

AN ABSTRACT OF THE THESIS OF

Charles A. Manion for the degree of Doctor of Philosophy in Mechanical Engineering
presented on June 14, 2017.

Title: Computational Design for Metal Organic Responsive Frameworks(MORFs)

Abstract approved: _____

Matthew I. Campbell Irem Tumer

Metal organic Frameworks(MOFs) that experience stimuli induced structural transformation could enable a whole new class of materials with remarkable properties. Photoactuating moieties in the structure could effect changes in the pore space or macroscale shape change enabling light driven gas separation and actuators. Here, we present a novel approach based on using graph grammars that makes the problem of discovering MOFs for specific applications amenable to powerful artificial intelligence tree search algorithms. We develop methods for rapidly evaluating MOFs for structural transitions. We will show how this approach may be used to computationally search for candidate MOFs that undergo reversible stimuli induced structural transformation.

©Copyright by Charles A. Manion

June 14, 2017

All Rights Reserved

Computational Design for Metal Organic Responsive Frameworks (MORFs)

by
Charles A. Manion

A THESIS

submitted to

Oregon State University

in partial fulfillment of
the requirements for the
degree of

Doctor of Philosophy

Presented June 14, 2017
Commencement June 2017

Doctor of Philosophy thesis of Charles A. Manion presented on June 14, 2017

APPROVED:

Co-Major Professor, representing Mechanical Engineering

Co-Major Professor, representing Mechanical Engineering

Head of the School of Mechanical, Industrial, and Manufacturing Engineering

Dean of the Graduate School

I understand that my thesis will become part of the permanent collection of Oregon State University libraries. My signature below authorizes release of my thesis to any reader upon request.

Charles A. Manion, Author

ACKNOWLEDGEMENTS

I would like to thank my major advisors Dr. Matthew I. Campbell and Dr. Irem Tumer for providing me with guidance and support during my study. Personal thanks Dr. Irem Tumer for encouraging me to do this research. Special thanks to P. A. Greaney for help with LAMMPS and encouraging me to be more interdisciplinary. I would like to thank Brady J. Gibbons, Kagan Tumer, and David H. McIntyre for agreeing to serve on my committee. I would also like to thank the W.M. Keck Foundation for funding and support.

Finally, I would like to thank my parents for encouragement and support.

TABLE OF CONTENTS

	<u>Page</u>
1 Introduction	1
1.1 Molecular Design	6
1.1.1 Drug Design	6
1.1.2 MOF and similar system design	9
2 Methods	16
2.1 Graph Grammars	16
2.2 Review of Molecular Dynamics	21
3 Designing Linkers with Specified Mechanical Properties	25
4 Fragment Characterization	32
4.1 Linker Population Summary	42
5 Pressure Switching MOFs	44
5.1 Grammatical Evolution	45
5.2 Evaluation	48
6 Photo switching MOFs	54
6.1 Photoisomers	54
6.2 Modelling Azobenzene	55
6.3 Evaluation	58
7 Monte Carlo Tree Search for molecules	65
7.1 Test Problem	67
7.2 Results	69
8 Conclusions and Future Directions	74
Bibliography	78

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
1.1 A linker of MOF-5 left and its corresponding unit cell right	2
1.2 A conceptual diagram of how a MORF might work	3
1.3 an example of how a MORF might be used for hydrogen gas storage	5
1.4 How a MORF might be used for gas separation	5
1.5 A compatible linker molecule can be found quickly by searching looking for molecules that have connection point vectors d' apart with angle α between them, taken from (1)	8
1.6 generation process for hypothetical MOFs taken from (2)	9
1.7 Abstract continuous representation(a), and its associated projection onto a real molecule(b), taken from (3).....	11
1.8 Search process for (4)	12
1.9 Generation process for metal organic polyhedrons (5)	13
2.1 A graph rule and its associated applications	17
2.2 Rules	18
2.3 Convex hull of molecule	20
2.4 A dihedral going through atoms i, j, k,l	22
2.5 An improper going through atoms i, j, k,l, taken from LAMMPS documentation	23
2.6 Lennard Jones potential normalized against ϵ and r_m	23
3.1 Illustration of search process	25
3.2 Linear force curve molecule(a) and its associated force displacement curve(b)	29
3.3 Exponential force curve molecule(a) and its associated force displacement curve(b)	29
3.4 Constant force curve molecule(a) and its associated force displacement curve(b)	30
3.5 A compliant constant force mechanism from (6)	31

LIST OF FIGURES (Continued)

<u>Figure</u>	<u>Page</u>
4.1 Directionally dependent stiffness indicatrix for MIL-53 under uniaxial stress (left) and uniaxial strain (right). For a pressure switching MOF we require deforming linkers rather than framework flexibility.	34
4.2 Example of a candidate linker, (a) relaxed structure, (b) maximally compressed structure, (c) the molecules' potential energy during deformation, and (d) the force displacement curves for compressing and stretching the molecule at 300K with a deformation rate of 1 m/s. The left column is for linker candidate 49694 which exhibits an elastic buckling instability, the right column is for candidate 25923 which undergoes a stress driven conformation change.	40
4.3 Example of the fitting of the model curve indicating the mechanical interpretation of the fitting parameters.	41
4.4 Number of molecules that behave as good pressure switches at a given pressure.	41
5.1 Seed molecule, green line denotes symmetry plane, arrow indicates rotation axis	44
5.2 Cone defined by carboxylate	47
5.3 Linker 4058	50
5.4 Linker 4032	50
5.5 Linker 3260	51
5.6 Average fitness	51
5.7 Synthesis route for an analog of the top performing linker	52
5.8 Pressure induced transitions of top performing candidates, zero pressure(left), 490 atm (right)	53
6.1 Examples of photoisomers (taken from (7))	54
6.2 Potential for the C-N=N-C torsion of azobenzene from (8) (9)	56
6.3 Switching function	57
6.4 Seed molecule, green line denotes mirror plane, arrow indicates rotation axis	60
6.5 1st highest performing candidate, with a volumetric strain of -0.4990	61

LIST OF FIGURES (Continued)

<u>Figure</u>	<u>Page</u>
6.6 2nd highest performing candidate, with a volumetric strain of -0.4857	61
6.7 3rd highest performing candidate, with a volumetric strain of -0.4691	62
6.8 3rd top performing candidate, with a volumetric strain of -0.4691	62
6.9 2nd top performing candidate, with a volumetric strain of -0.4857	63
6.10 1st top performing candidate, with a volumetric strain of -0.4990	63
6.11 Comparison of behavior during isomerization.....	64
7.1 Choose feasible carboxylate heuristic(left), make most axially distant option heuristic(right).....	68
7.2 Top Performing Linker(left) found through UCT and its force displacement curve	69
7.3 Reward for MCTS with iterations, green line is average	70
7.4 Top performing linker epsilon greedy	71
7.5 Reward for UCT epsilon greedy, green line is average	72
7.6 Top performing linker, changed default policy	73
7.7 Reward for UCT epsilon greedy changed default policy, green line is average	73
8.1 An azobenzene based MOF that contracts on two axes, but not the other ...	77

LIST OF TABLES

<u>Table</u>	<u>Page</u>
3.1 Force curve equations	28
3.2 Performance	28
4.1 Candidate behavior descriptors for predicting pressure switching linkers....	38
4.2 Sample of linker fragments that correlate with behavior descriptors.	39
4.3 Linker Database Summary	39
7.1 Performance Summary	71

Computational Design for Metal Organic Responsive Frameworks (MORFs)

1 Introduction

In this work we hope to develop a method to design a whole new class of materials that change shape when a stimulus such as light is applied. We believe our current means of designing materials are not suitable for doing this. We do not have very good processes for inventing new materials. In fact it is something of a stretch to say that we invent them at all, rather; we discover them. Historically this has happened in a haphazard and unplanned manner. The inventors (or rather discoverers) of various important materials like teflon (*10*), artificial mauve dye (*11*), cellophane (*12*), safety glass (*13*), etc did not originally set out to produce such things. When attempts are made at directed materials discovery they often proceed in a manner similar to how Edison discovered practical filaments for incandescent light bulbs (*14*); testing thousands of different random materials until we find one with the properties we want. Modern approaches simply do the same thing only faster, using either robots (*15*) or virtual simulations (*16*) to test material properties. What we would like to do is make this process more efficient, developing new methods to better target our efforts.

For this work, we consider the problem of designing molecular space frame type structures which are capable of shape change in response to external stimuli. We call such structures Metal Organic Responsive Frameworks or MORFs. One particularly interesting class of materials from a design perspective are Metal Organic Frameworks. Metal Organic Frameworks (MOFs) are a molecular space frame consisting of organic linkers, held together by metal ions. Linkers are organic ligands or molecules capable of binding to metal ions. An example of such a framework is shown in Figure 1.1b. An example of a linker is shown in Figure 1.1a and its corresponding metal organic framework is shown in Figure 1.1b. What makes them interesting is the amount of structural control they provide and the useful appli-

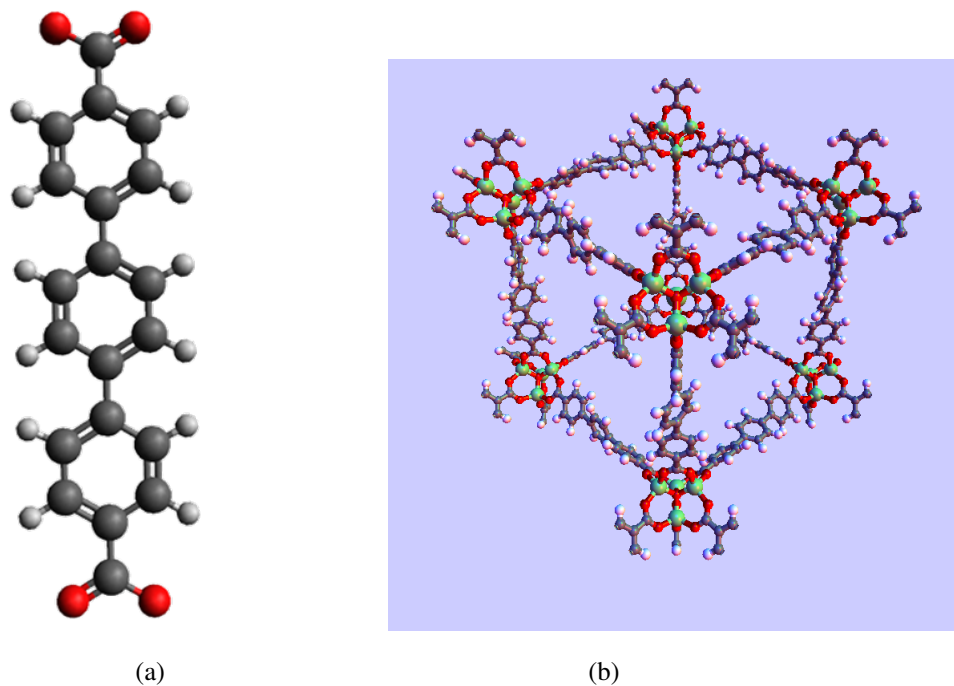


FIGURE 1.1: A linker of MOF-5 left and its corresponding unit cell right

cations they have. One advantage of MOFs, is that they are modular. One can make a wide variety of MOFs with the same network topology by substituting different linkers with the same shape. This is because metal ions and parts of the linkers will self assemble into defined structures known as secondary building units(SBUs). This allows us to control how linkers connect together. So long as our linkers have roughly the same shape, if we connect them together in the same manner we can make structure with the same network topology. Likewise, some linkers can be used to make MOFs with different network topologies by using different metal ions and/or changing the chemistry. It can be difficult to realize materials with desired crystal structures. Even if such structures are chemically and energetically favorable, they may be extremely difficult to synthesize. However, the chemistry of MOFs can make realizing different crystal structures almost easy. A great example of this is the polybenzene crystal structure. The polybenzene crystal structure was originally predicted as an allotrope of carbon in 1946 (17) (18) and as of yet no one has managed to synthesize it. Alezi et al. (19)

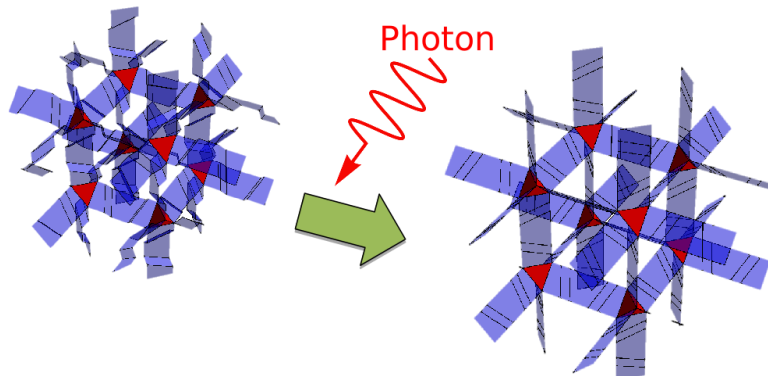


FIGURE 1.2: A conceptual diagram of how a MORF might work

were able to engineer a MOF with this, as of the time never before made, crystal structure. They started with the conceptual net structure of a MOF and determined from the net what shape the linkers should be, as well as the chemistry that should be used to get these linkers to assemble into the desired shape. In other words, MOF chemistry is sufficiently advanced that MOFs with crystal structures never before seen in nature can be engineered from the ground up.

The open framework structure of MOFs and other coordination polymers offers fantastic promise for applications that take advantage of the frameworks' surface area or enclosed pore space such as catalysis, gas storage, and separation. MOFs can have unusual properties which are not exhibited by other materials, such as negative thermal expansion (20), negative compressibility (21), and negative gas adsorption (22). The porous structures of MOFs have lots of surfaces available so gas molecules tend to stay inside the framework rather than out of it. This allows more gas molecules to be stored inside a MOF than in an equivalent empty volume at the same pressure. This has economically important applications such as allowing more natural gas to be stored in a vehicle's tank (23). One can also exploit these very same properties for gas separation, tuning pore size and adsorptive properties to target capture of certain gases.

MOFs could be made more useful and interesting if they could be designed with actuating moieties inside their structure. We call such a material a Metal Organic Responsive

Framework or MORF. There are a number of molecules capable (24) of changing shape and producing mechanical work in response to external stimuli. However, it can be difficult to exploit this ability to actuate. We propose that MOFs will provide a means of structuring actuable molecules in 3d space so that we can do useful things with them. We can harness the collective actuation of such molecules to effect pore size change or produce macroscale shape transformation. Changing pore size allows us to change adsorptive properties while macroscale shape transformation is useful for making actuators. There are a number of molecules that can turn light into mechanical work, but again it is difficult to obtain this mechanical work at the macroscale. Now if we could make a material that could actuate in response to light there are number of interesting applications. Light is nice, because it is easy to structure and project light. So if we can make sheets of photoactuating material we can put them under a projector and make them move around. It has been proposed that one could make a very large, yet lightweight, telescope using a sheet of photoactuable material with one one side coated in reflective material (25). One could use laser projectors to control how the sheet deforms and make a large telescope mirror with variable focus. For a number of applications, it is desirable to have a change in adsorptive properties. For gas storage, we want a state that preferentially adsorbs gas when being filled and desorbs it when it is desired for the gas to be released. These are conflicting properties, higher attractive forces inside the structure allow us to store more gas, but also prevent that gas from leaving the structure when we wish to use it. We can overcome this if we can alter the adsorptive properties. An illustration of how MORFs might be used for gas storage is show in Figure 1.3. Storage is not the only application that exploits a change in adsorptive properties. Another is gas separation. Gas separation is an important and difficult problem. We may wish to filter CO₂ on a space station so that astronauts may breathe, or in a more down to earth application concentrate oxygen from the air for people with lung diseases. We might make our MORF have an affinity for a gas we desire, pass our gas mixture over it then when it can no longer hold any more of the gas we desire to separate, close off gas flow and recharge it by actuating our MORF. This might be done directly with an ambient energy source such as sunlight. Figure 1.4 shows an

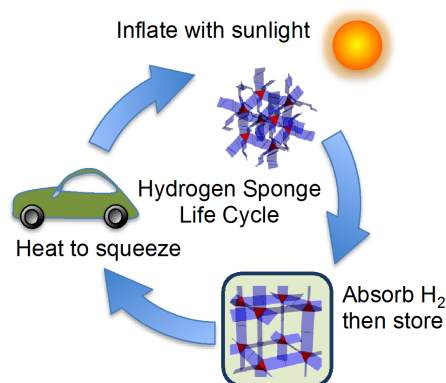


FIGURE 1.3: an example of how a MORF might be used for hydrogen gas storage

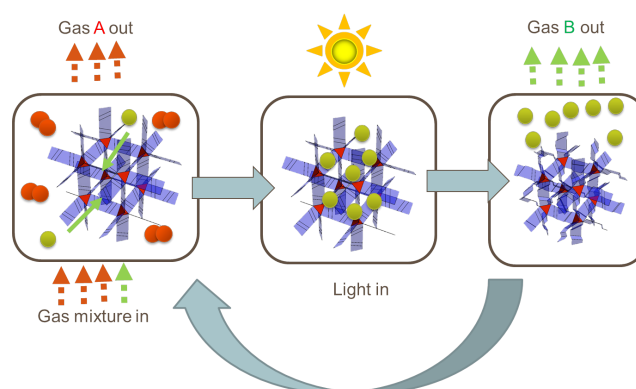


FIGURE 1.4: How a MORF might be used for gas separation

example of this. These are only a small sampling of the countless applications MORFs may have. It should be possible to realize these applications relatively rapidly due to the modular nature of MOF chemistry. As demonstrated in the section above, MOF chemistry has gotten to the point where MOFs can be designed to have a crystal structures that have never been observed before (19). The problem is no longer how we should make a structure, but what structure we should make. An approach is needed for designing these active MOFs. However, as far as I am aware, no one has developed methods for designing such a material. This work sets out to do just that, developing methods to rapidly design shape changing MOFs.

1.1 Molecular Design

Using computers to assist in finding valuable compounds has mainly been employed in the field of drug design, although there have been some recent efforts to computationally design MOFs. It is worth spending some time discussing drug design/development simply because it is currently the biggest application of computational molecular design methods.

There is no single method of drug design, much in the same way as there is no single method to invent new things. One of the biggest problems with drug design is that we do not understand the interactions between chemistry and biology all that well. When we first start testing the prototype of a mechanical system such as a vehicle, we rarely encounter any issues that require the whole vehicle to be completely redesigned from scratch. In drug design, this is not the case; 40% of drugs that reach phase III clinical trials fail (26). That is, drugs that have been tested in humans before, subsequently fail after being tested in larger populations and for longer time periods. 90% of drugs that are tested for the first time in humans fail. There is a need to continuously generate new candidates in this process, in order to compensate for the high failure rate. Due to the organic nature of the design process, it is best to discuss the design problem itself and the computational methods as they are used in different stages.

1.1.1 Drug Design

Drug design (27) (28) is the problem of finding a molecule that binds with a specific site on a protein or other biological element, subject to constraints, such as biotoxicity. This can be likened to the problem of designing a key to fit into a lock. Continuing with the lock analogy, drug design can be divided into two different domains: whether we know what the key looks like, or whether we know what the lock looks like. The first approach is called ligand-based design and the second approach is called structure-based design.

In the beginning of the drug design process, we typically have a specific target molecule whose behavior we would like to modify. Most often, drug designers do not know how to do this, so they will physically test whether or not any of potentially thousands of different compounds modify the behavior of the the target molecule. Based on how well each of

the tested compounds did at modifying the behavior, we can infer which substructures are related with high performance. We call the relationships between these substructures and performance metrics quantitative structure activity relationships (QSARs) (29).

We can use these QSARs to help pick new molecules to physically test for performance. In addition, these QSARs might be used to determine what the key looks like or how certain chemical groups with different functionalities should be positioned relative to each other in 3D space. The functionality in this case tends to refer properties such electric charge, hydrophobicity, hydrophilicity, and other bonding properties. Opposite electrostatic charges will of course tend to attract each other and like hydrophobic/philic regions will tend to attract each other. This is referred to as a pharmacophore (30).

It is also worth noting that determining pharmacophores and QSARs is often not an entirely automated process. One can computationally design molecules that meet the intended geometric requirements by placing fragments with the required properties at the required locations relative to each other, and then finding molecules that link these fragments together. This was done in the NEWLEAD system (31), where fragments at specified locations and orientations are provided as inputs. The PhDD system (1) expands upon this and starts from a more abstract representation of the pharmacophore. Portions of the pharmacophore are replaced with fragments that have the necessary properties at these locations, rotated around to avoid interfering with each other, and then molecules to link these fragments together are found.

One important concept in both of these systems is the concept of bonding vectors. A database of bridging fragments is maintained whose look up table consists of the vectors between the connection points. An example of this is shown in Figure 1.5. This allows all possible connection points from a database to be tested quickly. These approaches might be used in an iterative manner to obtain better estimates of the structure of a pharmacophore.

Once we understand where our molecule must bind, we can then use structure-based designed approaches which use the 3D structure of the biomolecule that we are targeting. (32) One way this is done is simply to test if molecules from a database can fit into the

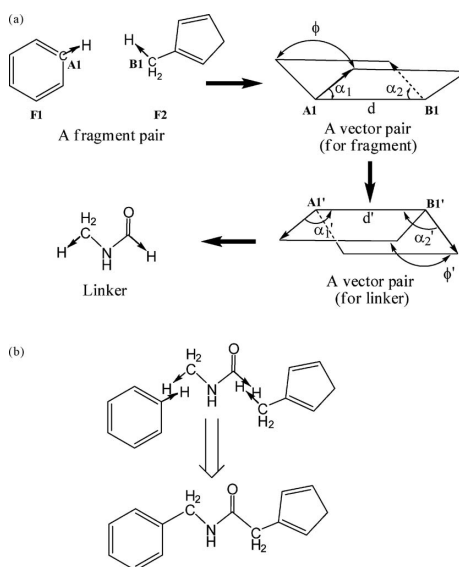


FIGURE 1.5: A compatible linker molecule can be found quickly by searching looking for molecules that have connection point vectors d' apart with angle α between them, taken from (1)

binding site. We can employ basically the same methods used to design molecules that meet pharmacophore requirements in structure based drug design. Instead of putting fragments at pharmacophore locations, we put fragments in locations in the binding pocket that we know we can bind to. We call these interaction sites.

Another is to start with a location on the binding site that we know we can bind to and iteratively add parts of other molecules, or fragments, in an attempt to get stronger binding. Systems such as LUDI (33), employ geometric rules to follow the contours of the binding site, choosing the next fragment, such that it maintains contact with the binding site. Contact in this case being that hydrogen bond acceptors on the fragment are put within range of hydrogen bond donors on the molecule, hydrogen bond donors on the fragment are put within range of hydrogen bond acceptors on the molecule, and that hydrophobic pockets are filled with similar hydrophobic fragments. It is worth noting that this is often done in an interactive process with the user specifying which interaction points on the molecule should be taken into account.

The SPROUT software (34) does the same thing albeit in a more automated manner. The user provides SPROUT with a number of target binding sites and SPROUT generates skeletons, or molecule representations that have hybridization type specified but not elements, that attach to all binding sites using a tree search method. An initial fragment is placed, and fragments are iteratively added to the structure towards the target binding until all binding sites are filled or the last placed fragment interferes with the binding site. From there, atoms are substituted into the skeleton to meet binding target connection requirements and hybridization requirements. This entire process uses the A* optimization algorithm (35) (36). Everything that is found to be feasible is evaluated and the highest scoring results are presented to the user.

1.1.2 MOF and similar system design

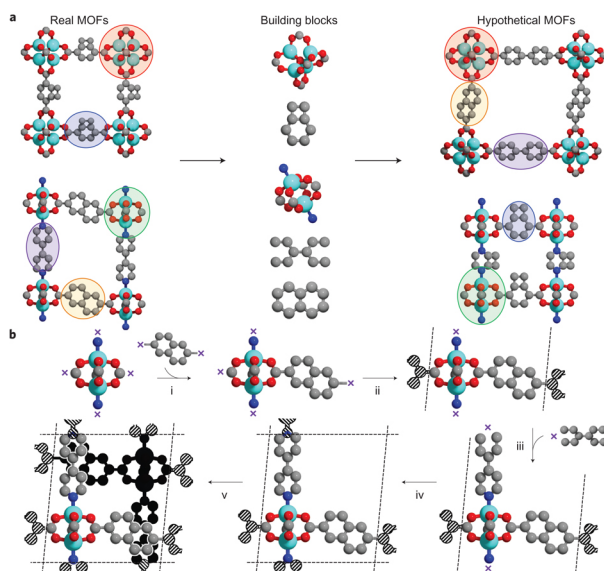


FIGURE 1.6: generation process for hypothetical MOFs taken from (2)

More recently, there have been computational efforts to design MOFs with specific desirable properties. Notably, Wilmer et al. (2) generated a database of 137,953 hypothetical MOF structures from a set of 102 building blocks and evaluated their properties. MOFs were generated in a 'Tinkertoy' like manner, where building blocks with the same shape were

substituted into one of four different crystal topologies to form new MOFs. Building blocks consisted of organic linkers, secondary building units (SBUs), and functional groups. The inorganic SBUs were placed first, then linkers, then another linker if possible and iterated through functional groups. Any generated MOFs with contacting atoms were discarded. This is a purely geometric process and energy minimized structures were not generated. They calculated geometric properties such as surface area, pore volume, and pore size distribution. They also estimated the methane storage capacity of the generated MOFs with grand canonical Monte Carlo (GCMC) holding the MOF structure static and using UFF for the atomic potentials.

From the results of this, they were able to identify a metal organic framework with high methane storage capacity which they synthesized and found to be in good agreement with the properties they predicted. They also used this database to derive structure property relationships for carbon dioxide separation and capture (37), structure property relations for MOFs that have high Xe/Kr selectivity (38), and identifying MOFs for high hydrogen storage capacity at room temperature (39). Gómez-Gualdrón et al. (40) generated more hypothetical MOF structures of PCU topology and used them to estimate the limits of methane storage in nanoporous materials, concluding that it will be difficult to meet ARPA-E methane storage goals.

Similar to the above work, Martin et al. (41) developed a method to design MOF-5 analogs with high methane storage capacity that use commercially available linker molecules. They queried an online database for all commercially available molecules that had two carboxylates, then they removed all molecules that were salts, had disconnected portions, or too many atoms. They then tested these molecules for rigidity by generating a number of conformer and rejected them if the distance between carboxylates varied by more than 0.65 angstroms. From these, they took the lowest energy conformer found and forced the carboxylates to be collinear and then tested if there were any collision between the carboxylates and other atoms in the molecule. They then put these linkers into MOFs with MOF5 secondary building units and PCU topology. Because some of their linkers found were asymmetric and

could adopt multiple different configurations inside a MOF, which could cause different gas adsorption properties to be measured, they had to generate as many distinct configurations of linkers as possible. Any configurations which weren't sufficiently distinct as determined by comparisons of pore network descriptors or had intersecting molecules were rejected. The MOFs were energy minimized and MOFs with SBUs that were far too deformed were removed, and methane adsorption properties for the remaining candidates were calculated with grand canonical Monte Carlo.

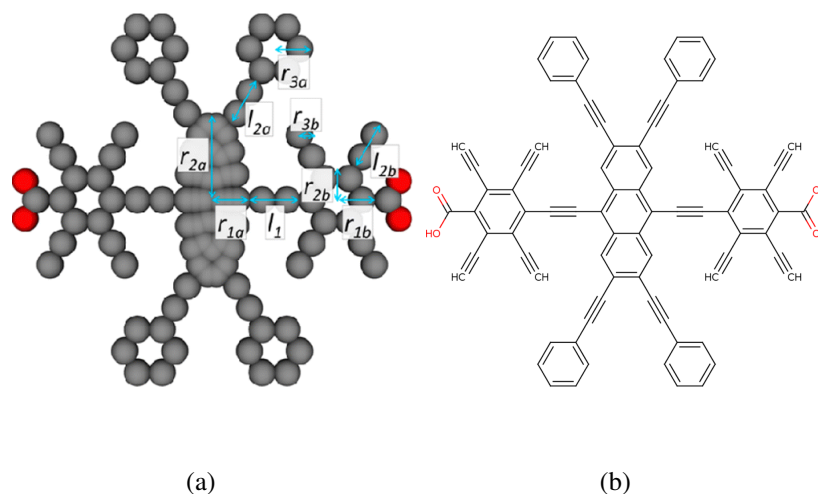


FIGURE 1.7: Abstract continuous representation(a), and its associated projection onto a real molecule(b), taken from (3)

Martin and Haranczyk (3), implement a method for designing MOFs based upon optimization an abstract continuous representation of linker geometry. An example of an abstract continuous representation of a ditopic linker is shown in Figure 1.7a. The authors claim that this representation allows smooth transitions through the space of artificial molecules and that these abstract representations can be projected to real molecules as shown in Figure 1.7b. This abstract representation is made of branching structures the authors call junctions. A junction is described with r_1 and r_2 which specify an ellipse at the center of the junction from which branches with length l_2 extend from to leaves with radius r_3 . Junctions are separated from other junctions with dimension l_1 . The ellipses, circles, and branches defined are discretized

into pseudo-carbon atoms of uniform size. These pseudo carbon atoms have the same properties as carbon atoms such as van der waals radius and van der waals, without any of the restrictions on bonding or structure real carbon atoms have. This allows properties such as surface area to be calculated and reasonable minimum energy configurations to be found once these linkers are put into a MOF of defined topology. They perform discretized steepest ascent, using the values of the neighbors of a given MOF to calculate the gradient and determine the new linker dimension and thus new MOF. They perform ten different optimization from different random initial linker conditions for each topology. They optimize for maximum surface area and show that they can obtain possible MOFs based on the abstract representation with surface areas reasonably close to that of the abstract representation. Downsides of this method include difficulty projecting down to real molecules, the need for different abstract representations for each different topology or different design problems. Bao et al. (4), devel-

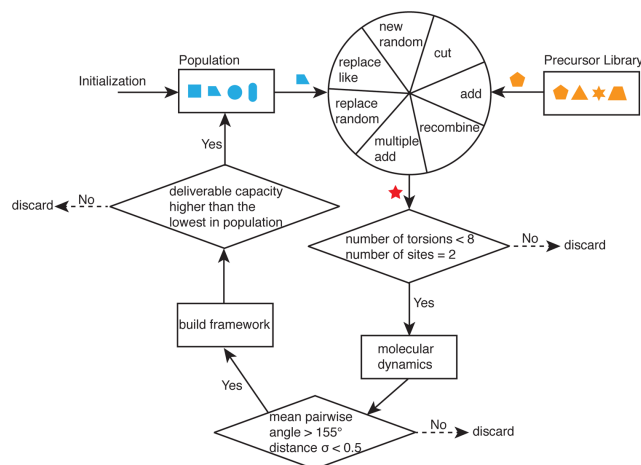


FIGURE 1.8: Search process for (4)

oped a method for finding MOFs with high gas storage capacities using a genetic algorithm. Candidate MOFs were described as a genome which specified which virtual chemical reactions and reactants were used to make a MOF linker and what framework topology that linker formed into. A key contribution of this work was a set of pre-evaluation metrics to determine whether or not a candidate linker should be considered for further evaluation. These pre-evaluation checks determine whether or not the linker has two carboxylates, if the linker has

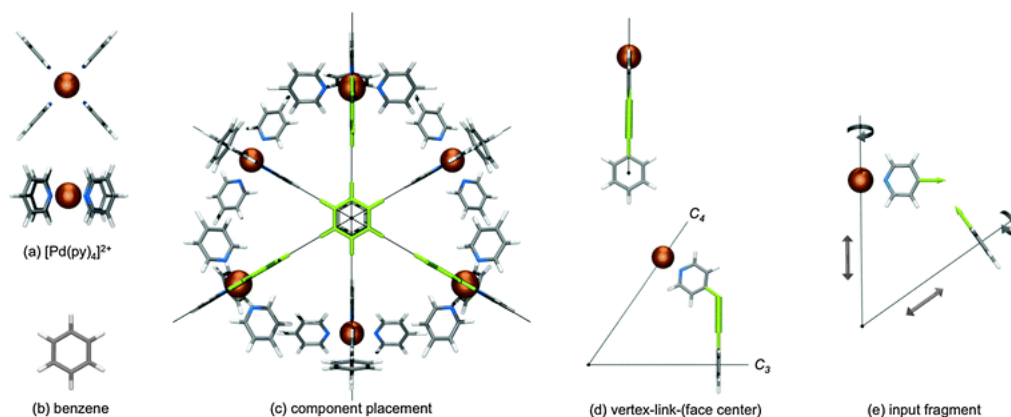


FIGURE 1.9: Generation process for metal organic polyhedrons (5)

more than a specified number of rotatable bonds, if the average angle between the carboxylates is within a certain range of values, and if the distance between the carboxylates varies too much. The rotatable bond number limits and distance variance checks are designed to prevent flexible linkers from being evaluated. The virtual chemical reactions used are based on so-called click chemistry. Using virtual chemical reactions has the advantage of ensuring that every generated molecule is likely to be synthetically feasible. One problem with implementing this for our work is that it is difficult to ensure that desired substructures are incorporated into the linker molecule. The ultimate end goal of our work is to be able to generate MOFs which are photo-active and thus, must include photo-active moieties into their structure. The problem with virtual chemical reaction is that it is difficult to ensure that a specific moiety is incorporated into the structure. If we use something akin to the genetic algorithm above, the genetic algorithm must expend effort finding a way to generate these desired moieties. For some photoisomers, this might be fairly complicated.

Young (5) developed a computational design methodology (42) (43) (44) for making metal organic polygons and polyhedra with specified shapes. Metal organic polygons and polyhedra are the non-periodic analogs of metal organic frameworks. Whereas a MOF consists of organic molecules held together by metal ions that form a structure that repeats along at least one axis, in metal organic polygons and polyhedra the linkers and metal ions form

together into a 2d or 3d shape. This might be a triangle, square, tetrahedron, prism, dodecahedron, icosahedron depending on what linkers and metal ions are used. The approach developed can only make shapes that have some degree of symmetry. The approach that was developed is somewhat similar to the one used to generate hypothetical MOFs and some drug design approaches. Vertices of a target shape are replaced with known metal organic complexes, stable substructures that organic molecules and metal ions will form into, and or defined organic fragments with proper symmetry and sufficient connection points to match the degree of the vertices. The user specifies which parts of the fragment and metal complexes will attach to edges of the polyhedron. The metal complexes and fragments are then translated and rotated relative to each other about symmetry axes of the polyhedron through a defined distance and rotation angle. At every step, the bonding vectors between the locations the user specified are used to select the appropriate molecule from a database capable of linking the two points together if any exist. Every linking molecule found allows a new polyhedron to be made.

Now these approaches so far are solving what are arguably static geometric problems. Drug design is based on finding the best shape so that a molecule fits into a binding site. Finding MOFs with high gas adsorption properties is essentially the problem with finding the best geometry to accommodate the most gas molecules. Our problem of finding MOFs that change shape is more challenging. Computational Design Synthesis offers some interesting possibilities for designing MOFs. Computational Design Synthesis is the process of using computers to synthesize new designs to solve a problem. This is typically accomplished by developing a method for computationally representing possible designs. This may include specifying components and how they can be connected together or describing a means for representing the geometry of three-dimensional shapes. We use this representation to generate new designs, which are evaluated for the desired performance parameters. Based on the success of these designs, we use the representation to generate new designs which are likely to have improved performance parameters. Typically this is done with an optimization algorithm over a number of iterations. Computational design synthesis has been employed to design

linkages (45), gearboxes (46), and conceptual designs for mechanical systems (47) .

-

2 Methods

2.1 Graph Grammars

Most systems that we care to design consist of discrete components connected to each other in a defined manner. For example, mechanical systems consist of gears, links, joints, motors, and other components connected to each other. One way of representing these systems is as graphs describing the connectivity of components. A graph is a method for abstracting the characteristics of the elements that make up a system. In a graph, nodes are defined to represent individual elements, and arcs are defined to represent connections or interactions between elements. Additional information is added to these nodes and arcs in the form of labels and variables. Labels provide string information to describe what the node represents and variables describe node numerical parameters. In the case of a mechanical gearbox, for example, we might specify that a node is a gear with a label and describe how many teeth it has with a variable. Additionally, we may wish to define characteristics for the entire system, and this information would be defined as a global label or variable. We might specify that our gearbox is to be made of a certain material and has a certain gear ratio.

Graph grammar rules are methods for transforming graphs by replacing subgraphs with new subgraphs. Rules consist of a subgraph to be found in a larger graph, (We refer to this as the “left-hand-side” in the language of graph transformations), and what that subgraph is to be changed into (the “right-hand-side”). Finding the locations of the left-hand-side in a larger graph is known as rule recognition. Transforming one of the found locations is referred to as rule application.

A graph grammar rule specifies a substructure and what that substructure is to be transformed in to. The act of finding this substructure is called rule recognition. Changing this substructure is called rule application. We call the substructure we want to find the left hand side and what this substructure is to be changed into the right hand side. These are denoted with L and R respectively. There is also the context, denoted with K, consisting of the parts that don't change when a rule is applied.

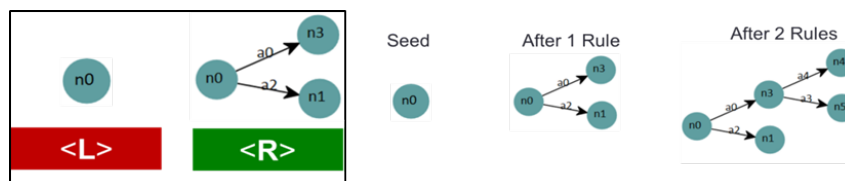


FIGURE 2.1: A graph rule and its associated applications

A very simple graph grammar rule is shown in 2.1. The rule finds a node (the blue circle) and adds two arcs (arrows) with nodes on them. Repeated applications of the rule on a seed graph consisting of a node, will lead to the formation of a tree-like structure as shown above. Graph Grammars can be used to represent feasible transformations constraints in systems. For example, that an wheel can only connect to an axle. While we could represent these constraints without graph grammars, graph grammars enable us to rapidly iterate.

GraphSynth (48) is a software tool which implements graph grammars developed by Matthew Campbell and the Design Engineering Lab. All graph grammars were implemented in GraphSynth.

One of the primary contributions of this work is a set of graph grammar rules for constructing molecules likely to be chemically feasible. In order to do computational design synthesis, we need some way to structure the design space in a way the computer can work with. In addition, it is also useful to prevent the generation of solutions that are not feasible or do not make sense. For example, creating a car with an engine connected to nothing or a molecule that is not thermodynamically stable. Graph grammars offer a means of accomplishing this, describing the space of potential solutions in the form of transformation operations on graphs. In organic chemistry, most molecules obey what is called the octet rule. The octet rule specifies that in a stable molecules, all atoms will have eight valence electrons with the exception that hydrogen only needs two valence electrons. Bonds between atoms are formed from shared valence electrons, with a single bond typically being made of two electrons. As carbon has four valence electrons while hydrogen has one valence electron so carbon can attain eight valence electrons by bonding with four hydrogen atoms.

Initially, the rules were based around the octet rule. Each node in the graph had infor-

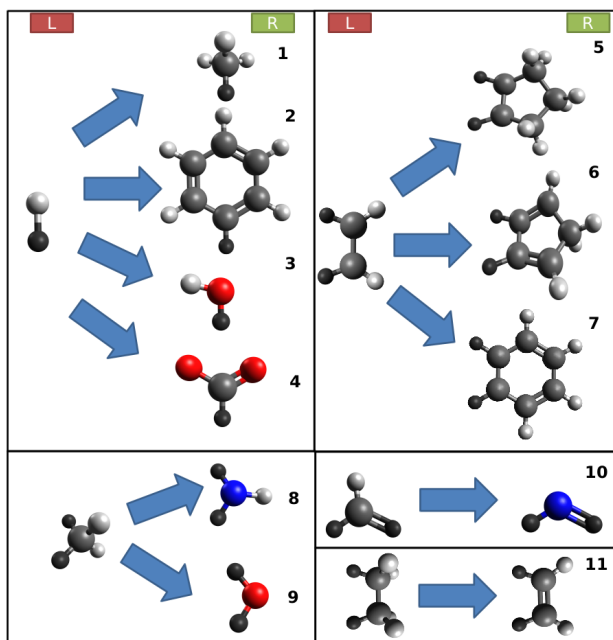


FIGURE 2.2: Rules

mation on how many valence electrons it had and how many valence electrons it was sharing. Rules utilized this information to determine if they could be applied to add new atoms or parts of molecules and updated the new valence electron count after applying. The problem with this was that this could not be purely implemented in graph grammars. Graph grammars can modify information on the node, but they do not perform mathematical operations such as sum. One could potentially implement graph grammar rules which could do this by making variants of the rules that only recognize if there are one electrons, two electrons, three electrons, etc, but this would be computationally wasteful. So these were implemented using Graphsynth's additional recognize and choose functions. The programmatic code of an additional function checks if a rule can be recognized and what to do after that rule is applied. However, using such things made adding new rules cumbersome. Every new rule required pieces of C# code to be associated with it for checking how many valence electrons certain atoms had and updating valence electrons on affected atoms. For this reason, this approach

was abandoned very early on. The rules were designed such that starting out with a seed molecule that obeys the octet rule, subsequent rule applications will always yield a molecule that obeys the octet rule.

In order to generate chemically feasible structures, a novel graph grammar based around hybridization is used. Hybridization is the number of hybridized orbitals that an atom has. Hybridization can provide information on how atoms are located with respect to each other. For example an sp^3 carbon atom, an example of which is shown in the right hand side of rule 1, will always have 4 single bonded atoms arranged in a tetrahedral shape. In addition, hybridization offers a useful and flexible basis upon which to manipulate chemical structures, because atoms with the specified hybridization will always have the same bonding configuration. Using hybridization ensures that every chemical structure generated will be chemically plausible. In order to avoid using a large fragment database, the rules are structured to define plausible fragments from previously made molecules. For example, instead of having fragments for every possible 6 member ring consisting of carbon and nitrogen, all of these possibilities can be enumerated by applying rule 6 repeatedly on a graph consisting of a benzene ring. One advantage of this is molecules can be incrementally improved. Chemical structures can be represented as a graph consisting of nodes which represent atoms and arcs which represent bonds between atoms. The type of atom a node represents is given by label on the node. In addition, nodes also have a label specifying the hybridization state they are in. The graph grammar rules also have additional conditions built into them to prevent chemically infeasible or undesirable structures from being produced. For example, an oxygen atom cannot be connected to another oxygen atom because such a structure is unstable. There are three basic types of rules, add molecule, fuse ring, and transmute. Add molecule rules replace a hydrogen with another molecule. Fuse ring rules add a ring molecule onto another molecule that is a ring. Transmute rules change a carbon atom into another type of atom or change its hybridization. The rules also have conditions on them to prevent the formation of chemically undesirable structures. Oxygen-oxygen bonds are not permitted to form as this would result in a highly unstable, and in some cases explosive, structure known as a perox-

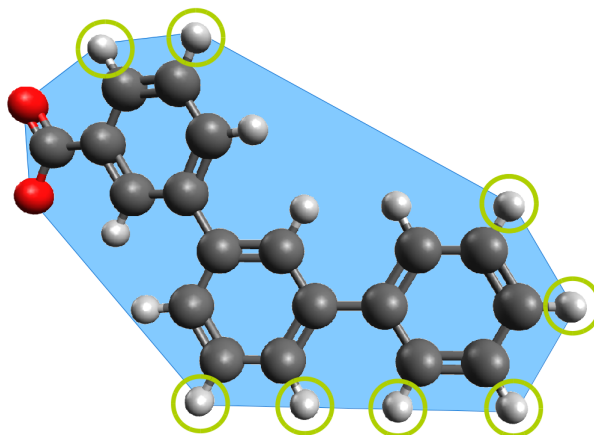


FIGURE 2.3: Convex hull of molecule

ide. In addition, nitrogen-nitrogen bonds are also not permitted, to prevent the grammar rules from making a structure entirely out of nitrogen. While there exist a number of structures with stable nitrogen-nitrogen bonds, it is better to exclude them than to allow them.

These grammar rules allow the generation of a wide variety of chemical structures containing carbon, nitrogen, and oxygen relevant for metal organic framework linkers, while ensuring the generated structures are likely to be chemically feasible. The rules can only add to a molecule or modify carbon atoms that are already in place. By only being able to add molecules or modify carbon atoms already in place, additional rule application will always yield a different molecule. This is a desirable property when carrying out a tree search. These rules cannot generate every possible structure consisting of hydrogen, carbon, nitrogen, and oxygen, but they can still generate a wide variety of structures. Allowable ring systems are rather limited. These rules cannot make anything other than 5 or 6 membered rings. In addition, these rules cannot make polycyclic structures where atoms are members of more than two rings. The rules cannot make structures that form macrocycles, or large ringed structures. However, being able to do such structures might not add much benefit and might be more difficult to synthesize. The rules could be modified to include such structures. We also cannot make disconnected structures such as rotaxanes.

A further advantage of the rule based approach for fast design is that the rules place atoms in approximately the correct geometric location, leading to faster energy minimization of the generated structures. When generating linkers, rules that add to the molecule are only applied on the convex hull of the molecule. This makes it less likely that a structure that interferes with itself will be built. Points on the convex hull point outward from the molecule as can be seen in Figure 2.3.

2.2 Review of Molecular Dynamics

A key part of this work is being able to model molecules. We refer to this as molecular dynamics. We model molecules as a system of particles which interact with each other through a set of interatomic potentials. A forcefield is a means of assigning interatomic potentials.

The Universal Force Field is a generic force field model that can provide reasonable inter-atomic potential coefficients for practically every element. (49) This is well suited for our domain as metal organic frameworks can involve a large swath of elements on the periodic table. MOFs have been made using elements as high in atomic number as 93, neptunium (50). UFF uses simple relations between atoms. Atoms are assigned atom types based upon what they are bonded to or manually specified. UFF was extended to better describe metal organic frameworks with UFF for MOF or UFF4MOF. The overall form of UFF4MOF is the same as UFF, but it includes atom types more suitable for reproducing the geometry of MOFs. UFF4MOF and UFF have been found to work fairly well for estimating the mechanical properties of MOFs (51). UFF considers five different types of interactions between atoms, non-bonded interaction, bonds, angles, dihedrals, and impropers. Total potential energy is given as:

$$E = E_{stretch} + E_{bend} + E_{torsion} + E_{improper} + E_{non-bonded} \quad (2.1)$$

$$E_{stretch} = \frac{K_{IJ}}{2}(R_{IJ} - R_0)^2 \quad (2.2)$$

Angles are interactions between three atoms where the angle between three atoms dictates energy of the interaction. Angles occur between two atoms that are bonded to the same atom. In general two bonds to the same atom will try to maintain an angle between them due to repulsion between bonding orbitals. In UFF, we have several different potentials for angles depending on the coordination environment of the angle. For the linear case we use:

$$E_{bend} = \frac{K_{IJK}}{n^2} [1 + \cos(n\theta)] \quad (2.3)$$

For all other cases we use:

$$E_{bend} = K_{IJK} [C_0 + C_1 \cos(\theta) + C_2 \cos(2\theta)] \quad (2.4)$$

With C_1, C_2, C_3 defined by:

$$C_2 = \frac{1}{4 \sin(\theta_0)} \quad (2.5)$$

$$C_1 = -4C_2 \cos(\theta_0) \quad (2.6)$$

$$C_0 = C_2 (2 \cos(\theta_0)^2 + 1) \quad (2.7)$$

Torsions or dihedrals are interactions between four atoms. Dihedrals can be taken as the

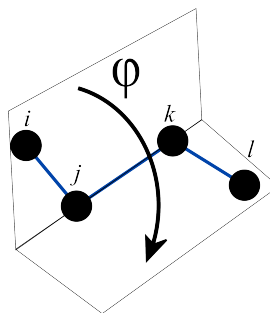


FIGURE 2.4: A dihedral going through atoms i, j, k,l

rotational stiffness of two bonds about a central bond. A diagram of a dihedral is shown in Figure 2.4. The energy of torsions in UFF is given by:

$$E_{torsion} = \frac{V_{IJKL}}{2} [1 - \cos(n\phi_0) \cos(n\phi)] \quad (2.8)$$

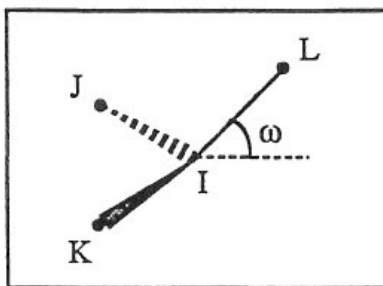


FIGURE 2.5: An improper going through atoms i, j, k, l, taken from LAMMPS documentation

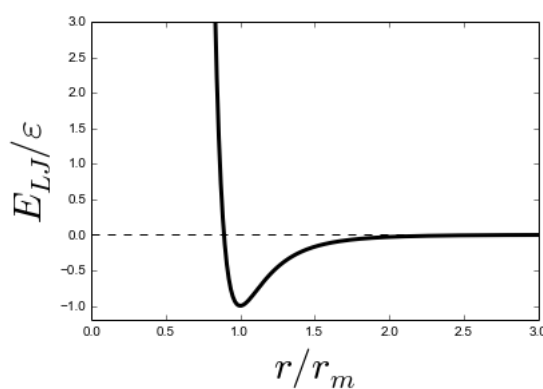


FIGURE 2.6: Lennard Jones potential normalized against ϵ and r_m

Improper are also interactions between four atoms where how far off all atoms are from being planar dictates the interaction energy. atom's distance from that plane. Figure 2.5 shows how this is determined. Atoms IJK define a plane and the angle the IL axis makes with that plane dictates interaction strength. The energy of improper in UFF is given by:

$$E_{improper} = K_{IJKL}[C_0 + C_1 \cos(\omega) + C_2 \cos(2\omega)] \quad (2.9)$$

Non-bonded interactions model attraction and repulsion forces between atoms such as Van der Waals forces. This is modelled in UFF with a Lennard-Jones potential:

$$E_{LJ} = \epsilon \left[\left(\frac{r_m}{r} \right)^{12} - 2 \left(\frac{r_m}{r} \right)^6 \right] \quad (2.10)$$

In this potential which is shown figure 2.6, there is minimum energy at r_m away from the atom with a depth of ϵ . Closer than the this point and energy approaches infinity, further

from this point and energy approaches zero. What this implies is that atoms will tend to attract each other to an equilibrium point. This is perhaps one of the greatest differences between design at the macroscale and design at the nanoscale. While bonds, angles, dihedrals have some analog at the macroscale, we do not have an analog for non-bonded interactions. It is rare to find mechanisms where attractive forces between parts plays a significant role. In order to save computation, these interactions are not taken into account if two atoms are beyond a certain cutoff distance. Unlike the name implies, in some force fields 'non-bonded' interactions are also applied between atoms that are bonded.

Unlike design at the macroscale, we must use discrete components and we are restricted in how we can connect these discrete components together. We cannot change the stiffness of a structure by making half an atom or using arbitrary amounts of bonds. The presence of non-bonded interactions is perhaps one of the greatest differences between design at the macroscale and design at the nanoscale. While bonds, angles, dihedrals have some analog at the macroscale, we do not have an analog for non-bonded interactions. It is rare to find mechanisms where attractive forces between parts play a significant role.

3 Designing Linkers with Specified Mechanical Properties

De novo design approaches (28), (27) have long been used in the pharmaceuticals industry to design molecules with pharmacological activity. De novo design approaches have also been used to design molecular cages (5). In both these problems, molecules are designed to have a specified shape. However, there has been little work on designing molecules to achieve specified mechanical functionality. In this work a method was developed for optimizing linker-like molecules to have specified force displacement curves.

The previously described graph grammar rules, except for rule 3, were used to construct linker-like molecules. All molecules are built from a seed molecule, shown in at the root of Figure 3.1, consisting of a carbon atom bonded to two oxygen atoms and a hydrogen atom. This is known as a carboxylate group/formate in chemistry and represents the part of the linker that bonds to metal ions. To make a linker-like molecule, a set of rules is sequentially applied to the seed graph. Force displacement curves for the molecules were measured by

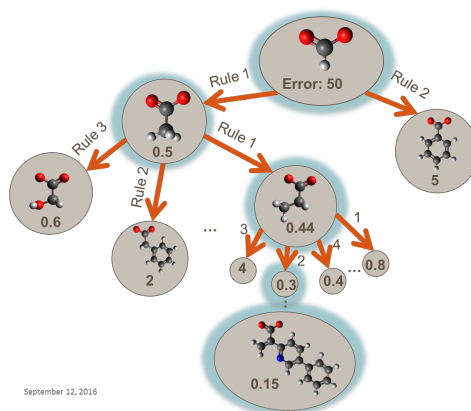


FIGURE 3.1: Illustration of search process

simulating them in LAMMPS (52) using the universal forcefield (49). Molecules generated by the above approach are often not in a minimum energy configuration, therefore it is necessary to carry out energy minimization. If a molecule is not in the minimum energy configuration it

might produce exceptionally high forces or lead to force curves that do not make sense. This minimum energy configuration was also used for further generation of molecular structures to prevent errors. To determine the force displacement curve portions of the molecule were moved relative to each other through a specified displacement and force was measured. After energy minimization the atom furthest from the carbon atom in the carboxylate group is found and specified as one of the end points of the molecule. The oxygen atoms in the carboxylate and the furthest atom are then fixed. Next the furthest atom is displaced 0.1 angstroms, the molecule is energy minimized, and the magnitude of the reaction force is measured. This is repeated until the atom has moved through the required displacement.

The value of this approach was demonstrated by using a simple tree search algorithm known as best first search to rapidly find molecules that have a desired stress-strain response. This approach is able to quickly and efficiently find molecules with a mechanical response close to the desired response. This took around 8-14 iterations to find molecules with desired properties, taking around 1-8 minutes to accomplish this. In addition an automated simulation and evaluation approach was developed in conjunction with this work. An example of a molecule produced by this method is shown in Figure 3.4a.

Many times the rules can produce molecular structures that have already been evaluated. Given that the simulation takes a relatively long time to run, it would be useful to compare the graph to be evaluated with every graph that has been evaluated and look up the evaluation function. The problem with this is storing all the graphs that have been evaluated would take up a large amount of memory and comparing graphs is a computationally difficult problem. To solve this, graphs were converted to the SMILES molecule representation format (53), (54), (55). SMILES is a format that enables practically any molecule to be represented as a string of characters, which is very memory efficient. Before a molecule is evaluated, it is converted to a SMILES string and compared with a list of SMILES strings for all molecules that have been evaluated. If the candidate molecule is found to be a molecule that has been previously evaluated the, corresponding force curve data is looked up in a table and used for further evaluation. If the candidate molecule is found to not have been evaluated,

then its SMILES string and force curve are stored in a file. Candidate molecules are simulated in LAMMPS (52) using parameters from the Universal Force Field (49). Before the simulation was carried out it was necessary to carry out energy minimization of the molecule. Energy minimization is a process that optimizes the positions of atoms in a molecule so as to minimize potential energy of interatomic potential functions. Molecules generated by the above approach are often not in a minimum energy configuration, therefore it is necessary to minimize them. This minimum energy configuration was also used for further generation of molecular structures to prevent errors. After energy minimization the atom furthest from the carbon atom in the carboxylate group is found and specified as one of the end points of the molecule. The oxygen atoms in the carboxylate and the furthest atom are then fixed. Next the furthest atom is displaced 0.1 angstroms, the molecule is energy minimized, and the magnitude of the reaction force is measured. This is repeated until the atom has moved through the required displacement.

Once the force curve is obtained the molecule is scored based upon how much it differs from the specified force curve using root mean square error (RMS). The scores of each molecule are used to guide the generation of new molecules. To guide the search process, best first search was used. To do this a list of evaluated candidates sorted by RMS error is maintained. The candidate with the lowest RMS error is removed from the list, so that it is not evaluated again, and all recognized graph grammar rules are applied to the candidate to generate new candidates. Out of these new candidates, candidates with more than 65 atoms are removed to depth limit the search. These new candidates are evaluated and added to the list. This process is repeated for a user specified number of iterations, which in all cases presented here is 30.

The above mentioned experiment was implemented on a laptop computer with a 2.4 GHz processor with 4 Gb of memory, each search process was run for 30 iterations with an atom limit of 65 atoms. To test designing molecules to have a specified force curve three different cases were considered a constant force curve, a linearly increasing force curve, and an exponential force curve. The equations for these force curves are shown in Table 3.1. The

TABLE 3.1: Force curve equations

Curve	Equation
linear force	$F = 0.1x$
constant force	$F = 0.02$
exponential force	$F = 0.01 * 2^{10x}$

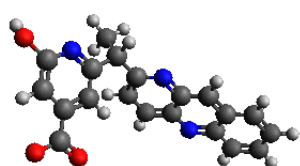
TABLE 3.2: Performance

Curve	Error	Iterations	Time to Solution(seconds)
linear	0.0134	14	502
constant force	0.0044	10	250
exponential	0.1532	8	42

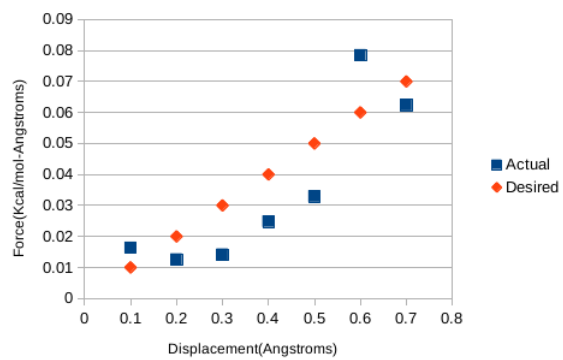
force displacement curves of optimized molecules were found to have RMS errors between the specified force curves ranging from 0.004 to 0.15 as shown in Table 3.2.

The linear force molecule is shown in figure 3.2a and appears to behave somewhat like a macroscale compliant mechanism containing stiff components that deform little and more flexible components that bend. The groups of fused rings are exceptionally rigid, while the methyl group was not. While it at first appears that the methyl group is acting as a revolute joint, the simulation of this mechanism shows that rotation is minimal and bending dominates. This molecule is not suitable for integration into a metal organic framework as much of the molecule is perpendicular to the carboxylate group.

In the exponential force curve molecule, which is show in Figure 3.3a, the two benzenes are compressed slightly off axis with respect to each other. Upon compression, this leads to behavior that produces an exponential force displacement curve with RMS error of

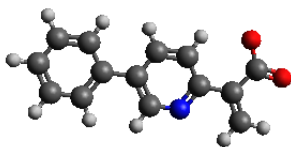


(a)

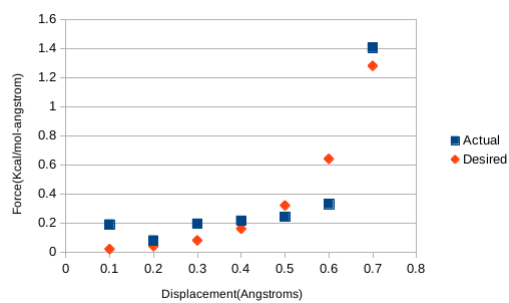


(b)

FIGURE 3.2: Linear force curve molecule(a) and its associated force displacement curve(b)



(a)



(b)

FIGURE 3.3: Exponential force curve molecule(a) and its associated force displacement curve(b)

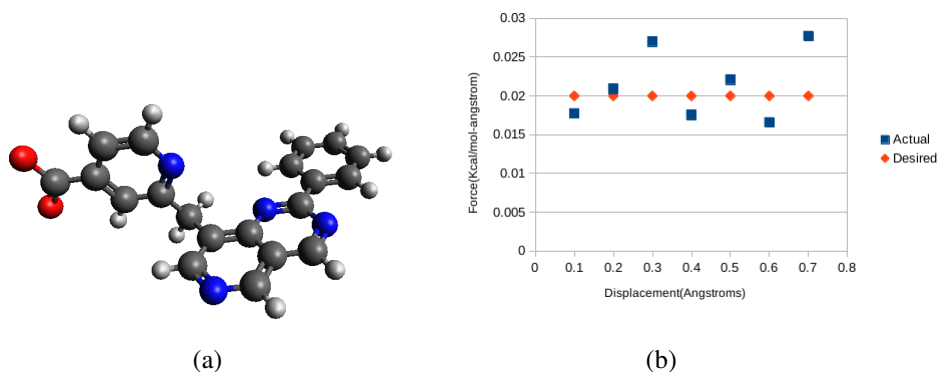


FIGURE 3.4: Constant force curve molecule(a) and its associated force displacement curve(b)

0.15. While this may not seem very close, consider that it is difficult to change the mechanical properties of individual bonds and atoms. The constant force curve molecule, shown in Figure 3.4a along with its force displacement curve shown in Figure 3.4b, appears to be similar in structure to the linear force molecule, however, its behavior is different. Unlike the linear force molecule, the two benzene rings connected to the methyl rotate significantly about their bonds to methyl group rather than bending as was the case with the linear force curve molecule. In both the linear force molecule and the constant force molecule, there is significant bending of the carboxylate group about the fixed oxygen atoms. The constant force molecule found seems to be analogous to a macroscale compliant constant force mechanism designed by Weight (6) which is shown in Figure 3.5. Each of the benzene rings might be acting like the rigid portions of the mechanism, while the rotational energy barriers about the single bonds between the benzene rings might be acting like the small length flexureal pivots.

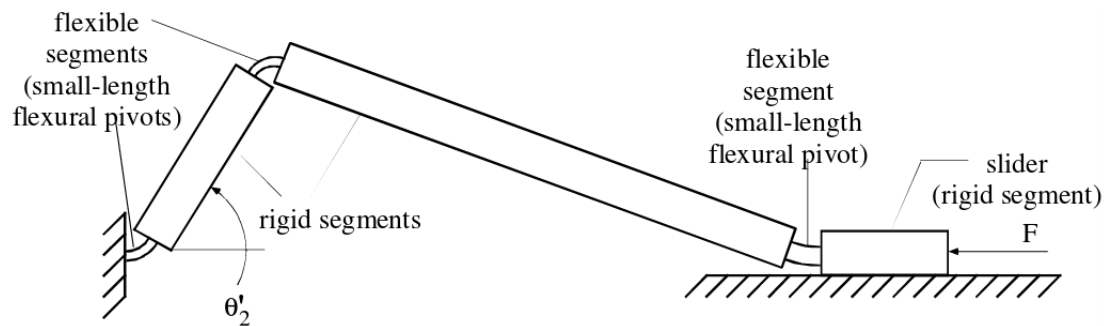


FIGURE 3.5: A compliant constant force mechanism from (6)

4 Fragment Characterization

In the aforementioned work, one optimized molecule is produced at the end of the search process. However, it can also be more useful to find features responsible for giving the molecule its desirable properties. Such features could be used to construct new desirable molecules faster. This work focused on developing a way to automatically infer what features make a good molecule for the hypothetical application of designing a molecule that of designing a linker for MOF that collapses at a specified pressure. This work was published as (56).

To generate linker-like molecules we randomly applied these graph grammar rules on a seed graph consisting of a formate molecule. This represents one of the carboxylic donor sites in our linker-like molecule.

Starting with this seed, the generation of a single linker-like molecule follows the following sequence of steps.

1. The replaceable subgraphs on the convex hull of the molecule are identified and a list of the permitted rule applications at each is compiled.
2. One of the permitted rule applications, excluding those that use rule 4 (termination of the linker-like molecule with a carboxylate), is selected randomly and applied.
3. The resulting molecule is annealed and relaxed in five separate molecular dynamics simulations and the the resulting lowest energy conformer carried forward to the next iteration.
4. Steps 1–3 are repeated to grow the candidate linker until some stopping criterion is reached. The stopping criterion could be based on properties of the linker, for example stopping after a desired molecule length or weight were reached, or when one particular grammar rule was selected for the second time (to ensure only one of a particular functional group). In this work many simple stopping criterion were used: stopping after

a set number of rule applications (typically 10), or stopping after a random number of rule applications (in range 2–10).

5. Finally, to ensure that a ditopic linker-like molecule is made, a carboxylate is added using rule 4. To attempt to make a geometrically feasible linker, only the option with the furthest distance along an axis defined by the first donor is chosen. To prevent making repeated molecules, SMILES strings were generated for every molecule and generated molecules matching SMILES strings of previously generated molecules were discarded. (53–55)

We used this sequence of steps to generate 59,000 unique linker-like carboxylic acids. Molecules generated may not necessarily meet the criteria to become a suitable linker (4) or might be difficult to synthesize. However, we can gain useful mechanical insight from molecules that are not (yet) easily synthesizable or geometrically infeasible. We view the role of our procedure being for identify strategies for linker flexibility. These strategies would be a starting point, a scaffold, or functional armature, that will be refined by synthetic chemists in a more interactive process of molecule design. Although, for purposes of automated linker design additional screening steps could be added to remove these molecules. Approximately 22178 of the molecules generated were found to meet Bao et al.'s (4) criterion for linker-likeness of an angle between carboxylate groups greater than 155 degrees.

For our example application of a pressure switching MOF we desire a MOF that exhibits strong elastic softening under compression. Ideally the material will transition abruptly from a stiff compressive response at low stress to being compliant in compression above a target threshold stress, and eventually stiffening again under very large compression. This process although inherently nonlinear should be reversible, and may or may not be accompanied by a jump in strain as described later. There are a number of MOFs which exhibit unusual compressive behavior including those such as MIL-53 which is extremely elastically anisotropic and will elongate along one direction under the action of a compressive hydrostatic pressure (21). Importantly this very compliant response of MIL-53 originates from articulation of its framework rather from the intrinsic behavior of its linkers. The result is

that the 'stiffness' of the materials is sensitive to the constraint from loading conditions. The directionally dependent stiffness of MIL-53 subject to a uniaxial stress is very different to that when subjected to a uniaxial strain (the former was first plotted by Coudert *et al.* (57) and using their computed elastic constants we have replotted it here in Fig.4.1 along with the response to an imposed uniaxial strain in order to illustrate this point). In our application

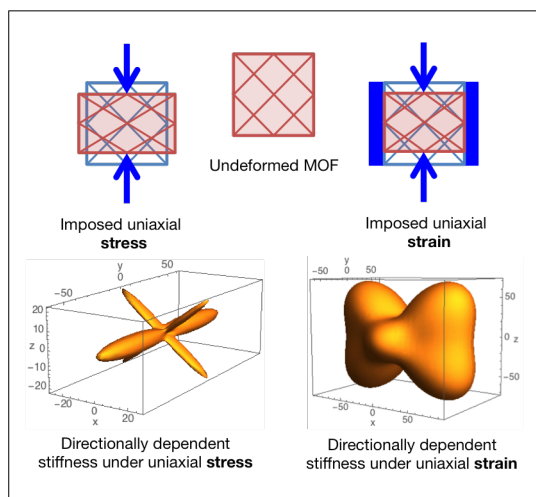


FIGURE 4.1: Directionally dependent stiffness indicatrix for MIL-53 under uniaxial stress (left) and uniaxial strain (right). For a pressure switching MOF we require deforming linkers rather than framework flexibility.

we seek a material that can undergo a large volume change under hydrostatic pressure. This imposes the following requirements for our linkers:

- Volume change must be brought about by linkers deforming into the free pore space.
- In order to create an abrupt change in stiffness the compressed linkers must either undergo a buckling instability or a transition into a different structural conformation. Buckling will provide a reversible transition with a continuous and smooth potential energy curve (see Fig.4.2). A transition to a different conformation involves a jump between potential energy landscapes — a process that can be hysteretic and accompanied by large stress drop and strain jump, and is potentially desirable for a pressure trip

applications.

- The large compressive deformation must be reached without generating large bending moments at the connection to SBUs which would drive shear or coupling to some other deformation mode of the framework.

The mechanical response of 59,000 candidate linkers were tested computationally through molecular dynamics simulations of the linkers being subjected to a cycle of displacement controlled compression and tension as shown in Figure 4.2.

The simulation sequence was as follows: (1) the molecule was relaxed to minimize forces to better than 10^{-8} eV/Å. (2) The hessian matrix for the relaxed structure was computed using the frozen phonon method. (3) The terminal carboxylate groups were then clamped and the system heated to 300 K and simulated in the canonical ensemble (NVT) for 50 ps to allow the system to equilibrate. (4) Still in the NVT ensemble, the linker was compressed by moving one terminal towards the other (along a straight trajectory between the centers of mass of each terminal group) at a fixed speed of 0.005 Å/ps, until a 10% compressive strain was reached. (5) The deformation direction was then reversed and the molecule extended to 5% tension, before (6) the deformation direction was reversed for a final time and the linker was restored to its original length. The deformation speed, while relatively fast (0.5 m/s) is much slower than the speed of sound in the linkers. The process of simulating deformation in the NVT ensemble gives the molecules time to explore configurational space near critical points and buckling transitions, and thus enables us to build a well sampled force displacement response for each linker. For all simulations the atomic interactions are modeled using the universal force field (49), and simulations performed using the LAMMPS molecular dynamics package (52) with a time step of 0.2 fs.

In our MD simulations we test the flexibility of linkers to deformation in which the terminal carboxylate groups are anchored rigidly to mimic the effect of the linker locked into position in between the MOF's SBUs. Deformation of a MOF could involve collective geometric relaxation through rotation of the SBU's. Our testing procedure does not capture these, so the deformation that we impose is overly conservative in terms of the constraint

that is imposed on the linker in a MOF. A second effect that could make the deformation of linkers in a MOF different from the deformation of the linker in isolation is steric interactions between adjacent buckling linkers. These effects could be screened for geometrically, but we have not included this screening in this work.

During deformation the candidate linker's potential energy and force on its clamped carboxylate terminals was averaged over 1 ps intervals. Post simulation, spline fits were performed to the potential energy curve and used to determine the potential energy minimum. This was often slightly removed from the initial clamped configuration of the linker; after being equilibrated at 300 K, the molecule often thermally contracted, expanded, or found an alternative lower energy conformation. For each linker the potential energy curve and force *vs.* strain curve were shifted to zero them on the potential energy minimum. Linkers with potential energy minima at 300 K that were more than 1% strain from the molecule's initial relaxed length were discarded from further consideration so as to remove any bias introduced by comparison of extension tests of differing strain ranges. Once this data had been created for force and potential energy *vs.* strain for both the compressive and tensile loading paths, two separate sets of analysis were performed to create linker performance metrics.

A set of behavior descriptors for describing the mechanical response of a molecule based on curve fit parameters to the stress strain curve of the molecule were developed. I utilized the simulation and construction method developed previously to randomly construct 59,000 candidate linkers and evaluate them. The same generation previously developed was utilized, except it was augmented by only applying rules on the convex hull of the molecule. This helps make molecules that are more linker like. Fragments, or parts of molecules, associated with each of the behavior descriptors were found by using the SUBDUE graph learning algorithm (58). In the graph learning approach fragments common to the most performant molecules that were not present in the the worst performing molecules. The top one hundred and bottom one hundred performing molecules were used as positive and negative samples for SUBDUE. In order to measure linker performance values for properties of direct relevance to pressure switching MOFs we performed a fits to the force *vs.* strain curves of the model

force/strain curve with the functional form:

$$f(\varepsilon; k_I, k_{II}, \varepsilon_c, \Delta\varepsilon, k_{II}') = \Theta(\Delta\varepsilon - \varepsilon + \varepsilon_c) \left(\frac{\sqrt{(-k_{II} + k_I)^2 (-\varepsilon_c - \Delta\varepsilon + \varepsilon)^2 + \Delta\varepsilon^2 (-k_{II} + k_I)^2}}{\Delta\varepsilon} - \Delta\varepsilon(-k_{II} + k_I) + k_I \varepsilon (1 - \varepsilon k_I') \right)$$

where $\Theta(x)$ is the Heaviside function. The shape of the model curve and an example of it is fit to a linker's force/strain curve is shown in Fig.4.3.

The fitting parameters k_I , k_I' , ε_c , $\Delta\varepsilon$, and k_{II}' are mechanically interpretable as, respectively: the stiffness of the molecule under zero stress; the non-linearity in the stiffness around zero stress; the strain at which softening occurs; the width of the softening transition; and the stiffness of the linker post buckling.

This learning task results in the top five molecular fragments that best differentiate high performing linkers from low performing linkers according to the minimum description length (MDL) scoring function. This scoring function rewards fragments that best compress the good linkers and penalizes fragments that best compress the bad linkers.

Next, the performance of each fragment is evaluated in the overall population of simulated linkers. A one tailed Mann-Whitney U Test (59) compares the behavior values of each linker that contains a fragment against the behavior values of each linker that does not contain that fragment. A statistically significant result for this test would indicate that linkers containing the fragment have significantly higher performance values than the other linkers. After a Bonferroni correction to account for the family-wise error rate of performing 65 statistical tests, the significance level for each test is reduced from 0.05 to 0.0009; a Mann-Whitney U Test p value of less than 0.0009 would be required to claim statistical significance for any single comparison. As a result, none of these single fragments found by *Subdue* can be claimed to have a statistically significant impact on linker behavior. In other words, we cannot outright accept that any of these fragments are certain to cause a specific linker behavior irrespective of the other substructures in the linker molecule. Alternatively, the Benjamini-Hochberg procedure (60) can be used to control the false discovery rate of significant fragments (the rate of false positives). By accepting a high false discovery rate of 40%, 21 linker fragments (shown

TABLE 4.1: Candidate behavior descriptors for predicting pressure switching linkers.

Simulation Type	Behavior Descriptor	Objective
Not Applicable	Hysteresis (Area between force-displacement curves)	maximize
Compression	Step height of piecewise linear fit of force-displacement curve	maximize
Tension	Step height of piecewise linear fit of force-displacement curve	maximize
Compression	(Strain at buckling) * (Stiffness at zero strain)	minimize
Compression	Width of the buckling transition region	minimize
Compression	Stiffness after buckling	minimize
Compression	(Stiffness after buckling) / (Stiffness at zero strain)	minimize
Compression	Nonlinearity in the stiffness around zero strain	minimize
Tension	(Strain at buckling) * (Stiffness at zero strain)	minimize
Tension	Width of the buckling transition region	minimize
Tension	Stiffness after buckling	minimize
Tension	(Stiffness after buckling) / (Stiffness at zero strain)	minimize
Tension	Nonlinearity in the stiffness around zero strain	minimize

as simple graphs in supplementary material) are identified as having a potentially significant effect on linker behavior.

A key result of this analysis is that while some linker fragments may increase the likelihood of a desired linker behavior, strong and clear correlations were not found. Nonetheless, the fragments with the highest relative p values (in this case with Mann-Whitney U test p values of approximately 0.029, 0.032, and 0.052) suggest new generative paths to create linkers that are most likely to possess desirable behavior characteristics, and thus warrant additional investigation. These fragments are shown in Table 4.2.

More broadly, these fragments provide a simple starting point to efficiently explore the space of linkers that behave like pressure switches. Furthermore, these results suggest that many of the behavior descriptors lack a significantly strong correlation to specific fragment substructures.

TABLE 4.2: Sample of linker fragments that correlate with behavior descriptors.

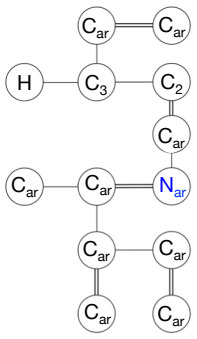
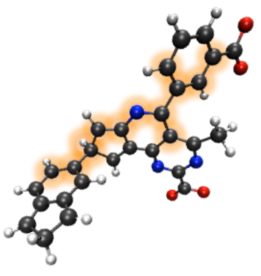
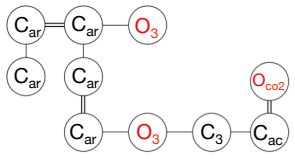
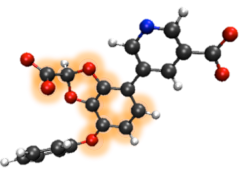
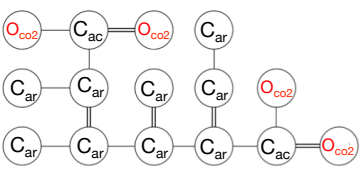

Behavior Descriptor	Fragment Graph	Fragment Example	U Test p-value
Step height of piecewise fit of Force-Displacement Curve (maximize, Linker in Tension)			0.0597
Stiffness after buckling (minimize, Linker in Compression)			0.0771
Nonlinearity in the stiffness around zero strain (minimize, Linker in Compression)			0.0367

TABLE 4.3: Linker Database Summary

Fingerprint Diversity (Average Pairwise Similarity) (61)	0.549
Percentage with Linker-like Pairwise Angle (4)	37
Average Branching Factor of Generative Algorithm at Depth 1	2
Average Branching Factor of Generative Algorithm at Depth 5	95
Average Branching Factor of Generative Algorithm at Depth 9	162
Average Length of Linkers	9.8461 Å
Population Size	59841

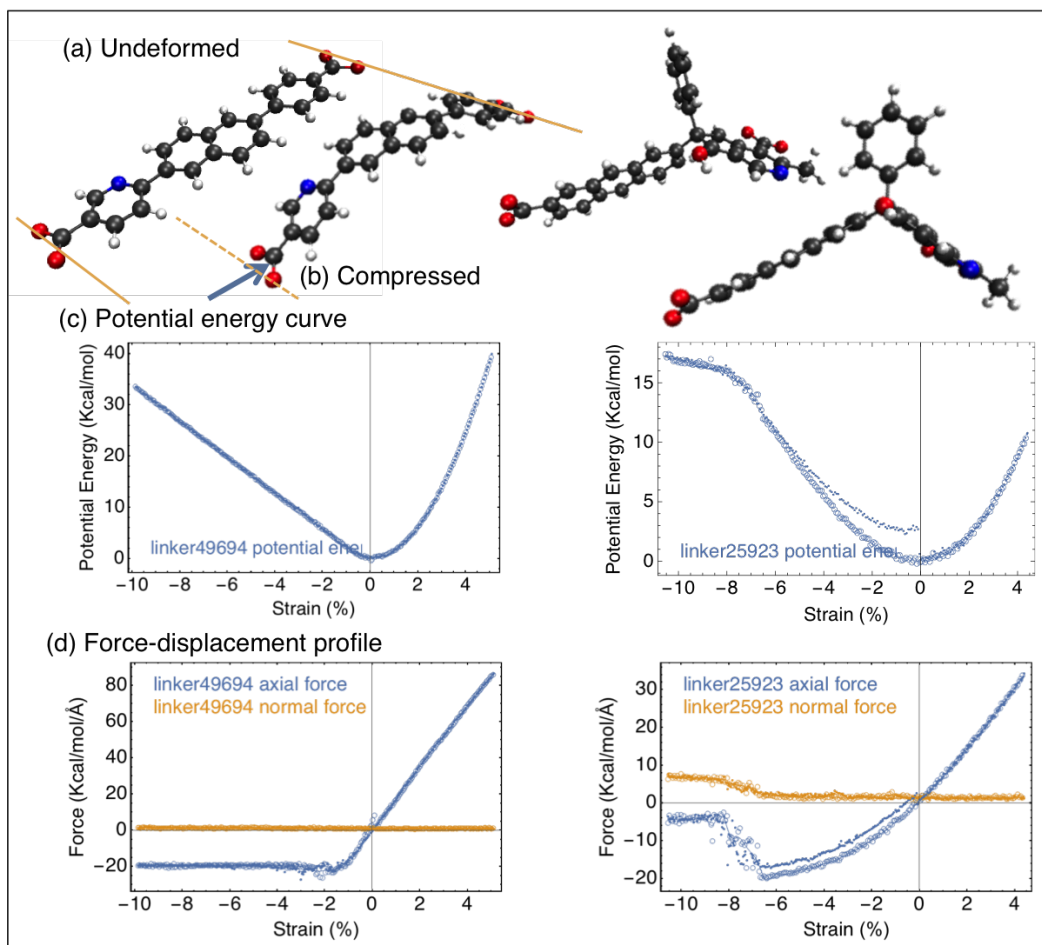


FIGURE 4.2: Example of a candidate linker, (a) relaxed structure, (b) maximally compressed structure, (c) the molecules' potential energy during deformation, and (d) the force displacement curves for compressing and stretching the molecule at 300K with a deformation rate of 1 m/s. The left column is for linker candidate 49694 which exhibits an elastic buckling instability, the right column is for candidate 25923 which undergoes a stress driven conformation change.

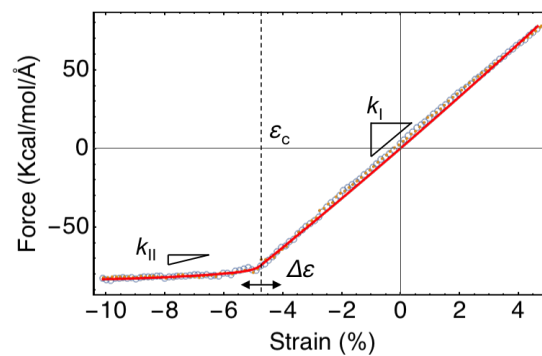


FIGURE 4.3: Example of the fitting of the model curve indicating the mechanical interpretation of the fitting parameters.

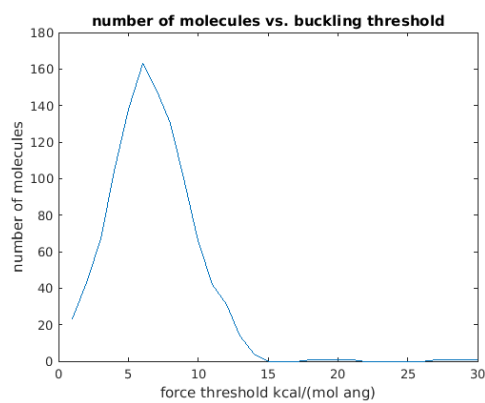


FIGURE 4.4: Number of molecules that behave as good pressure switches at a given pressure.

4.1 Linker Population Summary

Some general statistics describing the generated linker population are presented in Table 4.3.

The first of these statistics is the overall structural diversity in the linker-like molecule set. This diversity is assessed using OpenBabel’s fp2 fingerprint (62). Fp2 is based on the Daylight fingerprinting algorithm; it extracts all unique paths in the molecular graph between lengths 1 and 7 and then hashes each path into a bit array containing 2^{10} elements. The database structural diversity algorithm presented by Turner et al. (61) — which runs in linear time rather than the quadratic time necessary to do a full set of pairwise comparisons — is used to calculate the average similarity between all linker-like molecules. In doing so, the algorithm weights each fingerprint according to the reciprocal square root weighting scheme, and then finds the average pairwise cosine similarity (i.e., structural diversity) of the database. This average pairwise similarity is interpreted as the structural diversity of the database; a measure of how well the method generates a variety of linker-like structures.

Molecules that exhibit an angle between carboxylate groups greater than 155 degrees satisfy a criterion for geometric feasibility (4). The percentage of molecules with linker-like pairwise angle reflects what percentage of molecules meet this criterion. In one sense, this reflects the yield of molecules generated that are likely to act as MOF linkers.

The average branching factor at various depths describes the average number of rule choices found after a specified number of rule applications. Table 4.3 presents the branching factor at the application of the first, fifth, and ninth rules. These branching factors illustrate how the complexity of the search space grows as the linkers increase in size. As opposed to describing the linker population, this metric presents a coarse picture of the design space that each newly-generated linker populates.

Table 4.3 also provides the average length of generated linkers. This value is measured as the average length between carbon atoms in the carboxylate groups.

The ideal case for a pressure switch is a molecule that exhibits a step force displacement response under compression. We found the number of molecules that exhibited switching be-

havior closest to this ideal case at constant compressive threshold forces from 1-30 kcal/(mol Å). We consider molecules to be close to the ideal case if the root mean square error between the threshold force and the force displacement curve is less than an arbitrary value of 3 kcal/(mol Å). Figure 4.4 shows that the most molecules meet this criterion at a force threshold of 6 kcal/(mol Å), suggesting a small group of highly performant pressure switches.

5 Pressure Switching MOFs

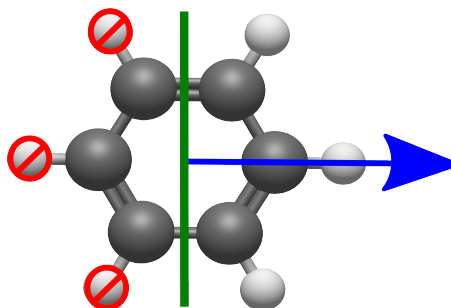


FIGURE 5.1: Seed molecule, green line denotes symmetry plane, arrow indicates rotation axis

As an intermediate goal towards the design of MORFs, a design methodology for pressure responsive MOFs was investigated. We define pressure responsive MOFs as MOFs that experience a change in properties with pressure. It is simpler to apply pressure to a simulated MOF and measure its response than it is to model the behavior of photoisomers or other actuating molecules. The rules described in the above section are used to construct MOFs with some complications. All rules except rule three were used, as this was found to be redundant with applications of rules 1 and 9. An error, also caused rule 3 to produce structures with peroxides. We only focus on the generation of symmetrical ditopic linkers, though this method might be extended to other types of linkers. It is easier to evaluate symmetric molecules. Asymmetric molecules in a real MOF might adopt a variety of different configurations. For example, all molecules facing toward a node, all molecules facing away from a node. If we only model a small number of unit cells as must in order to rapidly test candidates, we may not adequately capture these configurations. In addition, by using symmetric molecules we only need to optimize half a molecule.

Symmetric linkers are generated by applying rules on one side of a seed molecule and then performing 2-fold rotary reflection(180°) of the molecule. Rules are prohibited from

changing any atom of the seed molecule that is not a hydrogen atom and hydrogen atoms on one side of the mirror plane. This makes it so the seed molecule is only expanded on one side of the symmetry plane. Details of the seed molecule we used which was benzene are shown in Figure 5.1.

We refer to non-hydrogen atoms of the molecule as the core structure of the molecule. For this case, we use a seed molecule consisting of benzene. This is shown in 5.1. A green line denotes the mirror plane, the blue arrow indicates the rotation axis. Rules are prevented from applying on one side of the mirror plane by putting 'restrict' labels on the hydrogens on the corresponding side of the mirror plane. This makes it so the molecule is only built up on one side of the mirror plane. In addition, after a rule has been applied the rotary reflected form of the molecule is energy minimized and the convex hull of the rotary reflected form is calculated. As before, rule numbers 1-7, are only applied on the convex hull of the molecule.

5.1 Grammatical Evolution

In this approach, grammatical evolution was employed to optimize MOFs. Grammatical evolution is a type of evolutionary optimization method that evolves solutions according to a grammar (63), (64). Evolutionary optimization methods mimic the process of natural selection to find solutions that optimize a system. Whereas in a traditional genetic algorithm candidates are directly represented a binary string, in grammatical evolution this binary string is translated into a series of applications of grammar rules which modify a seed to make a candidate. In grammatical evolution candidate solutions have genomes which consist of a bit string. The bit string is translated into a sequence of integers by assigning a number of bits per integer. These integers are translated into sequential applications of grammar rules on a seed. At every step rules are recognized on the current system and the rule choice given by the current integer modulo the number of possible rule choices is applied. This continues until a terminal rule is applied or the end of the genome is reached. Grammatical evolution can work with different types of grammars and underlying representations, but for our case we use the

molecule graph grammar rules we developed previously.

While some grammatical evolution processes use crossover we do not use crossover. Crossover is constructing a new candidate by combining half of one genome from one candidate with half of the genome of another candidate. We found that we obtained better results without using crossover. Crossover is intended to improve performance by combining features of two highly performing candidates. What we found is that this does not happen and instead crossover makes a candidate that is essentially random after the point where the genomes have been combined. The problem is that the integer sequences are not meaningful between different candidates. Integers with the same value at the same point in the tree can refer to completely different rules, rule option numbers, and rule locations. A different rule application early in the sequence can completely change what the rule options refer to further down the sequence.

We use the previously described generation approach with rule choices being determined by grammatical evolution. Rules are applied until a carboxylate is added or the maximum number of rule choices is exceeded. This process does not always result in linkers suitable for placement into MOFs. Linkers that are unsuitable for MOFs are penalized by how far off they are from fitting in a MOF. In order for a linker to fit into a MOF it must have two carboxylate groups on each end. So grammar rules are applied until a carboxylate is placed or an atom limit is exceeded. However, not every candidate linker made can fit into a MOF, so candidates that do not form MOFs are penalized. In order for a linker to form a MOF, no atoms should be located behind the carboxylate as these may impede the formation of a MOF. The carboxylate needs to coordinate with metal ions to form into a MOF. If any part of the molecule is near where the carboxylate connects to metal ions, this may impede the formation of a MOF. To test for such scenarios, all atoms in the molecule except for those forming the carboxylate are tested for intersections with an infinite cone extending from the carboxylate. This cone has its apex at the central carbon atom of the carboxylate and has an angle of 120° which is the equilibrium angle of the oxygen atoms in the carboxylate. The carbon and oxygen atoms in the carboxylate can be used to define a cone, and if any other

atoms are found in this cone, the molecule is penalized and receives a cost of 1000.

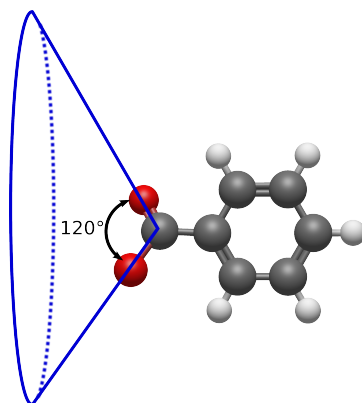


FIGURE 5.2: Cone defined by carboxylate

Bao et al. (4) suggests that the angle between carboxylates should be close to 180 degrees or at least more than 155°. A penalty cost was derived to penalize linkers that don't have geometry suitable to be in a MOF based on how far off they are from being good linkers. If the angle between the two carboxylates, θ , is less than 150°, then it receives a penalty cost based off of how far from 180 degrees it is:

$$angleCost = \begin{cases} 0 & \text{if } \theta \geq 150 \\ 10 \times (180 - \theta) & \text{if } \theta < 150 \end{cases} \quad (5.1)$$

Overall cost is:

$$totalcost = anglecost + carboxylatecost - objectivefunction \quad (5.2)$$

for candidate selection method, we used binary tournament selection. We use this rather than proportional selection because it is ill-defined how much better a molecule that cannot be put into a MOF is than one that can, however it is well defined simply to determine that a molecule qualitatively better than another.

Grammatical evolution may generate the same solution repeatedly. The same solution simulated again should yield the same results. To prevent expensive simulations from being repeated, we generate a canonical SMILES string for each generated linker. A canonical

SMILES string is a unique string representation describing the topology of a molecule. Linkers generated that match SMILES with a previously generated linker are assigned the costs of previously generated linkers. Linkers that encounter an error during simulation are assigned the maximum cost.

5.2 Evaluation

The objective function used to quantify the performance of the pressure collapsing MOFs was the change in solvent accessible surface area. It is desirable to obtain the greatest possible change in gas adsorption. Gas adsorption can be approximated with solvent accessible surface area. Specific solvent accessible area, or the solvent accessible surface area divided by the molecular weight, should be used as this allows for comparison between different MOFs. Higher gas adsorption per unit mass is better.

Once a valid linker is made, it is put into a MOF with a primitive cubic lattice net topology (PCU) and MOF5 secondary building units. An arrangement of 2x2x2 unit cells was used for simulation of the MOFs. It is necessary to use multiple unit cells to get a good estimate of MOF mechanical behavior as multiple unit cells may interact with each other during deformation. In addition multiple unit cells might deform in a non-periodic manner and this would not be captured using a single unit cell. Equal numbers of unit cells should be used along x, y, and z to prevent non uniform behavior due to a non-uniform number of unitcells on each axis. So 2x2x2 unit cells is the minimum we should consider. A MOF with PCU topology using MOF5 SBUs was chosen because such topology has a small unit cell with only three linkers and one secondary building unit. This is not a very common topology, as MOFs that form MOF5 secondary building units tend to have a different net topology where the MOF5 secondary building units are in alternating orientations along the chains of linkers. The unit cell of this topology has eight SBUs and thirty two linkers. Because of the need to simulate multiple unit cells using the realistic IRMOF topology adds tremendous overhead. In addition, this same topology has been used to generate hypothetical MOFs. (2)

Simulation was performed with LAMMPS using UFF4MOF (65) (66), an extension of the universal forcefield to MOFs. For each MOF a 2x2x2 unit cell arranged was used to more adequately model the environment inside a MOF. Each MOF was simulated in a triclinic NPT ensemble in which a Nosé-Hoover thermostat and barostat was used to regulated the temperature and pressure. The specific barostat used here maintains a specified pressure by changing the dimensions of a triclinic simulation box. Each MOF was allowed to reach equilibrium by simulating for 5 picoseconds at zero pressure. The triclinic NPT barostat settings were changed so that pressure was ramped up over the course of the simulation. At the start of the simulation the pressure was 0 atmospheres while at the end of the simulation this pressure was 490 atmospheres. The simulation took place over 10 picoseconds and used 1 femtosecond time steps. Temperature was held constant at 300 K through out the simulation. It should be noted that equilibration and simulation time periods are relatively short. This was done because we desire to simulate candidates rapidly rather than accurately. While we may not determine the exact properties, we can still determine that one candidate is qualitatively better relative to other candidates. The solvent accessible surface area was calculated with the Zeo++ software package (67) to get a rough estimate of gas adsorption. The procedure uses a probe atom and scans it over the surface of the MOF to determine surface area. We used a probe atom with radius 2.007 Å, which is equivalent to the size of a nitrogen dimer. Solvent accessible surface area was calculated for 5 frames at the end of each simulation and averaged. The objective function was to maximize the *change* in surface area between equilibrium and after pressure was applied.

Grammatical evolution was run for 189 iterations with a population size of 200. Genomes had a max size of 8 codons with 10 bits each. In addition, every iteration 10 randomly generated candidates were added to the population. 6658 unique molecules were generated, of these 223 were suitable for placement into frameworks. As can be seen from, 5.6, grammatical evolution exhibits convergence for this problem. The highest change in solvent accessible surface area was 4058 m^2/g . The top performing candidates seem to have features that serve as joints, allowing for a low rotational energy barrier to folding. In the top performing linker,

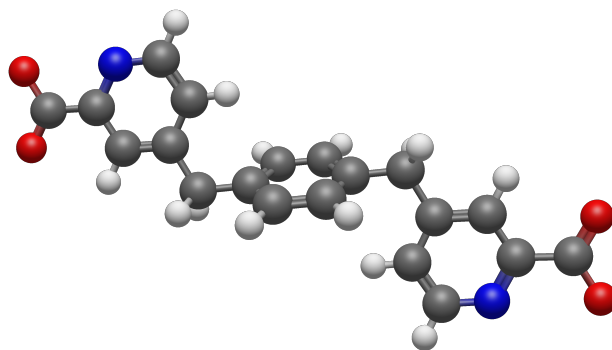


FIGURE 5.3: Linker 4058

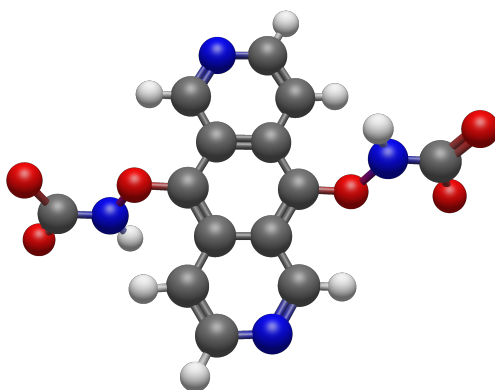


FIGURE 5.4: Linker 4032

5.3, these would be the methyl bridges. In addition a synthesis route was found for the top performing linker. It was found that an analog of the top performing linker has been synthesized (68). The analog shown at the bottom of Figure 5.7 differs from the found linker in the positions of the nitrogen atoms.

We did not maintain a list of the top performing candidates of all iterations and reintroduce these candidates into the gene pool. Consequently the best performing solutions were occasionally removed from the gene pool due to genetic drift. However, grammatical evolution is still functioning correctly as the average fitness continues to improve. The percentage of candidates with higher performance tends to increase with each iteration.

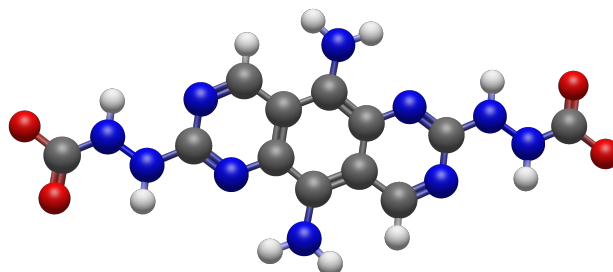


FIGURE 5.5: Linker 3260

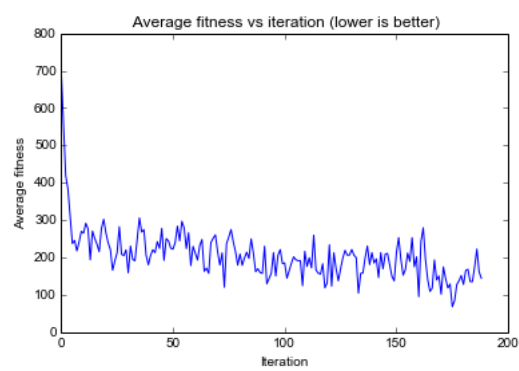


FIGURE 5.6: Average fitness

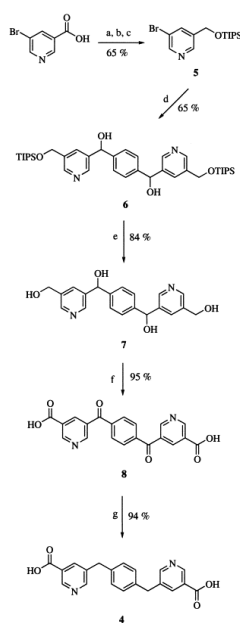


FIGURE 5.7: Synthesis route for an analog of the top performing linker

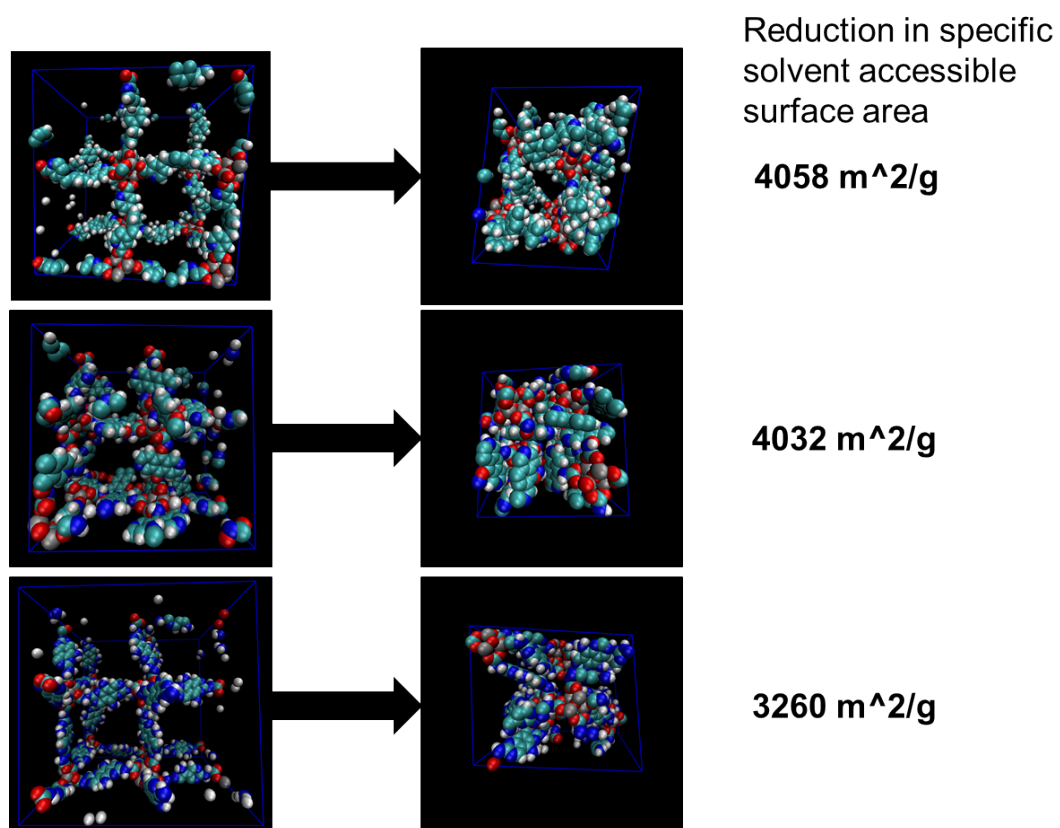


FIGURE 5.8: Pressure induced transitions of top performing candidates, zero pressure(left), 490 atm (right)

6 Photo switching MOFs

A similar approach was used to design photoswitching MOFs. The overall goal has been to design MOFs with actuating moieties in them.

6.1 Photoisomers

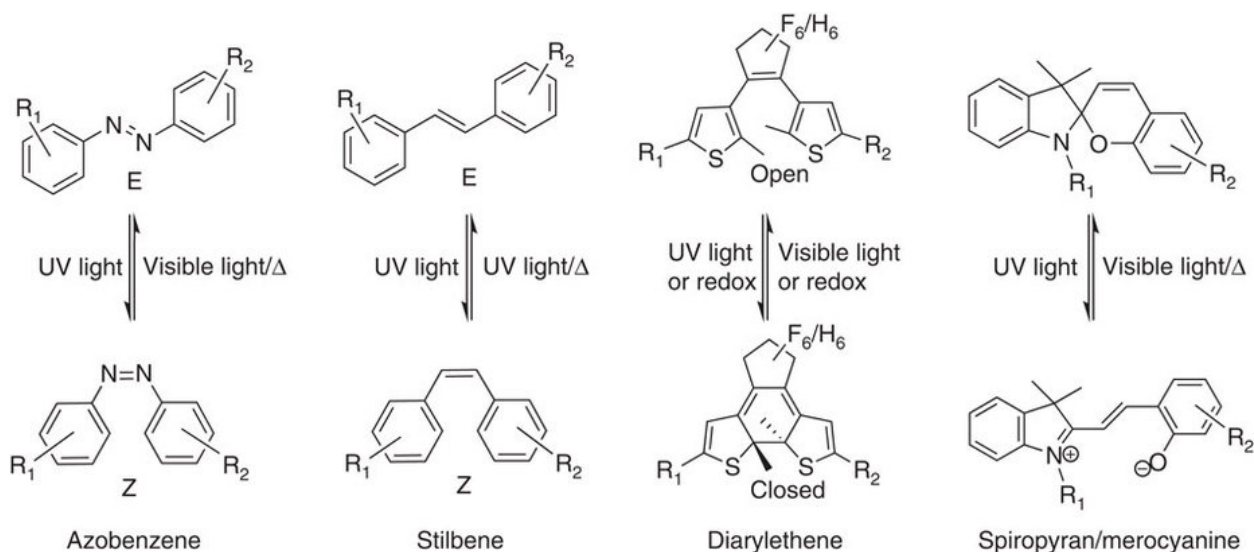


FIGURE 6.1: Examples of photoisomers (taken from (7))

One class of actuating molecules are photoisomers. These are molecules that isomerize, or change in structure or configuration without changing in composition, in response to visible light. Some examples of photoisomers are shown in 6.1. The structural change of photoisomers is due to light induced changes in the electronic structure of the molecules. Light may cause bonds to be broken or reformed as is the case with diarylethenes and spiropyran, or the new electron distribution might cause torsion about a bond as is the case with azobenzenes and stilbenes. To determine the forces caused by such changes we need some method for calculating the full excited electronic state at every time step of the molecular dynamics

calculation. Calculating even the ground state in and of itself is a time consuming calculation. Currently the best method for analyzing photoisomerization is non-adiabatic quantum molecular dynamics (69). However, this is a very computationally intensive and time consuming process even for the case of a single photoisomer. In order to model the behavior of a MORF, we must be able to model multiple molecules. In order to get results in a reasonable time we need some way to approximate the behavior of photoisomers.

6.2 Modelling Azobenzene

As there aren't any general approaches to modelling the behavior of photoisomers, it makes sense to choose a well understood photoisomer. Azobenzene is one of the most well studied and understood photoisomers and was first described at the very dawn of organic chemistry in 1834 (70). In addition, azobenzene has been synthesized into a number of molecules that are of sufficient complexity and function that we can arguably call them mechanical systems (71). So it makes sense to use azobenzene inside our molecules

A method to approximate the dynamics of azobenzene was needed. There have been a number of different approaches used to model azobenzene. Li (72) fit reactive forcefield parameters to approximate the excited and ground state potentials for azobenzene. By switching the potentials, isomerization to the trans or cis states could be modelled. Böckmann et al. (73) came up with an approach to model the isomerization of photoswitchable foldamers from trans-to-cis. To model the isomerization from trans-to-cis, a special driving potential was developed for the CNNC torsion that concatenates portions of the excited state with the ground state. It has approximately the same potential as the excited state from 180 degrees to about 90 degrees, from there this is connected with a cubic spline to the ground state potential at 20 degrees. Essentially it starts out in the excited state and the cubic spline models the 'fall' back to ground. Energy landscape of transition. These approaches are fine for modelling the behavior of molecules that contain a single azobenzene, but not for modelling populations of azobenzenes. In structures such as metal organic frameworks, the isomerization of one

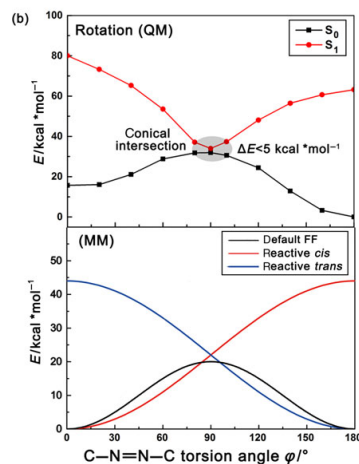


FIGURE 6.2: Potential for the C-N=N-C torsion of azobenzene from (8) (9)

azobenzene could hinder the isomerization of other azobenzene molecules. These approaches do not capture that. Tian et al. (8) (9) came up with an approach to model azobenzene in self assembling monolayers that solves these problem. Building on the work from Bckmann et al., they also use a concatenation of the excited state and ground state to describe the energy landscape of the transition for the CNNC torsion, however, unlike the previous approach they describe the nature of this transition in both directions with two different potentials. They also go through a process of changing what potential the CNNC torsion has with a probability proportional to torsion angle. Tian et al. concluded that the photoisomerization process is primarily driven by the central C-N=N-C torsion. Two potentials are used to describe the energy of this torsion, switching to the trans state and switching to the cis state. This is shown in 6.2. The cis-to-trans potential of the C-N=N-C torsion is given by:

$$E_{torsion} = K_{\phi} [1 + \cos \phi] \quad (6.1)$$

With $K_{\phi} = 38$ kcal/mol. The trans-to-cis potential of the C-N=N-C torsion is given by:

$$E_{torsion} = K_{\phi} [1 - \cos \phi] \quad (6.2)$$

with $K_{\phi} = 28$ kcal/mol, which Tian et al. found to produce a torsion angle closer to that determined experimentally. Modelling isomerization by changing the potentials of all C-N=N-C

torsions to trans-to-cis or cis-to-trans does not accurately reflect reality. Not all azobenzenes will switch at the same time. In fact, it's very unlikely that multiple azobenzenes will be simultaneously excited at the same time given that the rate kinetics are on the order of $10^{-4}/s$ (74). We need some way to capture the fact that some azobenzene molecules will have different potentials during the isomerization process. To model the dynamic nature of the isomerization, Tian et al. introduces a switching function that switches potentials with a probability proportional to the current torsion angle. Figure 6.3 shows a switching function for the cis-to-

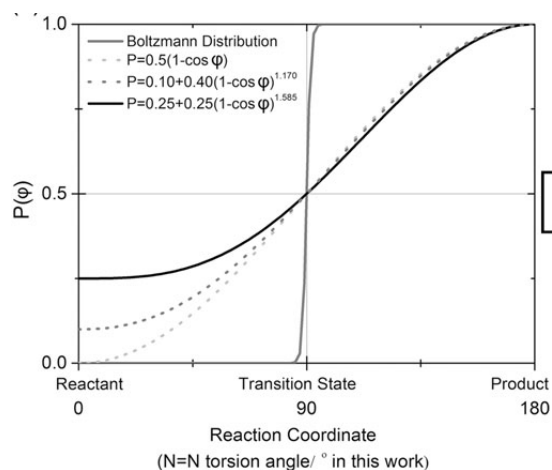


FIGURE 6.3: Switching function

trans isomerization reaction. The switching function relates the probability of a torsion angle having a specified potential to its current torsion angle. A molecule in the cis state, where torsion angle is close to zero degrees, starts out with a low initial probability of getting 'excited' and switching to the the cis-trans potential, as the torsion angle goes to 180 the probability of being in the trans state goes to one. This is for the the cis-trans reaction, the trans cis reaction is the inverse of this. Calculation of this probability and switching are only carried out after a set number of molecular dynamics timesteps. This can be seen as a something of a 'decay' period for the excited state. For the trans-cis isomerization starting out from the cis state, if the molecule switches to the cis-trans potential it is likely to continue to have this potential if during the decay period it rotates toward 180° . Likewise it is likely to return to the cis-trans potential if it does not move toward 180° .

For our switching function we used the following parameters:

$$P = 0.25 + 0.3 (1 - \cos \phi)^{1.17} \quad (6.3)$$

Values chosen differ from those used by Tian et al. as to slow down the rate at which the azobenzene molecules isomerized so that less were isomerizing at a time. Tian et al. also had biasing functions for the potentials for each of the reactions, wherein if an angle was above or below a specified value the molecule was guaranteed to have a specified potential. For example, for the trans-to-cis reaction, any dihedral angle smaller than 60 degrees was considered to be in the trans-to-cis state potential. These helped prevent back reactions and helped yields closer to that found in experiment. Our implementation did not use them so that we could allow back reactions to occur. The biasing functions could bias the reaction too much making MORFs that should not be able to change shape capable of changing shape. It is far more conservative to make MORFs that work in spite of back reactions than without them.

This approach was implemented as a *fix*, or a special command that applies in time stepping or minimization, in LAMMPS. Fixes are one way the functionality of LAMMPS can be extended. The fix allows the user to specify coefficients for potentials for unactuated and actuated states and a probability function for determining which potential should be used based on dihedral angle. This allows the user to specify which reaction should be occurring during a given simulation time period.

6.3 Evaluation

Originally we intended to optimize for maximum change in specific solvent accessible surface as we did with the pressure induced structural transition MOFs, but we decided against this for a number of reasons. Foremost, carrying out change in solvent accessible surface area analysis of the trivial case, where the linkers consist of azobenzene with carboxylates attached at both ends, yielded some strange results. We obtained an enormously high change in solvent accessible surface area. The solvent accessible surface area went to zero when the MORF was

actuated. Upon further analysis, it turned out that while *accessible* surface area went to zero, there was still a significant amount of inaccessible surface area on the inside of the MOF. The contraction of the framework caused the pores to become mostly closed off. For gas storage applications this is not desirable, as this would trap gas inside the framework rather than expel it during actuation. For such applications it would be necessary to develop a figure of merit for gas deliverability, or how much gas can be delivered upon framework contraction, which is the property we are trying to derive when use solvent accessible surface area. While solvent accessible surface area is useful for estimating the gas storage capacity of MOFs it is not suitable for MORFs because of aforementioned issues with pore closure.

In addition, there were concerns that the previously designed pressure switching MOFs were not very reversible. It is quite important that a MORF be reversible, a MORF that only actuates once is of limited utility. So it was decided that reversibility would be taken into account in the objective function. One of the ways to do this is to carry out multiple actuations back and forth from cis-to-trans. While properties based on surface area interesting they are computationally intensive to calculate as we must calculate the surface area of many consecutive frames to obtain the average surface area for each trial. For large MOFs this was taking a significant amount of time to calculate compared to the case of the pressure switching MOFs. Thus, we decided to optimize for maximum volumetric strain. High volumetric strain is useful for such applications as light driven actuators and could translate to high gas deliverable capacity. In addition, we can calculate volumetric strain quickly. This is still as interesting and useful of a problem as gas deliverable capacity and lets us demonstrate the approach.

So again, we desire high volumetric strain while still being reversible. We take the minimum volumetric strain of two consecutive trials of trans-to-cis isomerization. We perform a trans-to-cis isomerization, then attempt to return back with a cis-to-trans isomerization, and then perform a trans-to-cis isomerization. This way if the MORF obtains a very high volumetric strain in the first trial, but locks into place doing so, it will be penalized and recognized as a bad solution from the low volumetric strain measured in the second trial.

Two consecutive trials of trans-to-cis isomerization. Used 2x2x2 unit cell with PCU topology and MOF5 SBUs. To prevent this, the minimum volumetric strain of two trials was used as the objective function.

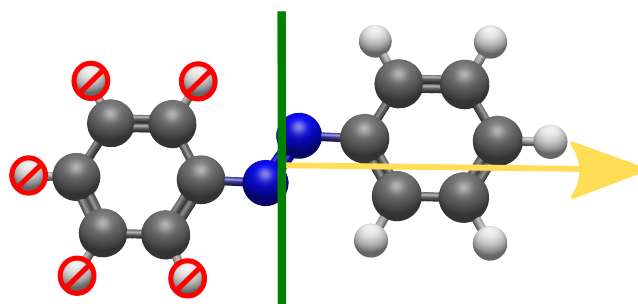


FIGURE 6.4: Seed molecule, green line denotes mirror plane, arrow indicates rotation axis

Azobenzene was put in the center of the linker. The seed graph used is shown in figure 6.4, where the green line denotes the mirror plane and the yellow arrow indicates the rotation axis. Rules were applied one side by only allowing rule application on one side of the mirror plane and then mirrored and rotated 180 degrees. This was done by putting 'restrict' labels, that force the rules not to recognize, and mirrored. By using azobenzene as the seed molecule, we can guarantee that every molecule we generate will always have azobenzene in it. To help prevent stagnation of the search process, a list of the top ten highest performing candidates was maintained and was always put into the gene pool. The population size was 200 candidates and 10 randomly generated candidates were added to the gene pool each iteration. Genomes used 10 bit codons with a max genome size of 8 codons. The same penalty functions described in the previous section were used. This was run for 182 iterations. 8392 unique molecules and 143 unique MOFs were evaluated.

The top performing candidate shown in 6.5, attained a volumetric strain of -0.4990. Top performing candidates attained almost -0.5 volumetric strain. The trivial case of attaching carboxylates to the end of an azobenzene molecule did fairly well with -0.35 volumetric strain. However, the trivial case does not fully isomerize. It appears to contract along two axes which prevents the other axis from isomerizing. The top three performing candidates

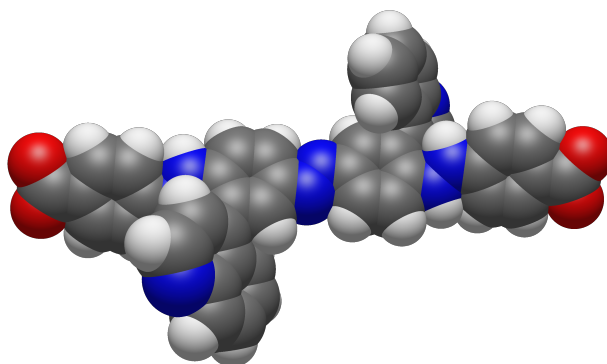


FIGURE 6.5: 1st highest performing candidate, with a volumetric strain of -0.4990

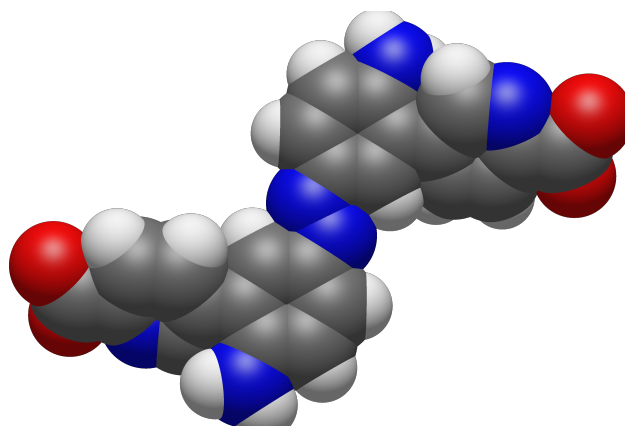


FIGURE 6.6: 2nd highest performing candidate, with a volumetric strain of -0.4857

appear to be completely different from each other. The top performing candidates seem to have analogous structures, all seem to have a rotational degree of freedom attached directly to the azobenzene, near a bulky group. In top performing candidate, the nitrogen-nitrogen single bond serves as the rotational degree of freedom and the two membered benzene ring groups serves as the bulky group. It was hypothesized that the rotational degree of freedom allows the molecule to fold, while the bulky group acts like a spring along this rotational degree of freedom to help the molecule spring back to its original configuration. It was hypothesized that these bulky groups prevent the MOF from collapsing. To test this, a version of the top

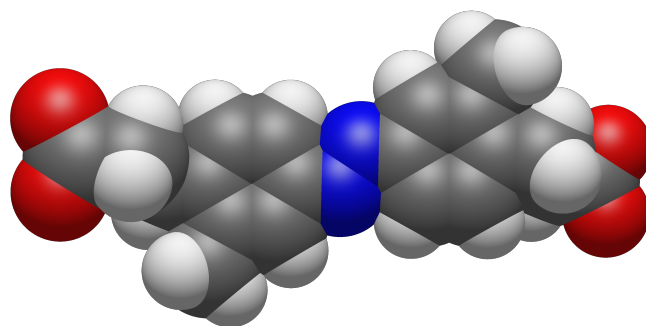


FIGURE 6.7: 3rd highest performing candidate, with a volumetric strain of -0.4691

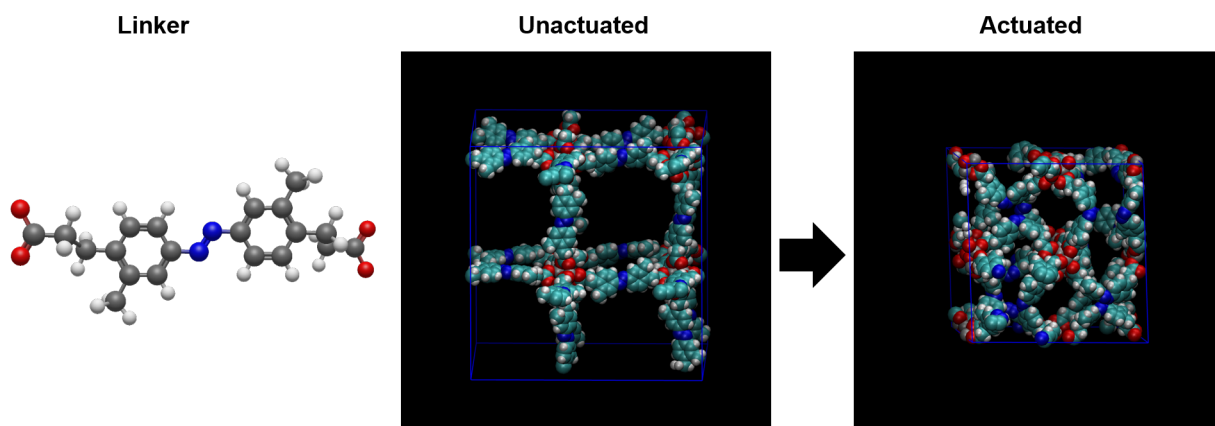


FIGURE 6.8: 3rd top performing candidate, with a volumetric strain of -0.4691

performing candidate without these bulky groups was created and tested. Unexpectedly, the new MOF does not collapse and exhibits worse performance than the original molecule as can be seen in the comparison of isomerization behavior shown in 6.11. While the original molecule achieves -0.499 volumetric strain, the new molecules achieves -0.220 volumetric strain. Performance is decreased by more than a half simply by the removal of these bulky groups. From the decrease in performance Another possible explanation is that non-bonded attractive forces between these groups and other linkers assist in causing the collapse of the

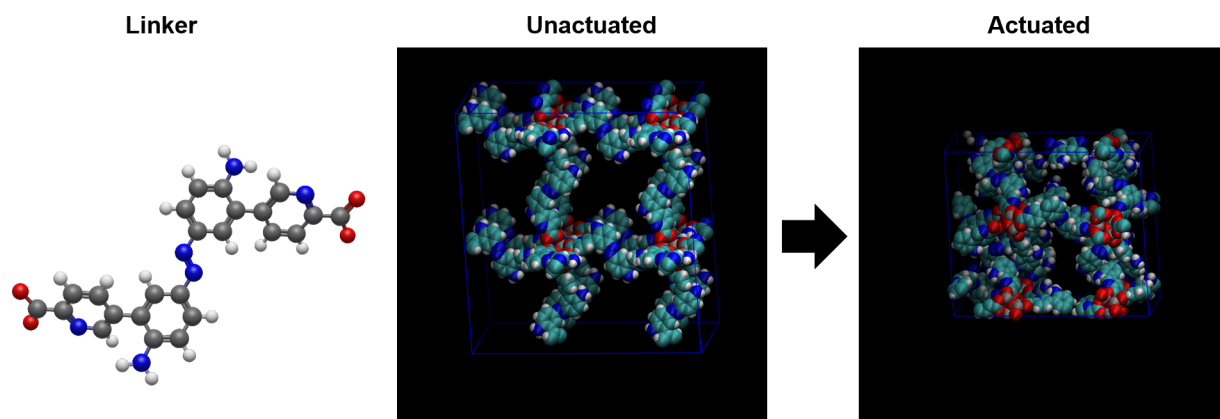


FIGURE 6.9: 2nd top performing candidate, with a volumetric strain of -0.4857

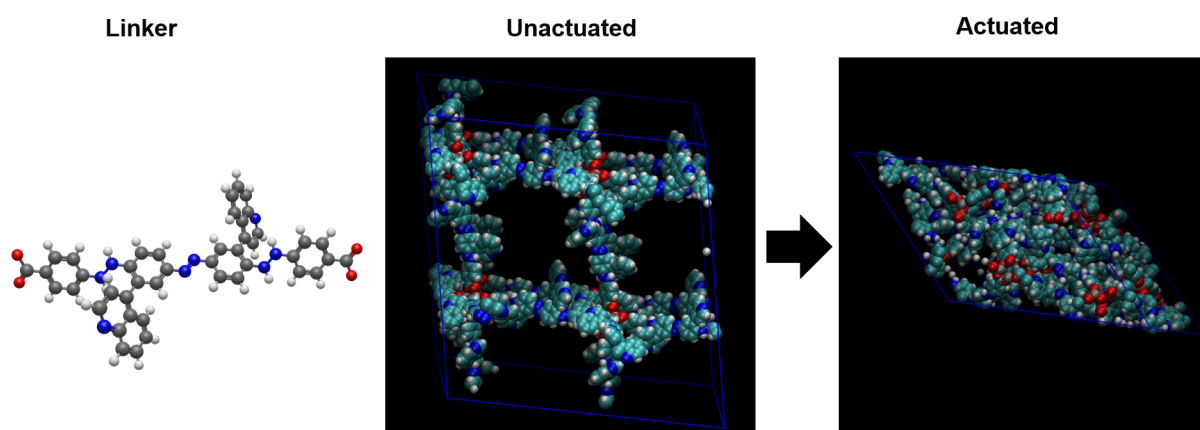


FIGURE 6.10: 1st top performing candidate, with a volumetric strain of -0.4990

unit cell. Although it would seem that these would impede the unit cell from expanding again. Determining exactly how this molecule works might be a research effort in and of itself. In any case this is a good example of just how non-intuitive this problem is and the evidence of the value of this design method over manual design. It is quite unlikely that a human would be able to come up with this solution.

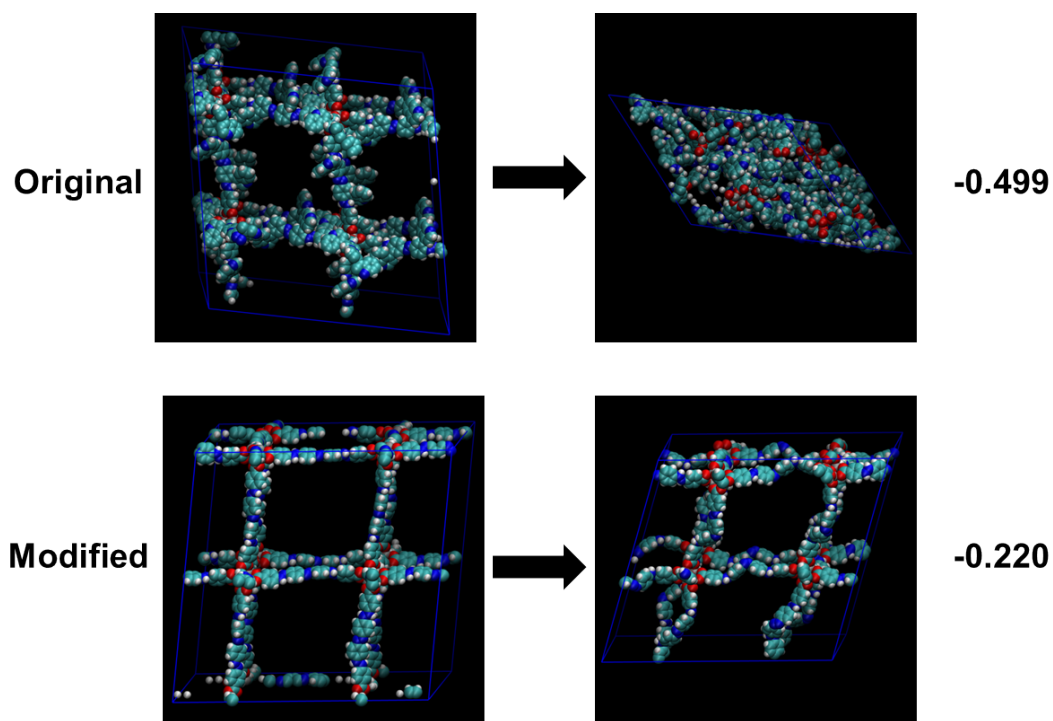


FIGURE 6.11: Comparison of behavior during isomerization

7 Monte Carlo Tree Search for molecules

Monte Carlo tree search was investigated using the force curve design problem. Monte Carlo tree search is a stochastic tree search method (75) that has been employed with great success to solving problems such as Go (76). Monte Carlo tree search offers some interesting possibilities because it does not require heuristics. One of the challenges with employing tree search algorithms such as A* to designing MOFs is that it is difficult to assign value to intermediate tree states. For path planning where we desire to find the shortest path between two point, we might use a heuristic such as euclidean distance to choose the next point to expand. A partially complete molecule might lead to a highly performing MOF or something incapable of forming a MOF. The only properties we care about are those of a completed linker, that is a molecule with at least two carboxylates. Not every state in the tree has this. It is not clear what heuristics should be used for designing MOFs.

Monte Carlo tree search algorithms involve four basic steps for each iteration:

1. Selection: starting at the root node descend through the tree by selecting a child node according some policy until a node that a terminal state or has unvisited children is found. We call the policy for selecting child nodes the tree policy.
2. Expansion: Add one or more of these unvisited children to the tree.
3. Simulation: From each of these children, attempt to estimate their value by performing a simulation. That is apply a policy, which we refer to as the default policy, until a terminal state is reached and determine the value of that terminal state. In games, this terminal state is an end game and the value of this state might be the computer's score or if the computer won.
4. Backpropagation: propagate the results of the simulations back up the tree, updating statistics on the child nodes and nodes taken to reach the child nodes. (75)

This process is carried out repeatedly until a computational budget is exhausted, a time limit is reached, or objective function change between iterations is at an acceptable level. Upper

confidence for trees(UCT) is a popular variation of Monte Carlo tree search developed by Kocis and Szepesvári (77) that has a tree policy which treats selection of new child nodes as separate n-armed bandit problems for each node. The n-armed bandit problem is a classic problem of an exploration-exploitation dilemma (78), of figuring out how to maximize profits from a set of slot machines with unknown probability distributions of payouts. Auer et al. (79) developed policy for balancing exploration and exploitation called UCB1 which chooses which arm to play based upon which has the highest upper confidence bound. The arm j which maximizes:

$$UCB1 = \bar{X}_j + \sqrt{\frac{2 \ln n}{n_j}} \quad (7.1)$$

is selected at each step. Where \bar{X}_j is the average reward of arm j , n_j is the number of times an arm has been played and n is the total number of plays so far.

Kocis and Szepesvári reformulated UCB1 as follows for selecting which node among the children of a parent node should be further explored. A child node j is selected with highest value of:

$$UCT = \bar{X}_j + 2C_p \sqrt{\frac{2 \ln n}{n_j}} \quad (7.2)$$

Each child node j represents a branch that has been visited n_j times that has an average reward \bar{X}_j . Where there have been n plays so far, corresponding to visits to the parent node. Typically \bar{X}_j is restricted to be in the range $[0,1]$ and C_p is a constant greater than zero. \bar{X}_j can be seen as an 'exploitation' term while the other term can be seen as an "exploration term" which balance each other out. The second term makes nodes that have not been visited much higher values, even giving nodes that have not been visited yet infinite values. \bar{X}_j gives nodes which did well higher values, but the search avoids getting stuck on these high performing nodes because their average value decreases each time they have been visited, eventually making their value less than nodes that have low average rewards and have been explored much. Even child nodes with low rewards are guaranteed to be chosen, but this might take a very long time. Different values of C_p enable us to change this balance between exploration and exploitation.

7.1 Test Problem

As a test problem we employed UCT to optimize a molecule to have a specified force strain curve. Our generation approach was basically the same as that used in the Fragment section, except instead of applying rules randomly we used UCT. As with before, we used a carboxylate as seed and our conditions for a graph being terminal were that the carboxylate rule was applied or the maximum allowable atom count had been exceeded. We used the same force curve evaluation method used by the above section. Our objective function was to minimize the normalized root mean squared error of the force-strain curve from a line with a slope of k 0.5 Kcal/mole-Angstrom:

$$NRMSE = \frac{\sqrt{\frac{\sum_{i=1}^n (k\varepsilon_i - f_i)^2}{n}}}{f_{max} - f_{min}} \quad (7.3)$$

Where f_i is axial force measured at strain ε_i for the set of n force strain measurements. We map the objective function $NRMSE$ to a reward in the domain [0,1] for a reward maximization problem with:

$$R = 1 - NRMSE \quad (7.4)$$

This domain is well bounded as the best possible $NRMSE$ is zero, corresponding to a molecule that exactly matches the desired stress strain behavior. In addition, molecules received 0 reward if they didn't have two carboxylates, the angle between the carboxylates was less than 150° , or either of the carboxylates was blocked. Such molecules have no value as they are not valid linkers. In addition, the existence of a molecule carboxylate with a pairwise angle less than 150° does not imply the existence of feasible molecules near said states. So it does not make sense to drive the search to such states. Every node in our tree stores visit count, total reward, and, the index of the option chosen to reach the node, the rule number chosen, a SMILES string, and a design graph. Reward is backpropagated up through the tree, by adding up the reward to each of the nodes total rewards along the path taken. We take \overline{X}_j to be the total reward for a node divided by the number of times that node has been visited.

While it seems fairly memory expensive to store a designgraph on every node in the tree it is worth doing so. After we apply a rule we must energy minimize the resultant molecule and calculate the convex hull which can be fairly costly to reevaluate completely for each molecule.

While the basic UCT algorithm uses a random default policy for doing simulations, we found that this does rather poor for this domain. Running 10,000 iterations with a random default policy only yielded three molecules which could be evaluated as linkers. The problem is that just applying rules randomly rarely results in linkers we consider feasible. The rules will rarely apply rules that add carboxylates, or if they do they won't put them in places that make sense.

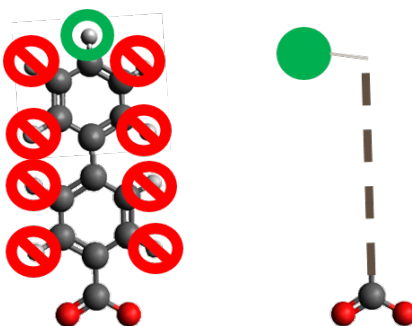


FIGURE 7.1: Choose feasible carboxylate heuristic(left), make most axially distant option heuristic(right)

To make it more likely that we generate feasible molecules the default policy was modified. At every step there was a 50% chance of applying rules randomly, the other half of the time rules were applied according to some heuristics. If any carboxylate rules were recognized, the pairwise angle between the option location and the seed carboxylate was calculated. From all of the locations where the pairwise angle was more than 150° , one option was chosen randomly and applied. If none were found then either rule 1, which adds a methyl group, or rule 2, which adds a benzene ring, was applied at the location most axially distant from the seed carboxylate. This can be summarized as make any carboxylates that are feasible, if you can't find any, build the molecule up away from the seed carboxylate.

7.2 Results

This was run for 50000 iterations, resulting in the top performing linker show in Figure 7.2. Upon further reflection, the target stress strain behavior was pathological and essentially

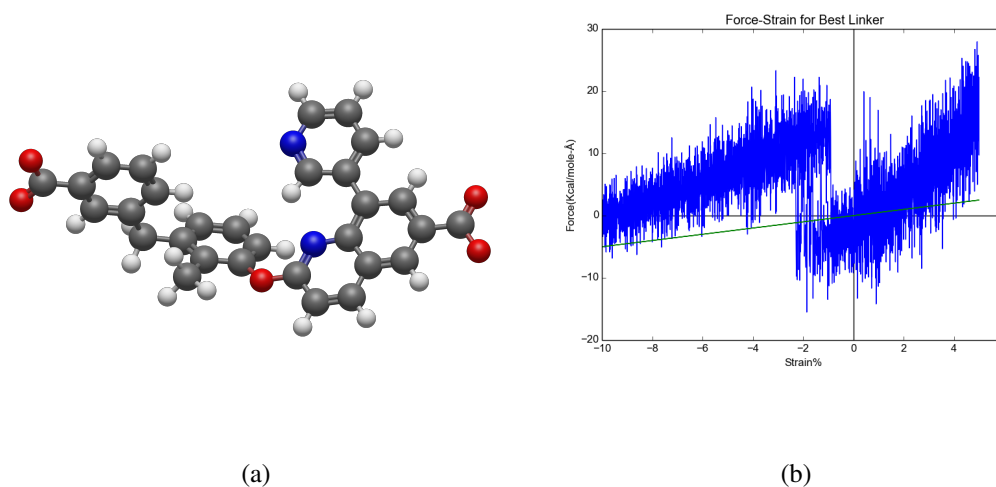


FIGURE 7.2: Top Performing Linker(left) found through UCT and its force displacement curve

requires that the linker have a stiffness close to zero. Despite this, MCTS still found a good solution. The linker has a normalized root mean square error from the desired force curve of 0.209.

While we obtain a linker with reasonable performance, it is not clear that the way this solution was obtained was much better than random. Overall search performance was incredibly bad. It tended to exploit the same solution over and over again. With one linker being made 47031 times out of the 50000 iterations. This can be seen in the behavior of the reward plot, shown in figure 7.3. In the beginning reward received is fairly spread out then it becomes a straight line indicating the same solution is being made over and over. Not only is the solution being made over and over again, it's not the best solution found. With the same solution being created this many times, even the relatively small cost of traversing the tree adds up. In addition, although 50000 iterations were ran, only 1732 unique linkers were

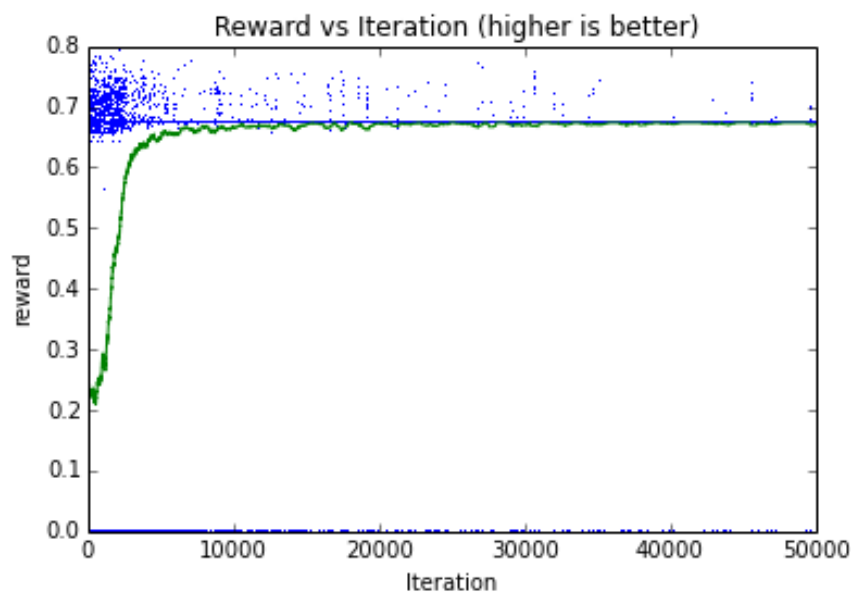


FIGURE 7.3: Reward for MCTS with iterations, green line is average

evaluated.

It has been recommended to use epsilon greedy selection rules for better performance (80). The selection policy becomes choosing child nodes that maximize:

$$Score = \overline{X}_j + \Gamma \frac{\epsilon n}{n_j} \quad (7.5)$$

We are able to again obtain a decent molecule, although there are still questions if the process by which this molecule was obtained was any better than random. We obtain a linker with 0.858 NRMSE from the desired force curve which is shown in Figure 7.4b. As shown in Figure 7.5, a larger spread in reward function was achieved and convergence to a single solution of moderate performance does not occur. Although now, there doesn't seem to be any convergence whatsoever. Average reward fluctuates around 0.45. On average it seems to be making a number of linkers that are very bad and have zero reward.

It was suspected that some of this behavior might have been due to the default policy adding carboxylates too aggressively, so the default policy was changed. So instead of applying the carboxylate rule every time a feasible location was found, there was a 50% chance of

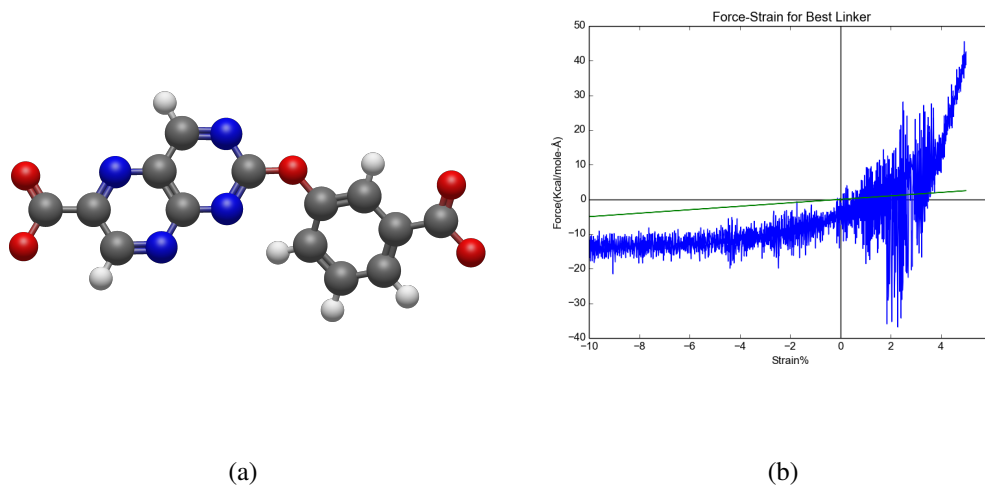


FIGURE 7.4: Top performing linker epsilon greedy

policy	best reward	nrms	iterations	best iteration	unique linkers
1	0.791	0.209	50000	2170	1732
2	0.858	0.142	7768	2781	1906
3	0.858	0.142	45232	10241	10107

TABLE 7.1: Performance Summary

doing so or building the molecule outward.

This also used epsilon greedy, and again found something of a reasonable molecule. The top performing molecule had 0.858 normalized root mean square error. Oddly this is approximately the same performance found in the previous policy. Stranger still, the molecules are completely different in topology. As can be seen from a plot of the reward in figure 7.7, we cannot say that performance is much better than random.

So the question is, why are we getting such terrible performance? The problem may be that UCT needs to be tuned more to get the desired performance to strike more of a balance between exploration and exploitation. Another may be the very nature of the UCT algorithm. Because UCT attempts to maximize average reward received going down a branch of the

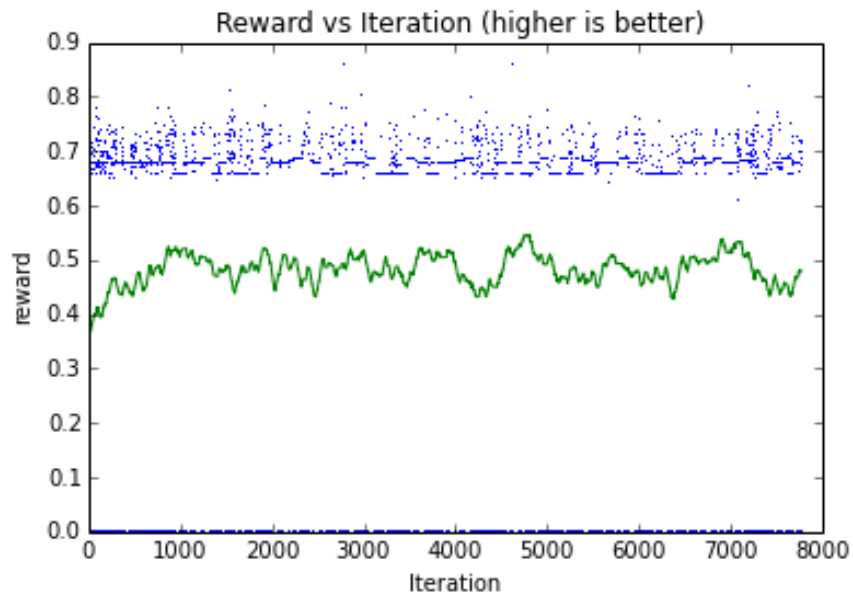


FIGURE 7.5: Reward for UCT epsilon greedy, green line is average

tree. Maximizing average reward might be why it is regenerating the same solution over and over again, because any other solution is likely to be much worse than a known solution. Average rewards are not that useful for this domain and it might be better to use the max reward obtained so far instead (80). There is also the possibility that there is an error in the implementation of UCT.

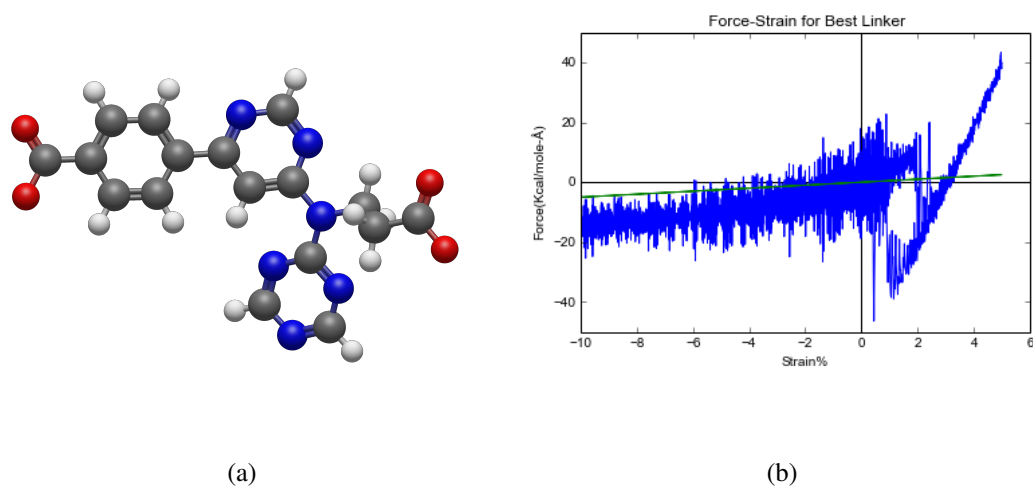


FIGURE 7.6: Top performing linker, changed default policy

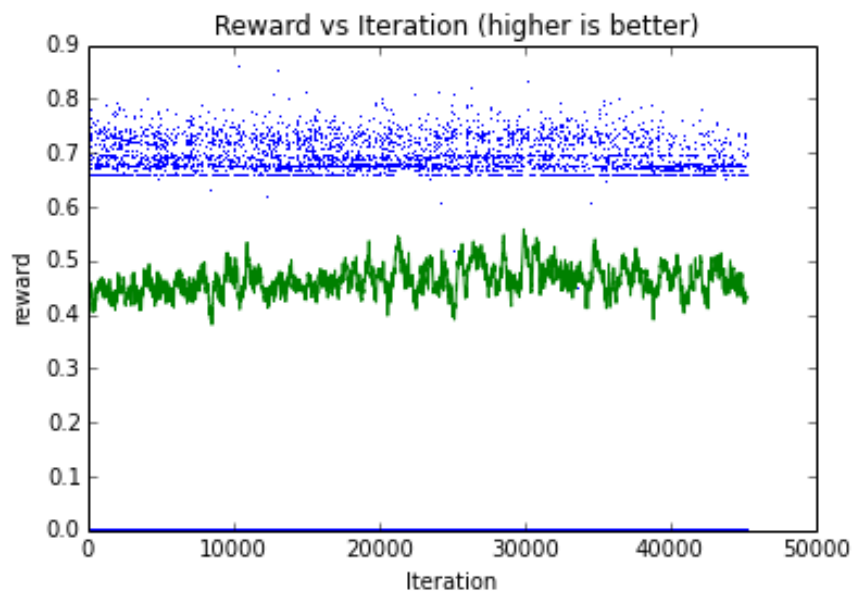


FIGURE 7.7: Reward for UCT epsilon greedy changed default policy, green line is average

8 Conclusions and Future Directions

To my knowledge, this work is the first to investigate the computational design of MOFs that change shape either passively or actively. We have developed a set of graph grammar rules capable of constructing a wide variety of molecular structures and demonstrated their value in helping to provide a solution to a number of problems. We demonstrated the value of this graph grammar approach by integrating the developed rules with tree search algorithms to search for linkers and MOFs with desired properties. We showed how they could be employed in a tree search to rapidly find molecules which specified force displacement curves. The rules were also employed to generate a database of potential flexible linker molecules which might have use in future MOF design. We developed methods where by the graph grammar rules developed could be used to make symmetrical linkers. Systems for automatically extracting which parts of a linker molecule are associated with performance related mechanical properties were investigated.

Metrics were developed for measuring the quality of metal organic frameworks which undergo dynamic shape change in a manner suitable for computational screening. A method for optimizing MOFs to have pressure tunable gas-sorption properties was developed. A synthesis pathway for a top performing linker molecule likely to form a MOF with pressure tunable gas-sorption properties was also identified. It would be worth attempting to synthesize this linker and integrate it into a MOF to see if the predicted behavior is attained.

We developed metrics for measuring the performance of photoactive MOFs based around reversibility. A method for simulating MOFs with azobenzene photoisomers integrated into their linkers was developed. We showed how this method could be used to rapidly test the performance characteristics of photoactuating MOFs. This was combined with a grammatical evolution search method to find candidate MOFs capable of photoactuation.

We demonstrated methods for optimizing the dynamic structural properties of MOFs, but so far only PCU topologies have been investigated. There is a need to examine topologies other than PCU. The PCU topology with MOF5, while amenable to simulation is not neces-

sarily realistic. This has not been known to occur in nature. In addition, the PCU topology currently simulated might be more flexible than the actual topology. The current framework for construction and simulation of MOFs allows the use of other topologies besides PCU, but they were not tested due to time constraints. Many of the mechanical properties we are interested in are driven by the topology the MOF has, so it makes sense to try MOFs with different topologies. However, just combining linkers and SBUs to make MOFs from a number of set topologies, as we do in the current approach, might not be enough to find MOFs with the properties we desire. Recently, Boyd and Woo (81) introduced a method for constructing framework materials of any topology from their net structure, or idealized periodic connection and topology graph. We could extend the approach Boyd and Woo presented with our grammar rules. Boyd and Woo presented a way for placing defined linkers and SBUs, but with our rules we could enumerate linkers. Our rules could be reformulated to modify the periodic graph which would allow for MOFs with novel topologies to be discovered.

Work should also be carried out to verify the behavior of optimized results with ab-initio methods. It is also worth comparing the results of the current work which used UFF4MOF, with other force fields which can supposedly realize more accurate phonon properties to see if the current force field reproduces the same qualitative behavior between MOFs. (51) showed that force fields such as UFF4MOF may not work that well for modelling phonon properties. However, for a system for designing MOFs, we do not necessarily need to accurately model phonon properties. It is far more important that the force fields used be qualitatively correct between different MOFs. As long as we are able to determine that one MOF has qualitatively better properties we can do optimization. This work also developed a well integrated automatic MOF construction and simulation tool kit. Monte Carlo tree search was explored for molecular design, although more work needs to be done to investigate the cause of its poor performance.

The reason I believe that Monte Carlo tree search and similar algorithms should be investigated further is that it might prove useful if one is doing tiered search. In tiered search we might decide to evaluate the same candidates multiple times, using evaluation methods of

ever increasing computational cost and accuracy. UCT, which treats node selection as an n -armed bandit problem, might be particularly good at this non-deterministic objective function. If the reward payout received from increasing simulation accuracy does not result in the same high reward, the search will shift away from that candidate to other areas of the tree. It is worth investigating new search methods capable of carrying out structure-property learning during the search process. If we can learn what structures are correlated with performance we can carry out search in a smarter manner.

This work has demonstrated the design of MORFs optimized for maximum volumetric change, however, there are more interesting properties one could optimize for. Variable pore size and gas adsorption are obvious properties one could optimize for, but a less obvious obvious property to optimize for is coupling between the actuation of one linker to another. Coskun and Stoddart et. al (82) originally proposed this idea, that MOFs could be used to hold molecular switches relative to each other in space such that 'coupling' constants between switches in the framework could be realized. They propose that this coupling between switches could be used to realize solid state active media that would form spatiotemporal patterns. Active media can form self-oscillating chemical waves, which can be used to construct chemical logic gates and have even exhibited behavior similar to cellular automata (83). Self oscillation might be useful for building unusual materials that self-pump a solvent through them via sequential oscillating contraction along one axis. Self oscillating behavior has already been observed in polymeric azobenzene systems (84). The ability to have cellular automaton like behavior might be exploited in photoactuating MOFs to produce spatiotemporal patterns smaller than the wavelength of visible light, potentially allowing control of motion and structure at very small scales.

We have found some evidence that we might be able to obtain these coupling constants. Figure 8.1 shows a found candidate MORF that contracts upon two axes, but not the other. (Note the rectangular unit cell shape and trans azobenzene) This indicates coupling behavior between linkers. Azobenzenes are able to contract along two axes, putting stress on the other axis preventing it from actuating. More work needs to be done to confirm this, but

this is a start to determining how coupling constants may be achieved.

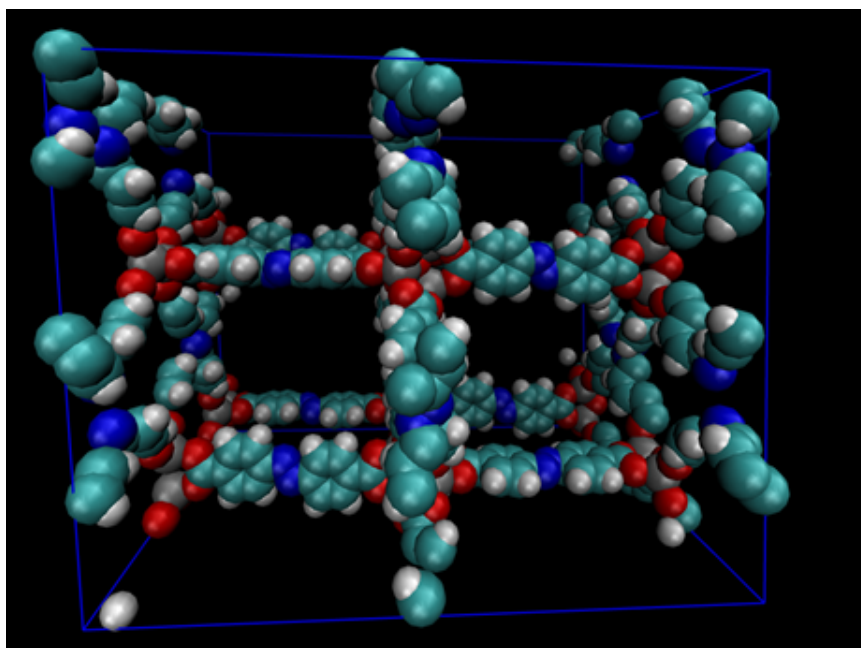


FIGURE 8.1: An azobenzene based MOF that contracts on two axes, but not the other

The approaches developed here can probably be generalized to other material systems. Trivially, the developed approach can be directly applied to design other framework materials such as Covalent Organic Frameworks (COFs) (85) or Supramolecular Organic Frameworks (86). The grammar rules themselves might prove useful in designing organic molecules. While the problem of designing molecules to have specific mechanical properties was mainly intended as a test problem, the approach developed may be applicable to design of molecular scale mechanical systems.

Bibliography

1. Q. Huang, L.-L. Li, and S.-Y. Yang, "PhDD: A new pharmacophore-based de novo design method of drug-like molecules combined with assessment of synthetic accessibility," *Journal of Molecular Graphics and Modelling*, vol. 28, pp. 775–787, jun 2010.
2. C. E. Wilmer, M. Leaf, C. Y. Lee, O. K. Farha, B. G. Hauser, J. T. Hupp, and R. Q. Snurr, "Large-scale screening of hypothetical metal-organic frameworks," *Nat Chem*, vol. 4, pp. 83–89, Feb. 2012.
3. R. L. Martin and M. Haranczyk, "Optimization-based design of metal–organic framework materials," *Journal of Chemical Theory and Computation*, vol. 9, no. 6, pp. 2816–2825, 2013.
4. Y. Bao, R. L. Martin, C. M. Simon, M. Haranczyk, B. Smit, and M. W. Deem, "In silico discovery of high deliverable capacity metalorganic frameworks," *The Journal of Physical Chemistry C*, vol. 119, no. 1, pp. 186–195, 2015.
5. N. J. Young and B. P. Hay, "Structural design principles for self-assembled coordination polygons and polyhedra," *Chem. Commun.*, vol. 49, pp. 1354–1379, 2013.
6. B. L. Weight, "Development and design of constant-force mechanisms," 2002.
7. X. Zhang, L. Hou, and P. Samorì, "Coupling carbon nanomaterials with photochromic molecules for the generation of optically responsive materials," *Nature Communications*, vol. 7, pp. 11118 EP –, Apr 2016. Review Article.
8. Z. Tian, J. Wen, and J. Ma, "Reactive molecular dynamics simulations of switching processes of azobenzene-based monolayer on surface," *The Journal of Chemical Physics*, vol. 139, p. 014706, jul 2013.
9. Z. Tian, J. Wen, and J. Ma, "Dynamic simulations of stimuli-responsive switching of azobenzene derivatives in self-assembled monolayers: reactive rotation potential and switching functions," *Molecular Simulation*, vol. 41, no. 1-3, pp. 28–42, 2015.
10. R. J. Plunkett, "The history of polytetrafluoroethylene: discovery and development," in *High Performance Polymers: Their Origin and Development*, pp. 261–266, Springer, 1986.
11. K. Hbner, "150 jahre mauvein," *Chemie in unserer Zeit*, vol. 40, no. 4, pp. 274–275, 2006.
12. J. Jewkes, *The Sources of Invention*. Palgrave Macmillan UK, 1969.

13. C. Panati, *Extraordinary Origins of Everyday Things*. HarperCollins, 2013.
14. P. Israel, *Edison: A Life of Invention*. Wiley, 2000.
15. W. Wong-Ng, Y. Yan, M. Otani, J. Martin, K. Talley, S. Barron, D. Carroll, C. Hewitt, H. Joress, E. Thomas, *et al.*, “High throughput screening tools for thermoelectric materials,” *Journal of Electronic Materials*, vol. 44, no. 6, pp. 1688–1696, 2015.
16. E. O. Pyzer-Knapp, C. Suh, R. Gómez-Bombarelli, J. Aguilera-Iparraguirre, and A. Aspuru-Guzik, “What is high-throughput virtual screening? a perspective from organic materials discovery,” *Annual Review of Materials Research*, vol. 45, pp. 195–216, 2015.
17. J. Gibson, M. Holohan, and H. Riley, “87. amorphous carbon,” *Journal of the Chemical Society (Resumed)*, pp. 456–461, 1946.
18. M. O’Keeffe, G. B. Adams, and O. F. Sankey, “Predicted new low energy forms of carbon,” *Phys. Rev. Lett.*, vol. 68, pp. 2325–2328, Apr 1992.
19. D. Alezi, I. Spanopoulos, C. Tsangarakis, A. Shkurenko, K. Adil, Y. Belmabkhout, M. O’Keeffe, M. Eddaoudi, and P. N. Trikalitis, “Reticular chemistry at its best: Directed assembly of hexagonal building units into the awaited metal-organic framework with the intricate polybenzene topology, pbz-mof,” *Journal of the American Chemical Society*, vol. 138, no. 39, pp. 12767–12770, 2016. PMID: 27615117.
20. S. S. Han, , and W. A. G. III*, “Metalorganic frameworks provide large negative thermal expansion behavior,” *The Journal of Physical Chemistry C*, vol. 111, no. 42, pp. 15185–15191, 2007.
21. P. Serra-Crespo, A. Dikhtiarenko, E. Stavitski, J. Juan-Alcaniz, F. Kapteijn, F.-X. Coudert, and J. Gascon, “Experimental evidence of negative linear compressibility in the mil-53 metal-organic framework family,” *CrystEngComm*, vol. 17, pp. 276–280, 2015.
22. S. Krause, V. Bon, I. Senkovska, U. Stoeck, D. Wallacher, D. M. Tbbens, S. Zander, R. S. Pillai, G. Maurin, C. Francois-Xavier, and *et al.*, “A pressure-amplifying framework material with negative gas adsorption transitions,” *Nature*, vol. 532, p. 348352, Jun 2016.
23. “BASF metal organic frameworks (MOFs): Innovative fuel systems for natural gas vehicles (NGVs),” *Chem. Soc. Rev.*, vol. 43, pp. 6173–6174, 2014.
24. B. L. Feringa and W. R. Browne, *Molecular Switches, second edition*. Wiley-VCH, 2011.
25. J. Ritter, “Niac phase I study final report on large ultra-lightweight photonic muscle space structures,”

26. D. Lowe, "I do hate to tell you this, but. . . ." *Science Translational Medicine In the Pipeline Blog*, Jan. 2017.
27. G. Schneider and U. Fechner, "Computer-based de novo design of drug-like molecules," *Nat Rev Drug Discov*, vol. 4, pp. 649–663, Aug. 2005.
28. G. Schneider, *De novo Molecular Design*. Wiley, 2013.
29. A. Hopfinger, "Extraction of pharmacophore information from high-throughput screens," *Current Opinion in Biotechnology*, vol. 11, pp. 97–103, feb 2000.
30. J. H. Van Drie, "Monty kier and the origin of the pharmacophore concept," *Internet Electron. J. Mol. Des*, vol. 6, no. 9, pp. 271–279, 2007.
31. V. Tschinke and N. C. Cohen, "The newlead program: a new method for the design of candidate structures from pharmacophoric hypotheses," *Journal of medicinal chemistry*, vol. 36, no. 24, pp. 3863–3870, 1993.
32. H.-J. Bhm, "Computational tools for structure-based ligand design," *Progress in Biophysics and Molecular Biology*, vol. 66, pp. 197–210, jan 1996.
33. H.-J. Böhm, "Ludi: rule-based automatic design of new substituents for enzyme inhibitor leads," *Journal of Computer-Aided Molecular Design*, vol. 6, no. 6, pp. 593–606, 1992.
34. V. J. Gillet, W. Newell, P. Mata, G. Myatt, S. Sike, Z. Zsoldos, and A. P. Johnson, "Sprout: recent developments in the de novo design of molecules," *Journal of chemical information and computer sciences*, vol. 34, no. 1, pp. 207–217, 1994.
35. V. Gillet, A. P. Johnson, P. Mata, S. Sike, and P. Williams, "Sprout: a program for structure generation," *Journal of computer-aided molecular design*, vol. 7, no. 2, pp. 127–153, 1993.
36. P. Mata, V. J. Gillet, A. P. Johnson, J. Lampreia, G. J. Myatt, S. Sike, and A. L. Stebbings, "Sprout: 3d structure generation using templates," *Journal of chemical information and computer sciences*, vol. 35, no. 3, pp. 479–493, 1995.
37. C. E. Wilmer, O. K. Farha, Y.-S. Bae, J. T. Hupp, and R. Q. Snurr, "Structure–property relationships of porous materials for carbon dioxide separation and capture," *Energy & Environmental Science*, vol. 5, no. 12, pp. 9849–9856, 2012.
38. B. J. Sikora, C. E. Wilmer, M. L. Greenfield, and R. Q. Snurr, "Thermodynamic analysis of xe/kr selectivity in over 137000 hypothetical metal–organic frameworks," *Chemical Science*, vol. 3, no. 7, pp. 2217–2223, 2012.

39. Y. J. Colon, D. Fairen-Jimenez, C. E. Wilmer, and R. Q. Snurr, "High-throughput screening of porous crystalline materials for hydrogen storage capacity near room temperature," *The Journal of Physical Chemistry C*, vol. 118, no. 10, pp. 5383–5389, 2014.
40. D. A. Gomez-Gualdrón, C. E. Wilmer, O. K. Farha, J. T. Hupp, and R. Q. Snurr, "Exploring the limits of methane storage and delivery in nanoporous materials," *The Journal of Physical Chemistry C*, vol. 118, no. 13, pp. 6941–6951, 2014.
41. R. L. Martin, L.-C. Lin, K. Jariwala, B. Smit, and M. Haranczyk, "Mail-order metal-organic frameworks (mofs): Designing isorecticular mof-5 analogues comprising commercially available organic molecules," *The Journal of Physical Chemistry C*, vol. 117, no. 23, pp. 12159–12167, 2013.
42. B. P. Hay and T. K. Firman, "Hostdesigner: A program for the de novo structure-based design of molecular receptors with binding sites that complement metal ion guests," *Inorganic chemistry*, vol. 41, no. 21, pp. 5502–5512, 2002.
43. B. P. Hay, C. Jia, and J. Nadas, "Computer-aided design of host molecules for recognition of organic guests," *Computational and Theoretical Chemistry*, vol. 1028, pp. 72–80, 2014.
44. "Hostdesigner, version 3.0." can be downloaded from the website <http://hostdesigner-v3-0.sourceforge.net>.
45. P. Radhakrishnan and M. Campbell, "A graph grammar based scheme for generating and evaluating planar mechanisms," in *Design Computing and Cognition 10* (J. Gero, ed.), pp. 663–679, Springer Netherlands, 2011.
46. A. C. Starling and K. Shea, "Virtual synthesizers for mechanical gear systems," in *ICED 05: 15th International Conference on Engineering Design: Engineering Design and the Global Economy*, pp. 679–693, Engineers Australia, 2005.
47. T. Kurtoglu and M. I. Campbell, "Automated synthesis of electromechanical design configurations from empirical analysis of function to form mapping," *Journal of Engineering Design*, vol. 20, no. 1, pp. 83–104, 2009.
48. M. Campbell, "Graphsynth," *Design Engineering Lab*, [Online]. Available: <http://designengrmlab.github.io/GraphSynth>, 2017.
49. A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard, and W. M. Skiff, "Uff, a full periodic table force field for molecular mechanics and molecular dynamics simulations," *Journal of the American Chemical Society*, vol. 114, no. 25, pp. 10024–10035, 1992.

50. N. P. Martin, J. Marz, C. Volkringer, N. Henry, C. Hennig, A. Ikeda-Ohno, and T. Loiseau, "Synthesis of coordination polymers of tetravalent actinides (uranium and neptunium) with a phthalate or mellitate ligand in an aqueous medium," *Inorganic Chemistry*, vol. 56, no. 5, pp. 2902–2913, 2017.
51. P. G. Boyd, S. M. Moosavi, M. Witman, and B. Smit, "Force-field prediction of materials properties in metal-organic frameworks," *The Journal of Physical Chemistry Letters*, vol. 8, no. 2, pp. 357–363, 2017. PMID: 28008758.
52. S. Plimpton, "Fast parallel algorithms for short-range molecular dynamics," *J Comp Phys*, vol. 117, pp. 1–19, 1995. <http://lammps.sandia.gov>.
53. D. Weininger, "Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules," *Journal of Chemical Information and Computer Sciences*, vol. 28, no. 1, pp. 31–36, 1988.
54. D. Weininger, A. Weininger, and J. L. Weininger, "Smiles. 2. algorithm for generation of unique smiles notation," *Journal of Chemical Information and Computer Sciences*, vol. 29, no. 2, pp. 97–101, 1989.
55. D. Weininger, "Smiles. 3. depict. graphical depiction of chemical structures," *Journal of Chemical Information and Computer Sciences*, vol. 30, no. 3, pp. 237–243, 1990.
56. C. Manion, R. Arlitt, M. I. Campbell, I. Tumer, R. Stone, and P. A. Greaney, "Automated design of flexible linkers," *Dalton Transactions*, vol. 45, no. 10, pp. 4338–4345, 2016.
57. A. U. Ortiz, A. Boutin, A. H. Fuchs, and F. m. c.-X. Coudert, "Anisotropic elastic properties of flexible metal-organic frameworks: How soft are soft porous crystals?," *Phys. Rev. Lett.*, vol. 109, p. 195502, Nov 2012.
58. N. S. Ketkar, L. B. Holder, and D. J. Cook, "Subdue: Compression-based frequent pattern discovery in graph data," in *Proceedings of the 1st international workshop on open source data mining: frequent pattern mining implementations*, pp. 71–76, ACM, 2005.
59. H. B. Mann and D. R. Whitney, "On a test of whether one of two random variables is stochastically larger than the other," *The annals of mathematical statistics*, pp. 50–60, 1947.
60. Y. Benjamini and Y. Hochberg, "Controlling the false discovery rate: a practical and powerful approach to multiple testing," *Journal of the Royal Statistical Society. Series B (Methodological)*, pp. 289–300, 1995.
61. D. B. Turner, S. M. Tyrrell, and P. Willett, "Rapid quantification of molecular diversity for selective database acquisition," *Journal of chemical information and computer sciences*, vol. 37, no. 1, pp. 18–22, 1997.

62. N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch, and G. R. Hutchison, "Open babel: An open chemical toolbox," *J Cheminf*, vol. 3, p. 33, 2011.
63. M. O'Neill and C. Ryan, "Grammatical evolution," *IEEE Transactions on Evolutionary Computation*, vol. 5, pp. 349–358, Aug 2001.
64. J. Brownlee, "Grammatical evolution." http://www.cleveralgorithms.com/nature-inspired/evolution/grammatical_evolution.html. accessed 2017 – 05 – 27.
65. M. A. Addicoat, N. Vankova, I. F. Akter, and T. Heine, "Extension of the universal force field to metalorganic frameworks," *Journal of Chemical Theory and Computation*, vol. 10, no. 2, pp. 880–891, 2014. PMID: 26580059.
66. D. E. Coupry, M. A. Addicoat, and T. Heine, "Extension of the universal force field for metalorganic frameworks," *Journal of Chemical Theory and Computation*, vol. 12, no. 10, pp. 5215–5225, 2016. PMID: 27580382.
67. M. Haranczyk, C. Rycroft, R. Martin, and T. Willems, "Zeo++: High-throughput analysis of crystalline porous materials, v0. 2.2," *Lawrence Berkeley National Laboratory, Berkeley*, 2012.
68. U. Gran, "Synthesis of a new and versatile macrocyclic nadh model," *Tetrahedron*, vol. 59, no. 24, pp. 4303–4308, 2003.
69. G. Granucci and M. Persico, "Critical appraisal of the fewest switches algorithm for surface hopping," *The Journal of Chemical Physics*, vol. 126, no. 13, p. 134114, 2007.
70. Demselben, "Ueber das stickstoffbenzid," *Annalen der Pharmacie*, vol. 12, no. 2-3, pp. 311–314, 1834.
71. E. Merino and M. Ribagorda, "Control over molecular motion using the cis-trans photoisomerization of the azo group," *Beilstein Journal of Organic Chemistry*, vol. 8, pp. 1071–1090, 2012.
72. Y. Li and B. Hartke, "Approximate photochemical dynamics of azobenzene with reactive force fields," *The Journal of Chemical Physics*, vol. 139, p. 224303, dec 2013.
73. M. Bckmann, S. Braun, N. L. Doltsinis, and D. Marx, "Mimicking photoisomerisation of azo-materials by a force field switch derived from nonadiabatic ab initio simulations: Application to photoswitchable helical foldamers in solution," *The Journal of Chemical Physics*, vol. 139, p. 084108, aug 2013.
74. M. Elbing, A. Baszczyk, C. von Hnisch, M. Mayor, V. Ferri, C. Grave, M. A. Rampi, G. Pace, P. Samor, A. Shaporenko, and M. Zharnikov, "Single component self-assembled

- monolayers of aromatic azo-biphenyl: Influence of the packing tightness on the sam structure and light-induced molecular movements,” *Advanced Functional Materials*, vol. 18, no. 19, pp. 2972–2983, 2008.
75. C. B. Browne, E. Powley, D. Whitehouse, S. M. Lucas, P. I. Cowling, P. Rohlfshagen, S. Tavener, D. Perez, S. Samothrakis, and S. Colton, “A survey of monte carlo tree search methods,” *IEEE Transactions on Computational Intelligence and AI in Games*, vol. 4, pp. 1–43, mar 2012.
 76. D. Silver, A. Huang, C. J. Maddison, A. Guez, L. Sifre, G. van den Driessche, J. Schrittwieser, I. Antonoglou, V. Panneershelvam, M. Lanctot, S. Dieleman, D. Grewe, J. Nham, N. Kalchbrenner, I. Sutskever, T. Lillicrap, M. Leach, K. Kavukcuoglu, T. Graepel, and D. Hassabis, “Mastering the game of go with deep neural networks and tree search,” *Nature*, vol. 529, pp. 484–489, jan 2016.
 77. L. Kocsis, C. Szepesvári, and J. Willemsen, “Improved monte-carlo search,” *Univ. Tartu, Estonia, Tech. Rep.*, vol. 1, 2006.
 78. M. Tokic, “Adaptive ϵ -greedy exploration in reinforcement learning based on value differences,” in *Annual Conference on Artificial Intelligence*, pp. 203–210, Springer, 2010.
 79. P. Auer, N. Cesa-Bianchi, and P. Fischer, “Finite-time analysis of the multiarmed bandit problem,” *Machine Learning*, vol. 47, no. 2, pp. 235–256, 2002.
 80. A. Sabharwal, H. Samulowitz, and C. Reddy, “Guiding combinatorial optimization with uct,” in *International Conference on Integration of Artificial Intelligence (AI) and Operations Research (OR) Techniques in Constraint Programming*, pp. 356–361, Springer, 2012.
 81. P. G. Boyd and T. K. Woo, “A generalized method for constructing hypothetical nanoporous materials of any net topology from graph theory,” *CrystEngComm*, vol. 18, no. 21, pp. 3777–3792, 2016.
 82. A. Coskun, M. Banaszak, R. D. Astumian, J. F. Stoddart, and B. A. Grzybowski, “Great expectations: can artificial molecular machines deliver on their promise?,” *Chem. Soc. Rev.*, vol. 41, pp. 19–30, 2012.
 83. B. de Lacy Costello, R. Toth, C. Stone, A. Adamatzky, and L. Bull, “Implementation of glider guns in the light-sensitive belousov-zhabotinsky medium,” *Physical Review E*, vol. 79, no. 2, p. 026114, 2009.
 84. K. Kumar, C. Knie, D. Bléger, M. A. Peletier, H. Friedrich, S. Hecht, D. J. Broer, M. G. Debije, and A. P. Schenning, “A chaotic self-oscillating sunlight-driven polymer actuator,” *Nature communications*, vol. 7, 2016.

85. S.-Y. Ding and W. Wang, "Covalent organic frameworks (cofs): from design to applications," *Chem. Soc. Rev.*, vol. 42, pp. 548–568, 2013.
86. J. Tian, T.-Y. Zhou, S.-C. Zhang, S. Aloni, M. V. Altoe, S.-H. Xie, H. Wang, D.-W. Zhang, X. Zhao, Y. Liu, and Z.-T. Li, "Three-dimensional periodic supramolecular organic framework ion sponge in water and microcrystals," vol. 5, pp. 5574–, Dec. 2014.