# Automated design of flexible linkers

**Article** *in* Dalton Transactions · December 2015

Impact Factor: 4.2 · DOI: 10.1039/c5dt03511b



**6 authors**, including:

Charles Manion

Oregon State University

**1** PUBLICATION **0** CITATIONS

SEE PROFILE

Matthew I. Campbell

Oregon State University

**127** PUBLICATIONS **1,338** CITATIONS

SEE PROFILE

Irem Y. Tumer

Oregon State University

**210** PUBLICATIONS **1,481** CITATIONS

SEE PROFILE

P. Alex Greaney

Oregon State University

**32** PUBLICATIONS **430** CITATIONS

SEE PROFILE

Dalton Transactions

PAPER

View Article Online
View Journal

Cite this: DOI: 10.1039/c5dt03511b

# Automated design of flexible linkers†

Charles Manion,‡[a] Ryan Arlitt,‡[a] Matthew I. Campbell,[a] Irem Tumer,[a] Rob Stone[a] and P. Alex Greaney*[b]

Received 8th September 2015,
Accepted 7th December 2015

DOI: 10.1039/c5dt03511b

www.rsc.org/dalton

This paper presents a method for the systematic and automated design of flexible organic linkers for construction of metal organic-frameworks (MOFs) in which flexibility, compliance, or other mechanically exotic properties originate at the linker level rather than from the framework kinematics. Our method couples a graph grammar method for systematically generating linker like molecules with molecular dynamics modeling of linkers' mechanical response. Using this approach we have generated a candidate pool of >59 000 hypothetical linkers. We screen linker candidates according to their mechanical behaviors under large deformation, and extract fragments common to the most performant candidate materials. To demonstrate the general approach to MOF design we apply our system to designing linkers for pressure switching MOFs—MOFs that undergo reversible structural collapse after a stress threshold is exceeded.

## 1 Introduction

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—with gas storage and catalysis being the most aggressively developed applications. However, there are many novel applications that could exploit the open framework and the extra mechanical freedom afforded the materials as a result. Vibrations of linker molecules into the free space surrounding them can lead to negative thermal expansion,[1] auxeticity,[2] and unusual thermal transport properties.[3,4] There are other exciting ways of exploiting linker deformation into the space surrounding it that could lead to other hitherto unexplored applications. One such class of materials that motivates our research is the development of photoactive MOFs that use photoisomerization to actuate linker folding into the cage space. We refer to these shape changing materials as metal–organic-responsive-frameworks or MORFs, and we envision them enabling many new applications such as self-squeezing gas storage sponges; active filtration and catalysis; and solar energy conversion and storage.

Central to designing shape-changing MORFs is designing articulated linkers that can maintain the connectivity of the framework as linkers are actuated stochastically and sequentially. Thus, our approach at this stage focuses on deformation properties originating within the linker rather than geometric freedom of a framework that enables it to skew—as is exemplified by the 'wine-rack' deformation MIL-53.[5] Geometric flexibility of frameworks arises from axial aligned hinges between linkers and the secondary building units (SBUs or nodes). Férey and others[6,7] have outlined an elegant set of requirements for geometric flexibility based on the assumptions that linkers act as relatively rigid linkages in a kinematic model of the framework. However, there are MOFs where flexibility arises from flexibility in the linkers, such as DynaMOF-100.[8] Our work expands on this approach with the objective of designing flexible or articulated linkers with tailored mechanical properties. This is applicable beyond designing linkers to be compatible with complex geometric shape-changes of photoisomerizing moieties in MORFs. There are many other processes that could arise from non-linear linker deformation. Some examples include highly non-linear stiffness yielding an intensity dependent speed for sound propagation; frequency of linker vibrations that are sensitive to stress yielding strong pressure temperature coupling and externally tunable thermal transport properties; or structural transitions driven by changes in vibrational entropy.

The novelty of the work that we present here is a completely general system for designing linkers suitable for any these exotic applications. To demonstrate this system we tackle the more prosaic problem of designing linkers to buckle under pressure for the creation of pressure switching MOFs. This task is not purely academic; such MOFs could provide an important safety or stress limiting function in microsystems, or linker based collapse could be coupled to other functional compounds resident in the cages of the MOF. Importantly we set our task as designing kinematically active linkers—linkers

[a]School of Mechanical, Industrial, & Manufacturing Engineering, Oregon State University, Corvallis, OR 97331, USA
[b]Department of Mechanical Engineering, University of California—Riverside, Riverside, CA 92521, USA. E-mail: agreaney@engr.ucr.edu

†Electronic supplementary information (ESI) available. See DOI: 10.1039/c5dt03511b

‡These authors contributed equally to this work.

that would be a scaffold for functionalization for further chemical activity, or a template for the design of totally different linkers. The remainder of the paper describes firstly our automated linker-like molecule invention scheme, then our computational prediction of linker properties, and our informatics method for relating structure to properties. We finish by describing the inferences that can be drawn from this combined approach.

## 2 Automated molecule design

Central to the design process is the formalization of the molecular design space in a way that is systematically searchable. This is achieved using a graph theoretic approach in which molecules are abstracted as graph representations of their bonding topology.

We have developed a set of graph grammar rules for automatically inventing candidate ditopic linker-like molecules. 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, (called 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 known as rule application. Graph transformations applied recursively on a seed graph can define a set of graphs (molecules in our case) that all heed a set of design constraints. Similar graph grammar approaches have been employed in engineering design automation to design gearboxes,[9] linkages,[10] and even conceptual models for complete mechanical systems.[11] The power of the approach is that the set of feasible designs are all linked to one another through a decision tree of grammar rule applications, which allows this feasible design space to be searched using artificial intelligence tree search algorithms (such as those used in computer chess).

Our graph grammar rules act on molecular graphs, or undirected graphs where every node represents an atom and every arc represents a bond between atoms. Node labels are used to specify atom type and hybridization, while arc labels are used to specify bond order. 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 that generated structures are likely to be chemically feasible. The graph grammar rules are made such that they only recognize locations where the hybridization is appropriate for the corresponding transformation to occur. 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 important for providing feasible transitions for tree search algorithms.

The rules are shown in Fig. 1, grouped by the left-hand-side on which they operate. Rules 1–4 replace a hydrogen bonded

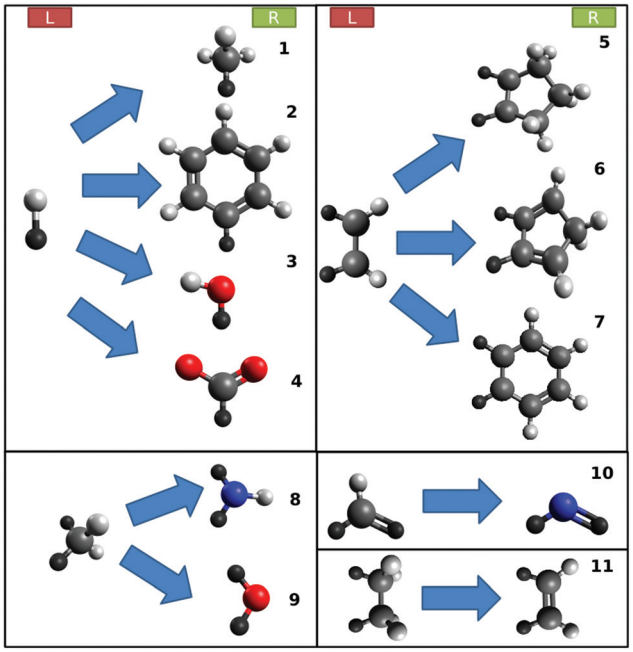



**Fig. 1** Graph grammar rules used to generate linker-like molecules, blue atoms represent nitrogen, gray atoms represent carbon, red atoms represent oxygen, and black atoms represent an atom of any type.

to one atom with another set of atoms. Rules 5–7 replace two hydrogen atoms connected to two carbon atoms that are bonded together with each carbon connected to another atom with a ring. Rules 8–11 change carbon atoms to have different hybridization or to be different types of atoms. Some rules are also subject to additional conditions to prevent undesirable molecules. Rules 3 and 9 do not recognize any atoms that connect to an oxygen atom or a carboxylate group. This prevents the rules from explicitly making structures with peroxides or carbonate groups. Likewise, rules 8 and 10 do not recognize any atoms that connect to nitrogen atoms—preventing the explicit formation of unwanted nitrogen–nitrogen bonds, which would be detrimental to MOF stability. Rule 4 does not apply at any atoms that connect to an oxygen atom to prevent the formation of carbonates. In order to encourage the rules to make more linker-like structures, rules 1–4 and 5–6 have been restricted to only recognize atoms on the surface of the convex hull of the molecule (see Fig. 2). This makes it so that molecules are built outward and helps prevent molecule with overlapping parts from being created.

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.

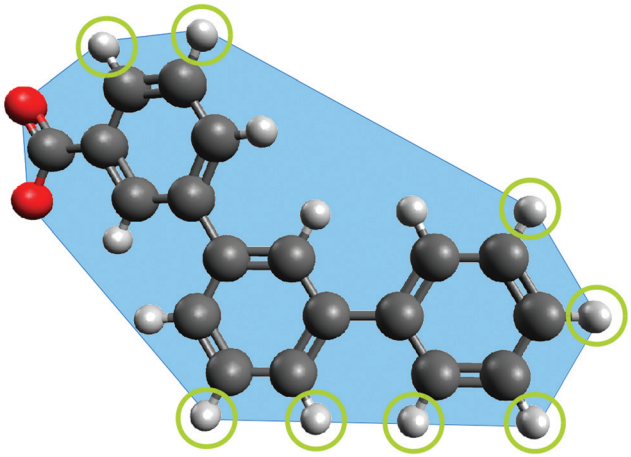


Fig. 2 An example of a molecule with its corresponding convex hull.

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 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 much simpler 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.[12–14]

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[15] 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 22 178 of the molecules generated were found to meet Bao *et al.*'s[15] criterion for linker-likeness of an angle between carboxylate groups greater than 155 degrees.

## 3. Computational prediction of linkers' mechanical response

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.[16] 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.*[17] and using their computed elastic constants we have replotted it here in Fig. 3 along with the response to an imposed uniaxial strain in order to illustrate this point). In our application 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

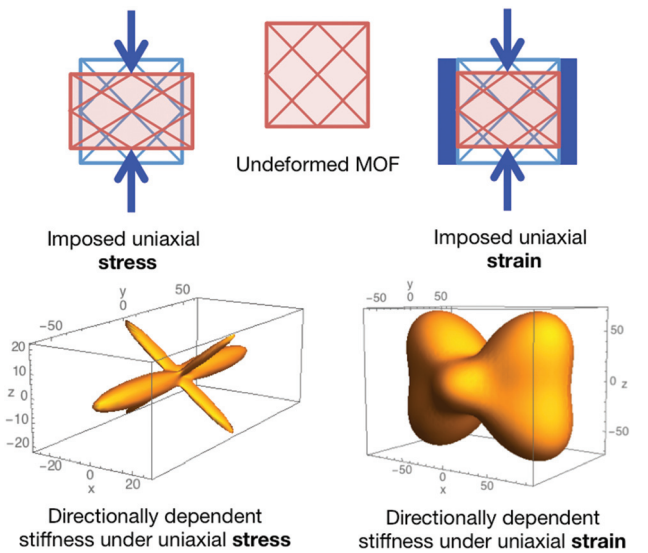



Fig. 3 Directionally dependent stiffnesss indicatrix for MIL-53 under uniaxial stress (left) and uniaxial strain (right). For a pressure switching MOF we require deforming linkers rather than framework flexibility.

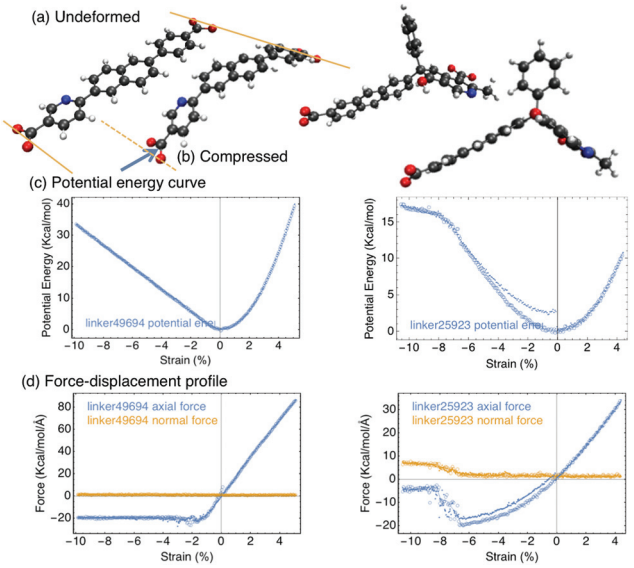



**Fig. 4** Example of a candidate linker, (a) relaxed structure, (b) maximally compressed structure, (c) the molecules' potential energy during deformation, and (d) the force displacement curves for compressing and stretching the molecule at 300 K with a deformation rate of 1 m $s^{-1}$. The left column is for linker candidate 49 694 which exhibits an elastic buckling instability, the right column is for candidate 25 923 which undergoes a stress driven conformation change.

will provide a reversible transition with a continuous and smooth potential energy curve (see Fig. 4). 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 Fig. 4.

The simulation sequence was as follows: (1) the molecule was relaxed to minimize forces to better than $10^{-8}$ eV $Å^{-1}$. (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 $ps^{-1}$, 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 $s^{-1}$) 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,[18] and simulations performed using the LAMMPS molecular dynamics package[19] 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.

### *Ad hoc* characterization of candidates for pressure switching applications

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:

$$f(\varepsilon; k_{\mathrm{I}}, k_I', \varepsilon_{\mathrm{c}}, \Delta\varepsilon, k_{\mathrm{II}}) = \Theta(\Delta\varepsilon - \varepsilon + \varepsilon_{\mathrm{c}})\left(\sqrt{(-k_{\mathrm{II}} + k_{\mathrm{I}})^2(-\varepsilon_{\mathrm{c}} - \Delta\varepsilon + \varepsilon)^2 + \Delta\varepsilon^2(-k_{\mathrm{II}} + k_{\mathrm{I}})^2} - \Delta\varepsilon(-k_{\mathrm{II}} + k_{\mathrm{I}})\right) + k_{\mathrm{I}}\varepsilon(1 - \varepsilon k_I'), \quad (1)$$

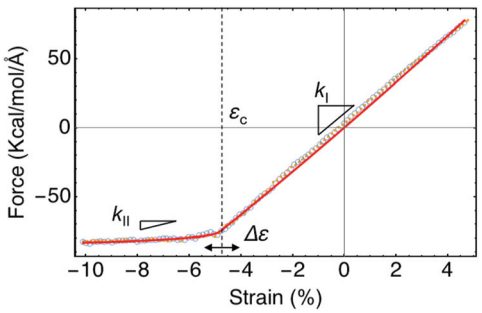



**Fig. 5** Example of the fitting of the model curve indicating the mechanical interpretation of the fitting parameters.

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. 5. The fitting parameters $k_I$, $k'_I$, $\varepsilon_c$, $\Delta\varepsilon$, and $k_{II}$ are mechanically interpretable as, respectively: as 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.

Our linker generation method produced a number of molecules that present unusual solutions to the pressure switching task—mechanisms that were not immediately or intuitively obvious before we began the design search. Of these a number exhibited bistable behavior and mechano-isomerization. One example is the linker candidate 25 923 shown in Fig. 4, which under compression switches conformation through rotation of the phenyl side group inducing a kink in the linker's backbone. As can be seen from the potential energy curve for candidate 25 923 (in Fig. 4) the conformation change takes the molecule onto a different potential energy surface, there is an abrupt drop in force but the transition is reversible and barrierless with negligible hysteresis. Other linkers exhibited mechanical induced switching between stable conformations. These have large hysteresis originating from a mechanism like throwing a two pole toggle switch. Under compression a side group is sterically pushed into a different configuration from which it reverts only when the molecule is put under tension. While these molecules' force displacement behavior are not highly suited for a pressure switch they provide stress induced switching of MOF's pore environment—a mechanism that could be exploited in many applications involving secondary guest species within the MOF framework.

## 4 Fragment mining

In both drug design and mechanical design, there is a well-established connection between a system's structure (*i.e.*, drug molecule or mechanical artifact) and its behavior (biological activity or dynamicity). In drug design this relationship is used to develop Quantitative Structure–Activity Relationship models (QSAR models), which can reduce the time needed to evaluate candidate molecules. This enables molecule designers to consider larger chunks of the design space, thus improving the overall outcomes of a drug design task. In this work, we investigate a similar relationship between structural features of MOF linker candidates and their mechanical behaviors. These mechanical behaviors are simulated under loading conditions that approximate those felt by a linker, as discussed in the prior sections. From these simulation results, we extract a set of linker fragments that correlate with a set of desired mechanical behaviors—in this case behaviors that indicate the functionality of a pressure-sensitive linker. Given this set of linker building blocks as a starting point, a linker designer could then generate linkers that are likely to behave like pressure switches or pressure trips.

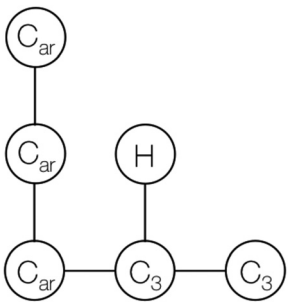
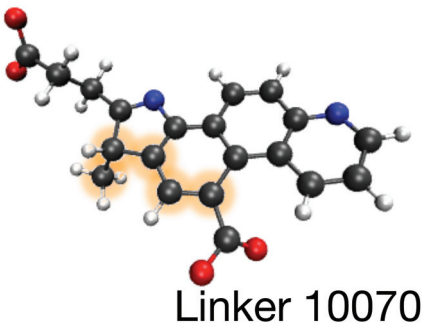
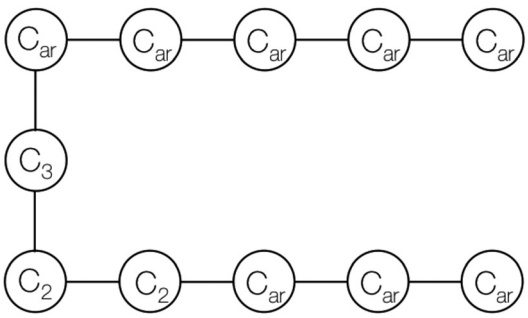
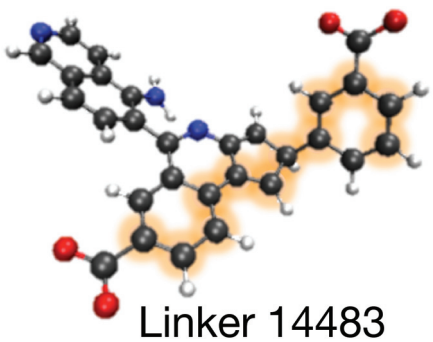
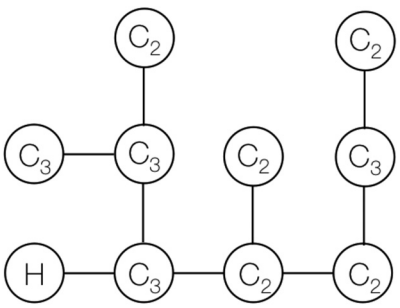
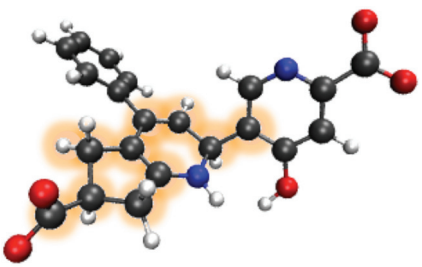
In order to identify appropriate fragments for a specific design problem, one must first establish a set of behavior targets, as discussed in the preceding sections. These behavior descriptors are selected specifically for their applicability to the pressure switch design problem. Different descriptors to address different problems can be extracted from the same simulation data. Next, the total candidate population is screened for linkers that most closely match those target behaviors. This new population undergoes a new search for their characteristic subgraphs using Subdue—a graph-based knowledge discovery system.[20] For each of 13 behavior descriptors given in Table 1, the top 100 and bottom 100 performing linker candidates are extracted from the simulation results to create a training set of exemplar molecules. For the pressure switch design problem, we perform supervised learning using Subdue to extract linker fragments that best describe each performance variable. 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[21] 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

**Table 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 |

**Table 2** Sample of linker fragments that correlate with behavior descriptors

| Behavior descriptor | Fragment graph | Example | *U* test *p*-value |
|---|---|---|---|
| Step height of piecewise fit of force–displacement curve (maximize, linker in tension |  |  Linker 10070 | 0.052 |
| Force at buckling (minimize, linker in compression) |  |  Linker 14483 | 0.029 |
| Nonlinearity in the stiffness around zero strain (minimize, Linker in compression) |  |  Linker 20977 | 0.032 |

Benjamini–Hochberg procedure[22] 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 as simple graphs in ESI†) 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 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. Given this result, it is likely that more holistic structural characteristics are required to adequately design linkers to match each of these target properties, similar to Lipinski's rule of five for drug design.[24]

## 5 Linker population summary

Some general statistics describing the generated linker population are presented in Table 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.[25] 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.*[23]—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.[15] 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.

Table 3 Linker database summary

| | |
|---|---|
| Fingerprint diversity (average pairwise similarity)[23] | 0.549 |
| Percentage with linker-like pairwise angle[15] | 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 | 59 841 |

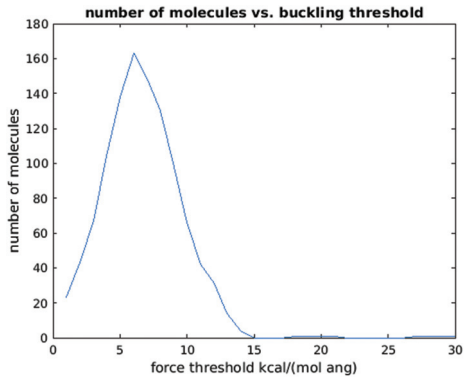



Fig. 6 Number of molecules that behave as good pressure switches at a given pressure.

The average branching factor at various depths describes the average number of rule choices found after a specified number of rule applications. Table 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 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 behavior closest to this ideal case at constant compressive threshold forces from 1–30 kcal (mol Å)$^{-1}$. 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)$^{-1}$. Fig. 6 shows that the most molecules meet this criterion at a force threshold of 6 kcal (mol Å)$^{-1}$, suggesting a small group of highly performant pressure switches.

## 6 Conclusions

This manuscript presents a method to automatically generate and screen MOF linkers for specific mechanical behaviors. The method also extracts the most probable topological fragments that lead to each mechanical behavior, providing a set of linker design shortcuts. Given the resulting linkers and linker fragments discussed in the preceding sections, the method has produced linkers with interesting flexible behaviors. The combination of this method, the linker candidates, and the

linker fragments contribute to the larger motivation of creating MORFs by providing a baseline library of flexible linkers.

There are several areas for future work. One of these areas will be to improve the predictive power of the structure-behavior relationship by constructing a Quantitative Structure–Property Relationship (QSPR) model. This model will relate structural descriptors—such as the presence of specific fragments—to behavioral descriptors that are important to flexible MOF design. Such a model would provide an effective preliminary screening tool for the construction of new flexible molecules in that it would bypass the need for large-scale simulation. It is also worth investigating whether the sequence of a grammar rules used to make a given linker could be used as structural descriptors for a QSPR model. This might enable a QSPR model to be built up as the design space is searched for molecules with desired behaviors. Other future work will extend the library of flexible linker fragments using additional behavior descriptors and Subdue search methods.

The primary thrust of future work will explore the computational and experimental synthesis of new flexible MOF and MORF structures using the sets of linkers and linker fragments that are likely to produce flexible behaviors. This undertaking will require additional research to create targeted materials for specific applications, along with strong attention paid to synthetic accessibility—a constraint that has been greatly relaxed in the work in this manuscript. Exotic new smart materials arising from such work have extraordinary potential in a boundless number of applications.

## Acknowledgements

The authors would like to acknowledge the useful inputs and discussions by Brady Gibbons, Bryan Maack, and Laura Oliveira from Oregon State University; as well as Jeffrey Rack from University of New Mexico. 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 S. S. Han and W. A. Goddard III, *J. Phys. Chem. C*, 2007, **111**, 15185–15191.
2 Z. A. Lethbridge, R. I. Walton, A. S. Marmier, C. W. Smith and K. E. Evans, *Acta Mater.*, 2010, **58**, 6444–6451.
3 B. Huang, A. McGaughey and M. Kaviany, *Int. J. Heat Mass Transfer*, 2007, **50**, 393–404.
4 L. Han, M. Budge and P. A. Greaney, *Comput. Mater. Sci.*, 2014, **94**, 292–297.
5 A. U. Ortiz, A. Boutin, A. H. Fuchs and F.-X. Coudert, *J. Chem. Phys.*, 2013, **138**, 174703.
6 G. Férey and C. Serre, *Chem. Soc. Rev.*, 2009, **38**, 1380–1399.
7 L. Sarkisov, R. L. Martin, M. Haranczyk and B. Smit, *J. Am. Chem. Soc.*, 2014, **136**, 2228–2231.
8 S. Mukherjee, B. Joarder, A. V. Desai, B. Manna, R. Krishna and S. K. Ghosh, *Inorg. Chem.*, 2015, **54**, 4403–4408.
9 A. C. Starling and K. Shea, ICED 05: 15th International Conference on Engineering Design: Engineering Design and the Global Economy, 2005, pp. 679–693.
10 P. Radhakrishnan and M. Campbell, *Design Computing and Cognition âĂŹ10*, Springer, Netherlands, 2011, pp. 663–679.
11 T. Kurtoglu and M. I. Campbell, *J. Eng. Des.*, 2009, **20**, 83–104.
12 D. Weininger, *J. Chem. Inf. Comput. Sci.*, 1988, **28**, 31–36.
13 D. Weininger, A. Weininger and J. L. Weininger, *J. Chem. Inf. Comput. Sci.*, 1989, **29**, 97–101.
14 D. Weininger, *J. Chem. Inf. Comput. Sci.*, 1990, **30**, 237–243.
15 Y. Bao, R. L. Martin, C. M. Simon, M. Haranczyk, B. Smit and M. W. Deem, *J. Phys. Chem. C*, 2015, **119**, 186–195.
16 P. Serra-Crespo, A. Dikhtiarenko, E. Stavitski, J. Juan-Alcaniz, F. Kapteijn, F.-X. Coudert and J. Gascon, *CrystEngComm*, 2015, **17**, 276–280.
17 A. U. Ortiz, A. Boutin, A. H. Fuchs and F. m. c.-X. Coudert, *Phys. Rev. Lett.*, 2012, **109**, 195502.
18 A. K. Rappe, C. J. Casewit, K. S. Colwell, W. A. Goddard and W. M. Skiff, *J. Am. Chem. Soc.*, 1992, **114**, 10024–10035.
19 S. Plimpton, *J. Comput. Phys.*, 1995, **117**, 1–19.
20 N. S. Ketkar, L. B. Holder and D. J. Cook, *Proceedings of the 1st international workshop on open source data mining: frequent pattern mining implementations*, 2005, pp. 71–76.
21 H. B. Mann and D. R. Whitney, *Annals of Mathematical Statistics*, 1947, 50–60.
22 Y. Benjamini and Y. Hochberg, *J. Royal Statistical Soc., Series B (Methodological)*, 1995, 289–300.
23 D. B. Turner, S. M. Tyrrell and P. Willett, *J. Chem. Inf. Comput. Sci.*, 1997, **37**, 18–22.
24 C. A. Lipinski, F. Lombardo, B. W. Dominy and P. J. Feeney, *Adv. Drug Delivery Rev.*, 2001, **46**, 3–26.
25 N. M. O'Boyle, M. Banck, C. A. James, C. Morley, T. Vandermeersch and G. R. Hutchison, *J. Chem. Inf. Comput. Sci.*, 2011, **3**, 33.