenschaften

des Karlsruher Instituts für Technologie (KIT)

genehmigte

DISSERTATION

von

Momin Ahmad

1. Referent: Dr. Manuel Tsotsalas
2. Referent: Prof. Christof Wöll
Tag der mündlichen Prüfung: 09.12.2020

Erklärung zur Selbstständigkeit

Ich versichere, dass ich diese Arbeit selbstständig verfasst habe und keine anderen als die angegebenen Quellen und Hilfsmittel benutzt habe, die wörtlich oder inhaltlich übernommenen Stellen als solche kenntlich gemacht und die Satzung des KIT zur Sicherung guter wissenschaftlicher Praxis in der gültigen Fassung vom 27.06.2017 beachtet habe.

Karlsruhe, den 01.03.2021, _____

Momin Ahmad

Contents

List of Figures

List of Tables

Kurzzusammenfassung

Deutscher Titel: High-Throughput Screening der CoRE MOF Datenbank für die Vorhersage von Quervernetzung in Metall-Organischen Gerüsten

Metall-organische Gerüststrukturen (engl. Metal-Organic Framework, MOF) sind poröse kristalline Materialien, die in den letzten Jahrzehnten immer mehr an Popularität gewonnen haben. Das liegt an erster Stelle an ihrer Vielseitigkeit, die eine Anwendung in unterschiedlichsten Anwendungsgebieten wie Gastrennung oder Gasspeicherung ermöglicht. Die Bausteine der MOFs sind zum einen die Metallzentren und zum anderen organische Linker, die zwei oder mehr Metallzentren miteinander verbinden. Durch diesen modularen Aufbau ist es möglich aus wenigen Bausteinen viele verschiedene neuartige kristalline Strukturen durch Variation zu erzeugen. Um MOFs als bioaktive Materialien nutzen zu können, ist es wünschenswert die Metalle herauszulösen, da eine etwaige toxische Reaktion durch die Metalle *in vivo* nicht gänzlich ausgeschlossen werden kann. Zudem ist die eingeschränkte Stabilität in wässriger Umgebung vieler MOF Strukturen ein weiterer Nachteil dieser Materialklasse für die Anwendung in der Biologie oder Medizin. Um die Vorteile der MOFs auch für solche Anwendungen nutzen zu können, wurde eine Strategie entwickelt in der die organischen Linker mit sekundären Linkern (cross-linker) quervernetzt werden (cross-linking). Dies erhöht die Stabilität des MOFs, sodass die Metalle anschließend problemlos entfernt werden können. Das Problem besteht darin die passenden cross-linker für eine bestimmte MOF Struktur zu finden. Zum einen muss der cross-linker ideal in die Struktur passen ohne sie zu verzerren, zum anderen sollte ein durchgängiges cross-linking der Struktur möglich sein, um Größen skalieren zu können. Der modulare Aufbau der MOFs hat aber auch zur Folge, dass die Zahl der theoretisch möglichen MOFs sehr stark zugenommen hat. Große Datenbanken mit zehntausenden solcher Strukturen sind entstanden. Den passenden MOF für einen gegebenen cross-linker zu finden gleicht nun der Suche nach der Nadel im Heuhaufen. In dieser Arbeit geht es gerade darum diese eine Nadel (oder mehrere) zu finden. Durch ein computergestütztes Screening soll eine MOF Datenbank durchleuchtet werden, um nach passenden MOF Kandidaten für einen gegebenen cross-linker zu suchen. Die Analyse und Auswertung dieser Datenbank gibt Aufschluss darüber, ob und wie eine MOF-Struktur sich quervernetzen lässt. Das Ergebnis des Screenings ist demnach eine Vorhersage einer MOF-Struktur, in der ein oder mehrere vorher bekannte cross-linker ideal hineinpassen. Die aus diesem Projekt vorhergesagten MOF cross-linker Kombinationen dienen als Vorlage für zukünftige experimentelle Arbeiten, deren Fokus auf die vielversprechendsten Kandidaten gelenkt wird.

1. Introduction

Metal-organic frameworks (henceforth MOFs) are highly tunable materials that offer a wide range of applications. Gas storage, gas separation, catalysis are only a few of many features that allow applications, such as batteries, fuel cells or membranes. A specific characteristic of these structures is their modular design. MOFs consist of metal nodes that are interconnected by organic linkers. The choice of the metal and the possibility to functionalize the organic linkers, as desired, allows large space for creativity. It is precisely this modularity that has allowed the number of MOFs to increase rapidly in recent years. In order not to lose the overview, databases of such MOF structures were created. The sheer volume of potential MOFs and their associated characteristics call for computer-aided methods for analyzing and evaluating such databases. A manual search of a MOF with certain properties and without experience is therefore doomed to failure from the outset. This is also the reason why high-throughput screenings of MOF databases are popular to address this problem[1].

An innovative idea is to use MOFs as carriers for biomolecules. This potentially enables them to act as drug carriers. Drug delivery is indeed one popular fields of MOF research[2]. However, some (not all) metals such as chromium are highly toxic and have to be strictly regulated inside the human body. But simply removing them destroys the desired structure of the MOFs. An interesting approach is to use Covalent-Organic frameworks (henceforth COFs) that contain purely organic building blocks with a crystalline structure[3]. Thus, they combine the best of both worlds when it comes to bio-compatibility. One possibility of synthesizing a COF is to first start with a MOF. Interconnecting the organic linkers by another secondary organic linker (henceforth cross-linker) further stabilizes the framework. This step of post-synthetic modification is called cross-linking. With this, removing the metals out of the structure is now possible without destroying the structure, as the cross-linker keep the structure together. In brief, a MOF was used as a template to synthesize a COF.

Although these concepts are quite simple, they are difficult to put into practice. The desire for highly-ordered COF structures requires suitable partners for cross-linking. For this, two main conditions must be met: (I) the cross-linker should fit exactly into the MOF and (II) should not interfere with the cross-linking of another cross-linker at another position in the same MOF. Identifying two structures, a MOF and a cross-linker, that fulfill both conditions is essential and the subject of this work. There are two ways to solve this problem. One is to screen a cross-linker database for a given MOF, second is to screen a MOF database for a given cross-linker. The first option fails, since there are (to the best knowledge of the author) no significantly

large cross-linker databases available. The overall probability of actually finding matching cross-linkers from a smaller database would be extremely low. Therefore, only option two remains.

In this work I introduce a high-throughput screening method of a MOF database to find matching MOFs for a given cross-linker. The results of the screening suggests possible MOF candidates for a specific cross-linker that allow an ideal fit. The predicted structures are an aid for experimental groups to synthesize novel COFs. Thus, I go beyond prior approaches such as *retrofitting*[4] that aims at enhancing their properties such as mechanical robustness/stability. In our case the more complex cross-linker binding is considered instead. Future work could use this work as an intermediate step to achieve highly-ordered COF structures.

This dissertation is divided into six chapters. Chapter two introduces the concept of MOFs in detail. In particular, their topology and synthesis are discussed, but also a theoretical model of MOF nucleation and growth is given. Further applications and hierarchical structures are addressed. In addition, a brief overview on challenges of MOFs is pointed out. The chapter concludes with the techniques of cross-linking and turning MOFs into COFs.

Chapter three focuses on the computational introduction with the main part being the method of simulating molecular structures via *Molecular Dynamics* (MD). A very crucial concept are force fields that describe the physics of the system to be simulated. Especially in connection with MOFs, such force fields face a major challenge in the key task of parametrization. One section describes how a molecular structure is optimized by minimizing its total energy. Lastly, the concept of a screening is explained where also the preprocessing of the data is mentioned. In short, this chapter highlights popular computational methods that were also used in this work.

Chapter four discusses the structure of the screening. The motivation of a high-throughput screening is emphasized before introducing MOF databases. The chapter continues with preliminary considerations of the screening that, among others, include the cross-linker lengths and identifying binding sites. The implementation of the screening method is the main point of this chapter. In the end of the chapter, a similar project is shown that preceded this screening project.

Chapter five is a continuation of chapter four. First, it explains the technical implementation of the work, then discusses the results of the high-throughput screening. Advantages and disadvantages of certain decisions are discussed and an outlook describes a possible continuation of this project.

Chapter six provides a summary and concludes this thesis. The appendix shows the screening algorithm as a pseudo code.

2. Metal-Organic Frameworks

MOFs are a novel class of porous materials with great surface to mass ratios ($^{cm^2}/_{g}$) that allow great tunability and promise a wide range of future applications[5]. Built up by a metal node and an organic linker (see figure 2.1), the possibilities of different structures are almost limitless. The number of papers published on MOFs over a decade is shown in figure 2.2 and affirms the increasing interest and potential of these materials. In this chapter the basic concepts of MOFs and their synthesis are discussed. Also a brief overview of applications is given later on. An important addition to MOFs, namely hierarchical structures and COFs, that are of great importance for this work, are also included in this chapter.

2.1. Introduction

MOFs are highly tunable porous structures that consist of metal nodes and organic linkers. The metal nodes are interconnected by organic linkers. Choosing the size and type of the building blocks determines the MOF crystal itself. For example, the pore size of a MOF can be increased by choosing an extended linker and keeping the same metal node. This pore is a key property with regard to interactions with guest molecules. Tuning the pore size and functionalizing the linkers has a great impact on the diffusion of these guest molecules[8]. Depending on the building blocks, namely an organic linker and a meta node, the length of the crystal lattice can go up to several nanometers[5]. The interaction of guest molecules with the pore depends on the property of the MOF. Thus, the great tunability of MOFs leads to many applications, such as gas storage or gas separation[5].

2.2. Topology of MOFs

The properties of the metal node, in specific the number of coordination bonds it can form, combined with the type of the organic linker (e.g. being linear, planar, tetrahedral and so on) defines the topology of the crystal structures. For example, a metal node with six coordination sites with two in each direction (x, y, z) can form a simple cubic crystal, with two metal nodes interconnected by a linear linker. An example is given in figure 2.3 where an organic linker *1,4-benzenedicarboxylic acid* (henceforth BDC) binds to the metal nodes Zn_4O and forms a cubic crystal structure with the topology **pcu**[9]. The blue marked area shows where the organic linker connects to the metal node. A three letter code given by the *Reticular Chemistry Structure Resource* (RCSR)[10] describes the topology (in this case **pcu**). Different geometric parameters, such as unit lengths (each direction) and angles of the unit cell define the type of topology.

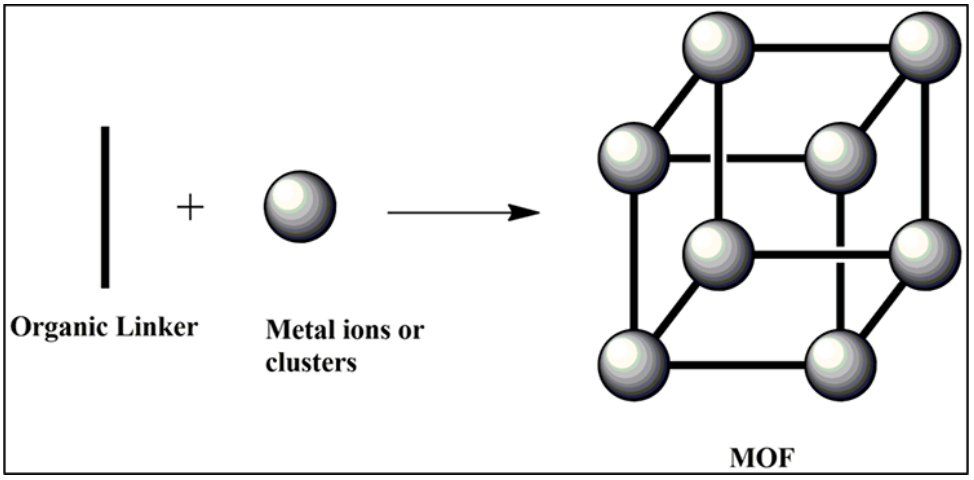


Figure 2.1.: A schematic presentation of the structure of MOFs. Taken from [6].

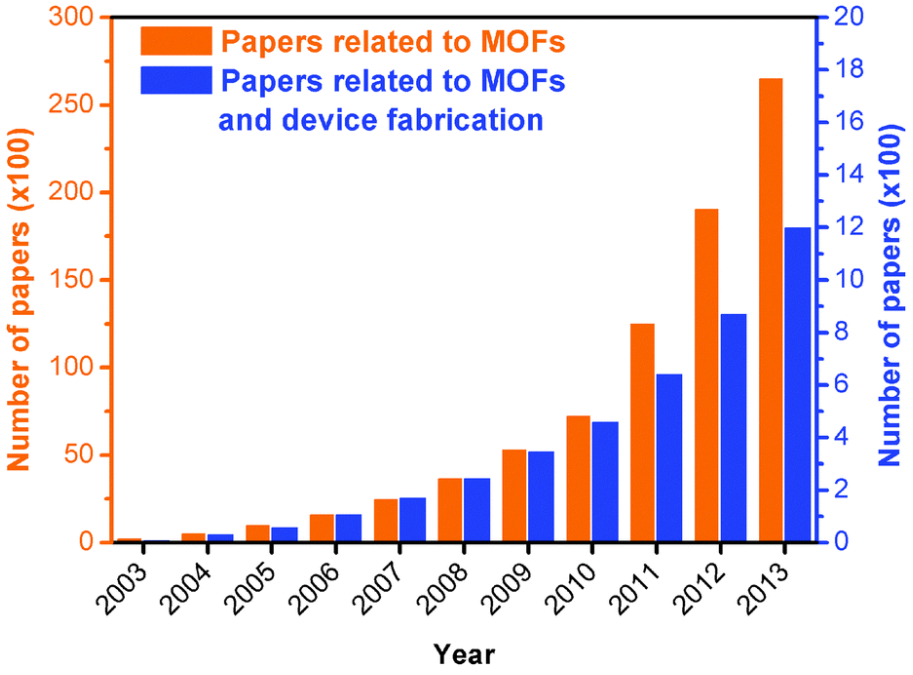


Figure 2.2.: A graph showing the number of published papers on MOFs over time. The number increases exponentially. Taken from [7].

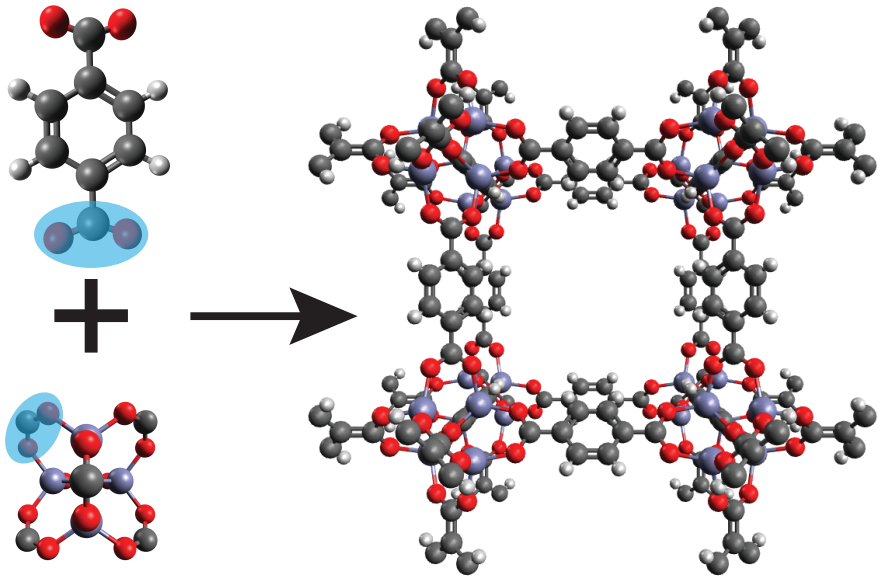


Figure 2.3.: A linear linker *1,4-benzodicarboxylic acid* (BDC) connected to a metal node Zn_4O results to the MOF *IRMOF-1* (or *MOF-5*). The topology is called **pcu**[9] and represents a simple cubic crystal structure. The blue marked area shows where the organic linker connects to the metal node. Color code: H (white), C (black), O (red), Zn (purple).

Different types of topologies can be seen in figure 2.4, where the same linker connects to different metal nodes to result in different topologies. In the case of MOF UiO-66, more organic linkers are able to connect to the metal node and form a more complex structure resulting in the topology **fcu**[11]. MIL-53(Al) and MIL-101(Cr) possess different structures. The first one has the topology **dia**[12], whereas the second one has the topology **mtn**[13].

In fact, many studies exist where investigations on theoretical synthesizable MOF structures were made to find exotic structures for various applications[1, 15, 16, 17].

2.3. Synthesis of MOFs

In general classical approach, MOFs are synthesized by the solvothermal method (section 2.3.1) resulting in crystalline powders. In recent years MOFs growth on various surfaces has been successfully demonstrated via liquid phase epitaxy/ layer-by-layer method (section 2.3.2).

2.3.1. MOF synthesis by solvothermal method

Key factor for the self-assembly of MOFs is to form dynamic bonds to correct faulty bonds that would cause disorder in the crystal structure[18]. Metal salt and an organic linker are mixed together in a solvent with high boiling point, such as *N,N-Dimethylformamide* (DMF), in a vial (see figure 2.5) and heated at certain temperatures. Following this approach, both metal salt and organic linker have to be soluble with the chosen solvent at the targeted reaction temperature. In addition, it is of utmost importance to adjust experimental parameters, such as temperature, pH value, concentration of the building blocks and type of solvent that are able

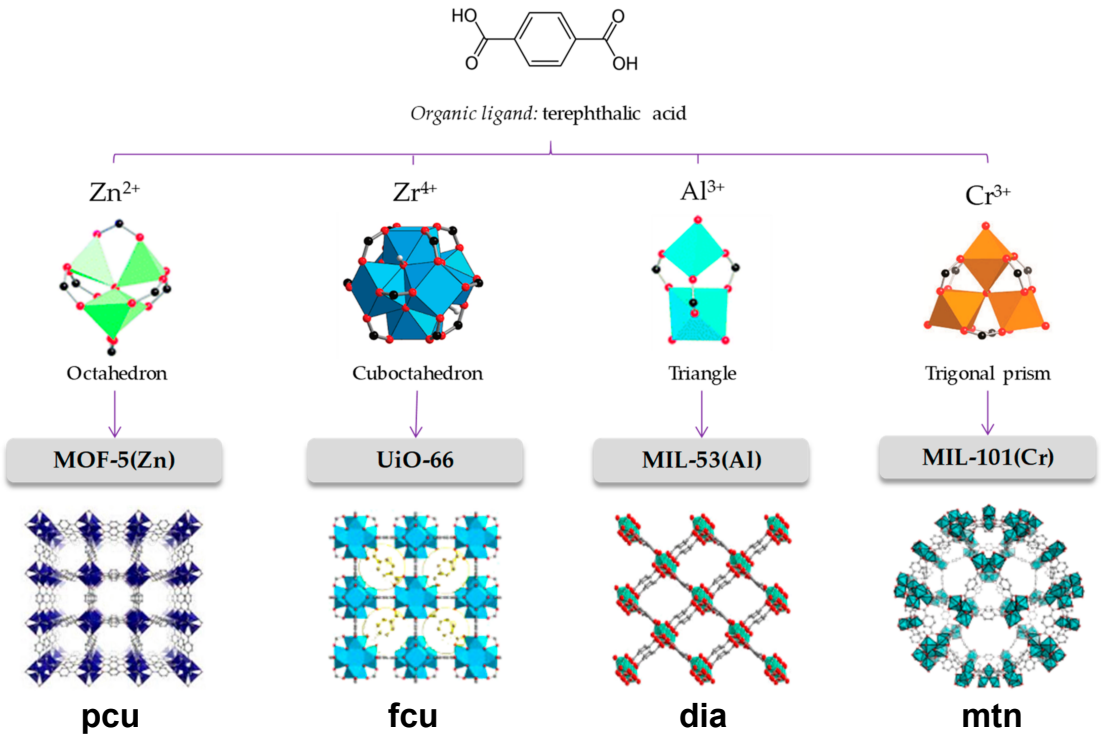


Figure 2.4.: The same organic linker combined with different metal nodes results in different topologies. MOF-5 possesses a **pcu** topology[9], UiO-66 an **fcu** topology[11], MIL-53 a **dia**[12] topology and MIL-101 possesses an **mtn** topology[13]. Taken from [14].

greatly influence the success of the synthesis. These parameters are not only able to change the topology of the crystal, but also the crystal size and phase purity[18]. Account must also be taken that by-products of the reaction during the synthesis may affect the outcome. For example, a metal chloride (MCl_x) mixed with a carboxylic acid linker creates at minimum a stoichiometric amount of HCl as byproduct. This strong acid is able to dissolve the forming MOF and, thus, slowing down the crystal growth. Alternatively, one can use metal acetylacetonate ($M(acac)_x$) whose byproduct acetylacetonate is less reactive compared to HCl[18]. In some cases, especially when dealing with strong metal-ligand bonds (i.e. Zr-MOFs), the help of a modulator is required to prevent rapid formation of disordered bonds that would result in an amorphous material[19]. Modulators do not contribute to the structure and are equally designed as linkers, such as benzonic acid or acetic acid. These modulators form dynamic bonds with the metal nodes to slow down the formation of structural bonds [18].

Today more and more chemists rely on computationally calculated data that assist in choosing the best conditions for a reaction and machine learning is one big catchword in the field of computational chemistry[20, 21, 22]. Neural networks trained with collected data from the past are able to predict ideal laboratory conditions. The online platform *MATERIALSCLOUD* possesses such a tool that may assist in synthesizing MOFs[23].

A downside of the MOF self-assembly is shown in figure 2.6. When the pore of a MOF gets so large that not only guest molecule find place in it, but also the metal

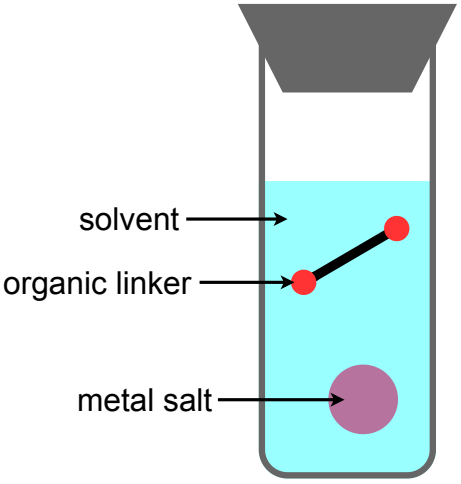


Figure 2.5.: Metal salt and organic linker mixed together in a solvent. Adjusting parameters, such as temperature or pH value has great influence on the synthesis.

node and the organic linker that were used to create the MOF, an interpenetrated MOF is formed. Since the topology is an important key factor of a MOF, one likes to control this feature. Keeping the self-assembly as the synthesizing method, one method is to specify the topology by introducing a surface on which the MOF is mounted, making it a SURMOF. A monolayer is connected to the surface, on which the metal node of the MOF is anchored, to mount the MOF onto the surface.

2.3.2. Layer-by-Layer synthesis

Especially, when the goal is to obtain thin MOF films the layer-by-layer method (or liquid phase epitaxy [25]) is to be preferred[18]. One advantage is the use of different linkers, because it is possible to replace one building block by another in each cycle. Switching between several different linkers is called *heteroepitaxy* and staying with a single linker is called *homoepitaxy*.

At first, a suitable surface is chosen where molecules are attached that are able to bind to the metal nodes of the MOF. This step is necessary, as the standard linker of the MOF may not bind to the surface. Adding the metal nodes and interconnecting them with horizontal organic linkers forms the base structure. From this point onwards, multiple cycles can be performed to grow the structure. In this process one can choose between different linkers to vary the crystal structure (see figure 2.7).

2.3.3. MOF nucleation and growth

During nucleation a transition is performed from a thermodynamic phase with high free energy to a new phase with organized structure and minimized free energy[27]. In classical nucleation theory the change of the free energy ΔG with a spherical nucleus radius r is described as

$$\Delta G = 4\pi r^2 \cdot \gamma - \frac{4}{3}\pi r^3 \cdot \Delta\mu \tag{2.1}$$

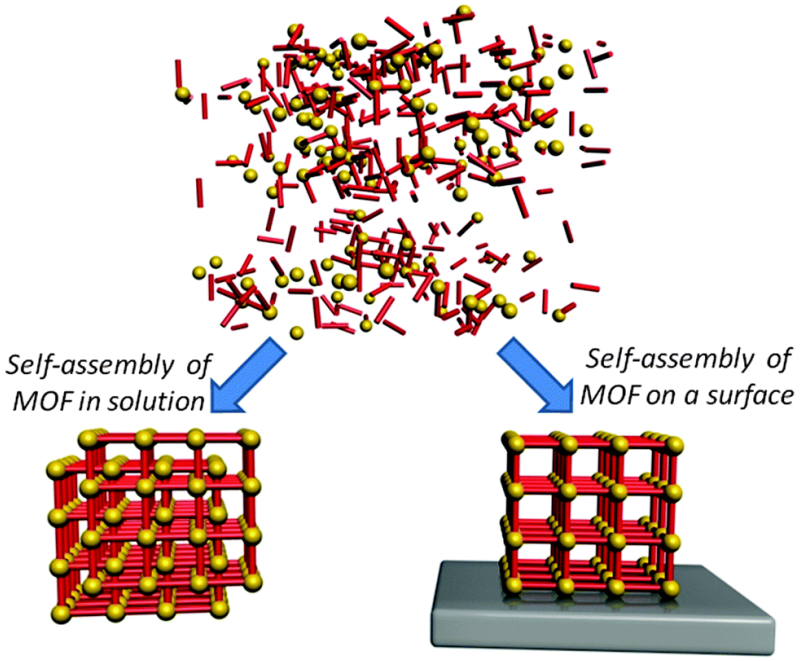


Figure 2.6.: The self-assembly of MOFs does not guarantee a clean crystal structure. Due to larger pore size the empty volume in the material is able to host not only guest molecules, but also the metal nodes and the organic linkers, generating (multiple) interpenetrated networks. [7, 24]. One possible solution is to mount the MOF onto a surface to specify the desired structure. Figure taken from [7].

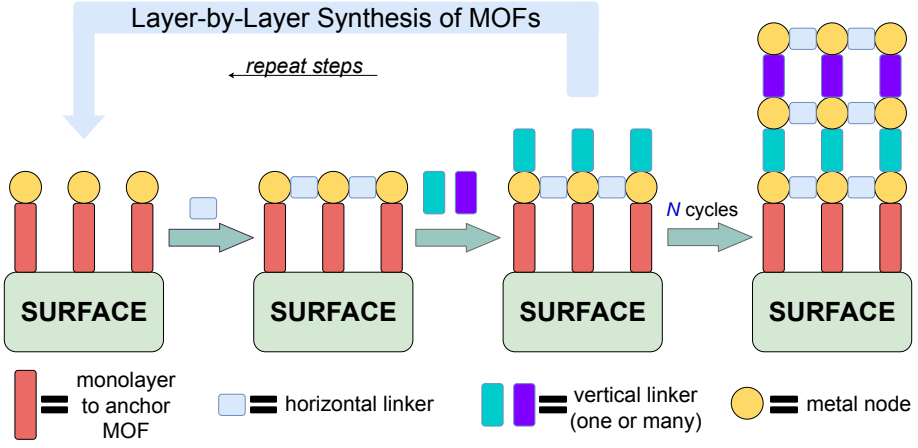


Figure 2.7.: Working steps of a layer-by-layer synthesis. First, a surface is chosen with molecules attached on it that connect to the metal nodes. After that, the metal nodes are interconnected by a horizontal linker. With this, the base is formed that then can be expanded by performing multiple cycles of using (different) vertical linkers and the metal node with its corresponding horizontal linker (note that horizontal and vertical organic linkers can also be the same for *homoepitaxy*). Reproduced from [26].

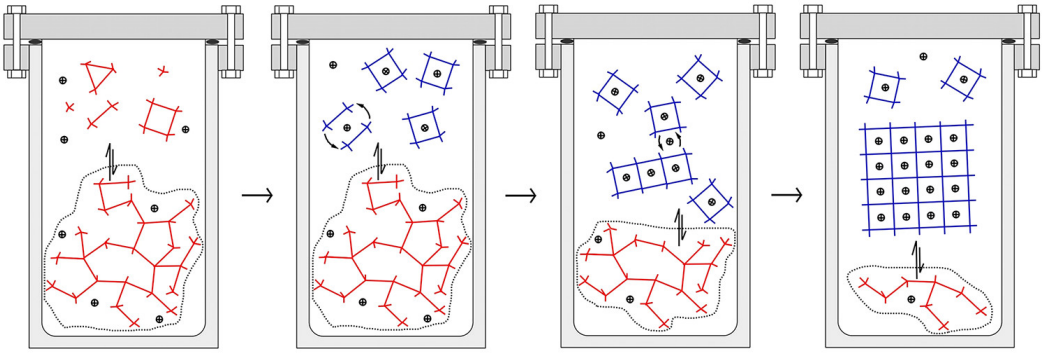


Figure 2.8.: Nucleation and growth model of a crystal structure with building blocks and secondary building units (partially reacted building blocks) as starting point. The black dots represent template cations to form the structure. Reproduced from [28].

with γ for the specific surface energy and $\Delta\mu$ as the difference of the chemical potential [28, 24]. The behavior of ΔG is strongly dependent on the radius. With larger r the volume term increases and results to a negative ΔG and smaller r leads to a positive ΔG.

The nucleation in a synthesis by the solvothermal method, where building-blocks in a solvent self-assemble to crystals under reactive conditions, can be described by a homogeneous nucleation[28]. One important factor is the high dependency of the nucleation on the concentration of the reactants. Controlling both, the nucleation and crystal growth, is crucial for the synthesis of MOFs (especially of those with higher hierarchy). Increasing the concentration towards a critical point (super-saturation) leads to instantaneous nucleation (see figure 2.9)[5]. Homogeneous nucleation is well described by the classical nucleation theory and, thus, can be used for MOFs. Although, in case of impurities in the substrate such as dust or bubbles the heterogeneous nucleation has to be considered. In figure 2.8 one example is shown where building blocks and secondary building units, which are partially reacted building blocks that can be used for the crystal formation, are distributed at the initial stage. The second steps is the formation of the unit cells of the MOF with the help of template cations. Thereby, as per classical nucleation theory, an interplay between the two terms from equation 2.1 takes place, where the difference of the chemical potential (volume term) needs to be greater than the change of the surface energy to result in a negative free-energy difference.

2.4. Applications of MOFs

MOFs offer a wide range of variety in terms of (future) applications. The possibility of post-synthetic modification even promises to widen the horizon. For example, MOFs can be modified in such way to enable them for biomedical applications[29]. Extensively researched are host-guest diffusion studies where the focus lies on the interaction of guest molecules, such as methane or carbon dioxide with the MOF. With this, applications like gas storage or gas separation can be described. Further possible applications include thermal/electrical conductive materials (insulation, electronics) and catalysis.

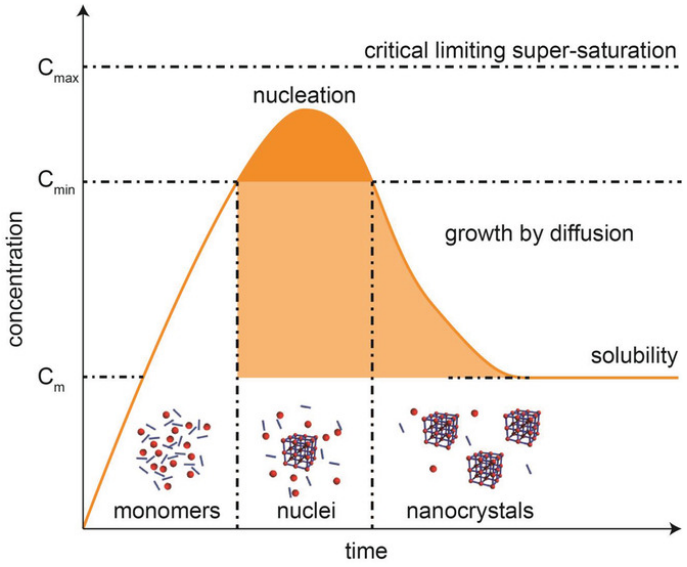


Figure 2.9.: Process of nucleation over time depending on the concentration of reactants. Taken from [5].

The porosity of MOFs enables guest molecules to interact with the material. Adsorption studies, for example, reveal the affinity of the MOF to bind different gases. If a MOF is able to bind great amounts of methane or carbon dioxide (or greenhouse gases in general), it could represent a prime candidate to solve environmental issues by challenging climate change[30].

Guest atoms (or molecules) can also be captured by the MOF. In some cases, they can be released again or stay captured in the material. For example, a gas mixture can be filtered selectively by using a MOF sensitive to only one type of gas. Capturing that gas and letting all other gases easily diffuse through the material is one way of filtering[31]. Therefore, a membrane out of MOFs gives an energy efficient solution[32]. Different diffusion rates allow to separate different kinds of gases (or even liquids). A direct application of MOFs is to act as fuel cells for hydrogen driven vehicles where MOFs allow the storage of greater amount of H_2 per volume at a certain pressure than any other material[33].

Doping MOFs in a way to obtain electroactivity is one way to use MOFs as conductors[34]. A very interesting application related to conductivity is to prepare MOFs as semiconductors. By choosing certain building blocks, the band gap of the MOF can be tuned, so that the material obtains semiconductive properties. An optical excitation of the electrons enables the MOF to work as sensors. Also the use of photovoltaic systems is feasible[35].

Designing MOFs for drug delivery allows them to carry biomolecules. The controlled storage and release of the biomolecules could greatly benefit the field of pharmacy[36].

2.4.1. Hierarchical MOF structures

To synthesize larger MOF structures, one has to carefully observe nucleation and crystal growth. Figure 2.10 shows three types of hierarchical structures. The architecture is important for crystal morphology, thus, the eventual form of the crystal. The porosity includes intended (intrinsic porosity) and unintended (defects)

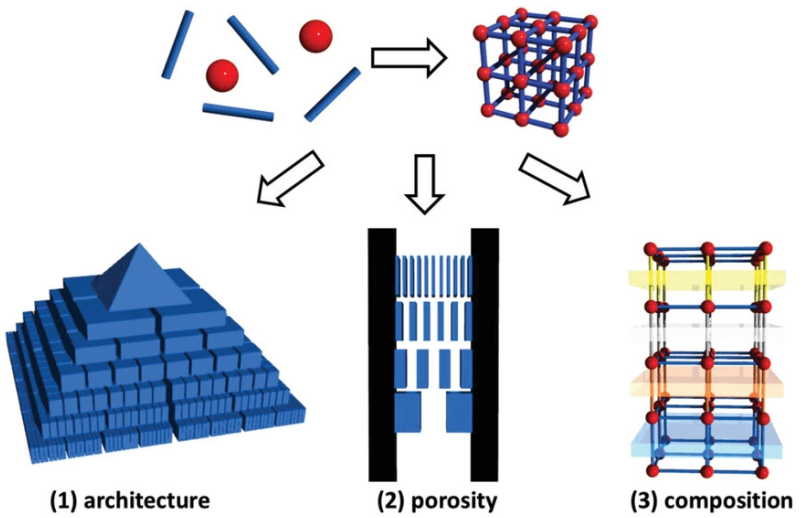


Figure 2.10.: Different types of hierarchical MOF structures. The architecture is significant for the crystal morphology. The porosity includes defects in MOFs controlling crystal properties such as material stability. Creative composition of MOFs offers a wide range of novel materials by combining two or more crystal structures. Taken from [5].

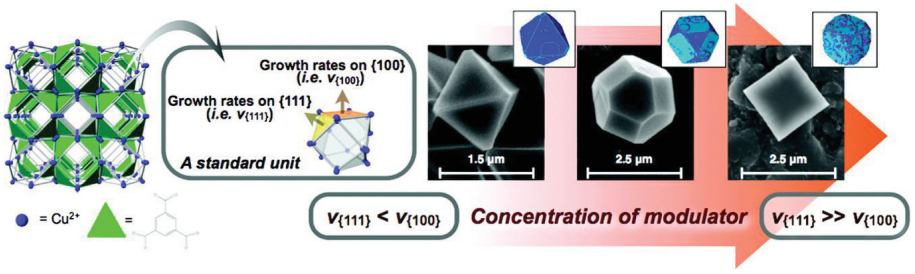


Figure 2.11.: The morphology of a crystal can be manipulated by choosing different concentrations of modulators. Taken from [5].

gaps. Higher hierarchical types of porosity leads to bigger pores such as mesopores (larger pores) and in general different dimensions of porosity in the same material. Hierarchical composition of MOFs is achieved by combining different types of MOF crystal structures in a single crystal. The three hierarchy types can be divided into three further subcategories, namely the primary, secondary and tertiary hierarchy. The primary hierarchy is in general the intrinsic property of the crystal that emerges during crystal formation, i.e. the structure of a MOF crystal itself. The secondary hierarchy represents an intermediate step, where for instance the porosity or defects in a crystal add up and form mesopores. The tertiary hierarchy is of higher dimensions where different composites of MOFs or different types of morphologies are formed.

In figure 2.11 one way of tuning the morphology of a MOF crystal is shown. By adding modulators at different concentrations, different effects can be achieved. These modulators stick to planes (given by the *Miller Indices*) depending on the concentration and, hence, define a desired direction of crystal growth. Depending on the direction, the crystal morphology changes.

An example of hierarchical porous MOF crystals is given in figure 2.12. Intentionally adding defective linker that do not fully interconnect the metal nodes and leave a gap

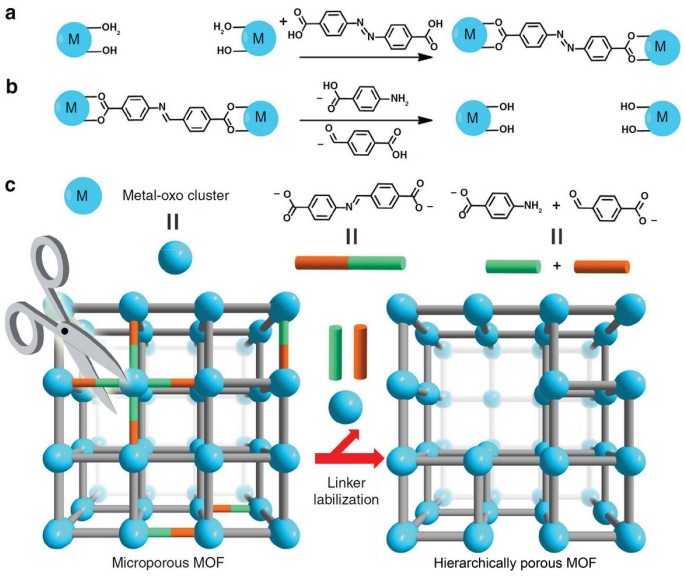


Figure 2.12.: By adding defective linkers (a, b) larger gaps can be realized increasing the size of the crystal (c). Besides the intrinsic porosity also bigger gaps are formed in the same crystal. Taken from [5].

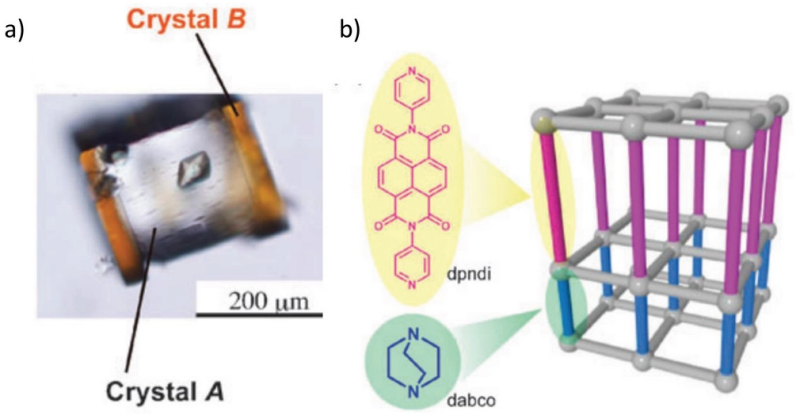


Figure 2.13.: Different organic linkers in the same crystal result in a MOF composition. Taken from [5].

between them yields a MOF with different sized pores. In addition to the primary category, the secondary category is also represented now. The intrinsic porosity in large scales represent a common crystal structure with equal sized pores. The additional gaps add up together and create mesopores.

A MOF composition is achieved by combining different building blocks in the same crystal. Figure 2.13 shows a MOF with two different organic linker that are on top of each other. Such structures can be synthesized by the layer-by-layer method discussed in chapter 2.3.2. An important application of MOF composition is the surface termination where terminating groups are connected to the outermost layer to form a surface.

As these hierarchical structures are not fully understood, computational approaches to analyze them have emerged. Modeling and simulating MOF structures with several types of hierarchies is very challenging, as enormous time and length scales have to be considered. In addition, there is also the computational effort, which consumes

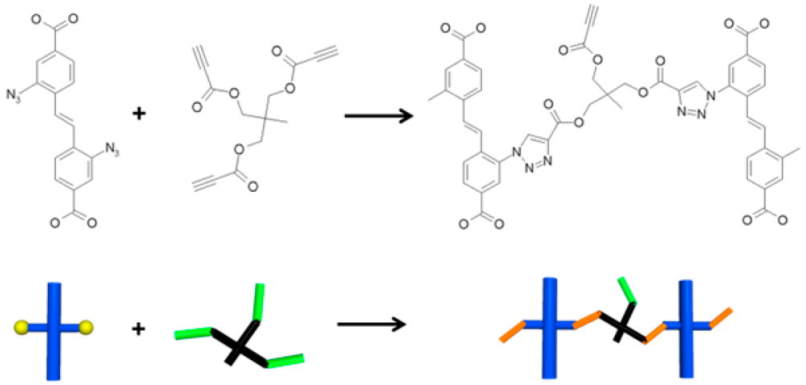


Figure 2.14.: Two organic linker are interconnected with a secondary organic linker. This bridging further stabilizes the MOF. Taken from [42].

an incredible amount of resources and therefore should not be underestimated. Approaches to solve these issues are given in[5].

2.4.2. Challenges

Challenges in synthesis of MOFs are mostly specific to the application. For example, in some cases the goal is to enhance mechanical stability, in other cases the key factor is to find the best suited components to achieve the desired properties. To name a more specific example, water stability is essential for most applications and is in general desired. MOF-5 is very sensitive under ambient water, thus, functionalizing the MOF may lead to less sensitivity towards water[37]. Another challenge is to control defects during MOF synthesis. Defects generate heterogeneous pore sized materials that may result in lower crystal stability and may change the uptake of gases at different pressures[38]. Controlling the defects still needs further investigation[39].

2.5. Cross-linking of MOFs & COFs

Cross-linking of MOFs is the main topic of this work. Using MOFs for biomedical applications is not straight forward, since certain metals are proven to be toxic for the human body. But removing the metals to obtain a COF obviously destroys the hardly achieved crystal structure. One possibility is to cross-link the organic linker with a secondary organic linker to further stabilize the structure. To use a MOF structure as template and turning it into a polymer has already been successfully realized[40]. A more general overview can be found in[41].

An example for cross-linking is given in figure 2.14, where two organic linkers (blue) are interconnected with a secondary organic linker (black/green). If this is applied in each direction, then all organic linkers are connected to each other and the metals can be dissolved out of the crystal. In this case, the MOF serves as a template so that the first step starts at a synthesized MOF.

Figure 2.15 shows an example for a MOF structure that is converted into a SURGEL. The term SURGEL is also common as a SURMOF is turned into a gel by this procedure. Here the MOF is able to form bonds with the secondary linker, thus, the cross-linking is possible. An important factor is to be able to cross-link all organic linkers in a chosen direction to achieve highly ordered structures. This means the

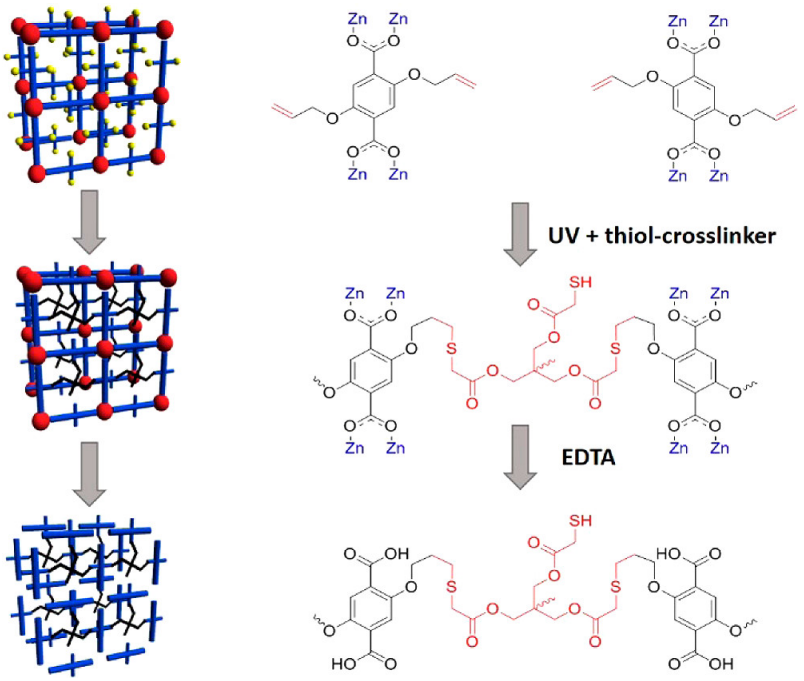


Figure 2.15.: In the first step a MOF is prepared to react with a secondary organic linker. The organic linker connects both linkers in a way that the cross-linking can be continuously pursued in each direction to achieve an ordered structure. After that, the metals can be dissolved through an acid treatment. Taken from [43].

cross-linking of two organic linkers should not prevent the cross-linking of other linkers at places that are symmetrically identical. In the end the metals can be dissolved out of the structure, for example, with an acid treatment. The product is then a material that possesses a crystalline structure but is fully built up by organic molecules.

As mentioned above, applications in the medical field are one possibility for MOFs and even more by using COFs. In figure 2.16 one part of a SURGEL is loaded with arabinose, which is a type of sugar and the other part is not. In order to see if the arabinose is able to diffuse through the SURGEL, bacteria is placed on top of it that emits fluorescence light when interacting with arabinose. The experiment shows that only the loaded part of the SURGEL is visible due to the fluorescence light proving the ability of arabinose to diffuse to the top.

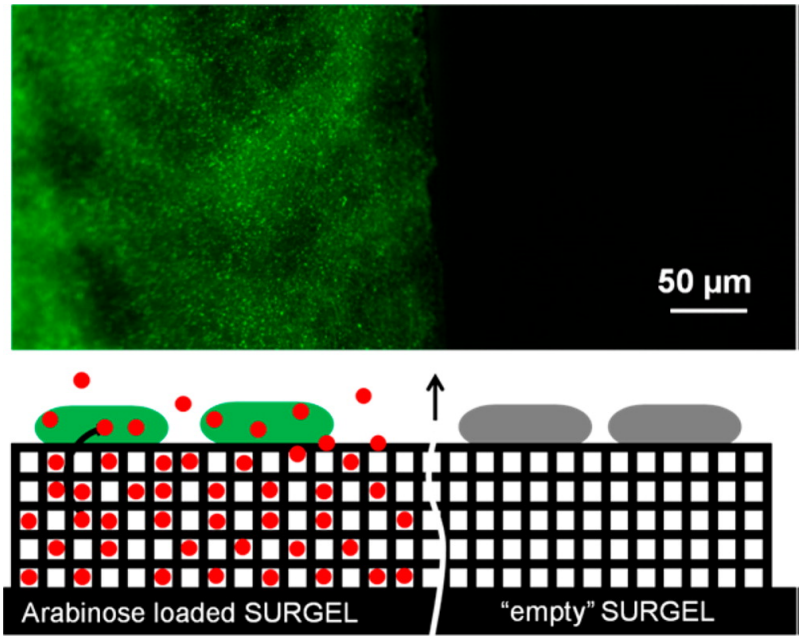


Figure 2.16.: Application of a SURGEL where one side is loaded with arabinose (type of sugar) and the other is not. On top of both sides are bacteria placed that emit fluorescence light if in contact with arabinose. This experiment proves the feasibility of such SURGELs to act as carrier for different molecules. Taken from [42].

3. Molecular Dynamics & Computational Methods

The last chapter mentioned the ever-increasing and large number of MOFs. Currently, the number is increasing so quickly, that it is getting more and more complicated to keep up with the literature. Manually looking through each and every MOF to find promising candidates for a specific application becomes incredibly time-consuming. Thus, it is logical to automatize as much analysis steps as possible to reduce unnecessary workload. The time saved by automation and by having less but better MOF cross-linker predictions can then be used for in-depth computational and/or experimental analysis. In an ideal case a screening algorithm should promise higher probabilities of success in finding desired MOF structures. This chapter gives a brief introduction on *Molecular Dynamics* (MD). For in-depth details see[44].

Very popular is also the theoretical calculation using *density functional theory* (DFT)[45]. In principle, this method calculates the electron density and the acting forces. Hence, this method of calculation is based on quantum mechanics and is not of classical nature as MD. However the computational effort is much higher than MD. This is because not every atom must be calculated individually, but the corresponding electrons. This leads to many calculations even with smaller molecules. Calculations of medium-sized molecules (500 - 1000 atoms) and larger structures therefore become very long and expensive. However, the advantage lies in the higher accuracy. The choice of the simulation method is therefore a question of the goal. In this work, simulations were only needed to confirm the screened results. A small optimization was enough to decide between a hit and miss. Hence, DFT was not needed.

3.1. Molecular Dynamics

Essential for the computational analysis are MD simulations. In an MD simulation a classical physical model is applied to the molecular structure in form of equations that represent the potential energy of the system. Subsequently, the potential energy of these equations is minimized to find stable structures. A classical model is based on Newtonian mechanics, where a force $\vec{F}$ is described as

$$\vec{F} = m \cdot \vec{a} \quad , \tag{3.1}$$

with m as the mass of the object and $\vec{a}$ as its acceleration. Note that the force is a vector and possesses a direction. An equivalent description of equation 3.1 is

$$\vec{F} = m \cdot \frac{d^2\vec{r}}{dt^2} \quad , \tag{3.2}$$

with the acceleration $\vec{a}$ being the second derivative ($\frac{d}{dt}$) of the position $\vec{r}$ over time t. As the force $\vec{F}$ is the gradient of the energy U, a force-energy relation can be established by

$$\vec{F} = -\vec{\nabla}U(\vec{r}) \quad , \tag{3.3}$$

with $\vec{\nabla}$ as the nabla operator in three dimensions

$$\vec{\nabla} = \begin{pmatrix} \partial/\partial x \\ \partial/\partial y \\ \partial/\partial z \end{pmatrix} \quad . \tag{3.4}$$

Solving the differential equations coming from 3.3 is the key task of MD simulations. First, the energy terms of $U(\vec{r})$ are discussed in chapter 3.1.1, then the methods of solving those equations in chapter 3.1.2.

3.1.1. Force-Fields

A force-field is the physical model that describes the potential energy $U(\vec{r})$ of the system and is a set of equations and corresponding parameters. Hence, the first step is to choose appropriate equations to fully describe a system in a way that the structure is represented. Many force-fields exist for one special purpose. For example the *AMBER* force-field focuses on organic atoms and molecules and is popular for protein based simulations (**A**ssisted **M**odel **B**uilding with **E**nergy **R**efinement)[46]. Other well known force-fields are *CHARMM* (**C**hemistry at **Har**vard **M**acromolecular **M**echanics)[47], DREIDING[48], OPLS (**O**ptimized **P**otentials for **L**iquid **S**imulations)[49] and UFF (**U**niversal **F**orce **F**ield)[50]. In particular, the latter is of great interest for MOFs. In general, energies come from bonded (atoms directly connected to each other) and nonbonded (e.g. van der Waals force) potentials. The sum of those potentials yields the total energy of the system

$$U_{total} = U_{bonded} + U_{nonbonded} \quad . \tag{3.5}$$

The bonded potentials consist of

$$U_{bonded} = \sum_{stretch} U(r)_{stretch} + \sum_{angle} U(\theta)_{angle} + \sum_{dihedral} U(\theta)_{dihedral} + \sum_{improper} U(\theta)_{improper} \tag{3.6}$$

and the nonbonded consist of

$$U_{nonbonded} = \sum_{vdW} U(r)_{vdW} + \sum_{el} U(r)_{el} \quad . \tag{3.7}$$

In the following sections the potential terms U_{bonded} and $U_{nonbonded}$ are discussed in more detail.

3.1.1.1. Bonded potentials

The force field formulas are mostly given as harmonic terms. With small deviations around the resting point they are exact enough and have only few (mostly one single) parameters. Bonded potentials are distinguished between two-body, three-body and four-body potentials. The number indicates the number of atoms contributing to the potential. The position of an atom in space is described by a vector $\vec{r}$. Out of N atoms an atom i has the following position in three dimensional space

$$\vec{r_i} = \begin{pmatrix} x_i \\ y_i \\ z_i \end{pmatrix} \quad . \tag{3.8}$$

This is the first parameter needed for most potentials.

Stretching potential

Knowing the position of the atom, a relation to other atoms can be concluded. For a simple bond between two atoms i and j the energy can be described in the form of a harmonic potential

$$U(r_{ij})_{\text{stretch}} = k_{ij}(r_{ij} - r_0)^2 \quad , \tag{3.9}$$

with the distance r_{ij} between the two atoms and r_0 as the relaxed position and the scaling factor k_{ij} that will be discussed later on. The unit of $U(r)$ is typically given in $^{\text{kJ}}/_{\text{mol}}$ (or in $^{\text{kcal}}/_{\text{mol}}$) and the unit of the distance is in general measured in Å, defining the unit of k_{ij} to $^{\text{kJ}}/_{\text{mol}\cdot\text{Å}^2}$. If r_{ij} is equal to the relaxed position, the energy for this specific bond equals to zero and is therefore in its minimum. Thus, this equation has the role to describe the stretching of a bond.

Figure 3.1 shows the course of the harmonic potential with a resting distance of $r_{ij} = 1.45\,\text{Å}$. If the distance of the atoms i and j is bigger, the resulting force pushes them back together and a smaller distance pushes them back apart. As the harmonic potential is not realistic (the energy goes to infinitive with increasing distance), another type of stretching potential can be used. This potential is known as the Morse potential with the form of

$$U(r_{ij})_{\text{morse}} = D_{ij}(1 - e^{-a(r_{ij}-r)})^2 \quad , \tag{3.10}$$

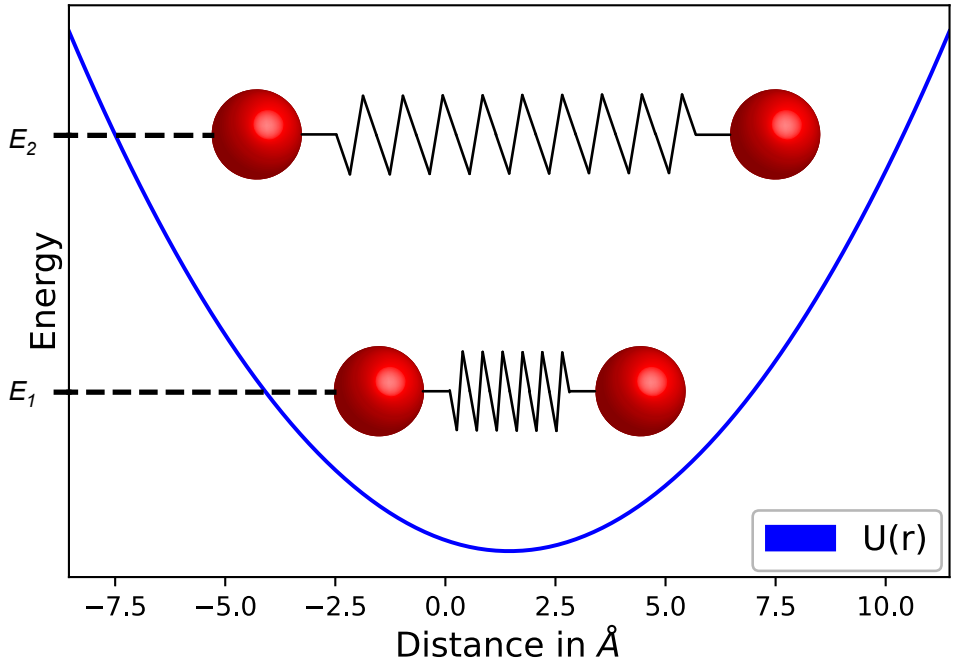


Figure 3.1.: Course of the harmonic potential of a simple bond. The equilibrium distance in this example is $r_{ij} = 1.45\,\text{Å}$. If the distance is bigger, the resulting force pulls the atoms back together. If the distance is smaller than the resting position, the force pushes them back apart.

with D_{ij} as the well depth of the curve (minimum in the figure) and a as the stiffness of the potential (see 3.2). One reason to use this potential is when large bond distances are expected in a simulation. In contrast to to the harmonic potential, the Morse potential manages to describe the longer distances better. But if only small deviations around the equilibrium are expected, then both potentials are approximately equal and good to use.

Angle potential

If an atom i is connected to two other atoms j and k, an angle has to be introduced as the parameter instead of the distance. Using only the distances $\vec{r}_{ij}$ and $\vec{r}_{ik}$ leaves one degree of freedom where the orientation of the atoms can be placed arbitrarily as long the bonded potential from the distances is minimized. Defining the structure by adding another potential with the distance of $\vec{r}_{jk}$ is rather troublesome, since k_{jk} is a parameter that is provided by the force-field and is used only for bonded connections. Instead of adding meaningless parameters, introducing the angle between $\vec{r}_{ij}$ and $\vec{r}_{ik}$ suffices. The potential is similar to the harmonic potential of the bond in form of

$$U(\theta_{ijk})_{angle} = \phi_{ijk}(\theta_{ijk} - \theta_0)^2 \quad , \tag{3.11}$$

with θ_{ijk} as the angle, θ_0 the equilibrium angle and ϕ_{ijk} again as the scaling factor. Figure 3.3 shows the potential and its similarity to the bonded potential.

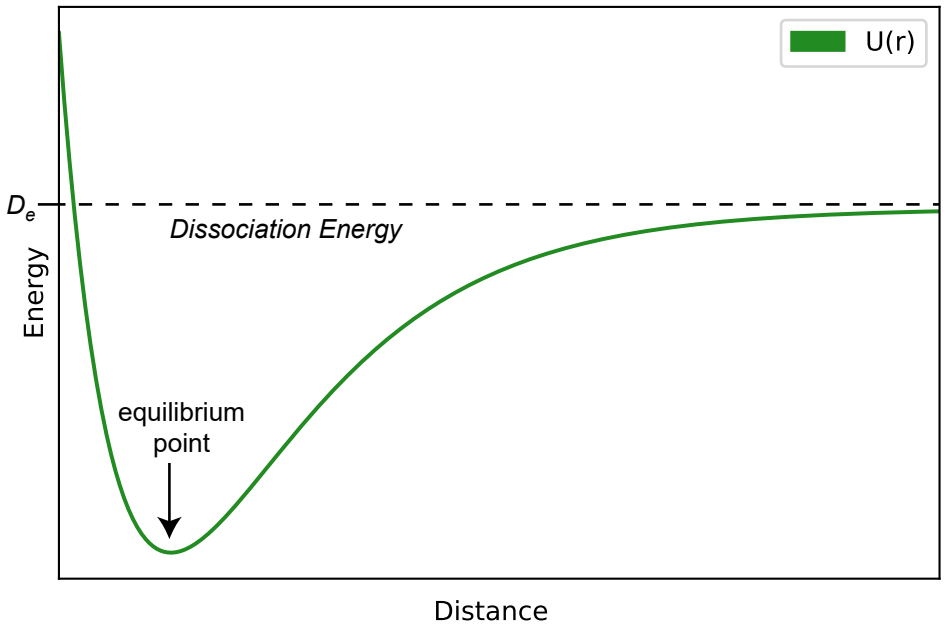


Figure 3.2.: Morse potential. The curve flattens out with larger distances and represents the behavior of a bond more realistically.

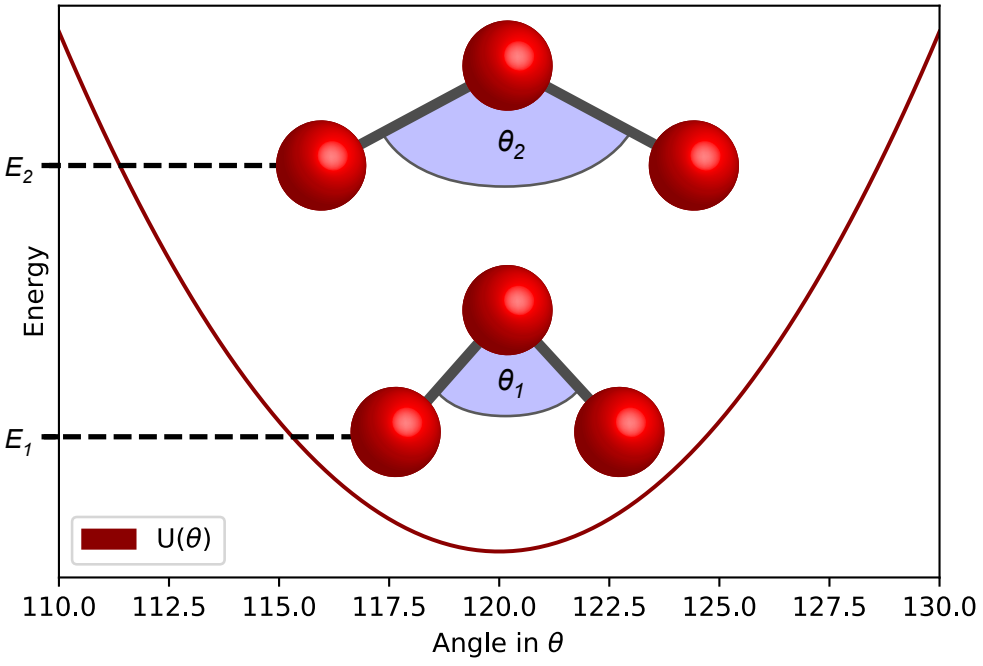


Figure 3.3.: Angle potential. The angle θ is formed between three atoms.

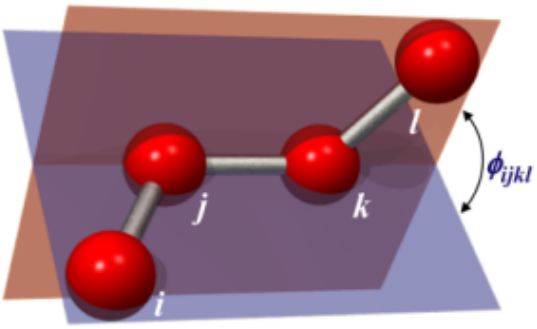


Figure 3.4.: Dihedral potential. The atoms i, j and k and atoms k, j and l form two planes that are inclined to each other at an angle ϕ_{ijkl}. The connecting line from j to k is the intersection line of the two planes, around which atoms i and k can freely rotate. To fix this degree of freedom, the angle ϕ_{ijkl} is introduced. Image taken from[51].

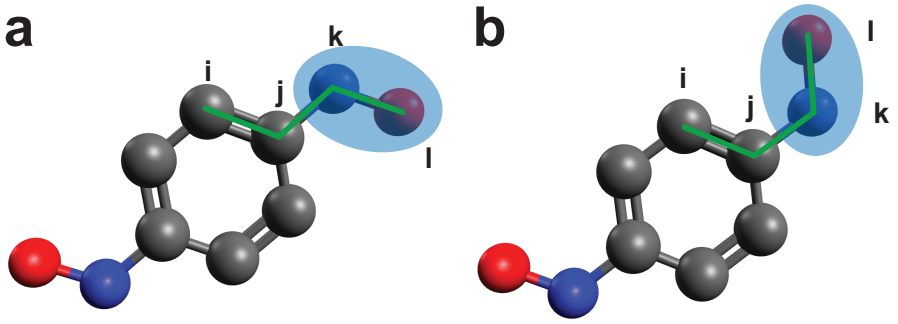


Figure 3.5.: **a**: Atoms i, j, k and l form two planes with an inclination of zero degrees. **b**: The inclination of the planes is 180 degree and represents another possible orientation of the molecule. This stereoisomerism is taken into account by the force-field. Color code: C (black), N (blue), O (red). Hydrogen omitted for visibility.

Dihedral potential

Considering another atom l in addition to the atoms i, j and k with the bond distances $\vec{r}_{ij}$, $\vec{r}_{jk}$ and $\vec{r}_{jk}$, further specifies the structure.

In figure 3.4 atoms i, j and k and j, k and l each form two planes inclined to each other at the angle ϕ_{ijkl}. The intersection line of the planes is the vector $\pm\vec{r}_{jk}$. If no further potential (and parameter) is introduced, then atoms i and l would be able to freely rotate without violating the bonded and angle potentials with θ_{ijk} and θ_{jkl}. Introducing ϕ_{ijkl} allows to specify a preferred position. The potential typically has the form of

$$U(\phi_{ijkl}) = k^d_{ijkl}(1 + \cos(n_{ijkl}\phi_{ijkl} - \phi_0)) \quad , \tag{3.12}$$

with k^d_{ijkl} as the scaling factor, (n_{ijkl} the multiplicity and ϕ_0 the equilibrium angle (phase shift angle for $n_{ijkl} > 0$). The multiplicity (n_{ijkl}) is a positive integer value that results in different energy minima and, thus, takes stereoisomerism into account. An

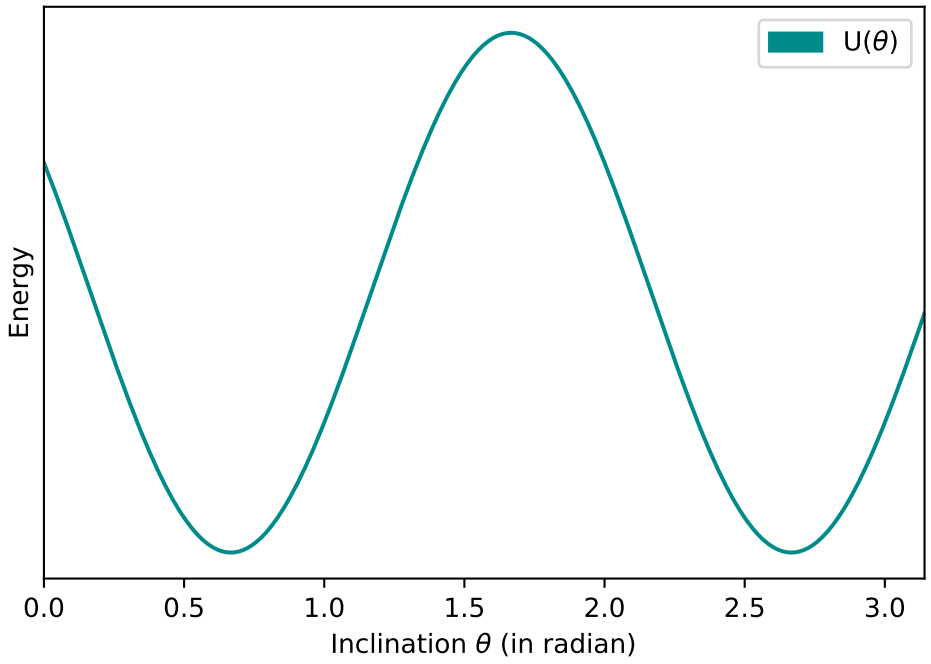


Figure 3.6.: The course of the dihedral potential is shown with multiple minima. This means that different orientations of the structure are possible and can indicate a symmetry.

example for that is given in figure 3.5. It shows an arbitrarily designed molecule where two different configuration are shown that are equally possible. Both configurations represent a planar molecule where the torsion is either zero or 180 degree. With this, the course of the potential shown in figure 3.6 should possess the same amount of energy minima in the range of one fully period as the number of possible configurations.

Improper potential

The improper potential is another potential with four atoms where atom i is connected to three other atoms j, k and l. In this case, the planes are formed by atoms i, j and k and j, k and l. This means that three atoms form a plane and the fourth atom is outside of the plane (see figure 3.7). The angle ϕ_{ijkl} again is the inclination of the two planes at each other. The form of the potential is in this case of a harmonic potential

$$U(\phi_{ijkl})_{improper} = k^i_{ijkl}(\phi_{ijkl} - \phi_0) \quad , \tag{3.13}$$

where k^i_{ijkl} is the scaling factor and ϕ_0 the equilibrium angle. This potential ensures that a planar structure is kept planar. The course of the potential is similar to the stretch potential in figure 3.1.

3.1.1.2. Nonbonded interactions

Besides bonded interactions, there also exist nonbonded forces such as the van der Waals force and the electrostatic force. These are forces between two atom i and j.

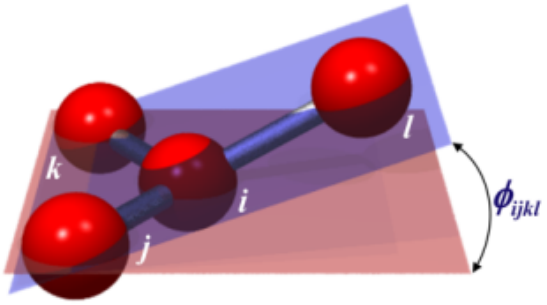


Figure 3.7.: Another form of torsion with four atoms. In this case three atoms form a plane while the fourth atom is outside. Image taken from[51].

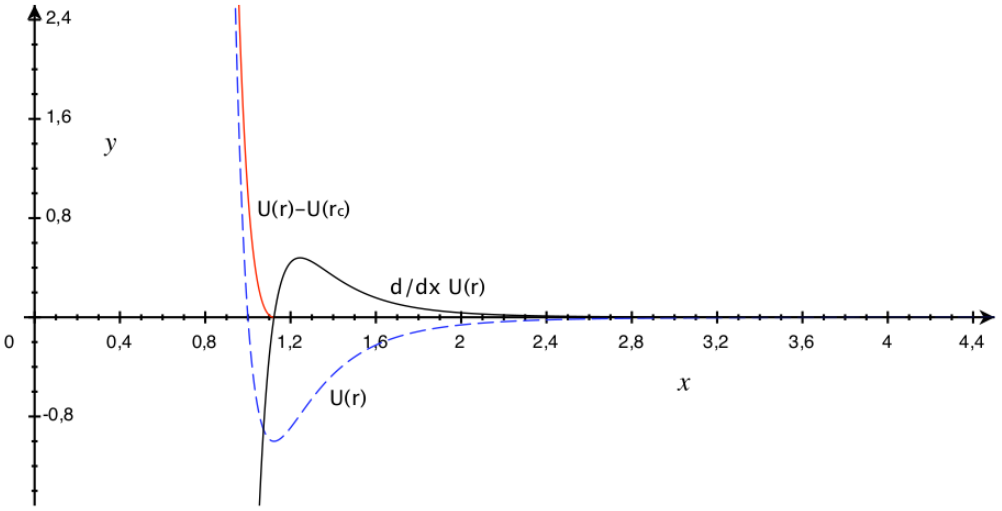


Figure 3.8.: The 12-6 Lennard-Jones potential (blue dotted line) acts repulsive (red line) at smaller distances. After the equilibrium distance it appears attractive between the two interacting particles. The black line shows the force. Image taken from[52].

Van der Waals interaction

The energy potential related to the van der Waals force has the form of

$$U(r_{ij})_{LJ} = D_{ij}\left(-2\left(\frac{r_{ij}}{r_0}\right)^6 + \left(\frac{r_{ij}}{r_0}\right)^{12}\right) \quad , \tag{3.14}$$

with the potential well depth D_{ij} and the distance r_{ij}. This potential is called the 12-6 Lennard-Jones potential [44], see figure 3.8. Repulsion and attraction are both part of this potential at different distances.

Electrostatic interaction

Another nonbonded interaction between atoms is the electrostatic interaction. This force is well known and is

$$U(r_{ij})_{el} = \frac{1}{4\pi\epsilon_0\epsilon_r}\frac{q_1 q_2}{r_{ij}} \quad , \tag{3.15}$$

with the permittivity of the vacuum ϵ_0, dielectric constant ϵ_r, the partial charges q_1 and q_2 and again the distance r_{ij}. The course of this potential is a simple $\propto {}^1/_x$.

In addition to these common potential terms, further potentials do exist for special cases. These often depend on the geometry or atom types. It should be noted that a periodic boundary conditions (will be explained in chapter 3.1.2) may interfere with the accuracy of the electrostatic potential. In short, the periodic boundary condition is a volume in which the molecule is located that is meant to be analyzed and everything outside of the volume is not part of the simulation. This affects long range electrostatic forces that have to be cut off to save computing time. Depending on the cut-off parameter, the influence of the electrostatic force may or may not be correct. An approach to solve this issue is given by the *Ewald summation* method[53].

3.1.1.3. Parameterizing a force-filed

The potential equations for most force-fields are more or less the same, however, the parametrization of the scaling factors is the key difference. In addition, it is of utmost importance to take different bond types into account. For example, a normal bonds behaves differently than a double or a triple bond. Even the type of hybridization is important, because a carbon atom in an aromatic bond is not the same as a carbon atom in a polymer chain (equilibrium distance). This leads to multiple parameters for a single atom with different types. All this has to be addressed by the force-field to achieve a high accuracy.

As mentioned above, the force-field *UFF* is often used for MOFs and widely used[54]. The reason for that is, that it covers a wide range of the periodic table in contrast to other force fields that aim a high accuracy on a smaller set of molecules and atoms. However, metal atoms in particular are not well described due to a lack of data and, thus, require additional investigation. An extensions of *UFF* with adjusted metal parameters is given by the force-field *UFF4MOF*[54]. In following, the parametrization of the parameters from the force-field *UFF* are presented. All equations and parameters (if not labelled otherwise) are shown from the original *UFF* work[50].

Bond radii

The natural bond length is calculated by

$$r_{ij} = r_i + r_j + r_{BO} - r_{EN} \quad , \tag{3.16}$$

with the individual radii r_i and r_j, the bond order correction r_{BO} (single, double, triple) and the electronegativity correction r_{EN}. Note that in this equation the electronegativity correction is subtracted and not added, contrary to what can be seen in the publication. There are some more errors that can be found in [55]. X-rays structures from molecules such as propane (propene, propyne), benzene and different

carboxylates were used to fit the individual radii of certain elements. The results were extrapolated for other elements, e.g. noble gases. The correction for the bond order is described as

$$r_{BO} = -\lambda(r_i + r_j)\ln(n) \quad , \tag{3.17}$$

with n as the bond order and $\lambda = 0.1332$ as the proportionality constant. The correction by the electronegativity is

$$r_{EN} = r_i r_j \frac{(\sqrt{\chi_i} - \sqrt{\chi_j})^2}{\chi_i r_i + \chi_j r_j} \quad , \tag{3.18}$$

with the electronegativities χ_i and χ_j.

Bonded constants

The scaling factors for the bonded interaction are derived as follows

$$k_{ij} = 664.12 \frac{Z_i^* Z_j^*}{r_{ij}^3} \quad , \tag{3.19}$$

with the atomic charges Z_i^* and Z_j^* (given by the force-field). The equilibrium angles were obtained from reference structures such as water and methyl vinyl ether for the angular potential. The values for the rest of the atoms is again extrapolated. The scaling parameter is calculated by

$$\theta_{ijk} = \beta \frac{Z_i^* Z_k^*}{r_{ik}^5} [3 r_{ij} r_{jk} (1 - \cos^2(\theta_0)) - r_{ik}^2 \cos(\theta_0)] \quad , \tag{3.20}$$

with β as

$$\beta = \frac{664.12}{r_{ij} r_{jk}} \quad . \tag{3.21}$$

The dihedral scaling parameter is

$$\theta_{ijkl} = \frac{1}{2} V_\phi \quad , \tag{3.22}$$

with V_ϕ as

$$V_{\mathrm{sp^3}} = \sqrt{V_j V_k} \tag{3.23}$$

for sp^3 hybridization and

$$V_{sp^2} = 5\sqrt{UjU_k}(1 + 4.18 \cdot \ln(BO_{jk})) \tag{3.24}$$

for sp^2 hybridization, again with the bond order BO_{jk}. The atomic parameter V_i and V_j are given by the force field and obtained by fitting low vibrational modes of ethylene, benzene and N,N-dimethylformaldehyde. The constants U_i and U_j for atoms are assigned by specific values depending on their group. The potential well depth D_ij for the van der Waals potential is calculated by

$$D_{ij} = (D_i D_j)^{1/2} \quad , \tag{3.25}$$

with the atomic van der Waals energies D_i and D_j. These are obtained from literature (or quantum mechanics calculations including approximations). For the electromagnetic interaction, only partial charges are needed and have to be calculated by other means (e.g. DFT).

A molecular structure with known atom positions can now be described to obtain a topology. This topology represents the physical model of the structure where each interaction is specified that contributes to the total energy of the system.

3.1.2. Structure Optimization

A given structure that is parameterized by a force-field can now be optimized by minimizing the total energy U_{total}. To do this, equation 3.3 has to be solved with the help of equation 3.2 to calculate the atomic forces. The atomic forces determine the atomic movements. If all forces cancel out each other or are below a minimal threshold, the energy is minimized and the structure is in its optimized form. The simulation itself happens in multiple time steps Δt, with one step usually being in the range of $\Delta t = 10^{-15}$ s. After each step all forces and other parameters, such as temperature or density, are calculated again to create a frame. The sequence of these frames displays a trajectory of all atoms, i.e. their position over time. It should be noted that computational resources limit such simulations, since increasing the number of time steps in a simulation leads to an proportional increase of computing time and data. A simulation of only 1 μs already requires 10^9 steps with a data footprint depending on the frequencies of saving trajectory snapshots. The simulation effort also increases with the number of atoms, because each atom contributes to the system and has to be calculates separately.

Integrators

If the force $\vec{F}(t)$ over time is known, a velocity step $\vec{v}(t)$ of the atom can be calculated by the *leap-frog integrator*

$$\vec{v}(t + \frac{1}{2}\Delta t) = \vec{v}(t - \frac{1}{2}\Delta t) + \frac{\Delta t}{m}\vec{F}(t) \quad , \tag{3.26}$$

with m as the mass[56]. With this, the position $\vec{r}$ can be calculated as

$$\vec{r}(t + \Delta t) = \vec{r}(t) + \Delta t \cdot \vec{v}(t + \frac{1}{2}\Delta t) \quad . \tag{3.27}$$

As can be seen, an integrator is needed to form a trajectory for each atom (position at a certain time), since equation 3.26 is based on equation 3.2. And the key task was indeed to solve the equation 3.2. Other integration algorithms can be found in [56].

For the simulation itself certain conditions have to be set. For example, the size of the simulation cell (in most cases a box) must cover every atom position at any time. Also, one must decide which thermodynamic quantity has to be kept untouched. Table 3.1 shows which parameters can be fixed and, in addition, a distinction is made between extensive and intensive parameters. The first choice is between the chemical potential μ and the particle number N, the second choice is between the volume V and pressure p and the third choice is between the energy E and the temperature T. The reason one can only vary one quantity from a column in table 3.1 is because they are directly dependent on each other.

The choice of each parameter depends on the system and the goal of the simulation. An open system with free energy exchange for example should not fix the energy but the temperature and a system where the particle number remains the same at all time should fix the number of particles and not the chemical potential. The microcanonical ensemble is widely used and is defined by fixing the particle number, volume and energy and is called a NVE simulation.

Table 3.1.: Thermodynamic quantities to set an ensemble. One parameter of each column has to be fixed to form a set of parameters that define the thermodynamic ensemble.

extensive quantity	particle number N	volume V	energy E
intensive quantity	chemical potential μ	pressure p	temperature T

There are two options how to deal with particles leaving the simulation cell. In case of fixed boundary conditions, the particles are lost and with periodic boundary condition they are able to re-enter the simulation from the opposite side (see figure 3.9). With a simulation cell of the size of a cubic box with the cell length L a particle underlies the following rule

$$\vec{r} + \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} L = \begin{pmatrix} x + L \\ y + L \\ z + L \end{pmatrix} = \begin{pmatrix} x \\ y \\ z \end{pmatrix} \quad . \tag{3.28}$$

Temperature and pressure coupling

In case of a NPT simulation, temperature and pressure coupling have to be applied. One method of temperature coupling is to use the Berendsen algorithm that connects the system to an external heat bath T_0

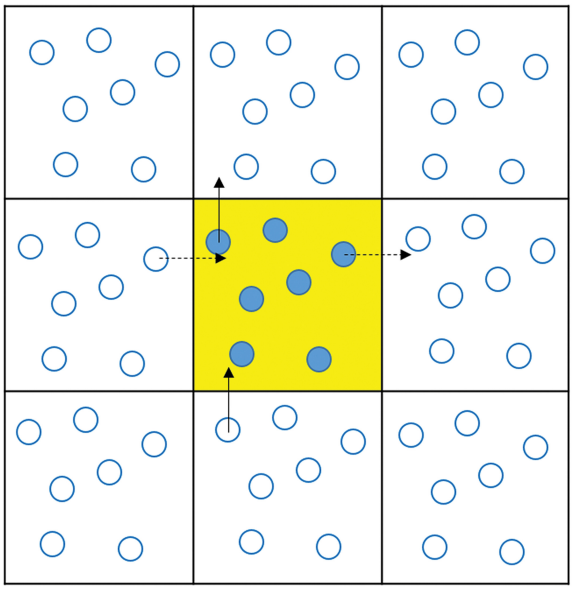


Figure 3.9.: Two-dimensional view of the periodic boundary condition. The middle cell is the simulation cell. When a particle exits the cell, it re-enters it again on the opposite side as if a new particle with equal properties entered from a neighboring cell under the same conditions. Image taken from[57].

$$\frac{\mathrm{d}T}{\mathrm{d}t} = \frac{T_0 - T}{\tau} \quad , \tag{3.29}$$

with τ as the time constant. To apply the new temperature to the system, the velocities of the particles are changed by λ as

$$\lambda = \left[1 + \frac{n_{TC}\Delta t}{\tau_T}\left\{\frac{T_0}{T(t - \frac{1}{2}\Delta t)} - 1\right\}\right]^{1/2} \quad , \tag{3.30}$$

with the relation for τ_T

$$\tau = \frac{2C_V\tau_T}{N_{df}k} \quad . \tag{3.31}$$

Here n_{TC} is the time step (of the system), C_V the total heat capacity of the system, k the Boltzman constant and N_{df} the total degree of freedom[56]. For the pressure coupling the Berendsen algorithm for pressure can be used. In this case, it is

$$\frac{\mathrm{d}P}{\mathrm{d}t} = \frac{P_0 - P}{\tau_P} \quad , \tag{3.32}$$

with P_0 as the reference pressure and τ_P the time step. The coupling is performed on the system when the scaling matrix μ is applied onto the simulation cell after every time step n_{PC}. The matrix μ is

$$\mu = \delta_{ij} - \frac{n_{PC}\Delta t}{3\tau_P}\beta_{ij}\{P_{0ij} - P_{ij}(t)\} \quad , \tag{3.33}$$

with β as the isothermal compressibility of the system[56].

3.1.3. Challenges of MOF MD simulations

A general issue is the availability and/or the accuracy of force-fields. The force-field *UFF* offers parameters for a wide range of atoms, but is not able to reproduce the exact model for all metal nodes. One possible solution for *UFF* is to fix the metals during the simulation[58]. This means all metal atom positions are fixed and are unable to move around. Another way is to use the extension *UFF4MOF* specially designed for MOFs[54]. Efforts are being made to create suitable force-fields for MOFs such as *MOF-FF*[59]. Even *ab initio* methods are investigated on, which derive parameters directly from theory[60].

Another challenge is the exact implementation of the parameterization of the MOF structure. Tools like the *lammps_interface* assign bonds, angles and torsion automatically but are far away from the perfect solution[58]. The only way to be one hundred percent accurate is to define each potential manually, which is a huge task even for small molecules and frustrating for medium sized structures $N \propto 10^2$.

Another major challenge is the limit set by the available hardware. Simulating larger systems on time scales sufficiently long are almost impossible to simulate in atomistic detail as the resource usage increases dramatically. To be able to simulate very large systems ($N \propto 10^5$ and bigger) compromises are required. One being the use of coarse-grained models. These models combine multiple atoms and molecules into one particle. The coarse-grained force-field then assigns parameters for these newly defined particles to reduce computing effort. However, it is still possible to simulate certain parts of the simulation cell at atomistic detail while the rest is handled as coarse-grained atoms[61]. A more in-depth analysis of these issues is given in[5].

Conventional force-fields as those mentioned above do not change the bonds of a given structure. This means, bonds are neither broken nor newly formed. Thus, simulating chemical reactions is not possible and requires a force-field that is able to do so or an outright different simulation method. One force-field that aims to include chemical reactions is *ReaxFF*[62]. There is however one issue (among others) where the force-field does not lead to expected results when the input structure has to be optimized first[63]. One reason for that is possibly the method for charge transfer used in *ReaxFF*, which has issues with long ranged interactions. But further progress still remains of great interest.

3.2. Database screening

Since the number of reported MOFs is steadily increasing, databases of MOF structures have emerged to collect them in one place. The *Cambridge Crystallographic Data Centre* (CCDC) has a section where about 70000 MOFs are collected[64]. Furthermore, the modular structure of MOFs allows many hypothetical MOFs[65]. In this work, the focus lies on the *Computation-Ready Experimental Metal-Organic*

Framework Database (CoRE MOF Database) where all experimentally realized MOFs, that are reported in literature, are collected[66]. More than 14000 MOFs are provided as .CIF files, a file format specially designed for crystalline structures (shown below)[67]. This file format easily allows to screen one file after another in form of a programmed analysis tool.

CIF FILE SYNTAX (MINIMALISTIC VERSION):

```
# LINE STARTING WITH '#' IS A COMMENT
#======================================================================

# CRYSTAL DATA

#----------------------------------------------------------------------

# FILE NAME
_example_mof                          EXAMPLE MOF STRUCTURE FILE

# UNIT CELL PARAMETER
_cell_length_a                        LENGTH 1
_cell_length_b                        LENGTH 2
_cell_length_c                        LENGTH 3
_cell_angle_alpha                     ANGLE 1
_cell_angle_beta                      ANGLE 2
_cell_angle_gamma                     ANGLE 3

# DEFINITION OF COLUMN ENTRIES
loop_
   _atom_site_label
   _atom_site_fract_x
   _atom_site_fract_y
   _atom_site_fract_z
   _atom_site_type_symbol
#ATOM AND ATOM POSITION
   Atom Label 1      X_1   Y_1   Z_1   Element Symbol 1
      :              :     :     :           :
      :              :     :     :           :
      :              :     :     :           :
   Atom Label N      X_N   Y_N   Z_N   Element Symbol N
```

By implementing the needed properties in an algorithm, one can quickly find promising candidates for desired applications. These candidates allow quicker investigation of interesting applications such as methane storage[65], hydrogen uptake[68] or kinetic separation[69]. Furthermore, designing more complex algorithms allows to screen for more than just one single physical property[70].

4. Screening Method

This chapter describes step by step the implementation of the screening method. The aim of this screening is to find suitable MOF structures for a given cross-linker. Suitable in this context means a geometrical fitting accuracy where the cross-linker connects to the MOF without deforming and is not deformed itself. Another important point is that it should be possible to continue the functionalization of the MOF in every direction to achieve highly ordered structures. Thus, the functionalization of one unit cell should not interfere with the functionalization of the next unit cell and so on.

To compare a MOF and the cross-linker both structures have to be known. While the MOF structure is obtained directly from the database, the cross-linkers were suggested by fellow chemist colleagues (the author of this thesis has a background in physics). The structure and topology of MOFs was discussed in chapter 2, so that only the specification of cross-linkers remains. A cross-linker has the task to connect to a MOF to further stabilize the structure. Depending on the type the cross-linker possesses two or more binding sites, which makes it either linear, planar or three dimensional.

Figure 4.1 shows two cross-linkers **a** and **b**. The linear cross-linker possesses two binding sites, the planar one possesses three. In this example the nitrogen atoms bind to the MOF. As the used force-field is UFF that is not able to simulate reactions, special measures have to be taken to implement a reaction free binding. This will be discussed in section 4.3.1.2.

4.1. Motivation

A cross-linker that fits exactly into the pore of a given MOF is quite to find. In addition, the increasing number of structures do not simplify this task. Staying up to date and also knowing the properties of each MOF for cross-linking is hardly feasible. The idea now is to outsource as much work as possible to automatized algorithms that filter a large number of MOFs. With a smaller selection of MOFs and a high rate for a perfect fit, the probability of success is assumed to increase significantly. Also, using a database whose structures are all experimentally realized helps for fast experimental cross-checking. If successful, the results of this work will allow to find many novel materials with tunable properties by using MOFs as templates to form COFs.

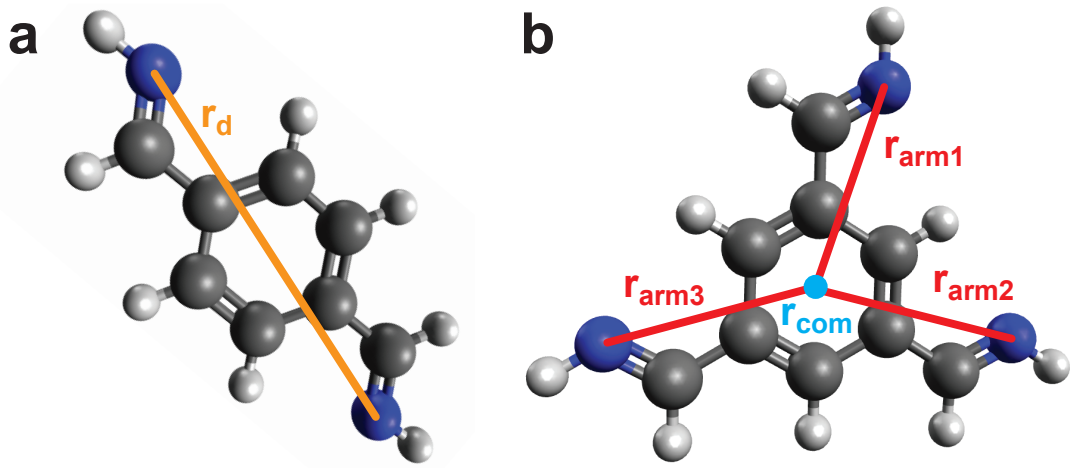


Figure 4.1.: a: Linear cross-linker with two binding sites. The length is defined by the distance of the binding sites. **b:** Planar cross-linker with three binding sites. Here there are three distances, each from the center of mass to the respective binding site. Color code: H (white), C (black), N (blue).

4.2. MOF database

In this work the *CoRE MOF Database* was screened. All structures are experimentally realized and reported in literature and allows to immediately confirm the results from the screening. From a technical point of view, it is also logical to screen a database of cross-linker for a given MOF structure. This procedure makes sense when the desired MOF structure is important, for example for future modifications. However, the problem is the absence of a database that contains a large number of cross-linkers. In this work, the focus was on a small set of cross-linker, thus, starting a new database out of these cross-linkers would be pointless. Such a small set only requires a sense of proportion and not a complex algorithm to find suitable cross-linkers. If, however, the goal is to screen through the cross-linker database multiple times with another large database of MOFs, the outcome would be the same as the outcome of this work. The only difference would be, that the screening method of this work is more computationally efficient and resource friendly and less prone to errors. If in future a cross-linker database is released, the screening of such database with one reference MOF structure should be also be considered.

4.3. High-Throughput screening

The screening algorithm is written in the programming language *python*[71] (version 3.7). The module (python library with functions) *pymatgen* was used to load, read and analyze the MOF structures[72]. Besides python standard modules, third party modules were used, such as *numpy* to implement the calculations[73] and *mpi4py* for code parallelization[74].

4.3.1. Screening preparation

The first question is how to compare the MOF with the cross-linker. The comparison happens by measuring the distance length of a cross-linker and then cross-checking with possible binding sites from the MOF. First, the length of a cross-linker has

to be defined. In case of a linear linker the length is just the distance[1] $\vec{r}_d$ from one binding site to the other. For planar type linker the distance has to be defined by other means, as multiple binding sites exist. Here it was decided to introduce the center of mass of all binding sites $\vec{r}_{com}$ with the mass m being equal for every atom. This is done as only the geometric center is needed for the distances. The formula reduces then to

$$\vec{r}_{com} = \frac{1}{N_{bs}} \sum_{i}^{N_{bs}} \vec{r}_i \quad , \tag{4.1}$$

with N_{bs} as the number of total binding sites of the cross-linker and $\vec{r}_i$ as the position of a binding site. With this, an additional value named *arm length* $\vec{r}_{arm}$ was introduced to finally define the distances

$$\vec{r}_{arm} = \vec{r}_{com} - \vec{r}_i \quad . \tag{4.2}$$

The name *arm* does not have any deeper meaning and was only chosen because most cross-linkers were nearly symmetrical. For a perfect symmetry, all arm lengths should be more or less equal

$$\vec{r}_{arms} \approx \vec{r}_{arm,1} \approx \vec{r}_{arm,2} \approx \ldots \approx \vec{r}_{arm,N} \quad . \tag{4.3}$$

The distances are shown in figure 4.1. Cross-linkers with multiple binding sites where two arm lengths are not equal require special treatment and are not considered in this work.

4.3.1.1. Deviation in cross-linker length

An important factor is the dihedral angle potential. It allows different configurations of the same molecule and also has an effect on the cross-linker length. Figure 4.2 shows the same cross-linker with two different configurations. The first configuration (**a**) results in a cross-linker length of $r_{d,1} = 7.141\,\text{Å}$, while the second configuration (**b**) has a length of $r_{d,2} = 7.450\,\text{Å}$. The question that now arises, if this effect has any impact on the screening. The initial idea was to find a perfect fit to form highly ordered structures. Even smallest deviation is supposed to pile up to large mismatches of cross-linker and MOF when going towards higher hierarchies. In this example, if the assumed perfect length of the cross-linker is $r_{d,2}$, then the error would be $\frac{r_{d,1}-r_{d,2}}{r_{d,1}} \approx 4\,\%$. Here the decision has to be made how big the deviation may be tolerated. If we take a direct line from one binding site to another (figure 4.2 **b**) and define it as r_d and compare it to a line where the binding sites do not face each other (figure 4.2 **a**) and define it as r_i, then the deviation between those two distances r_d and r_i is also dependent on the total length of the cross-linker. Thus, the longer the linker is, the smaller the effect of multiple orientation of a single linker. This

[1] Notice that $|\vec{r}| = r$.

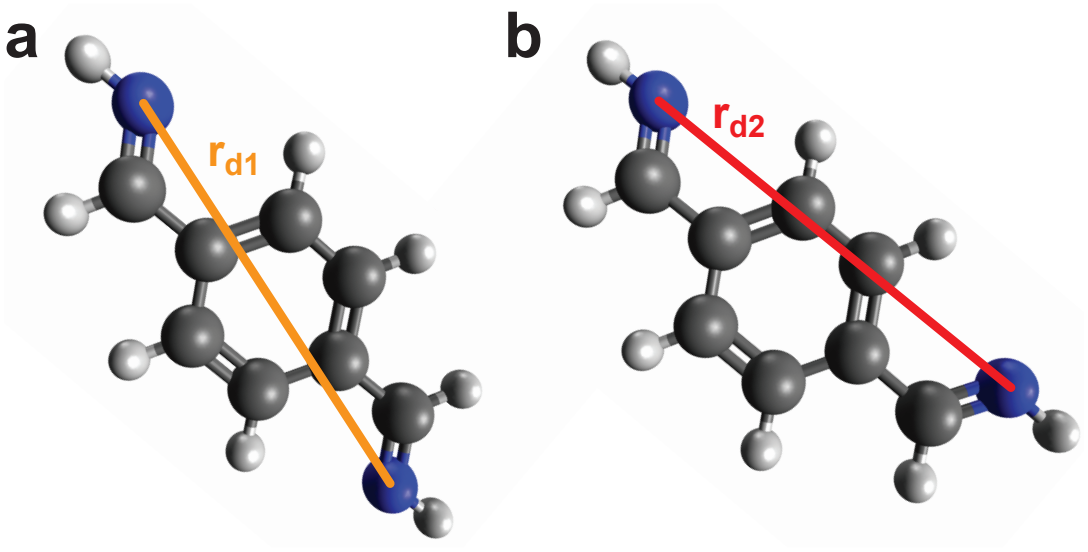


Figure 4.2.: Two configuration of the same cross-linker result in different lengths $r_{d,1} = 7.450\,\text{Å}$(**a**) and $r_{d,2} = 7.141\,\text{Å}$(**b**). Color code: H (white), C (black), N (blue).

conclusion is analog to cross-linkers with more than two binding sites where the total length is replaced by the arm length (from binding site to geometrical center).

The screening algorithm does not compare lengths equal to the input length as the data type of the input parameters has a high precision. It is almost impossible to find a suitable structure when for example the cross-linker length is $r_{cl} = 5.0\,\text{Å}$ and the algorithms calculates a fitting length for the MOF structure as $r_{MOF} = 5.0000001\,\text{Å}$. With this, an upper and lower limit $\pm r_{cl}$ is needed anyways and the deviation coming from multiple orientation of the cross-linker structure can be included. However, examining the fitting accuracy for small deviations will most certainly require further testing.

4.3.1.2. Cross-linker input structure

Another main point is to consider the inability of the used force-field to describe chemical reactions. What happens to the length of a cross-linker after synthesis? What if the reaction process is more complex where one can not easily characterize one single binding site? And are the cross-linker and MOF even compatible in the first place? Since MOFs have to be prepared for cross-linking and the files in the CoReMOF database are unmodified (clean) MOF structures, a general way of avoiding the process of reactions has to be found. In figure 4.3 the azide group of the linker (**a**) and the carbon carbon triple bond of the cross-linker in (**b**) reacts to a triazole shown in (**c**). This example shows that the approach mentioned above of measuring the cross-linker length runs into its limits. First, one has to find the exact distance from the binding site of the first linker to the binding site of the second linker to determine which cross-linker length fits into the MOF. And second, the cross-linker possesses multiple atoms that react with the MOF in each direction. Thus, it is impossible to even define a cross-linker length or a MOF distance.

The idea to solve this problem is to handle the cross-linker as already reacted and separate the cross-linker and MOF in a way where the MOF remains in its initial

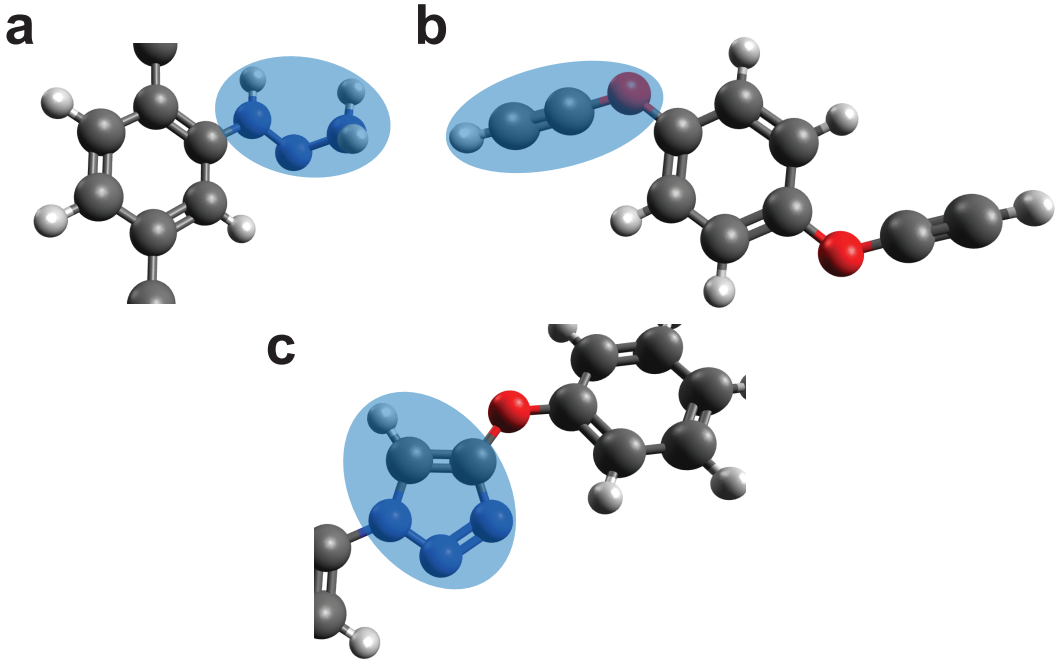


Figure 4.3.: The azide group of the linker in (**a**) and the carbon carbon triple bond of the cross-linker in (**b**) react together and form a triazole ring (**c**). The blue marked area shows all atoms that are involved in the chemical reaction and the product. Color code: H (white), C (black), N (blue), O (red).

state. This leaves the MOF untouched and allows again to define a cross-linker length for every reaction type. Another positive side effect is that the MOF cross-linker compatibility no longer matters, although possible matches found by the screening should be examined later on if they can actually be realized. Figure 4.4 shows the implementation of this idea by modifying the example shown in figure 4.3. The linker of the MOF (4.3 **a**) BDC remains unmodified and the cross-linker (4.3 **c**) keeps all other atoms that are involved in the chemical reaction. The "reaction" is simplified and in this case can easily be performed by replacing the binding site of the linker by the binding site of the cross-linker. With this, the placement of the cross-linker into the MOF is again a static process and it is possible to define a cross-linker length.

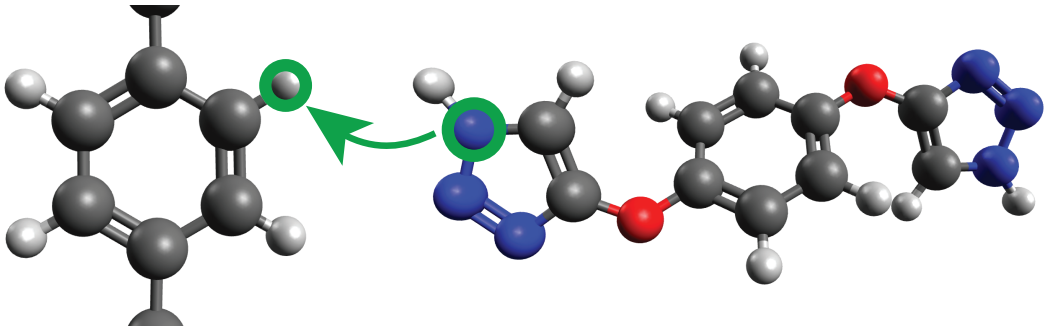


Figure 4.4.: Here the cross-linker takes over the reaction product, leaving the organic linker in its unmodified state. A "reaction" is performed by replacing the bindings site of the linker with the binding site of the cross-linker. Color code: H (white), C (black), N (blue), O (red).

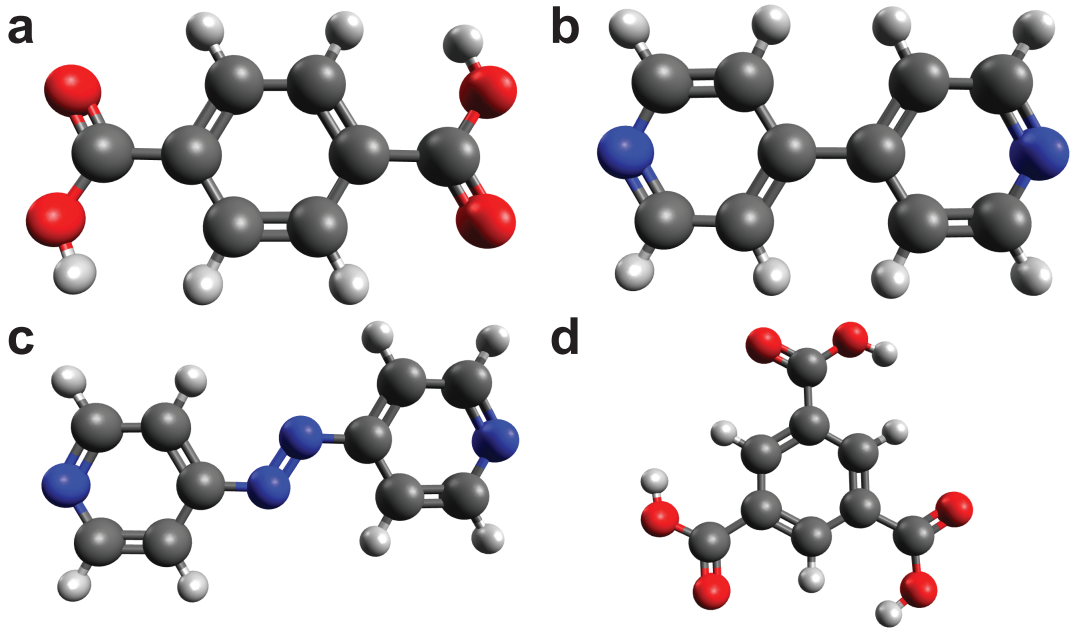


Figure 4.5.: Display of different linkers. (**a**): 1,4-benzenedicarboxylic acid, (**b**): 4,4'-bipyridin, (**c**): 4,4'-azopyridine, (d): benzene-1,3,5-tricarboxylic acid. Color code: H (white), C (black), N (blue), O (red).

4.3.1.3. MOF binding sites

Before any cross-linker lengths can be used for screening, one must first define where the cross-linker binds to the MOF. This is essential as there exist various linker with different geometries. Even distinguishing an organic linker from the MOF via an algorithm is a huge task on its own.

Four different linkers are shown in figure 4.5. Applying the idea from above, a cross-linker docks onto an aromatic carbon by replacing the hydrogen atom with its own binding site. These hydrogen atoms, henceforth *candidates*, are needed to calculate the distances that are then compared against the cross-linker lengths. In case of cross-linkers with two binding sites, the distance between two candidates yields the reference length of the MOF. For a higher number of binding sites, the same number of candidates are taken to calculate the relevant lengths (arm lengths).

The main task to find such candidates is to first identify them. All of the four linker in figure 4.5 possess hydrogen atoms but not every atom is suitable to represent a candidate. The hydrogen atoms attached to the oxygen atoms, for example, (figure 4.5 **a** and **d**) are replaced by the binding sites of the metal node of the MOF. Also in case of a complete MOF crystal, many other hydrogen atoms exist that are not involved in the reaction process such as hydrogen atoms bonded to metal nodes. Hence, the only solution is to find certain criteria that all candidates have in common. In the case of the four linker shown in figure 4.5, all candidates have one thing in common, namely being bonded to an aromatic carbon atom. As the structure file only contains the information of the element type and the position, it is impossible to name all existing bonds. The only thing that remains is to pick a possible hydrogen candidate and look for all atoms that are within a specific range (bond length). But this only finds all hydrogen atoms connected to a carbon atom. It must still be ensured that the found carbon atom is indeed part of a carbon ring structure. This

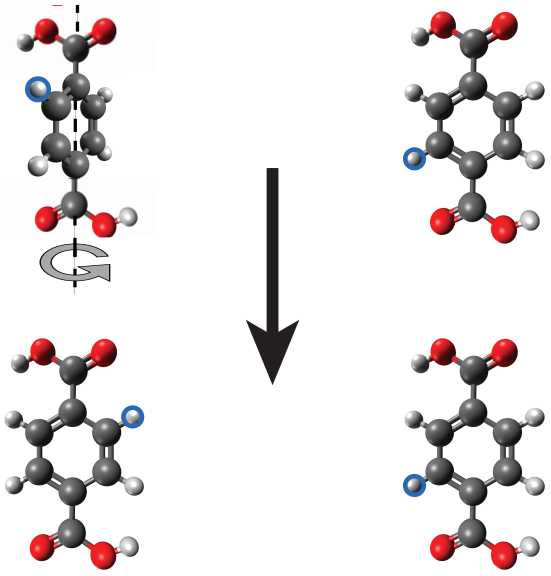


Figure 4.6.: Rotation of a linker to minimize the distance between the candidate hydrogen atoms that are bordered in blue color. The axis, shown as a black dotted line, is made up by the carbon atoms, that are only connected to three other carbon atoms. Color code: H (white), C (black), O (red).

results in recursive steps of neighbor atom checks. As these checks logically aim to find the linker structure itself, a general algorithm to identify all known linkers would grow very large, as each new structure has its own selection criteria. This inevitably leads to a significant increase in computing time. Again looking at figure 4.5, a general concept of searching candidate is not that easy, even when the linker structures are somewhat similar. Different atom types still require their own selection criteria. The algorithm has to find benzene ring type structures in figure 4.5 (**a**), in (**b**) a nitrogen is added, in (**c**) the rings are differently connected to each other and in (**d**) the linker is not linear. All this must be taken into account while scanning the MOF structure for candidates.

4.3.1.4. Linker orientation and screening filter

Another matter to be resolved is the orientation of the linkers that bind to the cross-linker. To avoid unfavorable conditions that forbid continuous cross-linking, further filters have to be applied. These filters have the task to exclude all combinations of candidates that cause any unwanted torsion or prevent continuity. Firstly, the candidates should face each other to avoid torsion. This can be realized by changing the position of a flexible linkers towards each other. And secondly, the line of sight of each candidate should approximately show towards the other candidate (show towards the geometric center in case of more than two candidates). This will likely ensures a minimal torsion of the newly formed bonds between linker and cross-linker.

Flexible linkers such as the one shown in figure 4.5 (**a**) are able to freely rotate around their own axis. This degree of freedom allows the candidates to change their positions. Picking two candidates and rotating them in a way so that their distance is minimized (see figure 4.6), turns them towards each other. Meanwhile, more than two candidates have to move in a way so that each distance from candidate to the arm length is minimized. It should be noted that not all linkers have the ability to

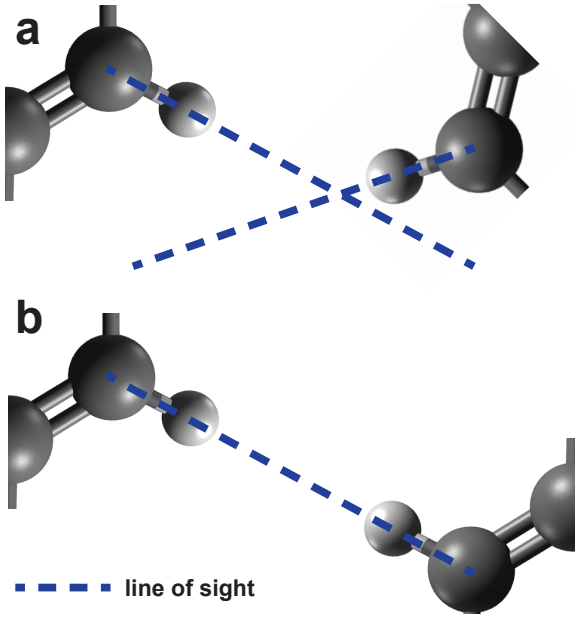


Figure 4.7.: Display of the line of sight of candidates. The line of sight of a candidate is the line coming from its neighbor atom that goes through the candidate itself. If two candidates are screened together, both line of sights should overlap (**b**) to increase the probability of a match. This is because the inserted binding sites of the cross-linker are then more likely to have less torsion. The more the lines deviate from each other the more torsion is to be expected (**a**). Color code: H (white), C (black).

rotate. The linker in figure 4.5 (**d**), for example, does not possess a rotation axis as the three binding sites to the metal node prevent it to move around. Thus, during the screening such linkers should not be able to provide candidates.

Turning the linkers towards each other does not necessarily mean that a clear line of sight exists from one candidate to another. Line of sight in this context means the direction of the bond between candidate and the atom it is connected to. This is shown in figure 4.7 by two orientations. The first orientation is when both lines do not overlap (**a**) and the second orientation shows an overlap of the lines (**b**). If, for instance, in the first case a cross-linker is inserted into the structure, under the condition that all other discussed criteria are met, then the bond between the MOF and cross-linker is expected to have an additional amount of torsion. Therefore, the force field will try to correct the angle that emerges between the different binding sites when neighboring atoms are included, which includes another set of constraints on the system. To avoid additional torsion, only structures with approximately overlapping line of sights are allowed to pass the filter.

4.3.2. Screening implementation

After discussing the basic ideas of the screening process, this chapter explains the detailed implementation. Figure 4.8 shows the general concept of the screening. It is divided into five major parts (some important minor steps still exist). The first step (**I**) is the candidate scan. In this work it was decided to focus only on MOF structures with BDC, since linker is able to freely rotate around. The algorithm for this part is discussed in 4.3.2.1. After the scan for candidates, all combinations of candidates

are screened to find a match. In case of a cross-linker with two binding sites and N candidates of the MOF structure, candidate c_1 is screened together with candidates c_2, c_3 ..., c_N and analogous with cross-linker with more than two binding sites. Two candidates c_i and c_j are then rotated around to find the minimum distance to each other (**II**). Section 4.3.2.2 describes this process. Before the candidate distances are compared against the cross-linker lengths, the line of sight is checked (**III**). The implementation of this step is discussed in section 4.3.2.3. The distance filter finally checks whether the cross-linker fits between two candidates (**IV**) and depending on the outcome the result is saved or not (**V**). It has to be noted, that this step allows to vary the strictness of the filter by tuning the upper and lower limits. This will be discussed in section 4.3.2.4.

To increase the speed of the algorithm, additional filters were implemented that exclude obvious candidate combinations that are not (or less) likely to result in a positive match. These filters are shown in section 4.3.2.5. At last, some computational hurdles are mentioned in section 4.3.2.6 to emphasize that even modern powerful hardware can be brought to its knees when dealing with a large number of files and multiple calculations per step.

4.3.2.1. Candidate selection

As mentioned above, the implemented algorithm for candidates search is only sensitive on BDC like linkers. Loading the MOF structure only provides the information on the element types and positions[2]. The main tasks to find candidates are:

1. Find all hydrogen atoms,
2. Remove all hydrogen atoms that are not connected to a carbon atom,
3. Check, if the carbon atom is an aromatic carbon,
4. Find the carbon atoms of the ring structure that build up the rotation axis.

The MOF structure files in the CoRE MOF database do not possess any information on the atom bonds. To identify the bonds, one must take empirical methods. In this work a neighbor atom was found if it was within a range r_{nd} of another atom. The equilibrium bond length of a C-C bond was estimated at $r_{nd} = 1.55\,\text{Å}$ after calculating the value from the force-field and cross-checking with literature[76]. This bond length also covers a C-H bond and higher bond orders, hence, it was used as a general distance between neighboring atoms. One can still use different bond lengths if needed.

Figure 4.9 shows the process of the candidate scan. First a hydrogen atom H1 is chosen. Then it is checked if the hydrogen atom is bonded to a carbon atom C3. If yes, then it has to be checked if the carbon atom is part of a benzene ring. I true, the carbon atom has to be connected to only two other carbon atoms C2 and C1, whereby C2 has to be connected to two other carbon atoms besides C3 and C1 has to be connected to only one hydrogen and two carbon atoms C4 and C3. If C4 again is only connected to three carbon atoms, the structure is assumed to be a BDC. With this, the hydrogen atom H1 is indeed a candidate and the carbon atoms C2 and C4 build up the rotation axis, which allows H1 to move around. C3 is also needed

[2]Cell parameters are also part of the data but in this case not necessary.

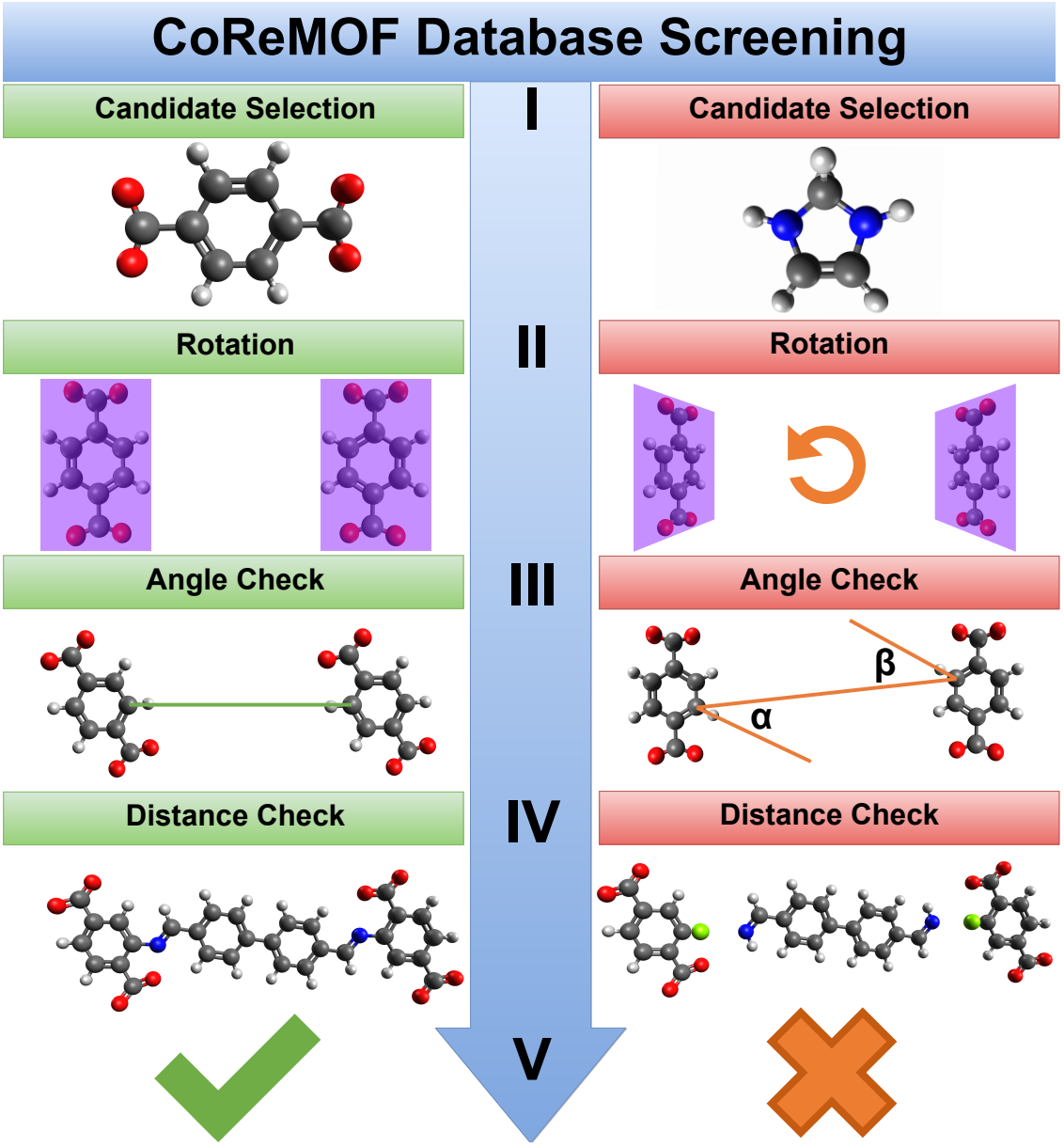


Figure 4.8.: General concept of the screening process. The screening can be divided into five steps. (**I**) The first steps is to filter out all structures that do not possess the desired linker structure. In this work the preferred linker was a simple BDC with the candidates being the hydrogen atoms bonded to the aromatic carbon atoms. (**II**) After collecting all candidates, they are checked if they are in a minimal distance to each other. This is done by rotating the linker around its axis. The axis is defined by the vector going through the carbon atoms of the BDC that are each bonded to three other carbon atoms. (**III**) The third step is a filter that removes all candidate combinations that are not in a direct line of sight to each other. (**IV**) The next filter finally compares the distance of the candidates with the length of the cross-linker (limits included). (**V**) If all criteria are met, the MOF structure including the candidates are saved in an output file, else the screened candidates are dismissed and the next candidates analyzed. Color code: H (white), C (black), N (blue), O (red), binding site (green). Reused from [75].

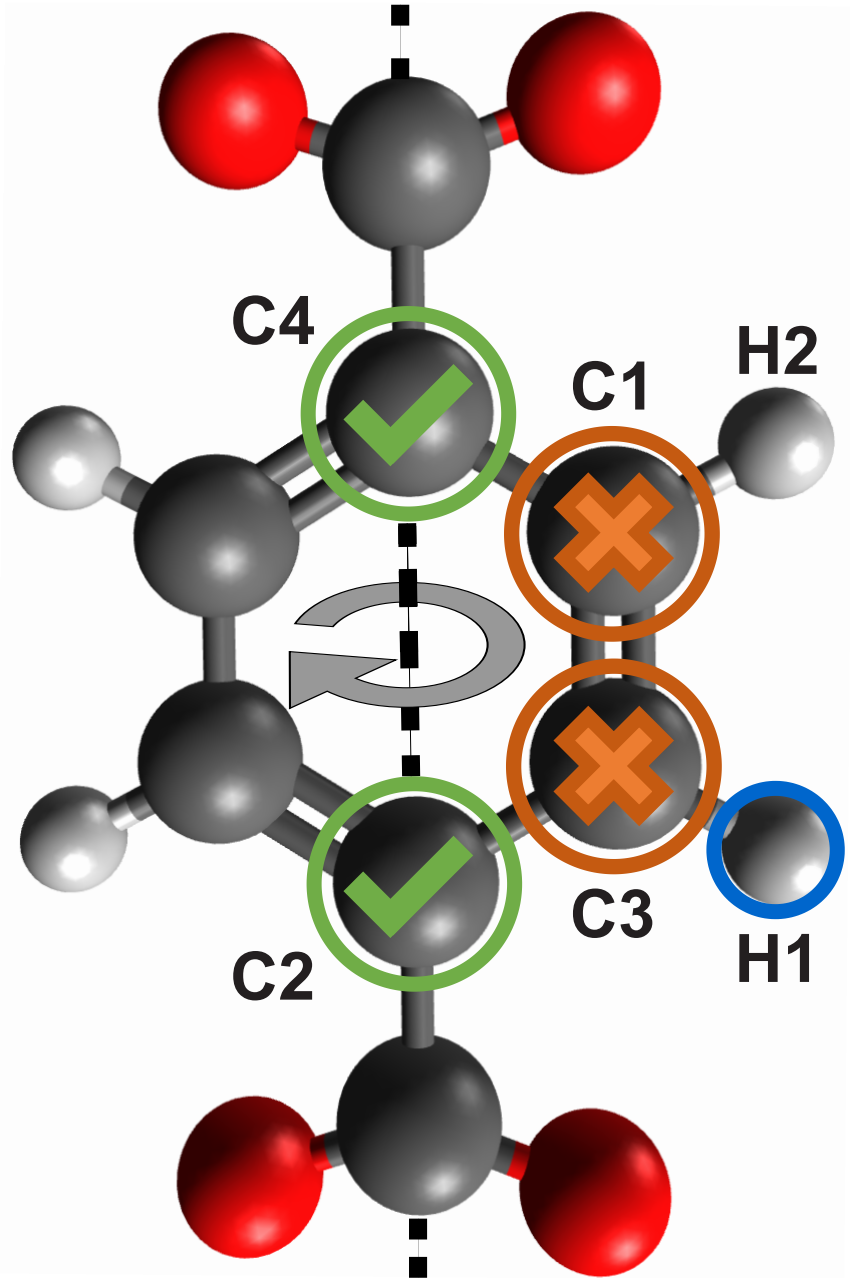


Figure 4.9.: Process of candidate scan. The algorithm first searches for all hydrogen atoms in the structures. After that, all hydrogen atoms that are not bonded to a carbon atom are sorted out. The next step is to verify if the carbon atom C3 bonded to the hydrogen H1 is part of a ring structure. If true, it has to be connected to two other carbon atoms C2 and C1 from which one has to be connected to three other carbon atoms (C2) and the other has to be connected to two carbon atoms C3 and C1 and a hydrogen atom H2. The last step requires that the carbon atom C1 is connected to another carbon atom C4 that is again connected to three other carbon atoms. If all criteria are met, hydrogen H1 is considered a candidate for the screening. Carbon atoms C2 and C4 are then necessary to define the rotation axis and carbon atom C3 is needed to check for the line of sight. This process is done for each hydrogen in a structure and in this case (BDC) all shown hydrogen atoms represent candidates. Color code: H (white), C (black) O (red). Reused from [75].

to calculate the angles for the angle filter that is applied before the distance check. Therefore, the candidate is stored in the following way:

$$\begin{aligned} \text{candidate list} = (&\text{candidate (H1)},\\ &\text{carbon bonded to candidate (C3)},\\ &\text{axis atom 1 (C2)},\\ &\text{axis atom 2 (C4)}) \quad . \end{aligned} \tag{4.4}$$

4.3.2.2. Rotation of linkers

To rotate the linker a rotation axis has to be defined. This was done in section 4.3.2.1 where two carbon atoms of the BDC linker $\vec{r}_1$ and $\vec{r}_2$ defined the rotation axis

$$\vec{r}_{axis} = \vec{r}_{r,2} - \vec{r}_{r,1} \quad . \tag{4.5}$$

The sign of the vector does not matter as it only defines the mathematical positive direction of the rotation. To rotate a vector $\vec{v}$ around an arbitrary axis $\vec{a}$ by the angle θ, one can use the Rodrigues formula[77]

$$\vec{v}_{rotated} = \vec{v} \cdot \cos(\theta) + (\vec{a} \times \vec{v}) \cdot \sin(\theta) + \vec{a}(\vec{a} \cdot \vec{v})(1 - \cos(\theta)) \quad . \tag{4.6}$$

However, some preparations have to be made before rotating the vector. The equation 4.6 rotates a vector around an axis that runs through the origin of the coordinate system. Therefore, the positions of the objects, that have to be rotated, should be shifted before applying the rotation and shifted back after rotating it around. As the vector always rotates around the origin of the coordinate axis, it is enough to shift the rotation objects $\vec{x}_i$ by the vector of any atom that builds up the rotation axis

$$\vec{x}_i^* = \vec{x}_i - \vec{r}_{r,1/r,2} \quad , \tag{4.7}$$

as $\vec{r}_{r,*} - \vec{r}_{r.*} = \vec{0}$. Since the angle for the rotation is unknown, which minimizes the distance of the candidate, it must be determined in a numerical way. The problem is that in case of two candidates two rotation angles must also be varied simultaneously. For more than three candidates it is sufficient to vary the angles one after the other. This is because the distance to each other is not minimized, but to the geometric center. The rotation steps are performed with an infinitesimal angle, which is $\frac{\pi}{N}$. The step size N can be freely chosen and determines the accuracy of the minimization. However, choosing a large number for N results in smaller steps and, thus, longer computing times. The individual steps for distance minimization of two atoms a_1 and a_2 are as follows:

1. calculate initial distance d_{init}
2. rotate first atom a_1
3. calculate temporary distance rotation d_{temp}
4. if $d_{temp} > d_{init}$: rotate in the other direction

5. if $d_{temp} < d_{init}$: update initial distance $d_{init} = d_{temp}$
6. rotate as long as $d_{temp} < d_{init}$
7. if $d_{temp} > d_{init}$: perform step 3 - 6 for second atom a_2
8. repeat until $d_{temp} > d_{init}$ for both atoms a_1 and a_2 .

In case of more than three candidates $a_1, a_2, \ldots a_N$, the algorithm is changed to:

1. calculate each initial distances $d_{init,1}, \ldots d_{init,N}$ to the geometric center
2. rotate first atom a_1 and calculate temporary distance $d_{temp,1}$
3. if $d_{temp} > d_{init}$: rotate in the other direction
4. if $d_{temp} < d_{init}$: update initial distance $d_{init} = d_{temp}$
5. rotate as long as $d_{temp} < d_{init}$
6. repeat for all other atoms .

After the rotation, the final distances are minimized and saved as reference for the distance check. Figure 4.10 shows again the rotation and in addition the distance between two candidates as a top down view. The dashed line is the trajectory of the rotation and the blue bordered hydrogen atoms represent the candidates. The area highlighted in purple indicates the rotation step that is performed in the direction of rotation. In this example, it is enlarged for a better view, but has to be kept small for better numerical accuracy (meaning $N \gg 1$). Also note that this example shows two nicely aligned linkers with parallel rotation axis. However, this is not always the case.

4.3.2.3. Angle filter

The angle filter calculates the angle between two vectors $\vec{r}_{cc}$ and $\vec{r}_{hc}$ and compares against the maximal allowed angle θ_{max}. The angle θ_{max} is defined at the beginning of the screening by the user. The vector $\vec{r}_{cc}$ is the direction from one carbon that is bonded to the first candidate to another carbon that is bonded to the second candidate. Again, in case of more than two atoms, the vector $\vec{r}_{cc}$ runs from the carbon atom from a candidate to the geometric center of all candidates. The vector $\vec{hc}$ is the direction of the carbon-candidate bond (see figure 4.11).

With the positions of the carbon atoms as a_C and the candidates positions as a_H the vectors are then

$$\vec{r}_{cc} = \vec{a}_{C,j} - \vec{a}_{C,i} \tag{4.8}$$

and

$$\vec{r}_{hc} = \vec{a}_{H,i} - \vec{a}_{C,i} \quad . \tag{4.9}$$

The angle φ between these two vectors can be calculated by

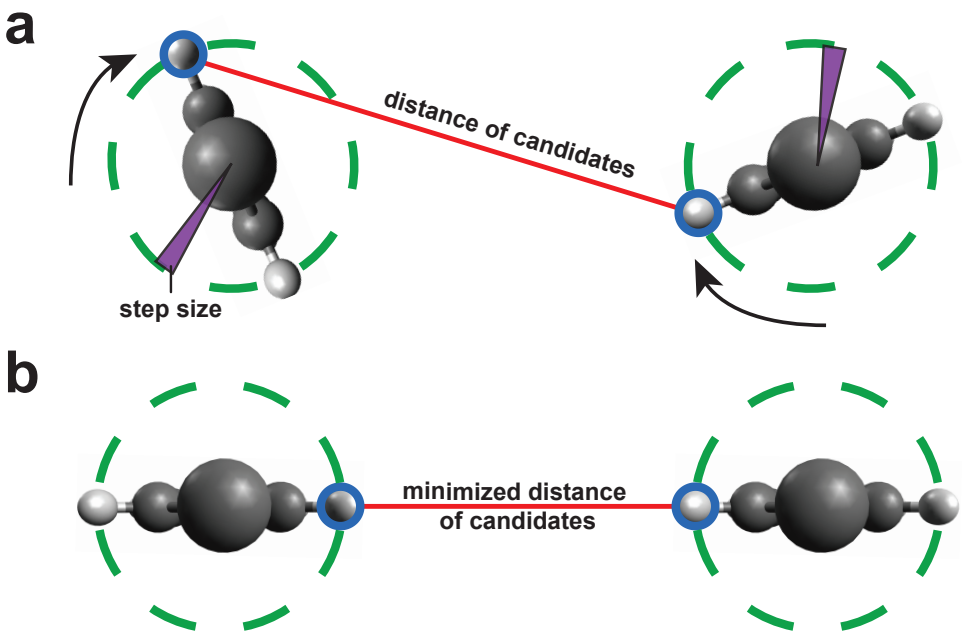


Figure 4.10.: Top down view of two linkers (**a**) before and (**b**) after rotation. The green dashed line shows the trajectory of the rotation and the blue bordered hydrogen atoms represent the candidates. The area highlighted in purple shows the rotation step. (**a**): Random position of linkers where the distance between candidates is not necessary minimal. (**b**): Minimized distance of candidates after rotation. Note that in this example the axes are parallel, but this is generally not always the case. Color code: H (white), C (black).

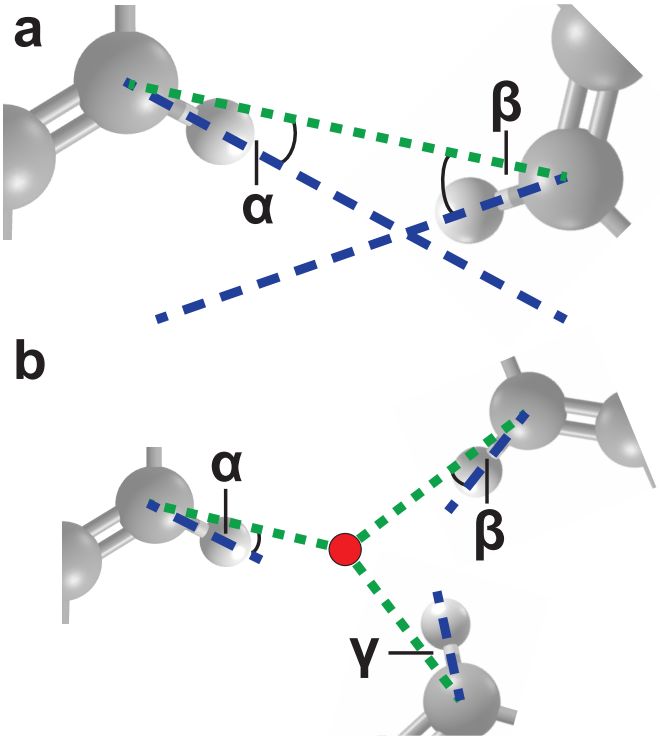


Figure 4.11.: Definition of the angle for (**a**) two and (**b**) three candidates. The geometric center in (**b**) is represented by the red point. Color code: H (white), C (black).

$$\varphi_i = \cos^{-1}\left(\frac{\vec{r}_{cc,i} \cdot \vec{r}_{hc,i}}{|\vec{r}_{cc,i}| \cdot |\vec{r}_{hc,i}|}\right) \quad . \tag{4.10}$$

In three-dimensional space, the angle can be considered as the opening angle of a cone. A sharp cone has a small opening angle and a blunt cone a large opening angle. The equation 4.10 can only result in a maximum angle of 180 degrees ($\{\varphi \in \mathbb{R} \mid 0 \leq \varphi \leq \pi\}$). However, that is no problem because the angle limit is likely to be smaller. .

4.3.2.4. Distance filter and screening results

The final filter compares the (minimized) distances of the candidates against the cross-linker length(s). Here comes the tolerance of the candidate distances into play. Firstly, this is necessary to obtain results at all, as an exact comparison is very unlikely to find matches, and secondly, the limits can be used to consider the flexibility of the cross-linker. The first point has already been explained above and is based on data types. The latter is motivated by the fact that the limits define a limit range that is accepted by the filter and, thus, considers the flexibility of a cross-linker. The broader the range, the more flexible the cross-linker. In the context of this work a rather small area was selected, since the goal was to find an ideal fit.

The limits are divided into two parts, the upper and the lower limit. Both are given as multiples of the cross-linker length r_{cl}. Choosing a limit of $X\%$, gives an upper limit l_{up} of

$$l_{up} = (1 + {}^{X}/_{100}) \cdot r_{cl} \tag{4.11}$$

and a lower limit of

$$l_{low} = (1 - {}^{X}/_{100}) \cdot r_{cl} \quad . \tag{4.12}$$

Thus, the distance filter checks if

$$l_{low} \leq r_{cl} \leq l_{up} \tag{4.13}$$

is true or not. If not, the algorithm jumps to the next set of candidates, and if true, the result is saved as:

HIT: Site **A** and Site **B** of MOF **STRUCTURE_FILE.CIF**
allows cross-linking length of **CL-LENGTH**
and rotation steps of **X** and **Y**!

In case of cross-linkers with more than two binding sites, the result output is changed accordingly. Figure 4.12 illustrates the distance filter. A cross-linker length (shown in orange) defines a limit range after applying the upper and lower limit. Every MOF candidate distance within this range generates a hit and is saved in the log file and all other candidates are dismissed. A wider limit range allows bigger length deviations and, thus, indicates a flexible cross-linker. Choosing a narrow limit range is advised for rigid cross-linkers.

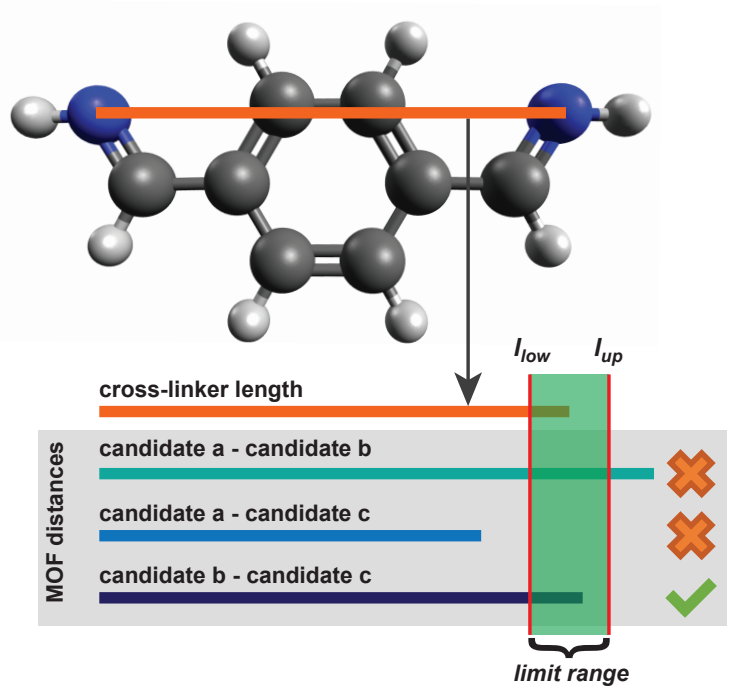


Figure 4.12.: Illustration of the distance filter. The orange line is the length of the cross-linker. Applying the upper and lower limits on this length creates a limit range. Every MOF distance that is within this range generates a HIT, larger and smaller candidate distances are dismissed. A narrow limit range indicates an inflexible cross-linker, a wider limit range a flexbile one. Color code: H (white), C (black), N (blue).

4.3.2.5. Additional filters

In order to reduce as much computing time as possible, candidates that obviously result in a mismatch, are sorted out before the first rotation step. The first filter removes all combinations of candidates that are part of the same organic linker. In this case the algorithm verifies if two or more candidate atoms share the same rotation axis.

The second filter is applied on cross-linkers with more than two binding sites. This filter checks, if the cross-linker has a symmetric structure, as all cross-linkers covered in this paper with more than two binding sites are sufficiently symmetric. For this, the arm-lengths are compared whether they are about equal.

The last filter was little used, which excludes too large and too small distances from the outset. For example, if the distances of the MOF candidates is not expected to get near the cross-linker lengths even after rotation, further calculations would be meaningless.

4.3.2.6. Computational remarks

Loading, analyzing and performing all sorts of calculations multiple times requires tremendous computing power. In this case, two factors contribute to a higher calculation time. On one hand, the number of structures to be examined, on the other hand the number of candidates per structure that contribute significantly to the calculation time. When dealing with n candidates, the algorithm has to calculate

$$\binom{n}{k} = \frac{n!}{(n-k)! \cdot k!} \quad , \quad \text{with } n \geq k \tag{4.14}$$

different (unique) combinations, where k is the number of binding sites. Since each candidate must be compared with every other candidate, the number of combinations logically increases with the number of binding sites[3]. If a cross-linker with four binding sites ($k = 4$) screens a single structure from the database that has $n = 300$ candidates (which is not uncommon), then the computing time consumed is about one hour, if one combination requires 1μs to execute. Screening thousands of structure files and using multiple cross-linkers as reference would lead to absurdly high computing times. Although the computing time of 1μs per combination was chosen arbitrarily, it still illustrates the urgency of finding a time-saving solution.

One possibility is to distribute the work to several processor cores. Thus, several calculations can be carried out simultaneously (parallelizing the algorithm). In comparison to this, in serial calculation with only one core every combination of the structure is screened one after the other. A full parallelization would take the workload of both cases, namely the number of structures and the number of combinations per structure, into account and outsource it to several cores. This work, however, only distributes the MOF structures to multiple processor cores, as it could be implemented rather easily. To use this screening algorithm on a supercomputer, it was MPI parallelized (**M**essage **P**assing **I**nterface).

4.4. Automatized cross-linker placement

When a HIT is found, one is interested in the final structure after functionalization. A smaller tool was written to directly compare a custom database of MOFs against a custom database of cross-linkers. The original purpose of this tool was to find matching MOF - cross-linker partners by trying every possible combination. Because the effort to create an own database was too high, I switched to the CoRe MOF database and the just presented screening method. However, a usage of this tool as front end for visualization purposes is quite conceivable, since the algorithms are very similar to each other. The only major difference is that instead of rotating the organic linker of a MOF the cross-linker is rotated and placed into the MOF.

The working steps are:

1. load MOF and cross-linker data from (custom) database
2. compare distances
3. if no match, jump to next structure
4. if match, attach cross-linker to MOF
5. output simulation input file for structure optimization .

The data of the MOF and cross-linker structure was stored in python dictionaries, which were prepared beforehand. The format of the MOF data was

[3] $\mathcal{O}(n^k)$, run time depends on total number of candidates and total number of binding sites.

```
# MOF DATA FILE

{
    # MOF ID
    "1": {
        "id": 0,
        # name of MOF
        "name": "cu_pw-l1",
        # file name of MOF simulation input file
        "file_name": "cu_pw-l1.gin",
        # number of atoms
        "n_atoms": 248,
        # number of binding sites
        "type": 4,
        # id, coordinates of binding sites
        "connectors": {
            "1": [ 10, 4.07114821, 0.00009331, -1.23626660 ],
            "2": [ 18, -0.00009330, 4.07114821, -1.23626660 ],
            "3": [ 72, 4.07114821, 9.52309331, -1.23626660 ],
            "4": [ 142, 9.52290670, 4.07114821, -1.23626660 ]
        }
    }
}
```

and the cross-linker structure was stored as

```
# CROSS-LINKER DATA FILE

{
    # id of cross-linker
    "1": {
        "id": 0,
        # name of cross-linker
        "name": "CL1",
        # number of binding sites
        "type": 2,
        # coordiantes of the center
        "center": [ 0.4888, -2.0986, -0.6162 ],
        # empiric value of flexibility
        # (similar to limit range)
        "flexibility": 1.2,
        # coordinates of binding sites
        "connectors": {
            "1": [ 27, 3.3891, -5.7574, 4.0577 ],
            "2": [ 23, -2.5633, 2.4490, 2.2202 ]
        },
        # id, atom type, force-field term, coordinates
        # of all atoms
        "atoms": {
            "1": [ 1, "C", "C_3", 0.4888, -2.0986, -0.6162 ],
            "2": [ 2, "C", "C_3", -0.4557, -2.6724, -1.693 ],
            ...
            "34": [ 34, "H", "H_", 3.6184, -2.5462, 1.8692 ],
```

```
28              "35": [ 35, "H", "H_", 3.2714, -3.5882, 0.4832 ]
29          },
30          # existing bonds
31          "bonds": {
32              "1": [ 2, 1, "" ],
33              "2": [ 2, 4, "" ],
34              ...
35              "35": [ 19, 34, "" ],
36              "36": [ 19, 35, "" ]
37          }
38      }
```

in .JSON (allows storage of python dictionaries) files. The comments (starting with a '#') explain the data. With this, the structure information could be easily read.

Comparing the MOF and the cross-linker was done in the same fashion as above, where distances (arm lengths) were compared. In case of a match, the cross-linker was attached to the MOF. First the (geometric) center of the cross-linker was shifted to the origin of the coordinate system to reset the position. Then it was rotated in almost the same way as described above in section 4.3.2.2. The only difference is the choice of the rotation axis. In case of a cross-linker with two binding sites, only one rotation has to be performed to adjust the position. The rotation axis is calculated by

$$\vec{r}_{rot,1} = \vec{r}_{MOF} \times \vec{r}_{cl} \quad , \tag{4.15}$$

with $\vec{r}_{MOF}$ as the direction from one binding site to the other and $\vec{r}_{cl}$ as the direction from the geometric center of the cross-linker to one binding site[4]. When dealing with cross-linkers with more than two binding sites, two rotation steps have to be performed. Figure 4.13 shows the two types of rotation.

The first rotation axis is similar to 4.15

$$\vec{r}_{rot,2} = \vec{r}_{MOF} \times \vec{r}_{arm} \quad , \tag{4.16}$$

with $\vec{r}_{MOF}$ the same as in equation 4.15 and $\vec{r}_{arm}$ again the direction from the geometric center to binding site (only when symmetric). The second rotation is motivated by the fact, that the planar cross-linker may be not in the same plane as the candidates of the MOF. To correct this, the surface normal vectors of both structures have to be parallel. The second rotation axis is then

$$\vec{r}_{rot,3} = \vec{r}_{MOF,n} \times \vec{r}_{CL,n} \quad , \tag{4.17}$$

where $\vec{r}_{MOF,n}$ is

$$\vec{r}_{MOF,n} = \vec{r}_{MOF,1} \times \vec{r}_{MOF,2} \tag{4.18}$$

and $\vec{r}_{CL,n}$ is

[4]All processed cross-linkers were symmetric around the geometric center.

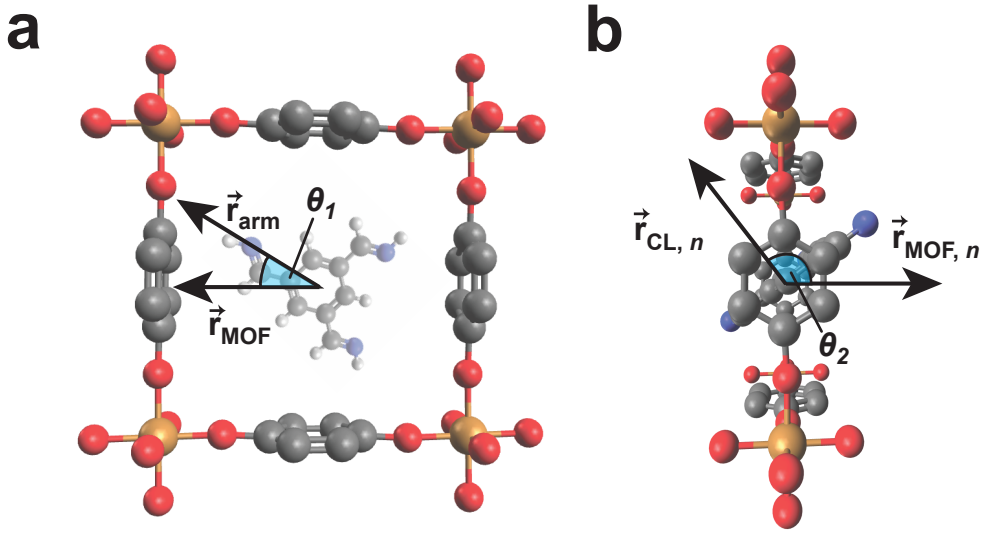


Figure 4.13.: Rotation axes of cross-linker and MOF. The geometric center of the cross-linker is placed on the geometric center of the MOF. Then, two rotations are performed to align the cross-linker. (**a**): The direction from geometric center to cross-linker binding site $\vec{r}_{cl}$ and geometric center to MOF binding site $\vec{r}_{MOF}$ creates the rotation axis $\vec{r}_{rot,1}$ via cross product. The rotation angle θ_1 is the angle between the vectors $\vec{r}_{cl}$ and $\vec{r}_{MOF}$. (**b**): The surface normal vector of the cross-linker plane $\vec{r}_{CL,n}$ and the MOF frame plane $\vec{r}_{MOF,n}$ are used to calculate the rotation axis via cross product. The angle θ_2 is again the angle between $\vec{r}_{CL,n}$ and $\vec{r}_{MOF,n}$. Color code: H (white), C (black), N (blue), O (red), Cu (gold). Hydrogen partially omitted for visibility.

$$\vec{r}_{CL,n} = \vec{r}_{arm,1} \times \vec{r}_{arm,2} \quad . \tag{4.19}$$

$\vec{r}_{MOF,1}$ and $\vec{r}_{MOF,2}$ are two different directions of MOF candidates ($1 \rightarrow 2$, $1 \rightarrow 3$). $\vec{r}_{arm,1}$ and $\vec{r}_{arm,2}$ are the directions from the geometric center to different binding sites. In all cases, the angle θ that is needed for the rotation is calculated by equation 4.10.

After rotation, the cross-linker position is inserted into the simulation input file including the atom types and the bonds. Finally, the structure is optimized and can be visualized.

In figure 4.14 one frame of the screened MOF is shown with all relevant distances included. The binding sites are the aromatic carbon atoms, with one binding site per side of the frame, resulting in a total of four binding sites. The cross-linkers used for the test run are shown in figure 4.15. CL1 possesses two binding sites, while CL2 possesses three.

Figure 4.16 shows a found hit where the cross-linker CL2 fits in the MOF. In (**a**) the binding sites can be nicely seen opposing each other. In (**b**) the cross-linker is now attached to the MOF and the structure is optimized. Overall little atom displacement can be noticed and a matching fit can be concluded.

The algorithm does not attach a cross-linker on a MOF, that does not fit. To show a mismatch, an unsuitable cross-linker was forcefully attached to the MOF. In figure

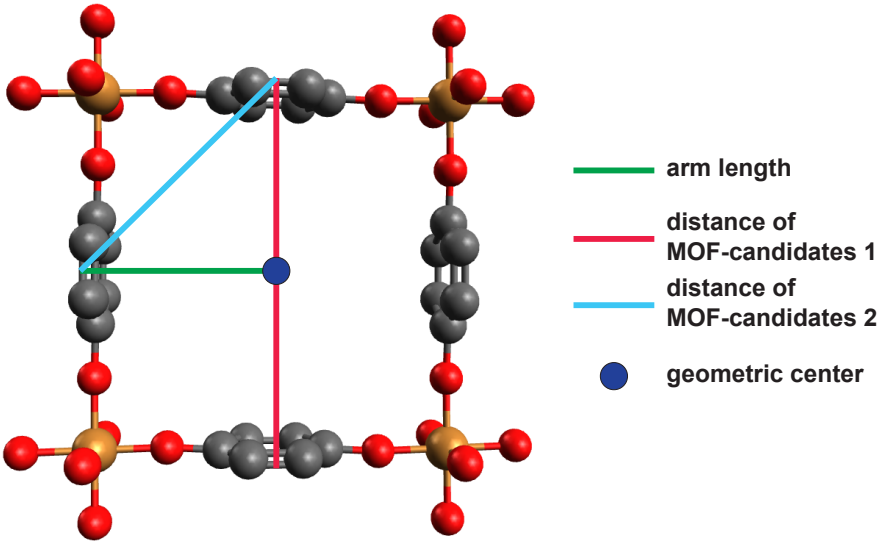


Figure 4.14.: A frame of a MOF from the custom database. The distances are relevant for the calculation of the rotation axis. Here, four binding sites are classified, that are aromatic carbon atoms. Each side has one binding site. Color code: C (black), N (blue), O (red), Cu (gold). Hydrogen omitted for visibility.

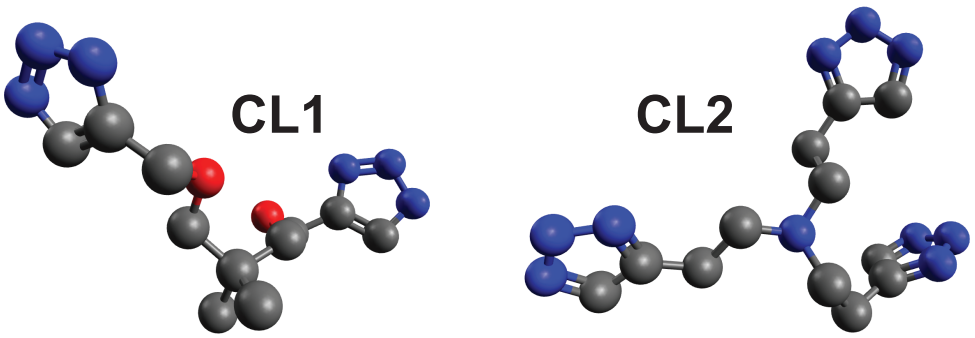


Figure 4.15.: Two different cross-linkers of the custom database. These two cross-linkers were compared against the MOF database for testing purposes. Color code: C (black), N (blue), O (red). Hydrogen omitted for visibility.

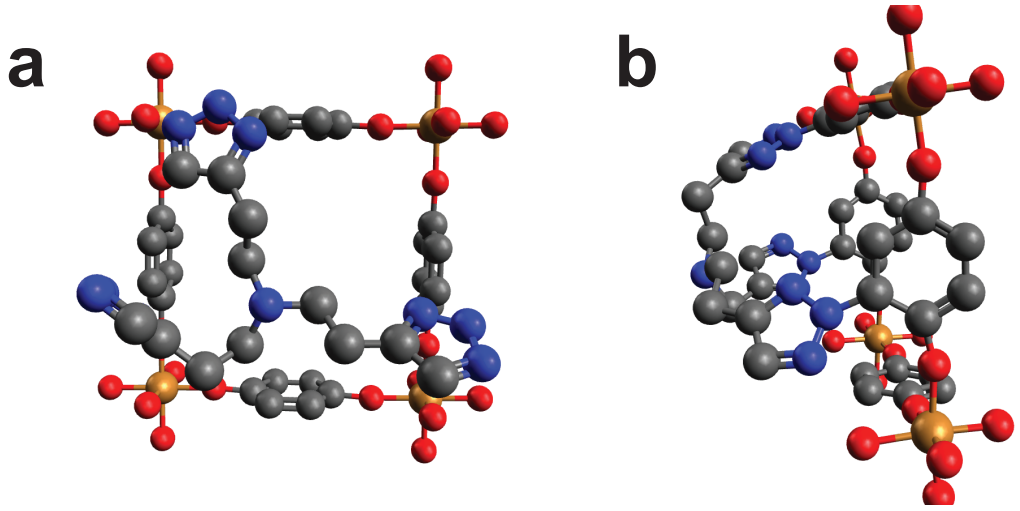


Figure 4.16.: Matching of cross-linker CL2 and the MOF frame. (**a**): Front view after rotating the cross-linker and before the structure optimization. The binding sites nicely oppose each other. (**b**): The cross-linker is now attached to the MOF and the structure is optimized. The overall structure is hardly deformed, indicating a positive match. Color code: H (white), C (black), N (blue), O (red), Cu (gold). Hydrogen omitted for visibility.

4.17 the cross-linker CL1 was attached to the MOF, despite being a mismatch. In **a** the binding sites are displaced. They would oppose each other as in figure 4.16, if the lengths would match. Connecting the binding sites and minimizing structure results in the expected deformation of the frame. The larger length of the cross-linker pulls the MOF apart. This kind of additional torsion underlines the mismatch.

All in all, this algorithm offers a faster and better screening method than the one described above. It offers to handle custom tailored databases with all needed parameters that can be added freely. However, the great effort that has to be made to create the database is in stark contrast to the better algorithm. In the beginning, only 24 different MOFs (four metal nodes and six organic linkers) were part of the database and in total six structures were part of the cross-linker database. Maintaining two different databases and adding further structures over a long period of time may be appropriate for this screening method, but is in no way feasible for a high-throughput screening with thousands of structures.

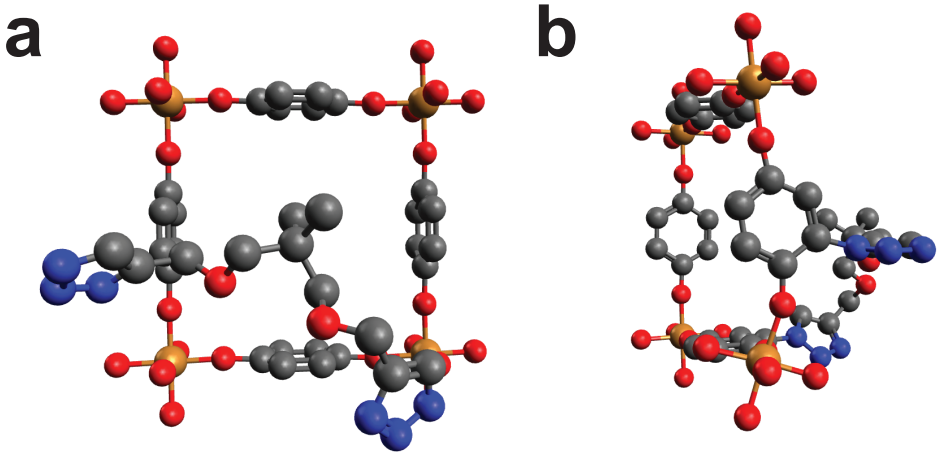


Figure 4.17.: Mismatch of cross-linker CL1 and the MOF frame. (**a**): Front view after rotating the cross-linker and before the structure optimization. Since the length of the cross-linker projects over the frame, the binding sites are not able to oppose each other. (**b**): After connecting the cross-linker to the frame and optimizing the structure, larger atom displacements can be observed. They result in a bigger deformation and indicate tension within the structure. Color code: C (black), N (blue), O (red), Cu (gold). Hydrogen omitted for visibility.

5. Results & Discussion

The high-throughput screening was performed on a local machine (12 × 3.20 GHz, Intel i7-8700, 32GB RAM) for linkers with two binding sites and on the supercomputer JUWELS (**Ju**elich **W**izard for **E**uropean **L**eadership **S**cience) for parallel code execution ($48 \times 2.70\,\mathrm{GHz} \cdot n$, n = number of computing nodes, Dual Intel Xeon Platinum 8168). Before screening the whole database, all structures without candidates were sorted out. This was achieved by only using the first part of the screening where candidates are searched. Whenever a structure was found that had any candidates, the respective structure file was copied to another directory. The method of finding candidates is described in section 4.3.2.1. This step reduced the number of MOF structure files from initially 12 724 to 2 795 structures. Therefore, the screening was only performed on the newly generated dataset to reduce computing time.

The screening on the local machine took roughly ten hours to complete. On the supercomputer the screening was distributed over 20 nodes. This makes a total of 960 cores. This leads to about three structure files per core for the screening. The maximum computing time allowed by the supercomputer facility was 24 hours, which were fully used. However, there were a few files that took longer than 24 hours and were therefore terminated prematurely. But the remaining data collected was sufficient for a successful evaluation.

A general distance of 1.55 Å was chosen for the bond length to find neighboring atoms. The upper and lower limit was set to 15 % of the cross-linker length. The maximum allowed angle was 20°. The step size was chosen to $N = 100$, which resulted in an angular increment of 1.8° per step. Though, a smaller step size of $N = 1000$ would be more appropriate, it was dismissed due to significantly increased computing time. This poses a problem on a supercomputer with finite resources.

5.1. Proof of concept

First of all, it had to be verified whether the algorithm delivered the desired results. This can be done by using a cross-linked reference structure that was already realized experimentally. A screening with the parameters of the cross-linker should find the unmodified MOF in the database. In figure 5.1 an example of such a MOF (SURMOF) is shown[78].

To use this structure as a benchmark, first the SURMOF and the cross-linker have to be created separately as structure files. The SURMOF frame was created with AuToGraFS[79] and the cross-linker with Avogadro[80]. The cross-linker is again in its already reacted form. Figure 5.2 shows the structure of the SURMOF frame in **a**

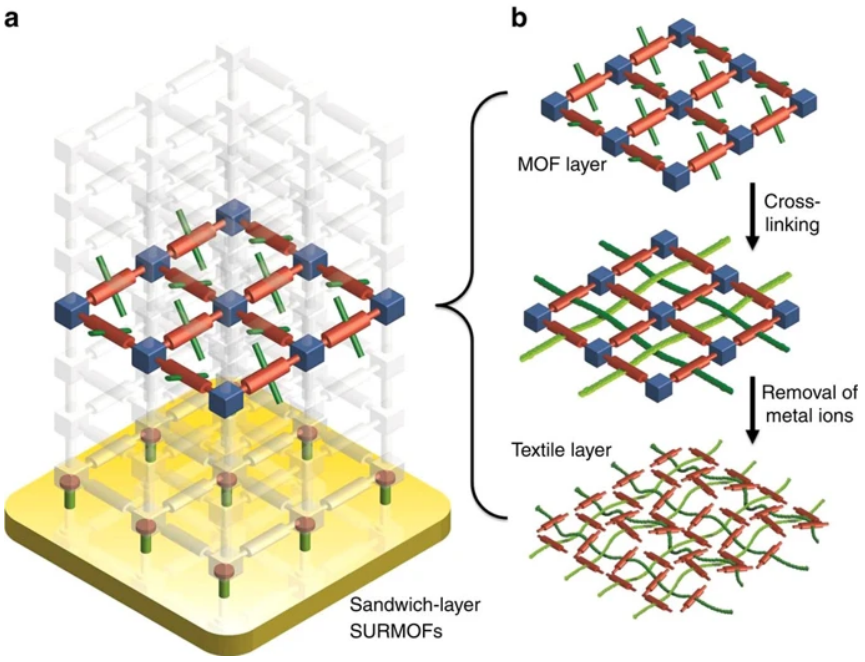


Figure 5.1.: (a): Structure of the SURMOF. (b): Process of cross-linking. Taken from [78].

and the structure of the cross-linker in (**b**). In (**c**) the final form after cross-linking (and before removing the metals) can be seen. It should be noted that two cross-linker are used for one (two dimensional) unit cell of the SURMOF. They connect the SURMOF in a crossed fashion as shown in figure 5.1.

The SURMOF frame was added to the database in the next step. A successful screening result shows a match of the cross-linker with the added SURMOF structure. But before the actual screening starts, the cross-linker must be defined. In figure 5.2 (**b**) the cross-linker is shown as a structure with two binding sites and a cross-linker length, thus, being a linear linker. The screening result should then show two hits while analyzing the added SURMOF frame, namely one for each linker in a single unit cell. In fact, there are 96 different possible hits due to symmetry (will be discussed below), as each benzene ring possess four eligible candidates. Two opposing benzene rings result in 16 different combinations after minimizing the candidate distance via linker rotation. One could now just screen the database with a linear linker, but that would result in collecting many unwanted hits where MOF structures similar to the SURMOF structure would get lost in the shuffle. The idea was now to view two crossed cross-linkers as one planar cross-linker with a geometric center and an arm length as length reference for the screening. As a bonus, the symmetry of the cross-linker structure is also included. A screening of the CoRE MOF database with the SURMOF frame structure including the parameters of the cross-linker was successful and confirmed the screening method.

5.2. Screening results

A series of different cross-linkers with different properties were used for the screening. Cross-linkers had two to four binding sites. All cross-linkers with more than two binding sites were symmetrical around their geometric center, therefore the lengths from the geometric center to one binding site were equal. All following figures

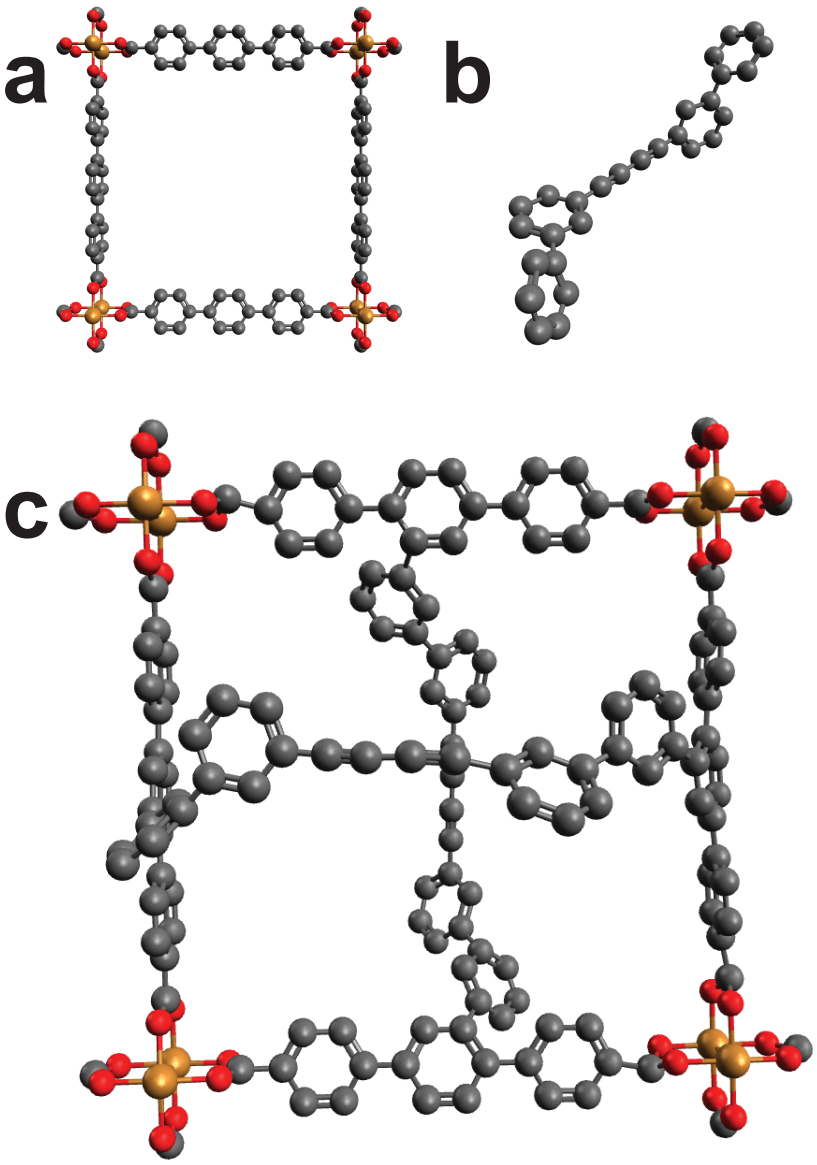


Figure 5.2.: (**a**): Frame of the SURMOF. (**b**): Cross-linker structure. (**c**): SURMOF after cross-linking. Color code: C (black), N (blue), O (red), Cu (gold). Hydrogen omitted for visibility. Reused from [75].

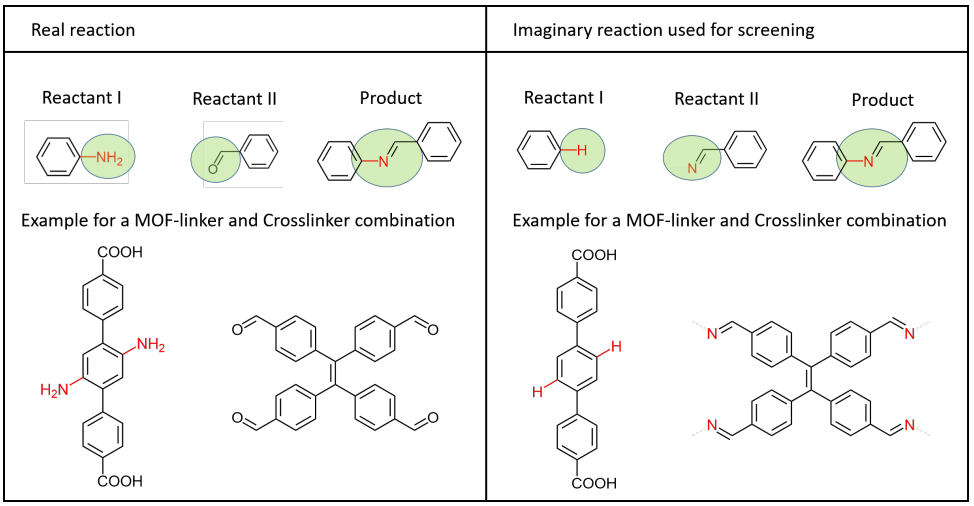


Figure 5.3.: Simplification of the linker and cross-linker binding sites for the screening. Taken from [75].

showing the results are divided into three or more parts. Part (**a**) of a result figure shows the unmodified MOF, part (**b**) shows the cross-linker and part (**c**) shows the position where the MOF is cross-linked. As already mentioned, all structures from the database have already been experimentally realized. The names of the MOFs can be used to look up their respective publication[81]. Furthermore, all coordinates of candidates and corresponding carbon atoms are defined as $\vec{h}_i$ for the candidates and $\vec{c}_i$ for the carbons. The index i is the number of the candidate. The number of coordinates for each candidate (and carbons) is equal to the number of binding sites of the cross-linker. All position values are rounded to two decimal, angle and deviation values are rounded to single decimal. The unit of the position vectors is in Å.

The reaction of the cross-linkers with the linkers is in most cases the same where an amino group reacts with an oxygen to form the bond and water as byproduct. Technically speaking, the nitrogen of the amino group connects to the binding partner of the oxygen in the cross-linker. Figure 5.3 shows the simplification of the cross-linkers.

Figure 5.4 shows the cross-linking of the MOF 'EKOPEA'[82]. The cross-linker interconnects two neighboring organic linkers. With this, a continuous modification can be undertaken towards higher length scales. The length of the cross-linker is $l_{CL1} = 2.51\,\text{Å}$ and the positions of the candidates are

$$\vec{h}_1 = \begin{pmatrix} 9.67 \\ 5.80 \\ 2.60 \end{pmatrix} \qquad \vec{c}_1 = \begin{pmatrix} 9.65 \\ 5.68 \\ 1.68 \end{pmatrix} \quad , \tag{5.1}$$

$$\vec{h}_2 = \begin{pmatrix} 9.57 \\ 5.26 \\ 5.20 \end{pmatrix} \qquad \vec{c}_2 = \begin{pmatrix} 9.59 \\ 5.36 \\ 6.13 \end{pmatrix} \quad . \tag{5.2}$$

The calculated distance is $l_{EKOPEA} = 2.66\,\text{Å}$ with a deviation of $\approx 6.0\,\%$. The angles

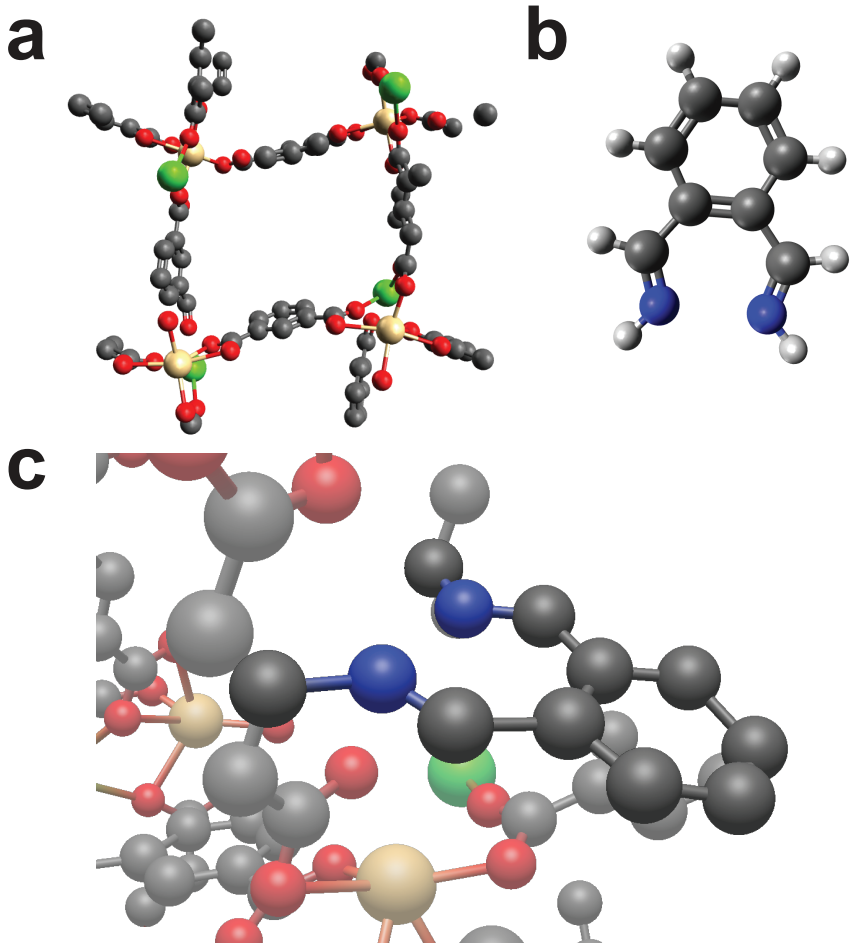


Figure 5.4.: (**a**): Unmodified MOF 'EKOPEA'. (**b**): Used cross-linker 'CL1'. **c:** Cross-linked structure with the cross-linker and the MOF binding sites being highlighted. Color code: H (white), C (black), N (blue), O (red), Ca (green), Cd (gold). Hydrogen omitted in (**a**) and in (**c**) for visibility.

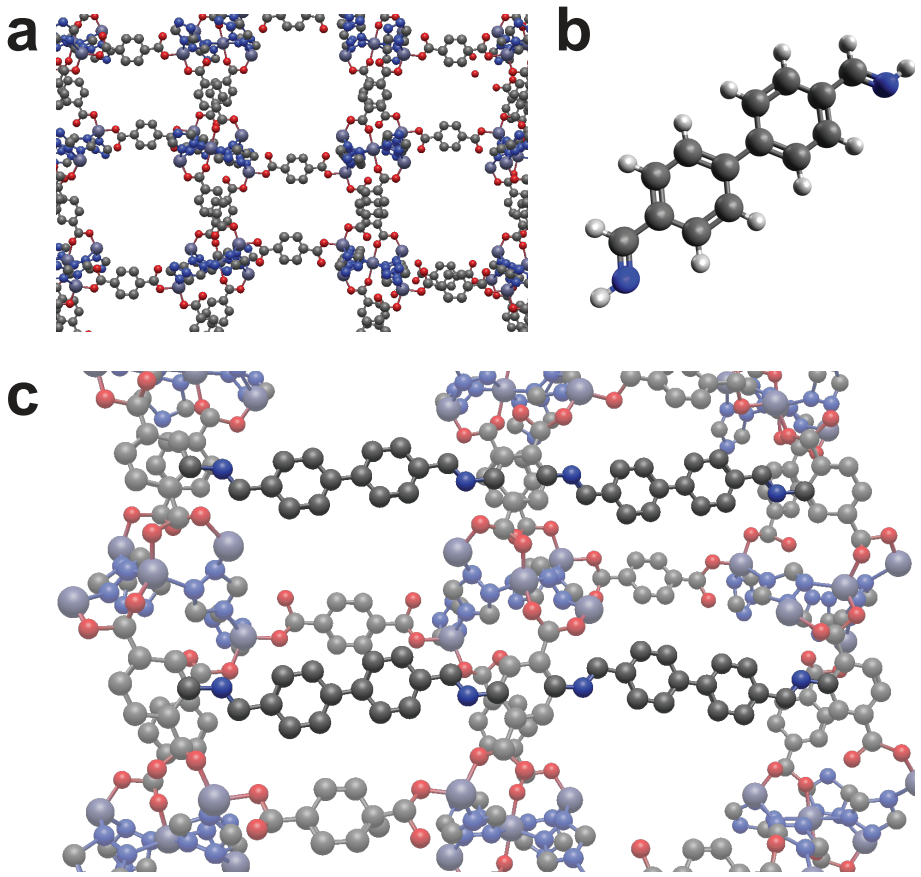


Figure 5.5.: (**a**): Unmodified MOF 'IBUBIT'. (**b**): Used cross-linker 'CL2'. (**c**): Cross-linked structure with the cross-linker and the MOF binding sites being highlighted. Color code: H (white), C (black), N (blue), O (red), Zn (purple). Hydrogen omitted in (**a**) and (**c**) for visibility.

are

$$\measuredangle h_1 c_1 c_2 = 2.9^\circ \qquad \text{and} \qquad \measuredangle h_2 c_2 c_1 = 2.7^\circ \quad .$$

The relatively small deviation of the distance and the very small angles indicate a very favorable position of the cross-linker CL1.

Another result for a linear linker is shown in figure 5.5 where the MOF 'IBUBIT'[83] is cross-linked. Here the length of the cross-linker is measured to $l_{CL2} = 11.54\,\text{Å}$. The cross-linker nicely fits into the gap between two organic linker. It can also be seen here that the cross-linking does not disturb the cross-linking at another place. Therefore, this result is an interesting candidate for a highly-ordered COF structure.

The positions in this case are

$$\vec{h}_1 = \begin{pmatrix} 0 \\ 2.35 \\ 5.83 \end{pmatrix} \qquad \vec{c}_1 = \begin{pmatrix} 0 \\ 1.42 \\ 5.80 \end{pmatrix} \quad ,$$
$$\vec{h}_2 = \begin{pmatrix} 0 \\ 15.05 \\ 5.83 \end{pmatrix} \qquad \vec{c}_2 = \begin{pmatrix} 0 \\ 15.98 \\ 5.80 \end{pmatrix} \quad .$$

The calculated length is $l_{IBUBIT} = 12.70\,\text{Å}$ and the deviation $\approx 10.0\,\%$. The angles amount to

$$\measuredangle h_1 c_1 c_2 = \measuredangle h_2 c_2 c_1 = 0.1^\circ \quad .$$

The calculated distance of the candidates is near the upper limit. However, an angle is almost nonexistent. Therefore, the cross-linker does not have to bend in any directions and can remain in its original state. These values underline once again the matching of the cross-linker and the MOF structure.

The next three MOFs 'RUYVEO'[84] (figure 5.6), 'RUYVIS'[84] (figure 5.7) and 'IXEJOM'[85] (figure 5.8) have all a ring-shaped structure in which a cross-linker with three binding sites fits in nicely. The distances are $r_{CL3} = 6.53\,\text{Å}$, $r_{CL4} = 8.98\,\text{Å}$ and $r_{CL5} = 7.34\,\text{Å}$. The interesting thing about 'RUYVEO' and 'RUYVIS' is, that their pore is formed by three organic linkers. The cross-linker binds to all three linkers so there are no loose ends. 'IXEJOM' is different as the cross-linker can only bind to three of the four linkers that build up the pore.

With more than two binding points we have to use the center of mass $vecr_{com}^{RUYVEO}$ again

$$\vec{r}_{com}^{RUYVEO} = \begin{pmatrix} 14.16 \\ 8.18 \\ 2.20 \end{pmatrix} \quad .$$

The other positions are

$$\vec{h}_1 = \begin{pmatrix} 18.12 \\ 14.30 \\ 2.20 \end{pmatrix} \qquad \vec{c}_1 = \begin{pmatrix} 18.62 \\ 15.17 \\ 2.17 \end{pmatrix}$$
$$\vec{h}_2 = \begin{pmatrix} 17.49 \\ 1.69 \\ 2.20 \end{pmatrix} \qquad \vec{c}_2 = \begin{pmatrix} 17.99 \\ 0.82 \\ 2.17 \end{pmatrix}$$
$$\vec{h}_3 = \begin{pmatrix} 6.88 \\ 8.54 \\ 2.20 \end{pmatrix} \qquad \vec{c}_3 = \begin{pmatrix} 5.88 \\ 8.54 \\ 2.17 \end{pmatrix} \quad .$$

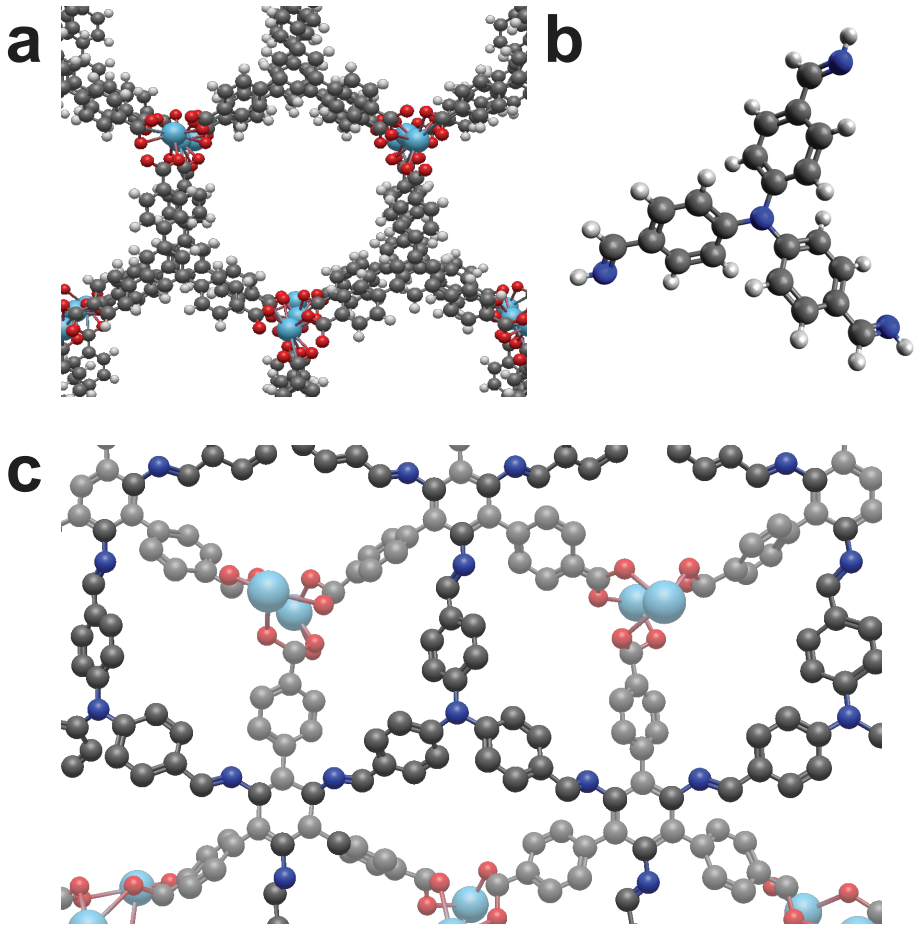


Figure 5.6.: (**a**): Unmodified MOF 'RUYVEO'. (**b**): Used cross-linker 'CL3'. (**c**): Cross-linked structure with the cross-linker and the MOF binding sites being highlighted. Color code: H (white), C (black), N (blue), O (red), La (light blue). Hydrogen omitted in (**c**) for visibility.

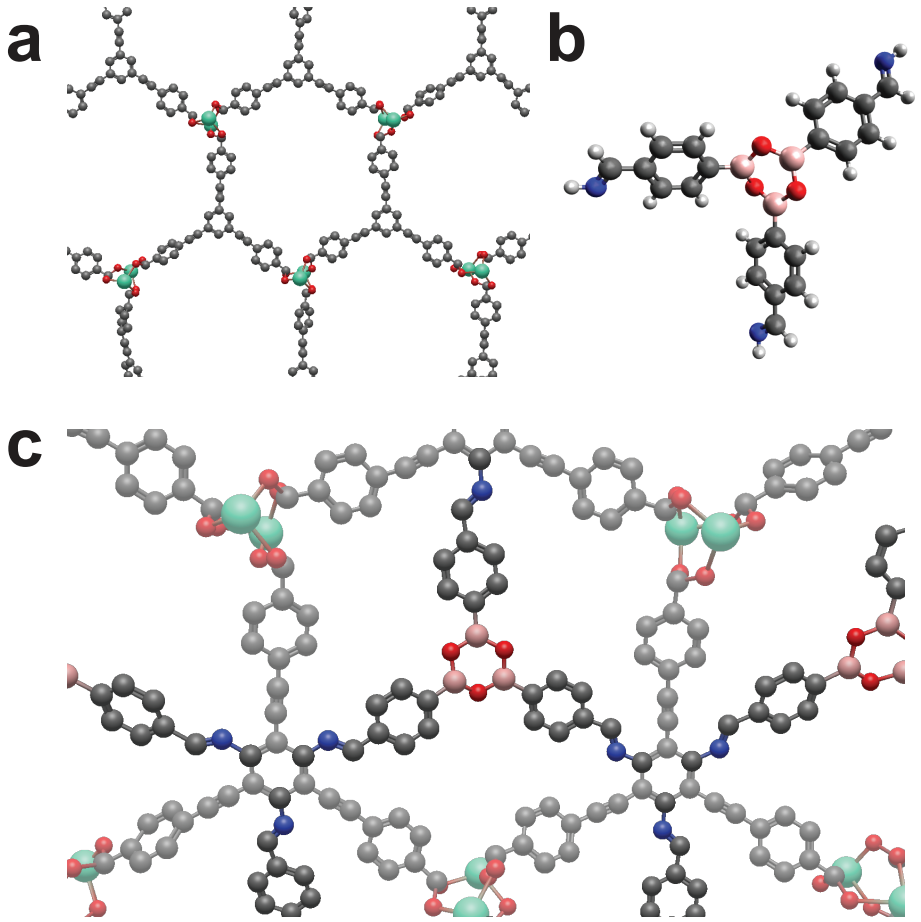


Figure 5.7.: (**a**): Unmodified MOF 'RUYVIS'. (**b**): Used cross-linker 'CL4'. Note that the cross-linker is not planar. (**c**): Cross-linked structure with the cross-linker and the MOF binding sites being highlighted. Color code: H (white), C (black), N (blue), O (red), B (pink), Eu (green). Hydrogen omitted in (**a**) and (**c**) for visibility.

The distance from each candidate to the center of mass is $l_{RUYVEO} = 7.29\,\text{Å}$ with a deviation of $\approx 11.6\,\%$. The angles result to

$$\measuredangle h_1 c_1 r_{com}^{RUYVEO} = \measuredangle h_2 c_2 r_{com}^{RUYVEO} = \measuredangle h_3 c_3 r_{com}^{RUYVEO} = 2.8° \quad .$$

Although the distance is a little larger than the cross-linker length, it still fits properly into the pore. Each angle is about equal to each other and is very small. Thus, a strain due to bending should not occur.

For 'RUYVIS' the candidate positions are

$$\vec{h}_1 = \begin{pmatrix} 23.05 \\ 18.29 \\ 2.20 \end{pmatrix} \qquad \vec{c}_1 = \begin{pmatrix} 23.53 \\ 19.156 \\ 2.22 \end{pmatrix}$$
$$\vec{h}_2 = \begin{pmatrix} 21.93 \\ 1.41 \\ 2.20 \end{pmatrix} \qquad \vec{c}_2 = \begin{pmatrix} 22.45 \\ 0.57 \\ 2.22 \end{pmatrix}$$
$$\vec{h}_3 = \begin{pmatrix} 7.879 \\ 10.819 \\ 2.20 \end{pmatrix} \qquad \vec{c}_3 = \begin{pmatrix} 6.89 \\ 10.80 \\ 2.22 \end{pmatrix} \quad .$$

The center of mass is

$$\vec{r}_{com}^{RUYVIS} = \begin{pmatrix} 17.62 \\ 10.17 \\ 2.20 \end{pmatrix} \quad .$$

The candidate to center of mass distance is in this case $l_{RUYVIS} = 9.76\,\text{Å}$ and a deviation of $\approx 8.7\,\%$. The angles are

$$\measuredangle h_1 c_1 r_{com}^{RUYVIS} = \measuredangle h_2 c_2 r_{com}^{RUYVIS} = \measuredangle h_3 c_3 r_{com}^{RUYVIS} = 4.7^\circ \quad .$$

It has to be mentioned that the force-field does not consider the cross-linker as a planar structure (can be seen figure 5.7). This case has to be reviewed in detail as four options arise. First option is that the cross-linker will keep its non-planar structure, thus, making a fit impossible. Second option is that the cross-linker allows to be bent into a planar form with a small strain on the connection to the MOF crystal. However, this will most likely result into deformation towards higher length scales. The third option covers the possibility that the cross-linker is perfectly fine with a planar form and will fit into the pore without any problem (desired case). And finally the last case questions the force-field itself. If the force-field is not suited to describe this kind of structures, errors naturally occur. Also, the calculation of the length turned out to be difficult due to the non planar structure of the cross-linker. The length was determined by the distance of a nitrogen from the center of mass of the carbon atoms that are bonded to boron atoms. With this, a length was approximated.

'IXEJOM' is a special case because the distances of the candidates to center of mass are not equal to each other. This is also due to the fact that the candidates together with their linkers are not in one single plane and a rotation of candidates is very likely. The position of the center of mass (after rotation!) is

$$\vec{r}_{com}^{IXEJOM} = \begin{pmatrix} 11.31 \\ 10.38 \\ 11.02 \end{pmatrix} \quad .$$

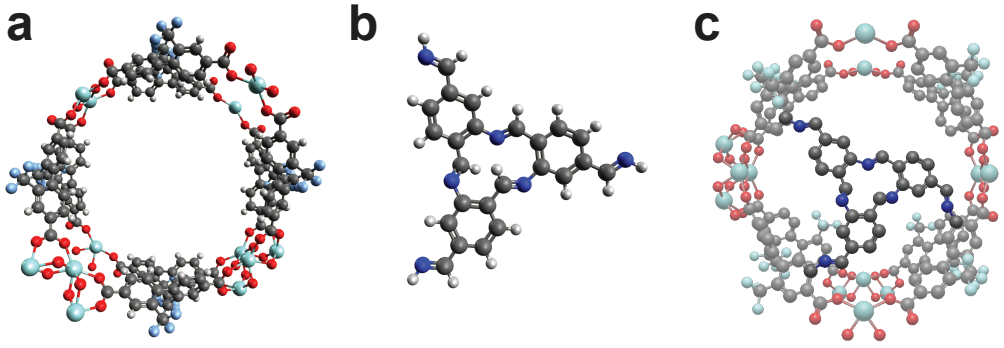


Figure 5.8.: (**a**): Unmodified MOF 'IXEJOM'. (**b**): Used cross-linker 'CL5'. (**c**): Cross-linked structure with the cross-linker and the MOF binding sites being highlighted. Color code: H (white), C (black), N (dark blue), O (red), F (blue), Zr (light blue). Hydrogen omitted in (**c**) for visibility.

The positions are

$$\vec{h}_1^* = \begin{pmatrix} 5.84 \\ 16.50 \\ 10.51 \end{pmatrix} \qquad \vec{c}_1 = \begin{pmatrix} 19.65 \\ 8.12 \\ 10.25 \end{pmatrix}$$
$$\vec{h}_2^* = \begin{pmatrix} 18.93 \\ 9.37 \\ 10.36 \end{pmatrix} \qquad \vec{c}_2 = \begin{pmatrix} 8.02 \\ 4.47 \\ 11.95 \end{pmatrix}$$
$$\vec{h}_3^* = \begin{pmatrix} 9.17 \\ 5.26 \\ 12.18 \end{pmatrix} \qquad \vec{c}_3 = \begin{pmatrix} 7.05 \\ 15.81 \\ 10.86 \end{pmatrix} \quad .$$

It has to be noted that these candidate positions are not the same position found in the original structure of 'IXEJOM'. These are the positions after inserting the cross-linker and optimizing the structures. Using the old positions results in incorrect values, since the rotation still has to be performed. That is why the hydrogen positions are marked as $\vec{h}_i^*$. The average distance from candidate to center of mass is $l_{IXEJOM} = 7.2\,\text{Å}$ with a deviation of $\approx 1.9\,\%$. The angles are

$$\measuredangle h_1^* c_1 r_{com}^{IXEJOM} = 16.6°, \quad \measuredangle h_2^* c_2 r_{com}^{IXEJOM} = 36.3°, \quad \measuredangle h_3^* c_3 r_{com}^{IXEJOM} = 27.3° \quad .$$

Here again the effect of the force-field is apparent. Even though the angle limit is 20 degrees and the results are greater than that, they are still correct. The screening algorithm finds a position for the candidates (if possible) where the distances as well as the angles are within the given limits. However, this position is not known or saved. The result only indicates that there exists a match between cross-linker and MOF by showing the candidates. The result after placing the cross-linker into the MOF and optimizing the structure is independent of the calculated values during the screening. The force-field positions every atom according to its defined potentials and parameters. There may be a solution where the cross-linker is placed in that way that the screening algorithm calculated, but it does not have to. If the optimal position

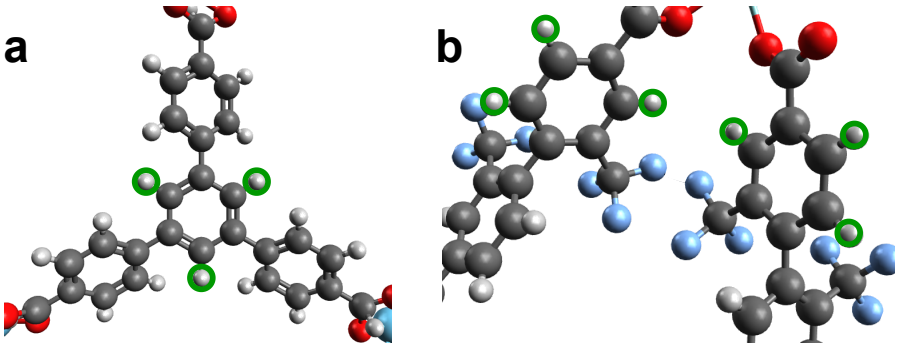


Figure 5.9.: Non-rotatable and rotatable linkers of ring-like structures.. The candidates are bordered with green color. (**a**): The linker of 'RUYVEO' (also similar to 'RUYVIS') yields three candidates. Although the linker is not rotatable, it is still accepted by the algorithm. (**b**): The linker of 'IXEJOM' also possesses a ring-like structure and candidates can be found here, but they are able to rotate. Color code: H (white), C (black), O (red), F (blue).

possesses distances and angles larger than the given limits during the screening step, the combination still might work and has to be checked in detail. This phenomenon is more likely to occur with increasing numbers of rotatable linkers (N binding points of cross-linker), since then there may be larger deviations from the original structure.

Another point is, that although 'RUYVEO' and 'RUYVIS' are recognized by the algorithm and provide candidates, they do not have the ability to rotate. Figure 5.9 shows the linker of 'RUYVEO' that is also similar to the linker of 'RUYVIS'. The selection criteria according to which the candidates are chosen are fulfilled. However, the rotation axis is incorrect, since there is none. But in the case of 'RUYVEO' and 'RUYVIS', this does not make any difference, as the candidates are already at a minimal distance to each other. 'IXEJOM' also possesses a ring-like structure but its linker are able to freely rotate. This special case is discussed in section 5.3.

The MOF 'YODWOF'[86] in figure 5.10 can be cross-linked by a linear linker. The length of the linker is $r_{CL6} = 7.32\,\text{Å}$. The calculated distance with the positions

$$\vec{h}_1 = \begin{pmatrix} 8.01 \\ 13.17 \\ 19.19 \end{pmatrix} \qquad \vec{c}_1 = \begin{pmatrix} 7.20 \\ 12.85 \\ 19.51 \end{pmatrix}$$
$$\vec{h}_2 = \begin{pmatrix} 13.17 \\ 13.17 \\ 24.36 \end{pmatrix} \qquad \vec{c}_2 = \begin{pmatrix} 12.85 \\ 12.85 \\ 25.17 \end{pmatrix} \quad .$$

results to $l_{YODWOF} = 7.31\,\text{Å}$ and a deviation of $\approx 0.2\,\%$. The angle

$$\measuredangle h_1 c_1 c_2^{YODWOF} = \measuredangle h_2 c_2 c_1^{YODWOF} = 68.0° \quad .$$

is rather interesting. It exceeds the angle limit of 20 degrees by a large amount. This can be explained again by the fact that the linker of the candidates is not rotatable

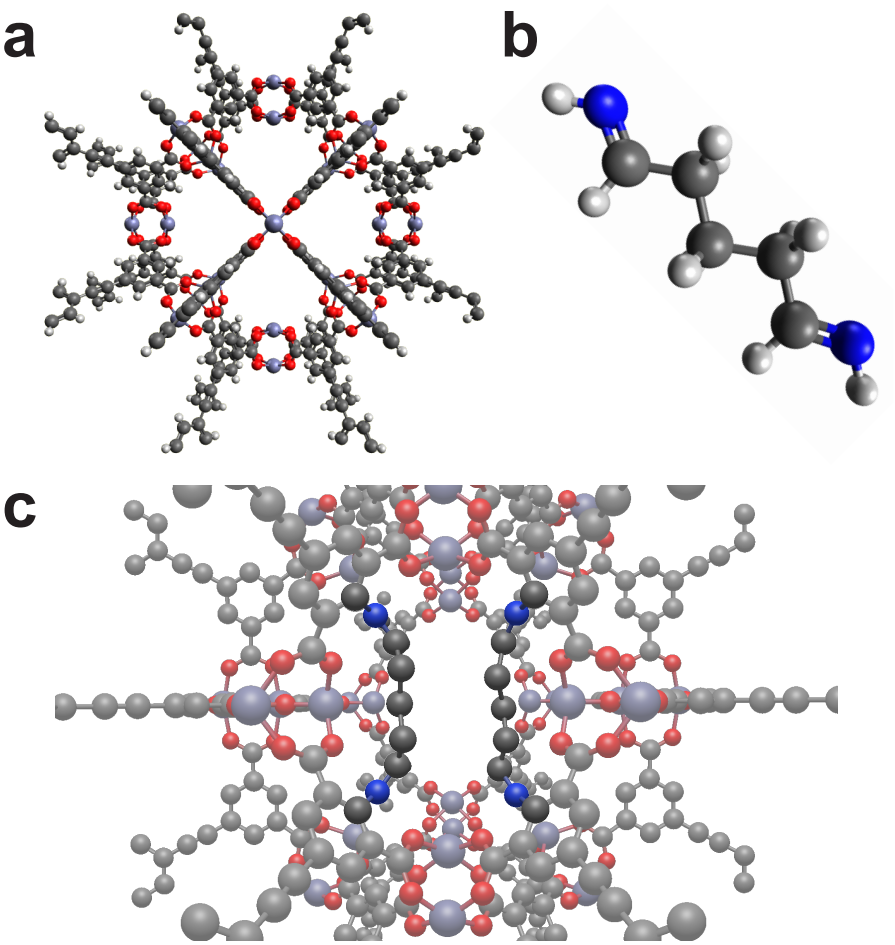


Figure 5.10.: (**a**): Unmodified MOF 'YODWOF'. (**b**): Used cross-linker 'CL6'. (**c**): Cross-linked structure with the cross-linker and the MOF binding sites being highlighted. Color code: H (white), C (black), N (blue), O (red), Zn (purple). Hydrogen omitted in (**c**) for visibility.

(similar to 'RUYVEO' and 'RUYVIS'). After rotating the candidates towards each other a much smaller angle is achieved (with a different distance than mentioned above). Thus, this structure was considered as a possible candidate by the screening algorithm and passed the filters. The large angle does not disturb the cross-linker at all and such results can still be included into the list of possible modified structures.

A cross-linker with four binding sites cross-links the structure 'QOWRAV12'[87] in figure 5.11. There are other similar MOFs, which are distinguished by a number after the name. The cross-linker has a length of $r_{CL7} = 6.91\,\text{Å}$. Horizontally viewed, the MOF consists of two halves connected by the metal. Here, the cross-linker binds these halves once again.

The center of mass is now calculated by four different candidates

$$\vec{r}_{com}^{QOWRAV12} = \begin{pmatrix} 13.49 \\ 6.37 \\ 13.49 \end{pmatrix} \quad .$$

The positions

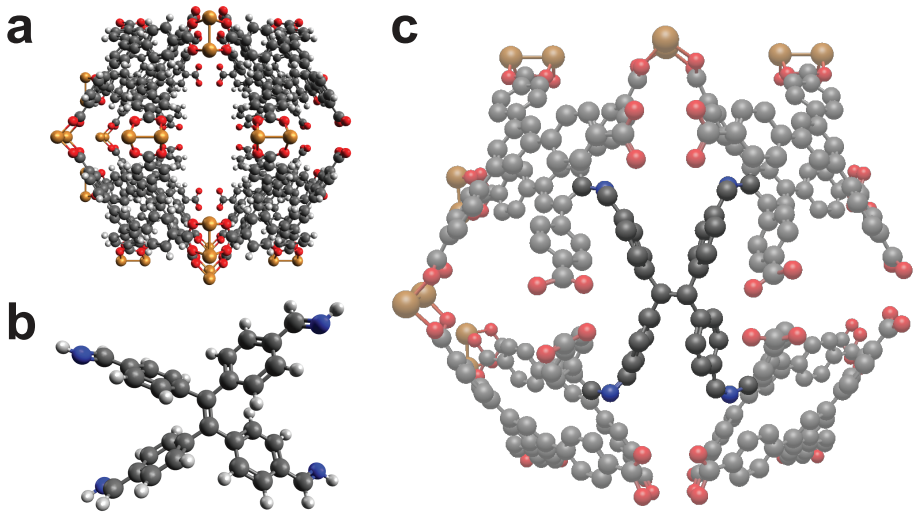


Figure 5.11.: (**a**) Unmodified MOF 'QOWRAV12'. (**b**) Used cross-linker 'CL7'. (**c**) Cross-linked structure with the cross-linker and the MOF binding sites being highlighted. Color code: H (white), C (black), N (blue), O (red), Cu (gold). Hydrogen omitted in (**c**) for visibility.

$$\vec{h}_1 = \begin{pmatrix} 17.36 \\ 6.37 \\ 7.52 \end{pmatrix} \qquad \vec{c}_1 = \begin{pmatrix} 18.09 \\ 6.96 \\ 7.64 \end{pmatrix}$$
$$\vec{h}_2 = \begin{pmatrix} 9.63 \\ 6.37 \\ 19.47 \end{pmatrix} \qquad \vec{c}_2 = \begin{pmatrix} 8.90 \\ 6.96 \\ 19.34 \end{pmatrix}$$
$$\vec{h}_3 = \begin{pmatrix} 9.63 \\ 6.37 \\ 7.52 \end{pmatrix} \qquad \vec{c}_3 = \begin{pmatrix} 8.90 \\ 6.96 \\ 7.64 \end{pmatrix}$$
$$\vec{h}_4 = \begin{pmatrix} 17.36 \\ 6.37 \\ 19.47 \end{pmatrix} \qquad \vec{c}_4 = \begin{pmatrix} 18.09 \\ 6.96 \\ 19.34 \end{pmatrix} \quad ,$$

result in a distance of $l_{QOWRAV12} = 7.12\,\text{Å}$ with a deviation of $\approx 3.0\,\%$. Here again the non-rotatable linker results in higher angles than the limits (see 'YODWOF'). The angles are

$$\begin{aligned} \measuredangle h_1 c_1 c_{com}^{QOWRAV12} &= \measuredangle h_2 c_2 c_{com}^{QOWRAV12} = \\ &= \measuredangle h_3 c_3 c_{com}^{QOWRAV12} = \measuredangle h_4 c_4 c_{com}^{QOWRAV12} = 65.20^\circ \quad . \end{aligned}$$

The next MOF in figure 5.12 is called 'UMODEH10'[88]. It has two places where the cross-linker with the length of $r_{CL8} = 3.73\,\text{Å}$ fits in. The first position is shown in (**c**) where it binds to the benzene rings parts of the linker. In addition, the unusual structure led to the examination of other places. Part (**d**) of the figure shows another position where this cross-linker nicely fits into. It should be noted, however, that

this position cannot be detected by the algorithm, since the hydrogen atoms here cannot be found by the candidate scan due to the missing ring structure.

Only the position in figure 5.12 part (**c**) will be discussed quantitatively, since it is the result of the screening. The center of mass results to

$$\vec{r}_{com}^{UMODEH10} = \begin{pmatrix} 5.59 \\ 9.37 \\ 7.47 \end{pmatrix} \quad .$$

The positions of the candidates are

$$\begin{aligned} \vec{h}_1^* &= \begin{pmatrix} 3.81 \\ 9.48 \\ 3.62 \end{pmatrix} & \vec{c}_1 &= \begin{pmatrix} 3.79 \\ 8.07 \\ 3.47 \end{pmatrix} \\ \vec{h}_2^* &= \begin{pmatrix} 7.87 \\ 11.67 \\ 9.08 \end{pmatrix} & \vec{c}_2 &= \begin{pmatrix} 8.99 \\ 10.93 \\ 8.63 \end{pmatrix} \\ \vec{h}_3^* &= \begin{pmatrix} 5.10 \\ 6.95 \\ 9.71 \end{pmatrix} & \vec{c}_3 &= \begin{pmatrix} 4.35 \\ 6.59 \\ 10.85 \end{pmatrix} \quad . \end{aligned}$$

The rotatable linker in this case again yields different lengths with different deviations. The positions of the candidates are presented after the optimization step, since the original non-rotated coordinates are not relevant. They range from $l_{UMODEH10} = (3.33 - 4.24)\,\text{Å}$ and a deviation of $\approx (3.0 - 13.0)\,\%$. The angles again are far beyond the limit range because of the optimization step where another position for the atoms is favored than the calculated position by the screening algorithm. The flexibility of the cross-linker still allows the cross-linking. The angles are

$$\begin{aligned} \measuredangle h_1^* c_1 c_{com}^{UMODEH10} &= 67.68^\circ \\ \measuredangle h_2^* c_2 c_{com}^{UMODEH10} &= 67.01^\circ \quad , \\ \measuredangle h_3^* c_3 c_{com}^{UMODEH10} &= 25.60^\circ \quad . \end{aligned}$$

Another weakness of the algorithm is shown in figure 5.13. The tetrahedral cross-linker with the length of $r_{CL9} = 6.63\,\text{Å}$ cross-links the MOF 'LAQNOJ'[89]. The problem here is, that the cross-linker is forced into a planar form, which is handled as unfavorable by the force-field. A comparison between the standalone cross-linker in (**b**) and the connected cross-linker in (**c**) clearly shows the difference. If such cases become more frequent (more data would have to be generated for the evaluation), then the algorithm should be further developed to dismiss such results.

The center of mass for 'LAQNOJ' is

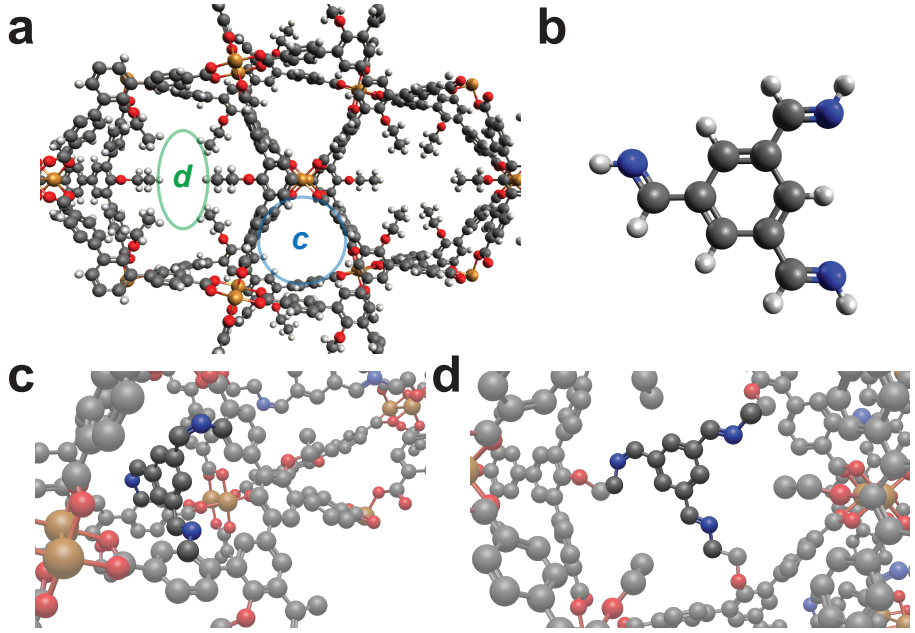


Figure 5.12.: (**a**): Unmodified MOF 'UMODEH10'. (**b**): Used cross-linker 'CL8'. (**c**): Cross-linked structure with the cross-linker and the MOF binding sites being highlighted. (**d**): Second part of the MOF with the cross-linker attached and highlighted. Color code: H (white), C (black), N (blue), O (red), Cu (gold). Hydrogen omitted in (**c**) and (**d**) for visibility.

$$\vec{r}_{com}^{LAQNOJ} = \begin{pmatrix} 9.29 \\ 9.29 \\ 20.52 \end{pmatrix} \quad .$$

With the positions

$$\vec{h}_1 = \begin{pmatrix} 9.29 \\ 16.30 \\ 20.52 \end{pmatrix} \qquad \vec{c}_1 = \begin{pmatrix} 9.29 \\ 17.22 \\ 20.52 \end{pmatrix} \quad ,$$

$$\vec{h}_2 = \begin{pmatrix} 9.29 \\ 2.30 \\ 20.52 \end{pmatrix} \qquad \vec{c}_2 = \begin{pmatrix} 9.29 \\ 1.37 \\ 20.52 \end{pmatrix} \quad ,$$

$$\vec{h}_3 = \begin{pmatrix} 2.30 \\ 9.29 \\ 20.52 \end{pmatrix} \qquad \vec{c}_3 = \begin{pmatrix} 1.37 \\ 9.29 \\ 20.52 \end{pmatrix} \quad ,$$

$$\vec{h}_4 = \begin{pmatrix} 16.29 \\ 9.29 \\ 20.52 \end{pmatrix} \qquad \vec{c}_4 = \begin{pmatrix} 17.22 \\ 9.29 \\ 20.52 \end{pmatrix} \quad ,$$

the length is calculated to $l_{LAQNOJ} = 7.00\,\text{Å}$ and a deviation of $\approx 4.32\%$. The angles are all zero ($\sim 10^{-5}$)

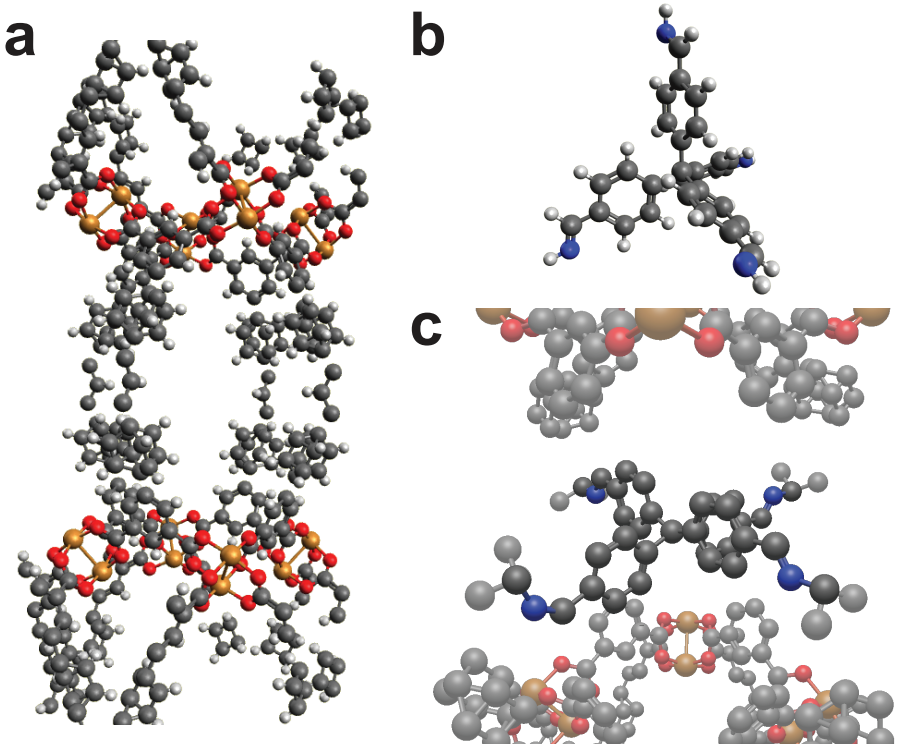


Figure 5.13.: (**a**): Unmodified MOF 'LAQNOJ'. (**b**): Used cross-linker 'CL9'. (**c**): Cross-linked structure with the cross-linker and the MOF binding sites being highlighted. The cross-linker is forced into a planar shape, which is handled as unfavorable by the force-field. This results in distortion of the structure. Color code: H (white), C (black), N (blue), O (red), Cu (gold). Hydrogen omitted in (**c**) for visibility.

$$\begin{aligned}\measuredangle h_1 c_1 c_{com}^{LAQNOJ} &= \measuredangle h_2 c_2 c_{com}^{LAQNOJ} = \\ &= \measuredangle h_3 c_3 c_{com}^{LAQNOJ} = \measuredangle h_4 c_4 c_{com}^{LAQNOJ} = 0.0^\circ \quad .\end{aligned}$$

With these values this MOF and especially with this combination of parameters is a prime candidate for planar cross-linker with four binding sites and a length of $\approx 7.0\,\text{Å}$.

The last MOF 'BIBXOB'[90] is cross-linked by a rather large cross-linker (see figure 5.14) with the length of $r_{CL10} = 19.08\,\text{Å}$. The cross-linker consists of two opened catenas that mechanically interlock with each other to form the cross-linking. This cross-linker and the one shown in figure 5.7 are more exotic structures and are intended for future projects from colleagues. The algorithm does not distinguish the time of cross-linking. Thus, even structures can be screened that have already a cross-linker built into it. Here the individual parts of the cross-linker are already part of the organic linker. They come together when the MOF is formed and due to their reactivity with other parts. Figure 5.15 shows the method of cross-linking during the synthesis of the MOF by using the examples from figure 5.7 and 5.14.

In case of a normal BDC the candidate would be a hydrogen. Here we have to calculate the distances after the optimization step, thus, using the position of the

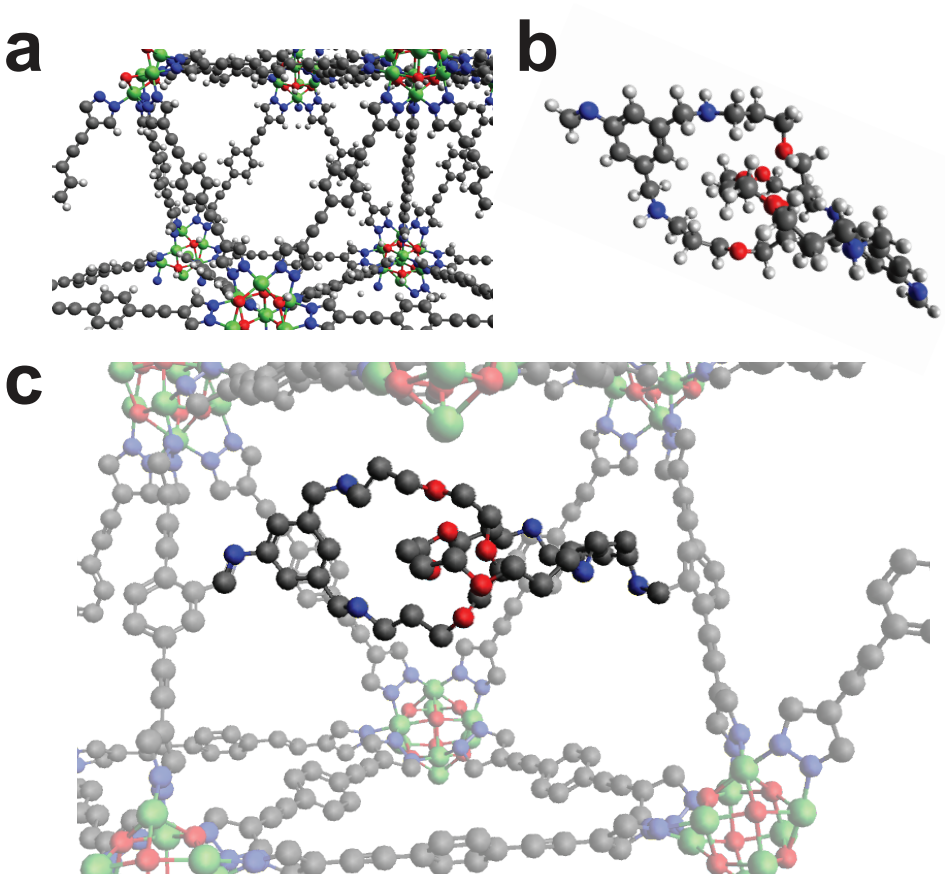


Figure 5.14.: (**a**): Unmodified MOF 'BIBXOB'. (**b**): Used cross-linker 'CL10'. (**c**): Cross-linked structure with the cross-linker and the MOF binding sites being highlighted. Color code: H (white), C (black), N (blue), O (red), Ni (green). Hydrogen omitted in (**c**) for visibility.

first carbon atom bonded to the carbon ring (again $\vec{h}_i^*$). The positions are

$$\vec{h}_1^* = \begin{pmatrix} 67.74 \\ 34.46 \\ 35.41 \end{pmatrix} \qquad \vec{c}_1 = \begin{pmatrix} 68.20 \\ 33.82 \\ 36.55 \end{pmatrix}$$
$$\vec{h}_2^* = \begin{pmatrix} 69.91 \\ 45.78 \\ 20.79 \end{pmatrix} \qquad \vec{c}_2 = \begin{pmatrix} 69.52 \\ 46.70 \\ 19.81 \end{pmatrix} \quad .$$

This results in a final length of $l_{BIBXOB} = 18.62\,\text{Å}$ and the deviation is $\approx 2.39\,\%$. Here again the angle is slightly higher than the limit because of the position of the linker after optimization. The angles are

$$\measuredangle h_1^* c_1 c_2^{BIBXOB} = 24.6° \qquad \measuredangle h_2^* c_2 c_1^{BIBXOB} = 20.7° \quad .$$

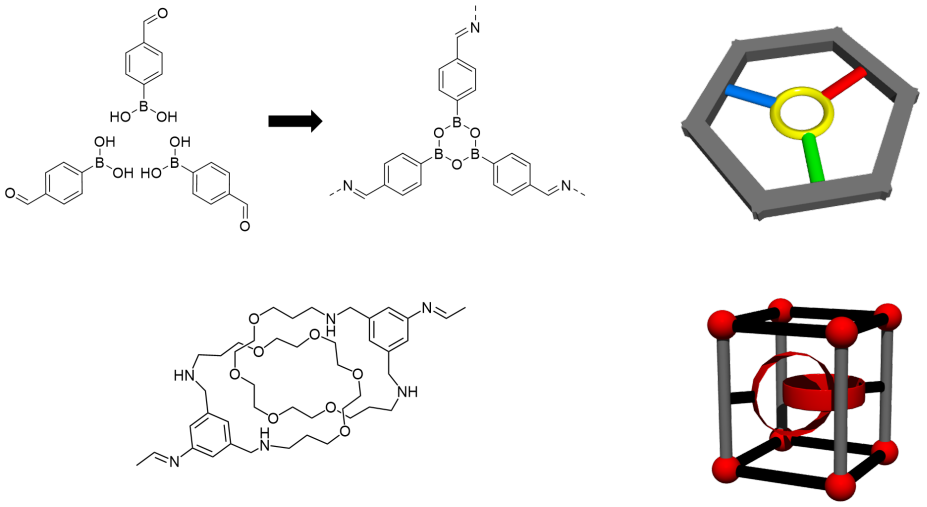


Figure 5.15.: Depiction of a different method of cross-linking. The first example shows where three cross-linker parts form a complete cross-linker. The second example shows two catenas interlocking each other to form a mechanical bond. Taken from [75].

5.3. False positives and true negatives

In section 5.2 some MOFs did not possess a rotatable linker and one case forced a tetrahedral cross-linker into a planar form (figure 5.13). It should be distinguished, whether non-rotatability is mandatory and whether further selection criteria should be added to take the orientation of the cross-linkers into account.

The terms *false positive* and *true negative* are frequently used in screening methods and statistics. Here, a false positive is a screening result that meets the selection criteria, but is unwanted. Vice versa, a true negative is a screening result that is wanted but is filtered out by any selection criteria. Whether or not a result is allowed, depends on the selection criteria. In brief, too strict selection criteria tend to lead to more true negatives and too loose selection criteria to higher false positives. In our case true negatives do not occur, since the goal of the candidate is to find ring structures. The candidate scan (see chapter 4, section 4.3.2.1) looks for hydrogen atoms and a rotation axis. Four out of six carbon atoms from a benzene ring are passed through by the algorithm, meaning about half of the ring is checked. Thus, in this case the selection criteria are necessary but not sufficient to find only BDCs. Necessary in the sense, that half of the benzene ring is always part of a whole ring, and not sufficient, because the second half is ignored. Hence, this speaks for loose selection criteria and the emergence of false positives instead of true negatives. In short, a benzene ring will always be found, since the algorithm identifies one half of it and passes it on. But whether the whole ring is indeed a BDC is not verified.

The original idea was to find MOFs whose organic linear linker has one or more carbon rings. The question now is, whether the selection criteria should be adapted to address this matter. Precisely, because of the rather loose selection criteria, many results were found that had not been considered before. Thus, some false positives still pose serious MOF and cross-linker mathings. If the selection criteria had been designed to strictly search for BDC linker only, interesting candidates such as 'RUYVEO' (figure 5.6), 'RUYVIS' (figure 5.7) and 'QOWRAV12' (figure 5.11)

would have never been found. Therefore, I suggests that the selection criteria should be somewhat looser to include results that have not been taken into account.

Nevertheless, the correct procedure for the future should be to implement any special case (e.g. non-rotatable linkers) allowed by the filters into the algorithm. This would turn the unintended positive results into intended positive results. However, one should not exaggerate and set the selection criteria too loosely, since otherwise the amount of output data strongly increases and its analysis becomes very difficult.

5.4. Distance study of cross-linkers with two binding sites

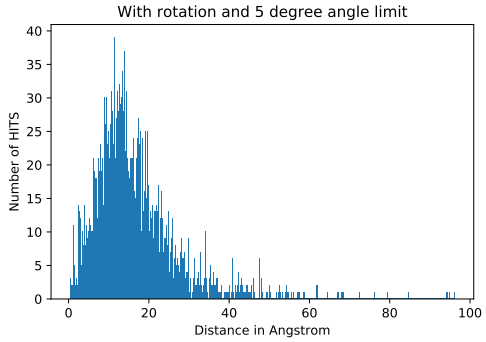
In order to get a feeling for the distances between two candidates, a screening was carried out without applying the distance filter, see 5.16. The values for the angle limits were 5°, 20° and 180°. Each with and without rotation of the organic linkers to check the influence of the angles and rotation. An angle of 180° means that every possible angle is allowed, since it is the maximum possible angle between two vectors.

First of all, the number of hits increases with the increase of the angle. This is obvious, because more and more unfavorable positions of candidates are allowed. But it becomes interesting when the values for a fixed angle of 5° or 20° with and without rotation are compared. The number of hits increases after the rotation of the linkers. This can be explained by the fact that symmetrical structures in particular yield not just one, but several hits after a rotation. As an example, one could imagine a simple cubic unit cell with one candidate on each of its edges. Assumed in their initial (stored) form, the candidates are placed unfavorably to each other with a relatively high angle between them. However, if two neighboring candidates were to be favorably placed after rotation, twelve different hits would result after the screening of this structure (each side surface of the cubic unit cell contributes two hits). A structure with several axes of symmetry would accordingly contribute several hits, provided that the angle is within the limits. If the angle does not play a role, then the rotation also does no longer play a role. For both configurations, i.e. no angle limit plus rotation and no angle without rotation, the number of hits is about the same. The algorithm neglects what happens to the candidates and simply measures their distance to each other. The number of hits should be equal, but there is a very small deviation. This could be due to the pre-filters (or human error), which do not allow certain combinations of candidates.

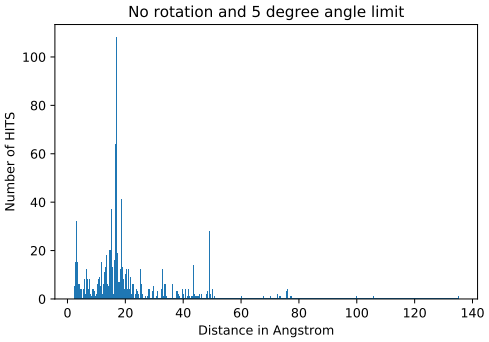
Another interesting point is the distribution of the distances themselves. A relative high amount many hits are located around the value 17 Å when looking at the range between 1 Å and 50 Å. The width of the peak indicates a great number of distances in the range from 5 Å and 22 Å for an angle limit of 5° and 20° degrees. With this information, the success of certain cross-linker lengths to find suitable MOF partners can be estimated in advance.

5.5. Future of the software

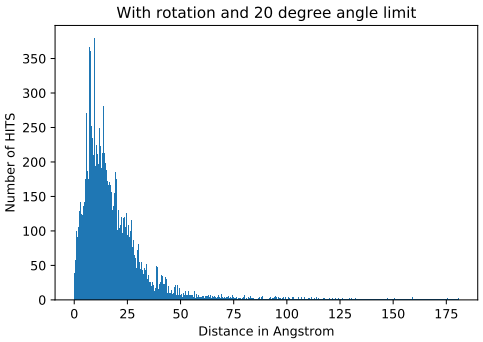
The primary goal of the software to enable high-throughput screening was fulfilled. However, attention was paid to functionality rather than user-friendliness. Therefore, knowledge of supercomputers infrastructure is necessary to carry out the screening. In addition, the cross-linkers were always placed by hand into the MOF, which took



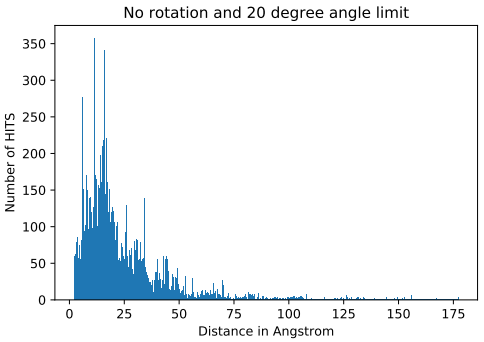
(a) Number of HITS: $n = 4011$



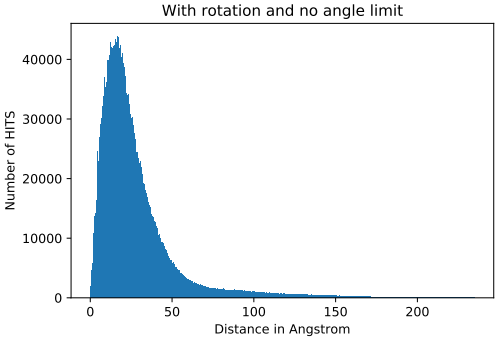
(b) Number of HITS: $n = 1479$



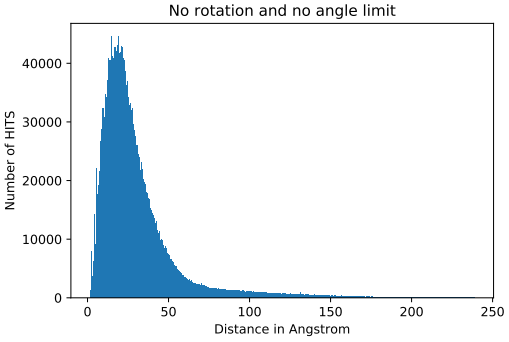
(c) Number of HITS: $n = 42383$



(d) Number of HITS: $n = 28391$



(e) Number of HITS: $n = 13002021$



(f) Number of HITS: $n = 13049162$

Figure 5.16.: All occurring distances between two candidates. Number of bins are ten times the maximum distance each. Taken from [75]

a lot of time and patience. In chapter 4.4 a solution for this issue is presented. What is still missing, is an interface between these two algorithms, so that the result of a screening can be used as an input for the automatic placement of the cross-linker. Such an interface should be feasible with relatively low effort. The candidates and their neighboring atoms are known and the binding sites of the cross-linkers can easily be included as input values. With this information it should be possible to place the cross-linker in the MOF. One difficulty that emerges, however, is the subsequent optimization of the structure. The structure files of the CoRe MOF database contain the positions of the atoms, but not their bonds. But this information is absolutely necessary because otherwise the simulation tool cannot generate an energy equation to physically describe the structure. The current optimization was done using a molecular editor that predicted the bonds itself. However, some bonds turned out to

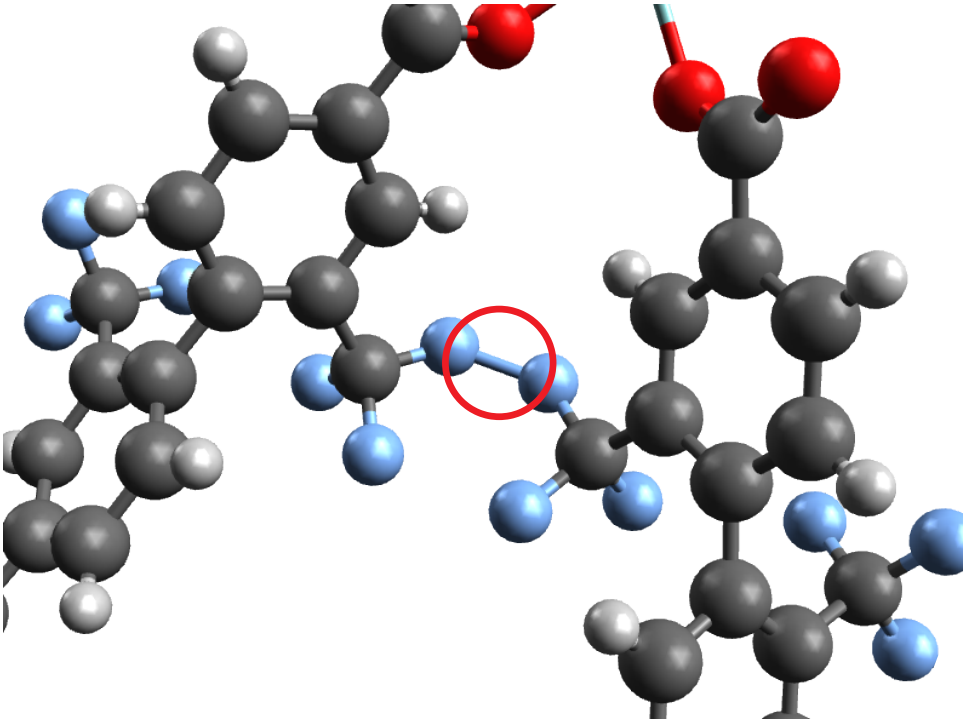


Figure 5.17.: Incorrect bond. The molecular editor *Avogadro* forms a bond between two fluorine atoms that is not meant to be. The red circle points out the incorrect placed bond. Color code: H (white), C (black), O (red), F (blue).

be wrong ('IXEJOM' was such a case, see figure 5.17) and in many cases the bonds may not be recognized at all. With human supervision such errors can be avoided. This however, makes the whole automatized screening process meaningless and a better solution must be found. The current version of the automatic cross-linker placement tool works with already created input files for the simulation software, which also include the bonds. An algorithm that translates all structures from the CoRe MOF database into a format in which bonds are also included would be a huge effort in itself, since automatic bond detection is still not fully developed. There exists one approach called *lammps_interface*[58] that could be used (with some customization) for the pipeline.

Another possibility would be to postpone the optimization step to the end. Before an optimization, one could automatically (and naively) insert the cross-linker into the MOF structure and visually display the result. The human filter in this case would then distinguish between good and bad matches, just like a doctor looking at radiographs. A subsequent investigation by MD would complete the process and either confirm or refute the prior decision. However, this approach includes human errors and with increasing interim results the probability of errors increases. Reducing the number of interim results by applying stricter filters in the screening algorithm would also lead to reduced unintended but good matches.

One technical aspect is the number of data generated after the screening. The more candidates a single MOF structure yields, the more combinations exist. If now some candidates fulfill the screening conditions, some other candidates that are symmetrically identical also fulfill the conditions. Figure 5.18 explains the symmetry concept. Assuming the green bordered atoms are investigated on. Then one bond

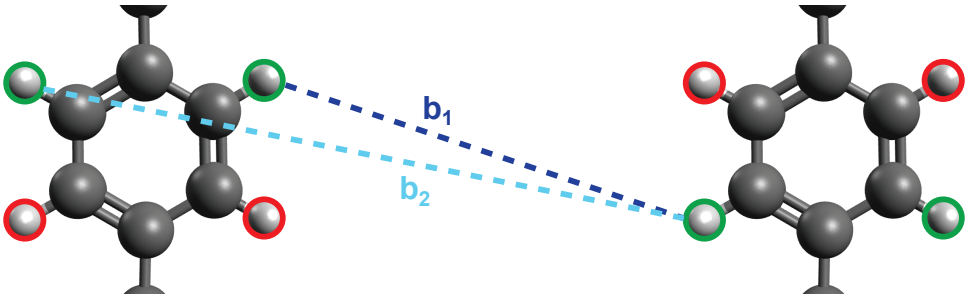


Figure 5.18.: Symmetrical bonds. If two candidates bordered in the same color fulfill the screening conditions, then multiple others also emerge due to symmetry. In this case the bond b_1 is one that passes the filters, but b_2 would also pass the filters after rotation. Color code: H (white), C (black).

b_1 may result in a positive candidate. Due to symmetry b_2 should also contribute a positive result after rotation. This example alone produces eight different results, which are technically identical. Such effects increase, when dealing with larger distances and higher numbers of binding sites.

The increase in data is not necessarily a problem, because the symmetry can be used in the data analysis. Looking for a certain number of hits from a single file might reveal an underlying symmetry of the screened MOF structure. If a single cross-linker produces exactly eight different hits (or other symmetrical multiples) in a MOF structure, where some candidates occur multiple times, the chance of a promising match increases. Thus, this idea should be considered when analyzing screening results.

A major benefit for future users with no background in programming would be in form of a *graphical user interface* (GUI). It simplifies the preparation of a screening and the overall user-friendliness. The screening algorithm is modular and allows the code to be extended. If in future the focus shifts from benzene rings to other linkers, then only the search for the candidates has to be modified or replaced. But it is also possible to screen several different linkers at the same time. All in all, the python code is relatively easy to understand and to modify for individual purposes.

This work only covered symmetrical cross-linkers with more than two binding sites. One approach to cover asymmetrical arm lengths is to list all occurring arm lengths of the cross-linker and distances of the MOFs. After sorting them based on their length, the largest lengths are compared, with the others following after. For example, a cross-linker with four binding sites has four arm length and the comparison ($\leftrightarrow$) then looks like

$$\text{comparison list} = \begin{pmatrix} l_{arm,1} \leftrightarrow l_{MOF,1} \\ l_{arm,2} \leftrightarrow l_{MOF,2} \\ l_{arm,3} \leftrightarrow l_{MOF,3} \\ l_{arm,4} \leftrightarrow l_{MOF,4} \end{pmatrix} \quad ,$$

with

$$l_{arm,1} \geq l_{arm,2} \geq l_{arm,3} \geq l_{arm,4}$$

and

$$l_{MOF,1} \geq l_{MOF,2} \geq l_{MOF,3} \geq l_{MOF,4} \quad .$$

Another important point to consider is the bond length between atoms. A carbon hydrogen bond lengths is different than a carbon carbon bond length or a carbon nitrogen bond length. This has an influence on the overall length of the MOF that is screened, since the force-field does not keep the binding site of the cross-linker (that replaced the hydrogen of the MOF) on the same position as the candidate. Especially for smaller cross-linker lengths this influence is stronger, since a tenth of an angstrom contributes significantly to the deviation of lengths (i.e. see MOF 'EKOPEA'). In this work this fact was neglected, since the main focus lied on the set up of a working screening algorithm. However, this should be considered in future updates of the algorithm.

Lastly, two programs of a future pipeline are already available. The screening itself and the automated placement of the cross-linker. What is still missing is the visual output of the results after using the cross-linker and the interfaces between these programs. For the purpose of visualization there are known software solutions, such as *PyMol*[91] or *VMD* (**V**isual **M**olecular **D**ynamics)[92]. The interface must ensure that the results of the previous program can be accepted and processed by the next program in the pipeline without problems.

6. Conclusion

The aim of this work was to find suitable MOFs for cross-linking in order to convert them into highly ordered COFs. The motivation was that these COFs consist only of organic material and are highly tunable through covalent bonds. The main problem was to find the ideal MOF structure for a specific cross-linker. Ideal in this sense meant that the cross-linker fits exactly into the MOF without deforming it. This should enable the future highly ordered COF to realize large scale structures in higher hierarchies. Since the number of MOFs is increasing rapidly and already is in the ten thousands, it is extremely difficult to keep track. An automated solution that compares a high number of MOFs with any given cross-linker was needed.

To achieve this goal, a high-throughput screening of the CoRe MOF database with more than ten thousand structures was conducted. The screening compared the length(s) of a given cross-linker with the distance(s) of the binding sites of the MOF. The screening itself was divided into several steps. At first, the connection points between MOF and cross-linker had to be defined. This step raised another problem, namely how to deal with chemical reactions. Binding of the cross-linker to the MOF takes place through a chemical reaction that can change the length of a cross-linker. But since an ideal fit of the cross-linker in the MOF was the prerequisite, the chemical reaction had to be considered. The solution to this issue was to consider the cross-linkers as already reacted to the MOF. Then, breaking the bond between MOF and cross-linker would result into two separate structures, each with an open binding site. If an atom a was attached to the open binding site of the MOF, cross-linking could be described as the atom of the cross-linker with the open binding site taking the place of the atom a. Thus, the binding partners were found, namely the atom a of the MOF and the atom with the open bond site of the cross-linker. This simplification is also advantageous, since the MOF in its unmodified form can be used for screening and has not to be altered. At this point it must be said that the atom a is always part of the organic linker, since covalent bonds were aimed for.

The second problem was to identify these atoms a in the MOF without knowing its exact structure. MOF structures with one or more benzene rings in the organic linkers were chosen, as they appear in many MOFs and are able to rotate around. Furthermore, the atom a could be exactly defined. All hydrogen, atoms whose binding partner was part of a benzene ring, represented the atom a. Thus, cross-linking could be achieved by replacing these hydrogen atoms by the binding atom of the cross-linker.

A special feature of these benzene rings is that they have an axis of rotation which the hydrogen atom a can rotate around. This lead to the next important point,

namely the position of hydrogen atoms $a_{1,\ldots,N}$ (N = total number of hydrogen atoms connected to aromatic carbons) to each other. A linear linker with two binding sites binds to the MOF by replacing the hydrogen atoms a_i and a_j with its own atoms that possess open binding sites. In order to enable continuous cross-linking, cross-linking at one point should not interfere with cross-linking at another point. Therefore, it was decided to minimize the position of the atoms $a_{1,\ldots,N}$ to each other. The underlying idea was, that in a symmetrical structure two adjacent linkers at minimum distance to each other should allow continuous cross-linking, if a cross-linker fits between them.

A series of filters, such as the symmetry filter or the angle filter, removed unfavorable combinations of the atoms a_i until finally the length of a cross-linker could be compared with the distance between two bonding sites. A positive result of this comparison revealed two structures, a MOF and a cross-linker, which were ideally suited to each other according to the selection criteria.

A point not to be neglected is that any automated task cannot be performed without the necessary resources. It was important to estimate the computational effort and the resources needed for an efficient screening. Especially when more than two binding sites had to be compared with each other, the computing effort and thus the computing time would increase considerably. In such a case, a serial screening process in which one combination is checked after the other would have taken several months. Therefore, it was of utmost importance to distribute the tasks to many processor cores. This meant that the screening algorithm had to run on a supercomputer that could provide the additional cores. The algorithm was parallelized in order to utilize these processor cores.

This work has shown ten selected examples of the screening results, whereas at least half of them represent serious predictions. Different cross-linkers types of cross-linker were chosen to for variety. The results, regardless of their experimental feasibility, confirm the intended functioning of the screening. Finding a few promising structures from such a large number of MOFs is a great advantage for experimental groups who can then devote themselves entirely to the synthesis of these predictions. Another major advantage is that all MOFs from the CoRe MOF database have already been synthesized and are published in the literature. For example, a commercially available cross-linker together with a matching MOF from the CoRe MOF database, for which synthesis instructions are available, allows quick action.

This work also addresses some improvements and functional extensions as a possible outlook of this project. Since in this work the cross-linkers were installed manually in the MOFs, it is obvious to automate the complete process from the beginning of the screening to the end where a cross-linked structure is obtained. The first step has already been taken by automating the placement of the cross-linker into MOF. But this was realized for a preliminary project of the high-throughput screening, hence both processes are not connected. The first task would be to develop an interface between the two processes to create a coherent program. The modular structure of the python algorithm allows easy modifications. For example, other linkers can be implemented quickly into the algorithm to increase the chances of positive hits.

On the whole, the objective of this theoretical work was achieved. The realization of the predicted structures must now be proven experimentally.

Appendix

A. Screening algorithm as pseudo code

This pseudo algorithm roughly describes the structure of the python code. Lines starting with '#' are comments.

```
import packages

# FUNCTIONS FOR THE SCREENING
def needed_functions(arguments)


# CALLED WITHIN THE SCREENING FUNCTIONS
def pre_filters(arguments)


# SEARCHS FOR CANDIDATES IN THE MOF STRUCTURE
def candidate_search(arguments)


# SCREENING FOR CROSS-LINKER WITH 2 BINDING SITES
def two_bond_screening(arguments)


# SCREENING FOR CROSS-LINKER WITH 3 BINDING SITES
def three_bond_screening(arguments)


# SCREENING FOR CROSS-LINKER WITH 4 BINDING SITES
def four_bond_screening(arguments)


# PARAMETERS
set_cross-linker_length = cl_length
set_arm_length = arm_length
set_angle_limit = angle_limit
set_distance_tolerance = tolerance
set_step_size = step_size
set_neighbor_distance = neighbor_distance
set_cross-linker_type = cl_type


# PARALLELIZATION
```

```
38 def parallelization(arguments)
39
40
41 # THIS FUNCTION IS EXECUTED FIRST
42 # AND CALLS ALL OTHER FUNCTIONS
43 def main():
44
45     # ITERATES THROUGH ALL STRUCTURE FILES
46     # OF THE DATABASE
47     for structures in database:
48
49     # LOADS A STRUCTURE FROM THE DATABASE
50     load structure from database
51
52     # SAVES ALL CANDIDATES FROM STRUCTURE
53     candidates = candidate_search(structure)
54
55     # CHECKS HOW MANY BINDING SITES THE CROSS-LINKER
56     # HAS AND EXECUTES THE SCREENING ACCORDINGLY
57     if cl_type == 2
58         results = two_bond_screening(candidates)
59
60     elif cl_type == 3
61         results = three_bond_screening(candidates)
62
63     elif cl_type == 4
64         results = three_bond_screening(candidates)
65
66     else:
67         continue
68
69     # SAVES THE RESULTS INTO A FILE
70     output(results)
71
72
73
74 # END OF CODE
```

Danksagung

Dass ich meine Promotion erfolgreich beenden konnte, wäre ohne Hilfe nicht möglich gewesen. Daher möchte ich abschließend noch die Personen danken, die mich durch diese schwierige aber aufregende Zeit begleitet haben:

- Dr. Alexander Schug und Dr. Manuel Tsotsalas für die Möglichkeit diese interessante Promotion anzutreten und die durchgehend sehr gute Betreuung.
- Dr. Manuel Tsotsalas zusätzlich für die Einführung in die Chemie und vor allem für die Geduld und das Verständnis mir gegenüber; als fachfremder Doktorand fühlte ich mich sehr gut aufgehoben.
- Prof. Wolfang Wenzel und Prof. Christof Wöll für die Wegweisung und für zusätzliche Ideen, die meine Arbeit positiv ergänzt haben.
- meiner Arbeitsgruppe am SCC – Ines Reinartz, Oskar Taubert, Jakob Rosenbauer, Marie Weiel-Potyagaylo, Arthur Voronin, Mehari Zerihun, Julian Herold und Fathia Idiris – für die tolle gemeinsame Zeit im Büro. Ich war zwar eher der Stille, habe dennoch die gemeinsame Zeit sehr genossen.
- die IFG-Arbeitsgruppe – Dr. Salma Begum, Dr. Qi An, Luo Yi, Yixuan Jia und Zeinab Hassan – für die lockeren und amüsanten Gruppenmeetings. Der Bordeaux-Trip war Klasse!
- Luo Yi zusätzlich für die gemeinsame Arbeit an verschiedenen Projekten und für die Zuverlässigkeit.

Naben meinen Kollegen darf natürlich meine Familie nicht fehlen. Mein besonderer Dank geht an

- meine Eltern Sarwat Mumtaz und Nasir Ahmad für die Unterstützung und Gebete, dass mich ich mich voll auf mein Studium konzentrieren konnte.
- meine Schwiegereltern Tayyaba Bhatti und Aziz Ahmad Bhatti für die Gebete und dafür, dass ich mich bei ihnen immer entspannen konnte.
- an meine Frau Naima Bhatti, die mich immer unterstützt hat und den Rücken freigehalten hat, damit ich mich in stressigen Zeiten voll auf die Arbeit konzentrieren konnte.
- meinen Bruder Qasid Ahmad für die vielen Diskussionen und vor allem für die Hilfe bei der Erstellung von Grafiken.

Darüber hinaus geht mein Dank an all diejenigen, die mich in irgendeiner Art unterstützt haben. Verzeiht mir wenn ich nicht jede Person einzeln nennen kann.

Lastly, I would like to thank Hazrat Mirza Masroor Ahmad Khalifat-ul-Masih the fifth^{atba} for his guiding and endless prayers.

Bibliography

[1] Yamil J Colón and Randall Q Snurr. High-throughput computational screening of metal–organic frameworks. *Chemical Society Reviews*, 43(16):5735–5749, 2014.

[2] Lei Wang, Min Zheng, and Zhigang Xie. Nanoscale metal–organic frameworks for drug delivery: a conventional platform with new promise. *Journal of Materials Chemistry B*, 6(5):707–717, 2018.

[3] Maria S Lohse and Thomas Bein. Covalent organic frameworks: Structures, Synthesis, and Applications. *Advanced Functional Materials*, 28(33):1705553, 2018.

[4] Eugene A Kapustin, Seungkyu Lee, Ahmad S Alshammari, and Omar M Yaghi. Molecular Retrofitting Adapts a Metal–Organic Framework to Extreme Pressure. *ACS Central Science*, 3(6):662–667, 2017.

[5] Yi Luo, Momin Ahmad, Alexander Schug, and Manuel Tsotsalas. Rising Up: Hierarchical Metal–Organic Frameworks in Experiments and Simulations. *Advanced Materials*, 31(26):1901744, 2019.

[6] Eram Sharmin and Fahmina Zafar. Introductory chapter: Metal Organic Frameworks (MOFs). In *Metal-Organic Frameworks*. IntechOpen, 2016.

[7] Paolo Falcaro, Raffaele Ricco, Cara M Doherty, Kang Liang, Anita J Hill, and Mark J Styles. Mof positioning technology and device fabrication. *Chemical Society Reviews*, 43(16):5513–5560, 2014.

[8] LI Hui, De-li Xiao, HE Hua, LIN Rui, and Peng-li Zuo. Adsorption behavior and adsorption mechanism of Cu(II) ions on amino-functionalized magnetic nanoparticles. *Transactions of Nonferrous Metals Society of China*, 23(9):2657–2665, 2013.

[9] Jiangtao Jia, Fuxing Sun, Qianrong Fang, Xiaoqiang Liang, Kun Cai, Zheng Bian, Huijun Zhao, Lianxun Gao, and Guangshan Zhu. A novel low density metal–organic framework with pcu topology by dendritic ligand. *Chemical Communications*, 47(32):9167–9169, 2011.

[10] Lars Öhrström. Let's Talk about MOFs—Topology and Terminology of Metal-Organic Frameworks and Why We Need Them. *Crystals*, 5(1):154–162, 2015.

[11] Tobie J Matemb Ma Ntep, Helge Reinsch, Jun Liang, and Christoph Janiak. Acetylenedicarboxylate-based cerium(IV) metal–organic framework with fcu topology: a potential material for air cleaning from toxic halogen vapors. *Dalton Transactions*, 48(42):15849–15855, 2019.

[12] Eugeny V Alexandrov, Andrey V Goltsev, Roman A Eremin, and Vladislav A Blatov. Anisotropy of Elastic Properties of Metal–Organic Frameworks and the Breathing Phenomenon. *The Journal of Physical Chemistry C*, 123(40):24651–24658, 2019.

[13] O.I. Lebedev, F Millange, C Serre, G Van Tendeloo, and G Férey. First Direct Imaging of Giant Pores of the Metal-Organic Framework MIL-101. *Chemistry of Materials*, 17(26):6525–6527, 2005.

[14] Priscilla Rocío-Bautista, Iván Taima-Mancera, Jorge Pasán, and Verónica Pino. Metal-Organic Frameworks in Green Analytical Chemistry. *Separations*, 6(3):33, 2019.

[15] Diego A Gomez, Jordi Toda, and German Sastre. Screening of hypothetical metal–organic frameworks for H_2 storage. *Physical Chemistry Chemical Physics*, 16(35):19001–19010, 2014.

[16] N Scott Bobbitt, Jiayi Chen, and Randall Q Snurr. High-throughput Screening of Metal–Organic Frameworks for Hydrogen Storage at Cryogenic Temperature. *The Journal of Physical Chemistry C*, 120(48):27328–27341, 2016.

[17] Benjamin J Sikora, Christopher E Wilmer, Michael L Greenfield, and Randall Q Snurr. Thermodynamic analysis of Xe/Kr selectivity in over 137 000 hypothetical metal–organic frameworks. *Chemical Science*, 3(7):2217–2223, 2012.

[18] Ashlee J Howarth, Aaron W Peters, Nicolaas A Vermeulen, Timothy C Wang, Joseph T Hupp, and Omar K Farha. Best Practices for the Synthesis, Activation, and Characterization of Metal–Organic Frameworks. *Chemistry of Materials*, 29(1):26–39, 2017.

[19] Andreas Schaate, Pascal Roy, Adelheid Godt, Jann Lippke, Florian Waltz, Michael Wiebcke, and Peter Behrens. Modulated Synthesis of Zr-Based Metal–Organic Frameworks: From Nano to Single Crystals. *Chemistry–A European Journal*, 17(24):6643–6651, 2011.

[20] Jerzy Leszczynski. *Computational Chemistry: Reviews of Current Trends: Volume 8*. World Scientific, 2003.

[21] Raghunathan Ramakrishnan, Pavlo O Dral, Matthias Rupp, and O Anatole von Lilienfeld. Big Data Meets Quantum Chemistry Approximations: The Δ-Machine Learning Approach. *Journal of Chemical Theory and Computation*, 11(5):2087–2096, 2015.

[22] KT Schütt, Michael Gastegger, Alexandre Tkatchenko, K-R Müller, and Reinhard J Maurer. Unifying machine learning and quantum chemistry with a deep neural network for molecular wavefunctions. *Nature Communications*, 10(1):1–10, 2019.

[23] Leopold Talirz, Snehal Kumbhar, Elsa Passaro, Aliaksandr V Yakutovich, Valeria Granata, Fernando Gargiulo, Marco Borelli, Martin Uhrin, Sebastiaan P Huber, Spyros Zoupanos, et al. Materials Cloud, a platform for open computational science. *arXiv preprint arXiv:2003.12510*, 2020.

[24] Osama Shekhah, Hui Wang, Markos Paradinas, Carmen Ocal, Björn Schüpbach, Andreas Terfort, Denise Zacher, Roland A Fischer, and Christof Wöll. Controlling interpenetration in metal–organic frameworks by liquid-phase epitaxy. *Nature Materials*, 8(6):481–484, 2009.

[25] Martin E Silvestre, Matthias Franzreb, Peter G Weidler, Osama Shekhah, and Christof Wöll. Magnetic Cores with Porous Coatings: Growth of Metal-Organic Frameworks on Particles Using Liquid Phase Epitaxy. *Advanced Functional Materials*, 23(9):1210–1213, 2013.

[26] Hea Jung Park, Monica C So, David Gosztola, Gary P Wiederrecht, Jonathan D Emery, Alex BF Martinson, Süleyman Er, Christopher E Wilmer, Nicolaas A Vermeulen, Alán Aspuru-Guzik, et al. Layer-by-Layer Assembled Films of Perylene Diimide-and Squaraine-Containing Metal–Organic Framework-like Materials: Solar Energy Capture and Directional Energy Transfer. *ACS Applied Materials & Interfaces*, 8(38):24983–24988, 2016.

[27] S Karthika, TK Radhakrishnan, and P Kalaichelvi. A Review of Classical and Nonclassical Nucleation Theories. *Crystal Growth & Design*, 16(11):6663–6681, 2016.

[28] Mary J Van Vleet, Tingting Weng, Xinyi Li, and JR Schmidt. In Situ, Time-Resolved, and Mechanistic Studies of Metal–Organic Framework Nucleation and Growth. *Chemical Reviews*, 118(7):3681–3721, 2018.

[29] Zhenqiang Wang and Seth M Cohen. Postsynthetic modification of metal–organic frameworks. *Chemical Society Reviews*, 38(5):1315–1329, 2009.

[30] Bin Wang, Lin-Hua Xie, Xiaoqing Wang, Xiao-Min Liu, Jinping Li, and Jian-Rong Li. Applications of metal–organic frameworks for green energy and environment: New advances in adsorptive gas separation, storage and removal. *Green Energy & Environment*, 3(3):191–228, 2018.

[31] Hao Li, Kecheng Wang, Yujia Sun, Christina T Lollar, Jialuo Li, and Hong-Cai Zhou. Recent advances in gas storage and separation using metal–organic frameworks. *Materials Today*, 21(2):108–121, 2018.

[32] Helge Bux, Christian Chmelik, Rajamani Krishna, and Juergen Caro. Ethene/ethane separation by the MOF membrane ZIF-8: Molecular correlation of permeation, adsorption, diffusion. *Journal of Membrane Science*, 369(1-2):284–289, 2011.

[33] Matthew T Kapelewski, Tomče Runčevski, Jacob D Tarver, Henry ZH Jiang, Katherine E Hurst, Philip A Parilla, Anthony Ayala, Thomas Gennett, Stephen A FitzGerald, Craig M Brown, et al. Record High Hydrogen Storage Capacity in the Metal–Organic Framework Ni_2(m-dobdc) at Near-Ambient Temperatures. *Chemistry of Materials*, 30(22):8179–8189, 2018.

[34] Sanjeev K Bhardwaj, Neha Bhardwaj, Rajnish Kaur, Jyotsana Mehta, Amit L Sharma, Ki-Hyun Kim, and Akash Deep. An overview of different strategies to introduce conductivity in metal–organic frameworks and miscellaneous applications thereof. *Journal of Materials Chemistry A*, 6(31):14992–15009, 2018.

[35] Cláudia Gomes Silva, Avelino Corma, and Hermenegildo García. Metal–organic frameworks as semiconductors. *Journal of Materials Chemistry*, 20(16):3141–3156, 2010.

[36] Ke Jiang, Ling Zhang, Quan Hu, Dian Zhao, Tifeng Xia, Wenxin Lin, Yanyu Yang, Yuanjing Cui, Yu Yang, and Guodong Qian. Pressure controlled drug release in a Zr-cluster-based MOF. *Journal of Materials Chemistry B*, 4(39):6398–6401, 2016.

[37] Jie Yang, Anna Grzech, Fokko M Mulder, and Theo J Dingemans. Methyl modified MOF-5: a water stable hydrogen storage material. *Chemical Communications*, 47(18):5244–5246, 2011.

[38] Aaron W Thornton, Ravichandar Babarao, Aman Jain, Fabien Trousselet, and F-X Coudert. Defects in metal–organic frameworks: a compromise between adsorption and stability? *Dalton Transactions*, 45(10):4352–4359, 2016.

[39] Jinxuan Liu and Christof Wöll. Surface-supported metal–organic framework thin films: fabrication methods, applications, and challenges. *Chemical Society Reviews*, 46(19):5730–5770, 2017.

[40] Takumi Ishiwata, Yuki Furukawa, Kouta Sugikawa, Kenta Kokado, and Kazuki Sada. Transformation of Metal–Organic Framework to Polymer Gel by Cross-Linking the Organic Ligands Preorganized in Metal–Organic Framework. *Journal of the American Chemical Society*, 135(14):5427–5432, 2013.

[41] Kazuki Sada, Takumi Ishiwata, and Kenta Kokado. Topochemical Polymerizations and Crystal Cross-Linking of Metal Organic Frameworks. In *Advances in Organic Crystal Chemistry*, pages 517–530. Springer, 2015.

[42] Manuel Tsotsalas, Jinxuan Liu, Beatrix Tettmann, Sylvain Grosjean, Artak Shahnas, Zhengbang Wang, Carlos Azucena, Matthew Addicoat, Thomas Heine, Joerg Lahann, et al. Fabrication of Highly Uniform Gel Coatings by the Conversion of Surface-Anchored Metal–Organic Frameworks. *Journal of the American Chemical Society*, 136(1):8–11, 2014.

[43] Sophia Schmitt, Stéphane Diring, Peter G Weidler, Salma Begum, Stefan Heißler, Susumu Kitagawa, Christof Wöll, Shuhei Furukawa, and Manuel Tsotsalas. Localized Conversion of Metal–Organic Frameworks into Polymer Gels via Light-Induced Click Chemistry. *Chemistry of Materials*, 29(14):5982–5989, 2017.

[44] Michael P Allen et al. Introduction to Molecular Dynamics Simulation. *Computational Soft Matter: From Synthetic Polymers to Proteins*, 23:1–28, 2004.

[45] Walter Kohn, Axel D Becke, and Robert G Parr. Density Functional Theory of Electronic Structure. *The Journal of Physical Chemistry*, 100(31):12974–12980, 1996.

[46] David A Case, Thomas E Cheatham III, Tom Darden, Holger Gohlke, Ray Luo, Kenneth M Merz Jr, Alexey Onufriev, Carlos Simmerling, Bing Wang, and Robert J Woods. The Amber biomolecular simulation programs. *Journal of Computational Chemistry*, 26(16):1668–1688, 2005.

[47] Bernard R Brooks, Charles L Brooks III, Alexander D Mackerell Jr, Lennart Nilsson, Robert J Petrella, Benoît Roux, Youngdo Won, Georgios Archontis, Christian Bartels, Stefan Boresch, et al. CHARMM: The biomolecular simulation program. *Journal of Computational Chemistry*, 30(10):1545–1614, 2009.

[48] Stephen L Mayo, Barry D Olafson, and William A Goddard. DREIDING: a generic force field for molecular simulations. *Journal of Physical Chemistry*, 94(26):8897–8909, 1990.

[49] William L Jorgensen, David S Maxwell, and Julian Tirado-Rives. Development and Testing of the OPLS All-Atom Force Field on Conformational Energetics and Properties of Organic Liquids. *Journal of the American Chemical Society*, 118(45):11225–11236, 1996.

[50] Anthony K Rappé, Carla J Casewit, KS Colwell, William A Goddard III, and W Mason Skiff. UFF, a full periodic table force field for molecular mechanics and molecular dynamics simulations. *Journal of the American Chemical Society*, 114(25):10024–10035, 1992.

[51] Maksim Kouza. Numerical Simulation of Folding and Unfolding of Proteins, 2013.

[52] Tutorial 1: Lennard-Jones Liquid. `http://espressomd.org/html/tutorials_html/01-lennard_jones/01-lennard_jones.html`. Accessed: 2020-05-08.

[53] Abdulnour Y Toukmaji and John A Board Jr. Ewald summation techniques in perspective: a survey. *Computer Physics Communications*, 95(2-3):73–92, 1996.

[54] Damien E Coupry, Matthew A Addicoat, and Thomas Heine. Extension of the Universal Force Field for Metal–Organic Frameworks. *Journal of Chemical Theory and Computation*, 12(10):5215–5225, 2016.

[55] MCCCS Towhee: UFF. `http://towhee.sourceforge.net/forcefields/uff.html`. Accessed: 2020-05-05.

[56] MJ Abraham, D Van Der Spoel, E Lindahl, B Hess, et al. Gromacs user manual version 2020.2. *Sweden: Royal Institute of Technology and Uppsala University*, 2020.

[57] Ratna S Katiyar and Prateek K Jha. Molecular simulations in drug delivery: Opportunities and challenges. *Wiley Interdisciplinary Reviews: Computational Molecular Science*, 8(4):e1358, 2018.

[58] Peter G Boyd, Seyed Mohamad Moosavi, Matthew Witman, and Berend Smit. Force-Field Prediction of Materials Properties in Metal-Organic Frameworks. *The Journal of Physical Chemistry Letters*, 8(2):357–363, 2017.

[59] Sareeya Bureekaew, Saeed Amirjalayer, Maxim Tafipolsky, Christian Spickermann, Tapta Kanchan Roy, and Rochus Schmid. MOF-FF–A flexible first-principles derived force field for metal-organic frameworks. *physica status solidi (b)*, 250(6):1128–1141, 2013.

[60] Sudi Jawahery, Nakul Rampal, Seyed Mohamad Moosavi, Matthew Witman, and Berend Smit. Ab Initio Flexible Force Field for Metal–Organic Frameworks Using Dummy Model Coordination Bonds. *Journal of Chemical Theory and Computation*, 15(6):3666–3677, 2019.

[61] Johannes P Dürholt, Julian Keupp, and Rochus Schmid. The Impact of Mesopores on the Mechanical Stability of HKUST-1: A Multiscale Investigation. *European Journal of Inorganic Chemistry*, 2016(27):4517–4523, 2016.

[62] Adri CT Van Duin, Siddharth Dasgupta, Francois Lorant, and William A Goddard. ReaxFF: A Reactive Force Field for Hydrocarbons. *The Journal of Physical Chemistry A*, 105(41):9396–9409, 2001.

[63] You Han, Dandan Jiang, Jinli Zhang, Wei Li, Zhongxue Gan, and Junjie Gu. Development, applications and challenges of ReaxFF reactive force field in molecular simulations. *Frontiers of Chemical Science and Engineering*, 10(1):16–38, 2016.

[64] Peyman Z Moghadam, Aurelia Li, Seth B Wiggin, Andi Tao, Andrew GP Maloney, Peter A Wood, Suzanna C Ward, and David Fairen-Jimenez. Development of a Cambridge Structural Database Subset: A Collection of Metal–Organic Frameworks for Past, Present, and Future. *Chemistry of Materials*, 29(7):2618–2625, 2017.

[65] Christopher E Wilmer, Michael Leaf, Chang Yeon Lee, Omar K Farha, Brad G Hauser, Joseph T Hupp, and Randall Q Snurr. Large-scale screening of hypothetical metal–organic frameworks. *Nature Chemistry*, 4(2):83, 2012.

[66] Yongchul G Chung, Emmanuel Haldoupis, Benjamin J Bucior, Maciej Haranczyk, Seulchan Lee, Hongda Zhang, Konstantinos D Vogiatzis, Marija Milisavljevic, Sanliang Ling, Jeffrey S Camp, et al. Advances, Updates, and Analytics for the Computation-Ready, Experimental Metal–Organic Framework Database: CoRE MOF 2019. *Journal of Chemical & Engineering Data*, 64(12):5985–5998, 2019.

[67] Sydney R Hall, Frank H Allen, and I David Brown. The crystallographic information file (CIF): a new standard archive file for crystallography. *Acta Crystallographica Section A: Foundations of Crystallography*, 47(6):655–685, 1991.

[68] Jacob Goldsmith, Antek G Wong-Foy, Michael J Cafarella, and Donald J Siegel. Theoretical Limits of Hydrogen Storage in Metal–Organic Frameworks: Opportunities and Trade-Offs. *Chemistry of Materials*, 25(16):3373–3382, 2013.

[69] Emmanuel Haldoupis, Sankar Nair, and David S Sholl. Efficient Calculation of Diffusion Limitations in Metal Organic Framework Materials: A Tool for Identifying Materials for Kinetic Separations. *Journal of the American Chemical Society*, 132(21):7528–7539, 2010.

[70] Andrew Tarzia, Masahide Takahashi, Paolo Falcaro, Aaron W Thornton, Christian J Doonan, and David M Huang. High-Throughput Screening of Metal–Organic Frameworks for Macroscale Heteroepitaxial Alignment. *ACS Applied Materials & Interfaces*, 10(47):40938–40950, 2018.

[71] Guido Van Rossum et al. Python programming language. In *USENIX annual technical conference*, volume 41, page 36, 2007.

[72] Shyue Ping Ong, William Davidson Richards, Anubhav Jain, Geoffroy Hautier, Michael Kocher, Shreyas Cholia, Dan Gunter, Vincent L Chevrier, Kristin A Persson, and Gerbrand Ceder. Python Materials Genomics (pymatgen): A

robust, open-source python library for materials analysis. *Computational Materials Science*, 68:314–319, 2013.

[73] Travis E Oliphant. *A guide to NumPy*, volume 1. Trelgol Publishing USA, 2006.

[74] Lisandro Dalcin. MPI for Python, 2012.

[75] Momin Ahmad, Yi Luo, Christof Wöll, Manuel Tsotsalas, and Alexander Schug. Design of Metal-Organic Framework Templated Materials Using High-Throughput Computational Screening. *Molecules*, 25(21), 2020.

[76] David R Lide Jr. A survey of carbon-carbon bond lengths. *Tetrahedron*, 17(3-4):125–134, 1962.

[77] Jian S Dai. Euler–Rodrigues formula variations, quaternion conjugation and intrinsic connections. *Mechanism and Machine Theory*, 92:144–152, 2015.

[78] Zhengbang Wang, Alfred Błaszczyk, Olaf Fuhr, Stefan Heissler, Christof Wöll, and Marcel Mayor. Molecular weaving via surface-templated epitaxy of crystalline coordination networks. *Nature Communications*, 8(1):1–8, 2017.

[79] Matthew A Addicoat, Damien E Coupry, and Thomas Heine. AuToGraFS: Automatic Topological Generator for Framework Structures. *The Journal of Physical Chemistry A*, 118(40):9607–9614, 2014.

[80] Marcus D Hanwell, Donald E Curtis, David C Lonie, Tim Vandermeersch, Eva Zurek, and Geoffrey R Hutchison. Avogadro: an advanced semantic chemical editor, visualization, and analysis platform. *Journal of Cheminformatics*, 4(1):17, 2012.

[81] Colin R Groom, Ian J Bruno, Matthew P Lightfoot, and Suzanna C Ward. The Cambridge Structural Database. *Acta Crystallographica Section B: Structural Science, Crystal Engineering and Materials*, 72(2):171–179, 2016.

[82] Jian-Di Lin, Shu-Ting Wu, Zhi-Hua Li, and Shao-Wu Du. Syntheses, topological analyses, and NLO-active properties of new Cd(II)/M(II) (M = Ca, Sr) metal–organic frameworks based on R-isophthalic acids (R = H, OH, and t-Bu). *Dalton Transactions*, 39(44):10719–10728, 2010.

[83] Jian-Wei Zhang, Man-Cheng Hu, Shu-Ni Li, Yu-Cheng Jiang, and Quan-Guo Zhai. Microporous rod metal–organic frameworks with diverse Zn/Cd–triazolate ribbons as secondary building units for CO_2 uptake and selective adsorption of hydrocarbons. *Dalton Transactions*, 46(3):836–844, 2017.

[84] Qingxia Yao, Antonio Bermejo Gómez, Jie Su, Vlad Pascanu, Yifeng Yun, Haoquan Zheng, Hong Chen, Leifeng Liu, Hani Nasser Abdelhamid, Belén Martín-Matute, et al. Series of Highly Stable Isoreticular Lanthanide Metal–Organic Frameworks with Expanding Pore Size and Tunable Luminescent Properties. *Chemistry of Materials*, 27(15):5332–5339, 2015.

[85] Shuai Yuan, Lanfang Zou, Haixia Li, Ying-Pin Chen, Junsheng Qin, Qiang Zhang, Weigang Lu, Michael B Hall, and Hong-Cai Zhou. Flexible Zirconium Metal-Organic Frameworks as Bioinspired Switchable Catalysts. *Angewandte Chemie International Edition*, 55(36):10776–10780, 2016.

[86] Dongwook Kim, Xinfang Liu, Minhak Oh, Xiaokai Song, Yang Zou, Devendra Singh, Kwang S Kim, and Myoung Soo Lah. Isoreticular MOFs based on a rhombic dodecahedral MOP as a tertiary building unit. *CrystEngComm*, 16(28):6391–6397, 2014.

[87] Yue Wu, Vanessa K Peterson, Emily Luks, Tamim A Darwish, and Cameron J Kepert. Interpenetration as a Mechanism for Negative Thermal Expansion in the Metal–Organic Framework $Cu_3(btb)_2$ (MOF-14). *Angewandte Chemie*, 126(20):5275–5278, 2014.

[88] Jiandong Pang, Caiping Liu, Yougui Huang, Mingyan Wu, Feilong Jiang, Daqiang Yuan, Falu Hu, Kongzhao Su, Guoliang Liu, and Maochun Hong. Visualizing the Dynamics of Temperature- and Solvent-Responsive Soft Crystals. *Angewandte Chemie International Edition*, 55(26):7478–7482, 2016.

[89] Florian Moreau, Daniil I Kolokolov, Alexander G Stepanov, Timothy L Easun, Anne Dailly, William Lewis, Alexander J Blake, Harriott Nowell, Matthew J Lennox, Elena Besley, et al. Tailoring porosity and rotational dynamics in a series of octacarboxylate metal-organic frameworks. *Proceedings of the National Academy of Sciences*, 114(12):3056–3061, 2017.

[90] Natalia M Padial, Elsa Quartapelle Procopio, Carmen Montoro, Elena López, J Enrique Oltra, Valentina Colombo, Angelo Maspero, Norberto Masciocchi, Simona Galli, Irena Senkovska, et al. Highly Hydrophobic Isoreticular Porous Metal–Organic Frameworks for the Capture of Harmful Volatile Organic Compounds. *Angewandte Chemie International Edition*, 52(32):8290–8294, 2013.

[91] Warren L DeLano et al. PyMOL: An Open-Source Molecular Graphics Tool. *CCP4 Newsletter on protein crystallography*, 40(1):82–92, 2002.

[92] William Humphrey, Andrew Dalke, Klaus Schulten, et al. VMD: Visual molecular dynamics. *Journal of Molecular Graphics*, 14(1):33–38, 1996.

[93] Long Jiao, Joanne Yen Ru Seow, William Scott Skinner, Zhiyong U Wang, and Hai-Long Jiang. Metal–organic frameworks: Structures and functional applications. *Materials Today*, 27:43–68, 2019.

[94] Tina M Nenoff. MOF membranes put to the test. *Nature Chemistry*, 7(5):377–378, 2015.

[95] Michael V Ushcats, Leonid A Bulavin, Vladimir M Sysoev, Vitaliy Yu Bardik, and Alexander N Alekseev. Statistical theory of condensation — Advances and challenges. *Journal of Molecular Liquids*, 224:694–712, 2016.

[96] Thomas P Senftle, Sungwook Hong, Md Mahbubul Islam, Sudhir B Kylasa, Yuanxia Zheng, Yun Kyung Shin, Chad Junkermeier, Roman Engel-Herbert, Michael J Janik, Hasan Metin Aktulga, et al. The ReaxFF reactive force-field: development, applications and future directions. *npj Computational Materials*, 2(1):1–14, 2016.