**Proceedings of the ASME 2015 International Design Engineering Technical Conferences &
Computers and Information in Engineering Conference
IDETC/CIE 2015
August 2-5, 2015, Boston, Massachusetts, USA**

DETC2015-46078

TOWARDS AUTOMATED DESIGN OF MECHANICALLY FUNCTIONAL MOLECULES

Charles A. Manion [*], Ryan Arlitt, Irem Tumer, Matthew I. Campbell, P. Alex Greaney
Design Engineering Lab
Department of Mechanical, Industrial, and Manufacturing Engineering
Oregon State University
Corvallis, Oregon 97333
Email: manionc@onid.oregonstate.edu

ABSTRACT

Metal Organic Responsive Frameworks (MORFs) are a proposed new class of smart materials consisting of a Metal Organic Framework (MOF) with photoisomerizing beams (also known as linkers) that fold in response to light. Within a device these new light responsive materials could provide the capabilities such as photo-actuation, photo-tunable rigidity, and photo-tunable porosity. However, conventional MOF architectures are too rigid to allow isomerization of photoactive sub-molecules. We propose a new computational approach for designing MOF linkers to have the required mechanical properties to allow the photoisomer to fold by borrowing concepts from de novo molecular design and graph synthesis. Here we show how this approach can be used to design compliant linkers with the necessary flexibility to be actuated by photoisomerization and used to design MORFs with desired functionality.

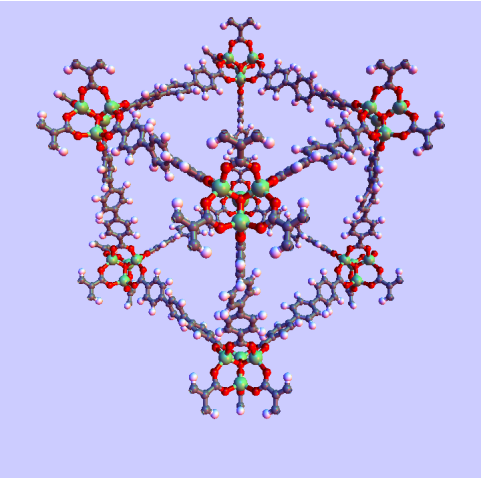

FIGURE 1. unit cell of MOF-15

INTRODUCTION

This paper seeks to engender a whole new class of photoresponsive materials capable of changing shape and porosity consisting of a Metal Organic Framework(MOF) with photo-folding molecules. We call this material a MORF, or Metal-Organic-Responsive-Framework. Such a material could have applications that are not currently possible, one of such applications is shown in Figure 2. By making a material that changes its porosity in response to light, a light controlled "gas sponge" can be created. Increasing pore size can be used to absorb gas into a framework and decreasing pore size can be used to release it. Designing such a material to have requisite macroscale material properties to facilitate an application like this is the problem of designing a 3-periodic compliant mechanism to meet specified force and motion requirements. Much as mechanism design at the macroscale involves building up new mechanisms from a catalog of joints, gears, and links, new shape changing metal organic frameworks can be constructed by choosing from a database of linkers, secondary building units, and framework

[*]Address all correspondence to this author.

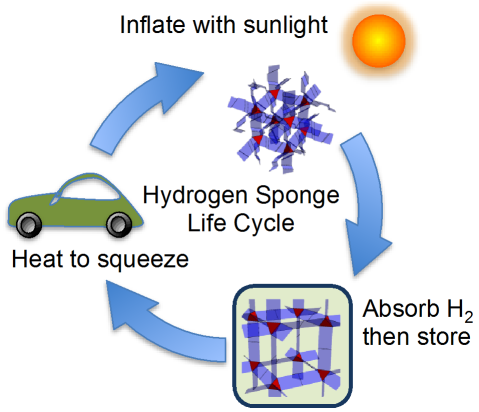

FIGURE 2. Envisioned Application

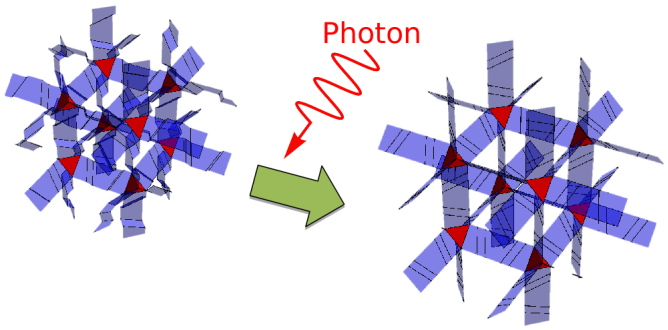

FIGURE 3. Metal Organic Responsive Framework

topologies. Design approaches used at the macroscale to synthesize planar mechanisms [1], gearboxes [2], and create conceptual designs [3] are well suited to the problem of searching through the design space by combining parts together. This paper presents a method for designing molecules to obtain a given force displacement curve and a novel graph grammar based approach for synthesizing molecules with concerns for chemical feasibility.

BACKGROUND

Metal Organic Frameworks are a class of highly porous materials consisting of organic linkers held together by metal ions. Figure 1 shows a single unit cell of the metal organic framework known as MOF-15. The chains of benzene rings are the linkers. The green atoms are metal ions which bind to the negatively charged carboxylate groups on these linkers to form what are called secondary building units or SBUs. Photoisomers are molecules that reversibly fold when exposed to certain wavelengths of light. Figure 3 shows a conceptual model of what one obtains when photoisomers and metal organic frameworks are combined, which we refer to as a metal organic responsive framework or MORF. This material can be thought of as a three dimensional compliant mechanism that repeats practically infinitely in all directions. One of the difficulties with integrating photoisomers into MOFs to make the material that we have proposed is that MOFs are very rigid and can prevent the photoisomers from folding. Photoisomers that have been integrated into MOFs often do not completely fold [4] as the framework constrains motion. Further complicating the problem, the process of photoisomerization is a random process, meaning that if the framework is designed improperly, photoisomers will prevent other photoisomers from fold. The mechanical properties of a MOFs are largely dictated by the mechanical properties of the linkers that constitute them, so in order to design shape-changing MOFs, linkers that facilitate the folding of photoisomers need to be developed.

The problem of designing molecules to have specified mechanical properties is very different than similar problems at the macroscale. At first it would appear that a molecule can be approximated as a linkage, and indeed this can be done for large molecules like proteins, however, we are considering much smaller molecules. At this scale rigid joint-like structures are difficult to construct. While a rotatable bond may resemble a macroscale revolute joint, rotatable bonds can experience significantly less bending. Unlike linkages at the macroscale, rotatable bonds are often subject to relatively large rotational energy barriers. At this scale molecules behave somewhat like compliant mechanisms. However, there are some key differences from compliant mechanism design at the macroscale. While material properties at the macroscale can often times be linear, interatomic potential functions are highly nonlinear. Disconnected atoms often experience can be attracted and repelled to each other by van der Waals forces. Interactions between unconnected components in this manner are rare at the macroscale. In compliant mechanism design at the macroscale, one can remove material to make regions of variable stiffness. However, at the nanoscale one cannot have half an atom or easily change the bond stiffness between two different atoms. In addition, the problem is further differentiated from macroscale design in that atoms can only bond to other atoms in ways that are chemical feasible.

There has been a large amount of work on designing molecules to have specified properties in the field of drug design. The drug design problem can be stated as finding molecule which places certain molecular features relative to each other so as to bind to a particular protein. This is similar to designing a key to fit a lock. Approaches to designing drugs most often consist of screening existing molecules to determine if they have the required properties or constructing new molecules. This technique of constructing new molecules to have desired properties is known as de novo design [5], [6]. De novo molecular design

approaches typically generate structures by combining smaller molecules known as fragments into larger molecules. One problem with this approach is that in order to get a large variety of structures, one needs a large number of fragments.

Another approach used to generate molecular structures is to carry out virtual reactions between the molecule being built up and a set of reactant molecules [7]. This has the advantage of ensuring only chemically feasible structures will be built and being able to immediately evaluate the expense of synthesizing a candidate molecule; however, this makes searching the space of chemical structures more difficult. Most often de novo design techniques use evolutionary algorithms to search through the possible space of molecules. While de novo design techniques have primarily been used for drug design they have also been used to design molecular tweezers [8], molecular polyhedra(non-periodic MOFs) [9], and MOFs [10]. However, all these approaches are focused on designing molecules to achieve a required shape and have not been used to design molecules to have specified mechanical properties. There has been a great deal of work focused on designing molecules that have certain shapes, but little work has focused on designing molecules with mechanical functionality. The most relevant work in this area is that by [11], where DNA origami techniques were used to construct nanoscale analogs of macroscale linkages. Much of this work is inapplicable to our domain and relevant only to the DNA origami domain. It is difficult to apply their DNA origami design techniques to make molecules capable of forming a MOF. In addition, the domain of MORF design involves smaller size scales. While the DNA origami mechanisms presented in [11] have dimensions in the hundreds of nanometers, the largest pore diameter found in a metal organic framework is 9.8 nanometers [12]. This makes it difficult to make structures similar to the joints they designed.

Design approaches used at the macroscale to synthesize planar mechanisms [1], gearboxes [2], and create conceptual designs [3] are well suited to the problem of searching through the design space of MOFs. These techniques use graph grammars as a means of efficiently generating candidate solutions combined with graph search to guide searching through the state space. A graph grammar is a method for transforming a subgraph found in a larger graph into a different subgraph. A transformation of one subgraph into another is called a graph grammar rule. The subgraph to be found is called the left hand side of a rule and the subgraph it is to be transformed into is called the right hand side. The act of finding the left hand side in a larger graph is called rule recognition. Using graph grammars for computational design synthesis has proven to be faster than genetic algorithms for some problems [13]. Graphs grammars are well suited to solve the problem of molecular design. Molecules can be represented as undirected labeled graphs with atoms represented by nodes and bonds between atoms represented with arcs. Atom type can be represented with a node label and bond order can be by an arc label. Graph grammars have also been used in computational chemistry systems as a means of describing chemical reactions [14], as a chemical reaction can be described as a transformation of chemical graph. This was employed by [15] as a means of carrying out chemical reactions to construct candidate drug molecules. However, the difficulty in applying this approach to our domain requires that all molecules we generate contain carboxylate groups so that they can bind to metal ions. This approach does not guarantee that generated molecules will contain carboxylate groups, meaning that a large amount of generated molecules must be discarded. In addition, this approach along with other de novo design approaches requires a large fragment database to produce a variety of molecules. Fragment databases are generally large and often contain thousands of different compounds. In addition, fragment databases often require domain knowledge to construct. In drug design, fragment databases are often produced by fragmenting drugs that have already been produced [6]. As our domain is completely new, this approach is impractical.

GRAPH SYNTHESIS APPROACH

To design candidate linker molecules one needs a way to produce chemically feasible structures. To accomplish this, a novel molecular graph grammar based around hybridization state is employed. The graph grammar rules used here are illustrated in Figure 4. These graph grammar rules make up one ruleset. In our system, molecules are represented with an undirected graph consisting of nodes which represent atoms and arcs which represent bonds. Each arc has a single label representing its bond order which can be either single or double. Each node has a label specifying which element it represents and every node that does not represent hydrogen has a label specifying its hybridization state. Hybridization is the number of hybridized orbitals that an atom has and can provide information on how atoms are located with respect to each other. For example an sp3 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 The graph grammar rules may also have additional conditions built into them to prevent chemically infeasible or undesirable structures from being produced. Two undesirable conditions are prevented by these

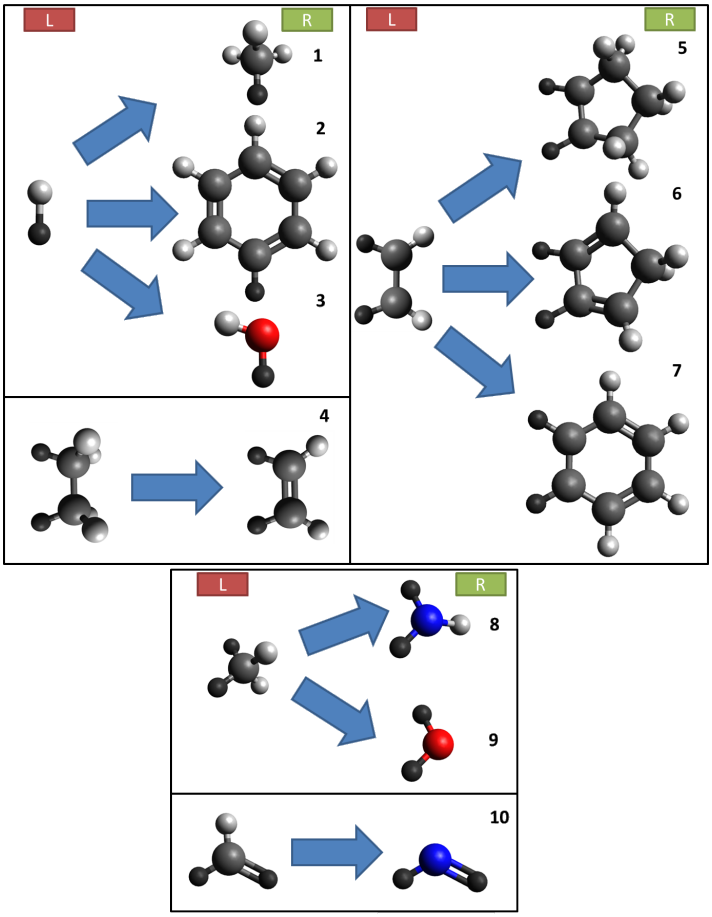

FIGURE 4. Rules: white atoms are hydrogen, grey atoms are carbon, blue atoms are nitrogen, and red atoms are oxygen. Black atoms can be any atom, with the exception that black atoms cannot be nitrogen when a black atom is bonded to a nitrogen atom or oxygen when a black atom is bonded to an oxygen atom.

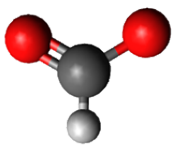

FIGURE 5. Initial Seed Molecule

rules, nitrogen to nitrogen bonds and oxygen to oxygen bonds. Both of these bonds are often unstable and difficult to synthesize so it is worth excluding solutions containing them. These cases are prevented with negative labels, that is node labels that must not be present in the left hand side of the rule. To make the problem more amenable to graph search, the graph grammar is designed such that rules can never "undo" the application of other rules. Such a scenario is undesirable as it leads to rules making states that have already been made and can lead to the search getting stuck in a loop. There are three basic types of rules, add molecule, fuse ring, and transmute all types of which are shown in Figure 4. One of the most versatile ways of building up molecules is to replace a hydrogen atom with another molecule. Hydrogen can only bond to one other atom with a single bond. So it is possible to make a chemically feasible molecule by joining the two molecules with a single bond at two locations that held hydrogen. These "replace hydrogen" rules replace a hydrogen with another molecule. Rules 1,2,3 in Figure 4 are "replace hydrogen" rules. In order to make molecules with connected rings, not one, but two hydrogens must be replaced with ring forming members. "Fuse ring" rules add a ring molecule onto an existing molecular ring by replacing two hydrogens with a connected set of atoms. Rules 5, 6, and 7 are "fuse ring" rules. "Transmute rules" change the hybridization state of atoms and their element type. One example of a "transmute rule" is replacing a sp2 hybridized carbon atom with with a hydrogen attached to it with an sp2 hybridized nitrogen atom removing the hydrogen in the process. "Transmute rules" can change an atom into a different type of atom, as in rules 8, 9, and 10 or modify atoms hybridization states as in rule 4. All "transmute rules" also remove hydrogen atoms. Removing hydrogen atoms helps to further define the molecule, by removing possible choices. While this ruleset is not universal, it is not capable of making every possible organic molecule, it is capable of making a large variety of molecules applicable to our domain. In addition, universality is not desirable for this application. While this ruleset cannot make structures similar to that found in graphene, such structures are difficult to work with chemically and do not provide much benefit in terms of controllable mechanical properties.

All molecules are built from a seed molecule, shown in Figure 5, consisting of a carbon atom bonded to two oxygen atoms and a hydrogen atom. This is known as a carboxylate group in chemistry and represents the part of the linker that bonds to metal ions. For the time being symmetric ditopic linkers with the photoisomer located in the center of linker are assumed. Because the linker is symmetric, we can simplify things by only constructing half of the linker. All graph grammar rules were implemented in GraphSynth [16].

METHODOLOGY

The simplest design problem in MORFs is designing the linkers which must have mechanical properties that permit photoisomer folding. In order for a photoisomer to fold, the force applied on the photoisomer by the linker must be less than the force the photoisomer applies as it folds. In addition, we should be able to design linkers to not only allow a photoisomer to fold, but to extract the maximum amount possible of mechanical work

from the photoisomer. Therefore, we should be able to design linkers that follow a certain force displacement curve. Displacement being the path the photoisomer takes as it fold. For the purposes designing a linker to allow a photoisomer to move force magnitude is more important than force direction.

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 [17], [18], [19]. SMILES is a format that enables practically any molecule to be represented as a string of characters, which is very memory efficient. As an example, the molecule shown in Figure 9 can be represented by the following SMILES string "OC(=O)c1cc(nc(c1)O)[C@H](C)c1ccc2c(n1)cc1c(n2)cccc1." 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.

To find the force displacement curve of a candidate linker we employ molecular statics to measure reaction force by the linker as it is subjected to the photoisomer's displacement. Molecular statics is a means of atomistic simulation where a system of atoms is subjected to a changing force or displacement and the static equilibrium positions of said atoms are found at every simulation step. Molecular statics has the advantage that it is less computationally intensive than molecular dynamics, which is useful for our need to evaluate large numbers of candidates quickly. Molecular statics was implemented by carrying out energy minimization repeatedly while displacing the ends of the molecule. In the simulation the carboxylate group is fixed and reaction forces are measured as it is subject to the motion of the photoisomer. For this scenario we assume a photoisomer that displaces 0.7 angstroms along a straight line path. We assume this photoisomer attaches to the furthest atom from the carbon atom in the carboxylate group and that the photoisomer displaces in the direction of the initial position of this carbon atom.

Candidate molecules are simulated in LAMMPS [20] using parameters from the Universal Force Field [21]. The Universal Force Field is a generic force field that can provide reasonable interatomic potential system coefficients for practically every element. This is well suited for our domain as metal organic frameworks can utilize a great deal of different metallic elements. 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.

RESULTS

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 1. The 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 2.

The linear force molecule is shown in figure 9 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.

TABLE 1. FORCE CURVE EQUATIONS

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

TABLE 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

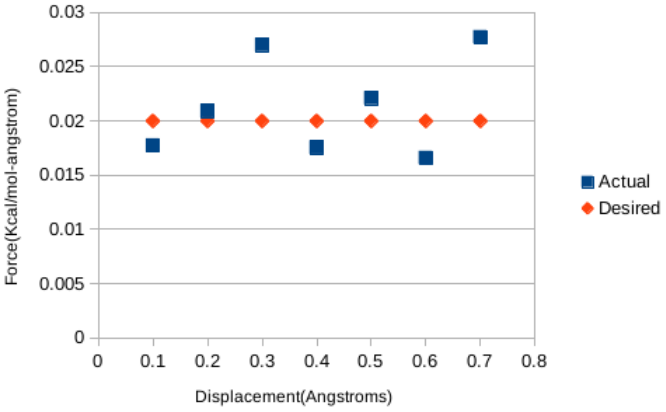

FIGURE 6. Constant Force Displacement Curve

The constant force curve molecule, shown in Figure 7 along with its force displacement curve shown in Figure 6, 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. In the exponential force curve molecule, which is show in Figure 11, the two benzenes are compressed slightly off axis with

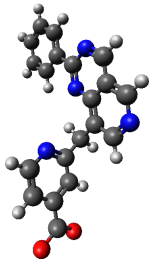

FIGURE 7. Molecule optimized to have a constant force curve

respect to each other. Upon compression, this leads to behavior that produces an exponential force displacement curve with RMS error of 0.15. While this may not seem very close, consider that it is difficult to change the mechanical properties of individual bonds and atoms.

CONCLUSION

A new approach to designing mechanically functional molecules has been developed that opens the door to designing angstrom scale compliant mechanisms. A new graph grammar for generating chemically feasible molecular structures has been developed. A new approach for designing molecules to have specified force displacement curves has been developed.

FUTURE WORK

One potential problem is that the molecules designed here might have multiple stable conformations, or conformers, and the force displacement curve could be different for these other conformations. Therefore, candidate molecules should be checked for the existence of stable conformers. However, it is likely that the conformations the linker can take will be restricted when it exists in a framework, so this may not be as much of a problem. A more urgent issue that needs to be addressed is ensuring that linkers can form into a structure that repeats in three dimensions. Linker design is only one aspect of MOF design. Linkers might exhibit different individual or collective behavior when integrated into a MOF. The current simulation may not necessarily reflect the behavior of a real photoisomer and the ”photoisomer” considered here is purely hypothetical. Instead of specifying a displacement path for the molecule to be deformed and measuring the forces to get and estimate of what the photoisomer would do, it would be better to somehow simulate the photoisomer. This does, however, require that better empirical models of photoisomers be developed. At the time of writing, the force application point is specified to be the atom furthest from a fixed atom. It would make more sense to specify the force application point as this would better describe the design problem. One problem with the current approach is that it only finds one solution for the re-

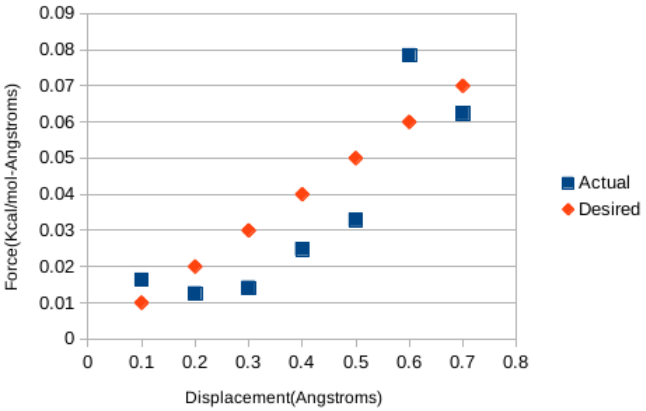

FIGURE 8. Linear Force Displacement Curve

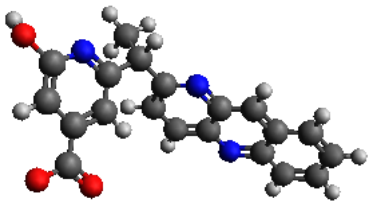

FIGURE 9. Linear Force Displacement Curve

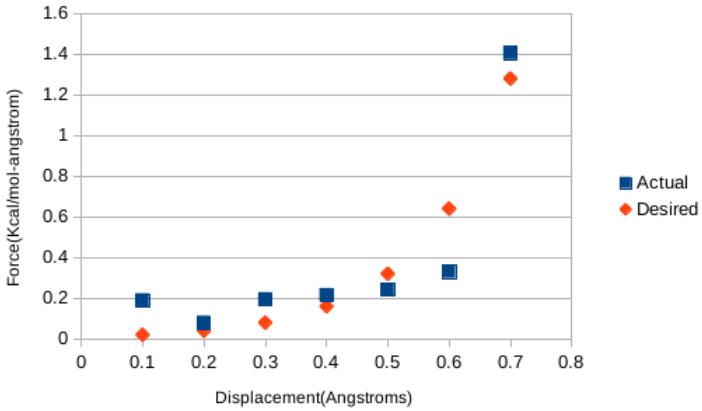

FIGURE 10. Exponential Force Displacement Curve

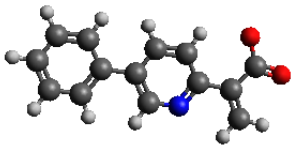

FIGURE 11. Molecule Optimized to have an exponential force curve

quired force curve. It is desirable to generate a diverse array of candidate molecules. The modeling approaches used here are not perfect and better modeling approaches are computationally intensive. Therefore, it would make sense to adapt the techniques presented here to be used to find molecules that are worth investigating with better modeling approaches.

ACKNOWLEDGMENT

The authors would like to acknowledge the useful inputs and discussions by Alex Greaney, Brady Gibbons, Rob Stone, Bryan Maack, and Laura Oliveira from Oregon State University; as well as Jeffrey Rack from Ohio State University. This research is funded by the W.M. Keck Foundation. All views expressed in this article are those of the authors and do not necessarily represent the views of the W.M. Keck Foundation.

REFERENCES

[1] Radhakrishnan, P., and Campbell, M., 2011. "A graph grammar based scheme for generating and evaluating planar mechanisms". In *Design Computing and Cognition 10*, J. Gero, ed. Springer Netherlands, pp. 663–679.

[2] Starling, A. C., and Shea, K., 2005. "Virtual synthesisers for mechanical gear systems". In ICED 05: 15th International Conference on Engineering Design: Engineering Design and the Global Economy, Engineers Australia, p. 679.

[3] Kurtoglu, T., and Campbell, M. I., 2009. "Automated synthesis of electromechanical design configurations from empirical analysis of function to form mapping". *Journal of Engineering Design,* ***20***(1), pp. 83–104.

[4] Lyndon, R., Konstas, K., Ladewig, B. P., Southon, P. D., Kepert, P. C. J., and Hill, M. R., 2013. "Dynamic photoswitching in metalorganic frameworks as a route to low-energy carbon dioxide capture and release". *Angewandte Chemie International Edition,* ***52***(13), pp. 3695–3698.

[5] Schneider, G., and Fechner, U., 2005. "Computer-based de novo design of drug-like molecules". *Nat Rev Drug Discov,* ***4***(8), Aug., pp. 649–663.

[6] Schneider, G., 2013. *De novo Molecular Design*. Wiley.

[7] Hartenfeller, M., Zettl, H., Walter, M., Rupp, M., Reisen, F., Proschak, E., Weggen, S., Stark, H., and Schneider, G., 2012. "Dogs: Reaction-driven ¡italic¿de novo¡/italic¿ design of bioactive compounds". *PLoS Comput Biol,* ***8***(2), 02, p. e1002380.

[8] Huang, H., and Drueckhammer, D. G., 2006. "A modular molecular tweezer designed using caveat". *Chem. Commun.*, pp. 2995–2997.

[9] Young, N. J., and Hay, B. P., 2013. "Structural design principles for self-assembled coordination polygons and polyhedra". *Chem. Commun.,* ***49***, pp. 1354–1379.

[10] Wilmer, C. E., Leaf, M., Lee, C. Y., Farha, O. K., Hauser, B. G., Hupp, J. T., and Snurr, R. Q., 2012. "Large-scale screening of hypothetical metal-organic frameworks". *Nat Chem,* ***4***(2), Feb., pp. 83–89.

[11] Marras, A. E., Zhou, L., Su, H.-J., and Castro, C. E., 2015. "Programmable motion of dna origami mechanisms". *Proceedings of the National Academy of Sciences,* ***112***(3), pp. 713–718.

[12] Deng, H., Grunder, S., Cordova, K. E., Valente, C., Furukawa, H., Hmadeh, M., Gndara, F., Whalley, A. C., Liu, Z., Asahina, S., Kazumori, H., OKeeffe, M., Terasaki, O., Stoddart, J. F., and Yaghi, O. M., 2012. "Large-pore apertures in a series of metal-organic frameworks". *Science,* ***336***(6084), pp. 1018–1023.

[13] Patel, J., and Campbell, M. I., 2010. "An approach to automate and optimize concept generation of sheet metal parts by topological and parametric decoupling". *Journal of Mechanical Design,* ***132***(5), Apr., pp. 051001–051001.

[14] Reisen, F. H., Schneider, G., and Proschak, E., 2009. "Reaction-mql: Line notation for functional transformation". *Journal of Chemical Information and Modeling,* ***49***(1), pp. 6–12. PMID: 19090785.

[15] Mann, M., Ekker, H., and Flamm, C., 2013. "The Graph Grammar Library - a generic framework for chemical graph rewrite systems". *ArXiv e-prints*, Apr.

[16] Campbell, M. I., 2014. The official grapsynth site.

[17] Weininger, D., 1988. "Smiles, a chemical language and information system. 1. introduction to methodology and encoding rules". *Journal of Chemical Information and Computer Sciences,* ***28***(1), pp. 31–36.

[18] Weininger, D., Weininger, A., and Weininger, J. L., 1989. "Smiles. 2. algorithm for generation of unique smiles notation". *Journal of Chemical Information and Computer Sciences,* ***29***(2), pp. 97–101.

[19] Weininger, D., 1990. "Smiles. 3. depict. graphical depiction of chemical structures". *Journal of Chemical Information and Computer Sciences,* ***30***(3), pp. 237–243.

[20] Plimpton, S., 1995. "Fast parallel algorithms for short-range molecular dynamics". *J Comp Phys,* ***117***, pp. 1–19. http://lammps.sandia.gov.

[21] Rappe, A. K., Casewit, C. J., Colwell, K. S., Goddard, W. A., and Skiff, W. M., 1992. "Uff, a full periodic table force field for molecular mechanics and molecular dynamics simulations". *Journal of the American Chemical Society,* ***114***(25), pp. 10024–10035.