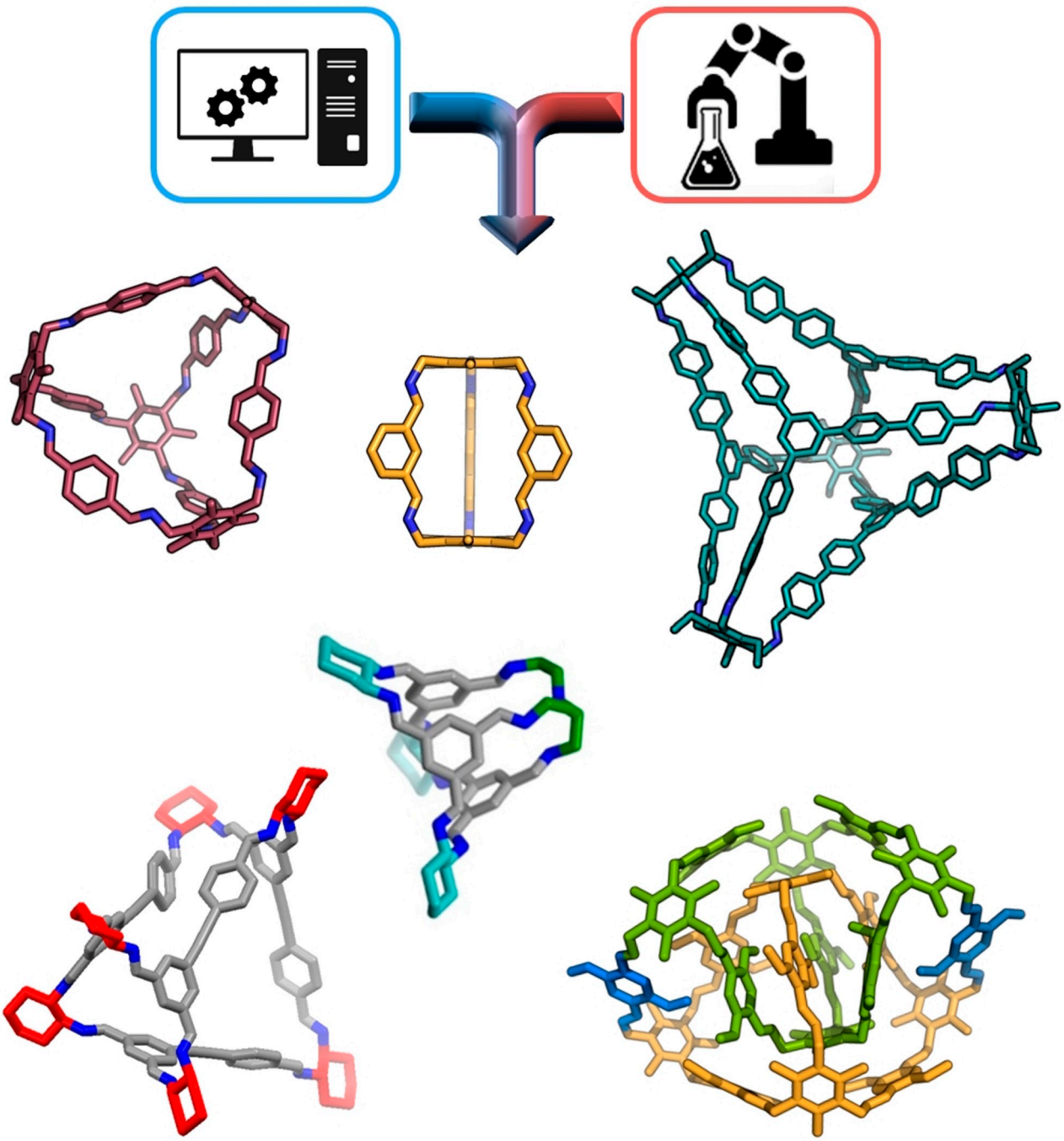
doi.org/10.1002/cplu.202000445

High-Throughput Approaches for the Discovery of Supramolecular Organic Cages

Rebecca L. Greenaway*[a] and Kim E. Jelfs[b]



© 2020 The Authors. Published by Wiley-VCH GmbH

The assembly of complex molecules, such as organic cages, can be achieved through supramolecular and dynamic covalent strategies. Their use in a range of applications has been demonstrated, including gas uptake, molecular separations, and in catalysis. However, the targeted design and synthesis of new species for particular applications is challenging, particularly as the systems become more complex. High-throughput computation-only and experiment-only approaches have been developed to streamline the discovery process, although are still not widely implemented. Additionally, combined hybrid workflows can dramatically accelerate the discovery process and lead to the serendipitous discovery and rationalisation of new supramolecular assemblies that would not have been designed based on intuition alone. This Minireview focuses on the advances in high-throughput approaches that have been developed and applied in the discovery of supramolecular organic cages.

Introduction

Supramolecular synthesis and dynamic covalent chemistries are powerful strategies for the assembly of complex molecular organic materials, such as macrocycles,[1] cages,[2–4] knots,[5] and catenanes,[6] from multiple precursors in a one-pot reaction. Recent examples include Fujita's metal-organic Goldberg polyhedra formed from 144 components,[7] braided molecular knots by Leigh,[8] and Mastalerz's quadruply interlocked giant organic cage catenane formed through the formation of 96 dynamic covalent bonds.[9] However, the targeted design of such molecules can be challenging, especially as the systems become more elaborate. Even when the target molecules are assembled using just two components, there is still the potential to form a wide range of products. This problem becomes increasingly more complex when additional components are employed, with the potential to form different self-sorted structures or statistical distributions/mixtures. Therefore, in general, the discovery of new molecular organic materials and supramolecular assemblies is extremely time consuming.

There is an increasing drive to design molecular materials *a priori* using computational prediction to guide the discovery of useful, and increasingly complex, molecules. To accelerate this discovery process, computational chemists are therefore developing design and prediction tools that help to target materials with particular functions. There has also been a revolution in synthetic materials chemistry with the development of high-throughput automated synthesis and parallel analysis of such materials. These two approaches could be highly complementary, but so far, they are largely disconnected.

[a] *Dr. R. L. Greenaway*
Department of Chemistry, Molecular Sciences Research Hub
Imperial College London
White City Campus, Wood Lane, London W12 0BZ (United Kingdom)
E-mail: r.greenaway@imperial.ac.uk
[b] *Dr. K. E. Jelfs*
Department of Chemistry, Molecular Sciences Research Hub
Imperial College London
White City Campus, Wood Lane, London W12 0BZ (United Kingdom)

This article is part of a Special Collection on "Supramolecular Chemistry: Young Talents and their Mentors". More articles can be found under https://onlinelibrary.wiley.com/doi/toc/10.1002/(ISSN)2192-6506.Supramolecular-Chemistry.

© 2020 The Authors. Published by Wiley-VCH GmbH. This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

While computation can enable the high-throughput screening of millions of potential compounds, which is not viable on an experimental timescale, even with high-throughput automation, one potential bottleneck is the selection of synthetically viable precursors, and also synthetically viable target compounds. One solution would be to solely base modelling on commercially available precursors, but this severely limits the scope of the materials targeted. Further, while high-throughput automation can be used to screen a broad synthetic space, screening thousands of potential combinations, when applied blindly, is inefficient and inelegant, and is often biased as it is based on chemical intuition or on small iterative changes to known compounds. In principle, the discovery of functional materials can be rapidly accelerated by combining computational modelling with state-of-the-art automation. For example, computational modelling can help to guide experimental efforts. To date, however, even with advances in computational modelling to assist in the design of supramolecular and organic materials,[10,11] this approach is still relatively uncommon.

In this Minireview, using examples from our own work in the area, we will summarise recent advances in the computational and experimental toolkits available for the high-throughput discovery of supramolecular assemblies, with a focus on organic cages. We will then discuss how computational modelling can be combined with state-of-the-art automation in hybrid workflows, exploiting a highly collaborative feedback loop to accelerate the discovery process.

High-Throughput Approaches for Materials Discovery

In the past few years, there has been emerging interest in the development of high-throughput computation-only[12] and synthesis-only[13] workflows targeting the discovery of materials with particular applications. For example, these methods have been successfully applied to organic,[14–19] hybrid,[20–23] and inorganic materials,[24] ranging from discrete molecules[17,18] to polymers,[14,15] frameworks,[20–23] and liquids,[16] and includes targeted screening of their solubility,[16] mechanical properties,[14] porosity,[20,22–24] optical band gaps,[19] and their potential to form host-guest complexes.[17] However, while there is this handful of studies on the high-throughput synthesis of materials such as metal-organic frameworks, zeolites, and organic polymers, the use of robotic synthesis for

© 2020 The Authors. Published by Wiley-VCH GmbH

supramolecular self-sorted organic materials and assemblies is rare, which is partly due to the difficulties in predicting the product outcome and the difficulty raised in characterising libraries of supramolecules.

Supramolecular Organic Cages

Porous organic cages are discrete molecules that contain permanent molecular cavities accessible through windows.[25] Typically, these molecules are formed from the combination of two or more multi-functionalised components using reversible dynamic covalent chemistries. This combines the error-correction mechanism of supramolecular synthesis with more robust covalent bonding. These multi-component self-assembly reactions have the potential to form different topologies (Figure 1a),[26] undergo different types of self-sorting to a range of different species (Figure 1b),[4,27] or alternatively, form oligomers, polymers, or framework materials. Due to this variety of different possible reaction outcomes, traditionally, their discovery is a function of synthetic labour, trial-and-error, carrying out iterative changes to known materials, intuitive design strategies, or in some cases, is purely due to serendipity.

Due to their discrete nature, porous organic cages are soluble and their packing can be directed in the solid-state to form interconnected pore networks, making the materials porous.[25,28] This solution processability and porosity has meant that they have been used in a range of applications including molecular isomeric[29] and chiral[30] separations, rare gas separations (Xe/Kr),[30] and processed into molecular-sieving thin-films and membranes for gas separations.[31] They have also been used as a fluorescent sensor for the detection of explosives,[32] and are ideal molecular species with which to generate porous liquids – a new class of functional materials combining the mobility of a liquid with the properties of a porous solid.[33]

Computational Toolkit

Computational modelling can be used for both *a priori* prediction and *a posteriori* rationalisation. While our ultimate aim is *a priori* prediction to identify the most promising systems for experimental validation, these capabilities must be built upon the foundations of reliably modelling these systems through validation of *a posteriori* rationalisation.

Porous organic cage molecules can form in a variety of different topologies based on the connectivity of the components which, based on known systems to date, can number anywhere from 5 up to 20[9,34] components (Figure 1). This complexity means that the manual construction of models of the possible reaction outcomes is laborious. We have developed open-source software called *stk* (supramolecular toolkit) for the assembly of porous organic cages (and other supramolecular materials) in a variety of topologies.[35,36] The first challenge once one has a hypothetical model of a porous cage is for the molecular structure to be representative of the true conformation(s) of the cage. This would typically involve finding the global energy minimum or low-energy molecular conformations through an approach to sample the potential energy landscape of the molecule, such as conformer search algorithms or molecular dynamics simulations with conformer sampling. These calculations would generally be performed at the force-field level, necessary due to the computational expense of modelling these systems that typically have many hundreds of atoms. In some cases, to reliably distinguish between different low-energy conformations, additional calculations to refine the relative energies with density functional theory (DFT) may be required.[26]

An additional challenge for molecular structure prediction originates from the fact that many of these large systems can contain many flexible components that allow for the structural collapse, and associated loss of the internal molecular pore of the molecule.[37] This is known as a lack of "shape persistence" – the loss of an internal cavity in the absence of solvent



Becky Greenaway is a Royal Society University Research Fellow at Imperial College London, UK. Becky's research focuses on porous liquids and the use of high-throughput automation in the synthesis of molecular organic materials and supramolecular assemblies. After completing her PhD on palladium-catalysed cascade cyclisations at the University of Oxford, she worked as a postdoctoral researcher at Liverpool before starting her independent research in 2019. She joined Imperial College in 2020.



Kim Jelfs is a Senior Lecturer and Royal Society University Research Fellow at the Department of Chemistry at Imperial College London, UK. Kim specialises in the use of computer simulations to assist in the discovery of supramolecular materials. After a PhD modelling zeolitic crystal growth at UCL, she worked as a postdoc at the University of Liverpool, before beginning her independent research at Imperial College in 2013. She was awarded a Royal Society of Chemistry Harrison-Meldola Memorial Prize in 2018 and a Philip Leverhulme Prize in Chemistry in 2019.
Becky has received support and advice from Kim since commencing her postdoctoral studies in 2013. This contribution arose from their ongoing collaboration on merging high-throughput prediction and experimental techniques in the discovery of molecular organic materials formed using dynamic covalent chemistries.

© 2020 The Authors. Published by Wiley-VCH GmbH

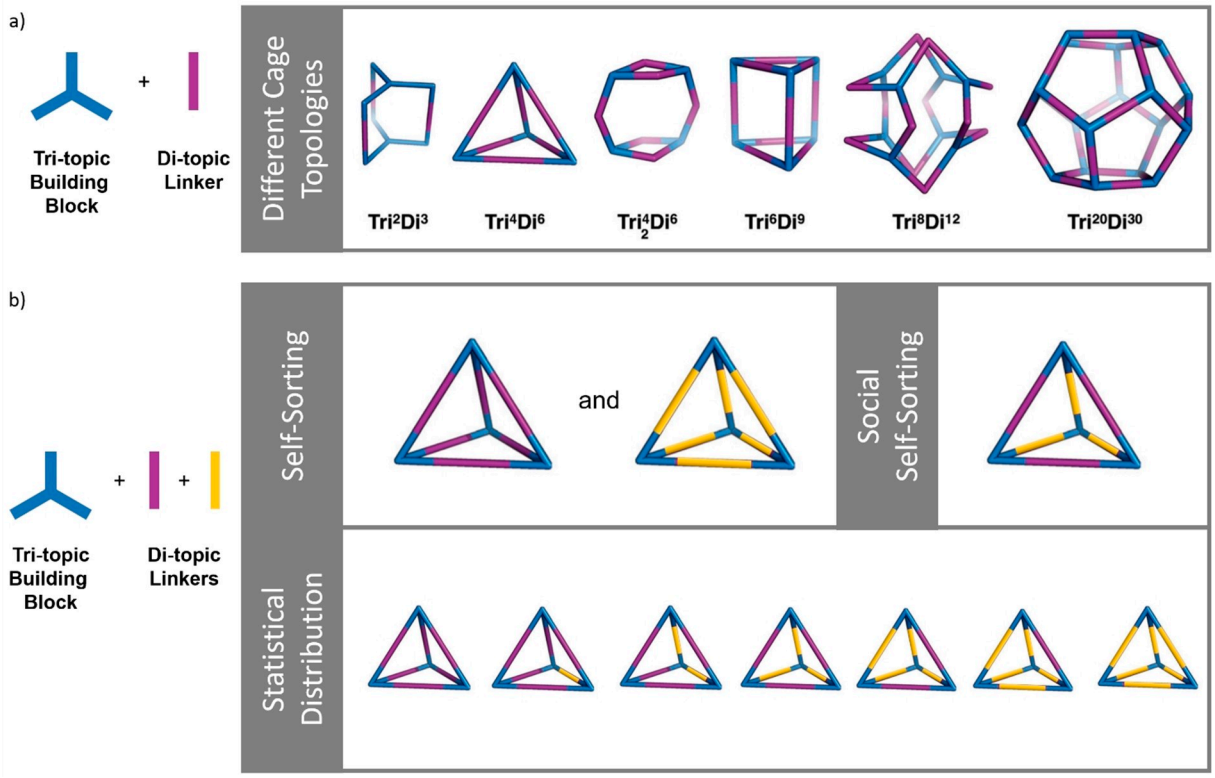


Figure 1. Example schematic representation of the different potential topologies that can be formed from (a) a tri-topic building block (blue) and a di-topic linker (purple or yellow),[26] and (b) the different self-sorting or statistical outcomes when a mixture of di-topic linkers is incorporated. Reproduced (adapted) from reference [26] with permission from the Royal Society of Chemistry.

molecules. While 'collapsed' conformations of organic cages are relatively easy to simulate, as they represent the low energy conformations that are reliably found by conformer searching approaches, the higher energy conformations that correspond to the 'open', solvated, molecule are more challenging. These open conformations are often many hundreds of kJ mol^{-1} above the collapsed conformations, supported by hundreds of solvent molecules within the cavity, and as such, are not found via typical conformer searching algorithms. However, the open conformations of non-shape persistent conformations can be relevant, both to allow comparison of different reaction outcomes and because several applications of porous molecules, such as sensing, take place in the solution phase. We have previously developed a constrained molecular dynamics approach to discover the open conformations of non-shape persistent molecules, without any prior knowledge of the molecular conformation or void size.[38] This works by effectively reproducing the scaffolding effect of internal solvent molecules and sampling the low energy *open* conformations of the molecules, reproducing experimental solvate single-crystal X-ray diffraction structures.

Although predicting the molecular conformation of porous organic cages *a priori* is valuable, there is an additional significant challenge in predicting the outcome of synthesis that comes from the fact that there are typically a whole family of topologies that can form from a given precursor pairing. These outcomes can be extremely sensitive to precursor choice, such as the addition of a single CH_2 group doubling the molecular mass of the resultant system, and dramatically modifying the system's properties.[37] However, as porous organic cages are typically synthesised using reversible dynamic covalent chemistry, there is an opportunity to predict the outcome – the reaction product should often be the thermodynamic product of the reaction. Thus, if the relative energies of the different possible outcomes are compared, the experimentally observed molecule *should* be the lowest energy assembly. This computational approach is demanding due to the fact that (a) reliable relative energies are required, typically requiring DFT calculations on many dozens of molecular conformations (each with hundreds of atoms), and (b) often the larger molecules are non-shape persistent, requiring the simulations mentioned above to discover the open conformations of the molecules that will initially form in solution. Overall, this approach has been shown to be able to predict and/or rationalise topological outcomes with some success, particularly in cases where there are significant energy differences between outcomes (typically over 15 kJ mol^{-1}) that are less likely to be influenced by factors such as solvent stabilisation that is not considered, but it is by no means foolproof.[26,39–41]

The final stage in structure prediction, if one has predicted the molecular reaction outcome and conformation, is the prediction of the crystal packing of the molecule in the solid-state. This is an important step for full *a priori* prediction of the solid-state properties of the material, as the crystal packing will

© 2020 The Authors. Published by Wiley-VCH GmbH

influence features such as the pore topology and surface area, and thus properties such as separation performance. Here, there has been considerable success in the application of crystal structure prediction (CSP) techniques, originally developed for pharmaceutical polymorph prediction. CSP works by sampling 100,000s of possible crystal packings, before identifying the low-energy conformations that are most likely to be observed, often via DFT calculations. CSP has been used in porous molecular systems to correctly predict the experimentally-observed crystal structure,[42] preference for enantiopure or racemic packing,[42] pore-size,[43] as well as identifying *a priori* the most promising molecules for targeted synthesis of properties via energy-structure-function maps.[44–46] We recently combined the whole workflow of structure prediction, from reaction outcome, molecular structure to crystal packing for the first time.[47]

With a computational model of the molecular or solid-state structure, it is then possible to perform a variety of calculations to determine the properties of the material, such as pore-topology, surface area, guest uptake and selectivity.[48] In the solid-state, these are standard simulations that have been widely applied in the field of porous framework simulations.[49] Given that the prediction of solid-state structure is laborious, it is valuable to be able to estimate some properties of the individual molecule from computationally cheaper molecular simulations.[50] We have developed open-source python software, *pywindow*, that analyses the molecular cavity size, number of windows, and window size, for porous molecular materials.[51,52] The software can be used to calculate the 'pore limiting envelope' – the range of pore sizes observed over a molecular dynamics simulation – to allow consideration of the importance of flexible breathing motions on porous organic cages to explain their experimentally observed sorption behaviour.[53] An approach based on creating 'eigencages' that encode the shape of a cage cavity has also proven useful for predicting cage sorption properties.[53]

With the above computational tools in-hand, there is the opportunity for larger scale screening of porous organic cages. The automated assembly of systems allows for 100,000s of possibilities to be tested, for example as we have conducted for identifying promising nested-cage systems.[17] The latter is an example of a brute-force screening approach, where all possibilities are enumerated and tested. However, the size of the search space for organic cage precursors is enormous, so it is not possible to test every combination. For this reason, we previously adapted our *stk* software to include an evolutionary algorithm (EA) that mimics 'survival of the fittest' in nature to efficiently explore the search space for organic molecules by swapping and modifying cage precursors before testing against a targeted property.[18] We have demonstrated the application of the EA for both evolving towards a specific pore size[18] and for discovering cages that are promising C_{60} encapsulants.[54] Finally, to accelerate the exploration of vast spaces of possible cage molecules and their properties, we have also created a database of $>66{,}000$ cages and used supervised machine learning to create an algorithm that can predict the shape-persistence of cages in a matter of seconds rather than hours or days. We provided a web-based tool that allows for easy prediction of shape persistence of a system to an experimental chemist, prior to any attempted laboratory synthesis.[55]

Experimental Toolkit

Alongside the use of computational tools, high-throughput synthetic screening using robotics and automation has seen increased usage in the discovery of both new materials and small molecules, albeit not to the same extent. For example, the synthesis of peptides,[56] oligonucleotides,[57] and oligosaccharides,[58] and a range of pharmaceutical and natural-product molecules,[59–62] has been automated, but there are relatively fewer examples of automating the synthesis of organic materials and supramolecular assemblies.[46] In relation to discrete organic molecules, automation has been used to determine the optimal reaction conditions required to synthesise a target compound,[62] discover new chemical transformations using 'accelerated serendipity',[63] discover new molecules, evaluate potential catalysts for couplings,[59] and multi-step scale-up of target compounds.[64] However, the majority of these approaches have focused on using either solid-phase synthesis or flow chemistry to automate synthesis,[60,65,66] which can limit the number of molecules that can be screened at once with reactions typically run in sequence.

In relation to supramolecular organic cages, alongside conventional batch synthesis,[67] a range of enabling and automated technologies has been reported for their formation. This includes the use of flow reactors,[68] microwaves,[69] mechanochemistry including ball-milling[70] and twin-screw extrusion,[71] and automated robotic platforms.[16,39,47,72] The latter, in particular, is useful for high-throughput screening, with the former methods suitable for scale-up.

While the use of robotic synthesis for supramolecular assembly is still rare, we have developed an experimental high-throughput workflow for the synthesis, characterisation, and property screening, of supramolecular organic cages formed using dynamic imine condensations (Figure 2). The workflow utilises Chemspeed automated platforms for liquid dispensing of precursor stock solutions into reactors, with the resulting reactions vortexed at different temperatures and aliquots taken for reaction monitoring to ensure the reactions have sufficient time to equilibrate to the products. Prior to translating the chemistry onto a synthetic platform, the conditions were optimised to avoid potential issues, for example, an excess of amine ensured full conversion to the desired products to overcome liquid handling inaccuracies, and high dilution was required to avoid polymer formation which can be problematic causing needle and line blockages. To address the difficulties in characterising these libraries, both solution and solid high-throughput analyses were carried out. First, solution based characterisation methods were utilised including Fourier-transform infrared spectroscopy (FTIR), ^{1}H NMR spectroscopy, high-resolution mass spectrometry (HRMS) and, in particular, diffusion NMR, which gives a fast and direct measure of the size of the reaction product in solution, to determine what assembly

© 2020 The Authors. Published by Wiley-VCH GmbH

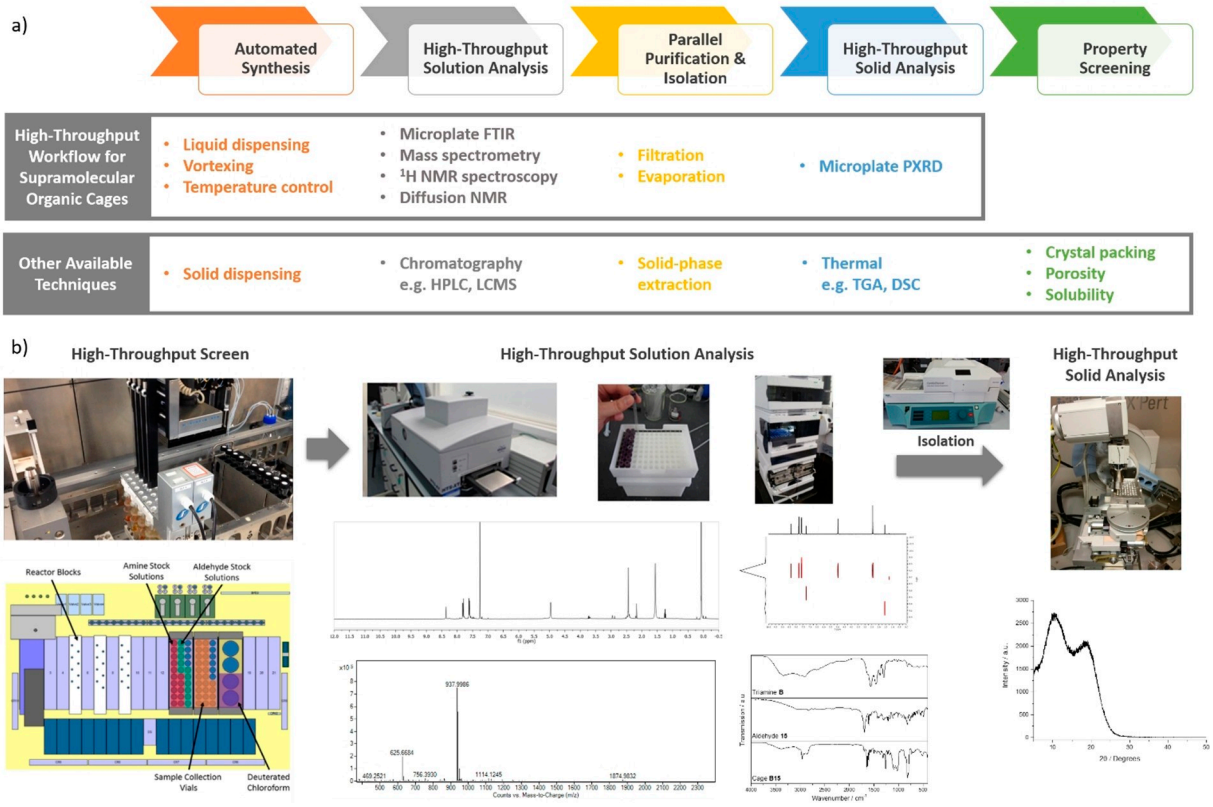


Figure 2. (a) General representation of the high-throughput experimental workflow incorporating synthesis, characterisation, and property screening; (b) Overall high-throughput experimental workflow for the discovery of supramolecular organic cages: synthetic screening carried out on a Chemspeed platform with an example deck layout (left), high-throughput solution based characterisation methods including ^{1}H NMR spectroscopy, high-resolution mass spectrometry, FTIR spectroscopy, and diffusion NMR (middle), and high-throughput solid based characterisation methods focusing on PXRD. Representative examples of the spectra for cage **B15** (see Figure 3) are shown. Reproduced (adapted) from reference [39] with permission from Springer Nature.

had formed prior to any isolation attempts. This is particularly important for organic cages, as there is the potential for the structure to collapse on solvent removal if they are not shape-persistent, which can lead to decomposition. Next, the reactions were filtered to remove any insoluble precipitate, the solvent was removed in parallel from the samples using a Combidancer evaporator, and alongside repeated solution-based analyses to determine stability to isolation, high-throughput powder X-ray diffraction (PXRD) can be used to confirm the crystallinity of the samples.

Overall, this high-throughput experimental workflow has been successfully applied to the synthesis of two-component binary organic cages of differing topologies (Figure 3),[39,72] the formation of three-component statistical distributions of vertex-disordered cage mixtures,[73] and multi-component self-sorted structures.[47] Additionally, this workflow could have other techniques incorporated at each stage, and be developed to link with property screening, such as crystal packing, porosity, and solubility (Figure 2a, bottom box). For example, Cooper and co-workers demonstrated how robotic crystallisation screens can be carried out,[46] and how the gas uptake ability of solid porous materials can be rapidly screened by monitoring the temperature change on exposing the samples to different gases after being evacuated under vacuum.[74] Alternatively, the workflow can be expanded for the discovery of new porous liquids by investigating the solubility of organic cages in cavity-excluded solvents,[16] and screening any highly soluble combinations for porosity. This greatly expanded the number of known porous liquids, and enabled the structure-property relationships that affect key parameters, such as gas uptake and viscosity, to be investigated for the first time. Additionally, porous liquids with up to three times the pore concentration were discovered that could store up to 55% more xenon than the previously reported best example.

Hybrid Workflows

The materials discovery process can be accelerated and streamlined by merging high-throughput automation and large-scale computational screening, but there are relatively few examples of putting a feedback loop between these two approaches into practice, in a combined hybrid workflow.[75] Recent examples include a combined approach in the discovery of inorganic photoelectrocatalysts,[76] metal oxides,[77] thin films,[78] organic polymer photocatalysts,[74] and porous molecular crystals.[46] Our own work has focussed on combining state-of-the-art automation with computational modelling in hybrid workflows for the synthesis of supramolecular organic cages (Figure 4).[39] Not only did this accelerate their discovery, it led to the synthetic

© 2020 The Authors. Published by Wiley-VCH GmbH

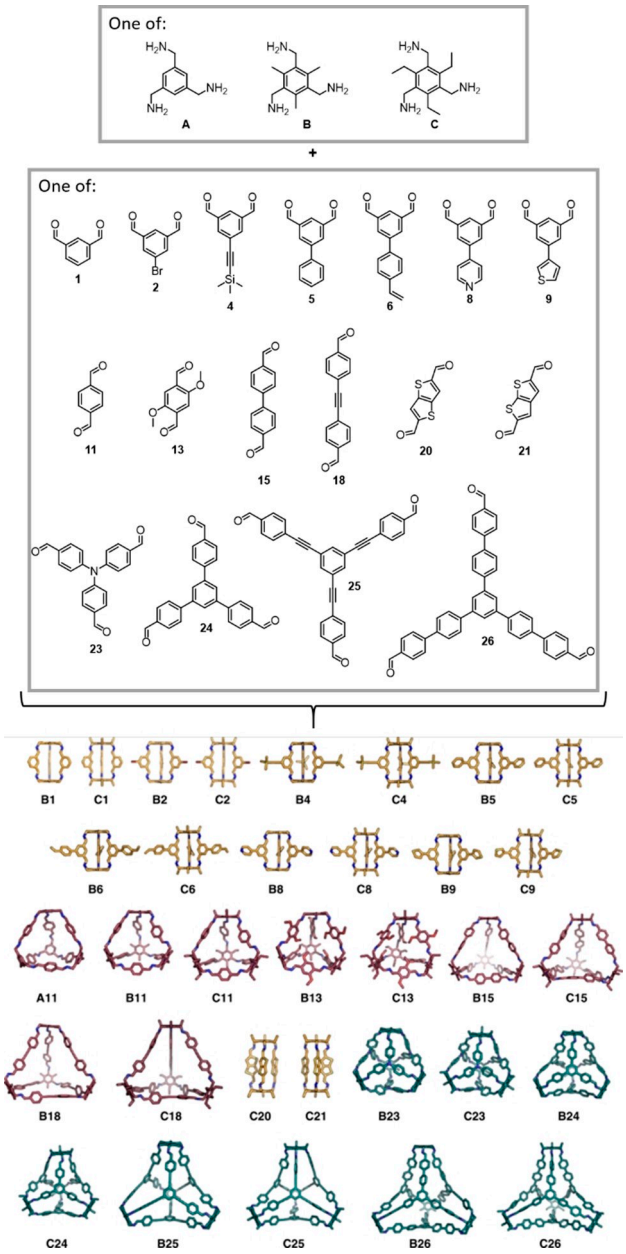


Figure 3. Examples of different organic cages formed from a binary mixture of precursors using high-throughput synthesis: capsular **Tri^{2}Di^3** cages in yellow, tetrahedral **Tri^{4}Di^6** cages shown in maroon, and **Tri^{4}Tri^4** tetrapods shown in teal.[39] The cages are named based on the combination of precursors incorporated into the structure. Reproduced (adapted) from reference [39] with permission from Springer Nature.

realisation of a computationally predicted structure, alongside the serendipitous experimental discovery of a new topology, validating the power of robotic screening methods to uncover structures that are not designed or predicted at all.

To our knowledge, there are no other studies that combine computation and robotic synthesis in this way for supramolecular organic systems. Although recently, the autonomous discovery of supramolecular co-ordination architectures has been reported by Cronin and co-workers.[79]

There are multiple ways in which computation can aid and inform experiment in a hybrid workflow: *a priori* prediction prior to experimental realisation enables materials with a specific pre-determined function to be targeted, and *a posteriori* rationalisation of an experimentally realised structure can enable a greater understanding of the material, elucidate structure-activity relationships, and aid in the discovery of potential applications. Additionally, a combination of both prediction and rationalisation can be used. Here, we present a representative example of each using examples from our own work (Figure 4).

A Priori Prediction and Experimental Realisation

In our largest combined computational and experimental high-throughput study to date, 78 building block combinations to form organic cages were attempted using robotic automation and high-throughput characterisation techniques.[39] These building block combinations had been selected by an experimental chemist as having promise for cage formation. Calculations predicted the expected topology for each of three families of precursor pairings, grouped by the topicity combinations of the precursors, suggesting a strong topology preference in two families, and a weak preference in the third. These topology preferences were experimentally confirmed, with the third family presenting a mixture of two topologies. Based on the predicted topologies, the DFT formation energies of all 78 of the combinations were then calculated. These formation energies were then compared to the experimental results, where it was found that 33 cages were cleanly formed in the one-pot synthesis, a 42% hit rate. For a single precursor combination, the calculated formation energies could not reliably predict whether a cage would form, and indeed the lack of an experimental hit is not necessarily definitive that the combination would not form under the right conditions. However, the formation energies were very useful in identifying which series, *e.g.* all cages involving a particular precursor, consistently gave less favourable formation energies. In a future integrated study, this would allow us to rule out costly and time consuming syntheses with less promising precursors, and to increase our hit rate. Finally, computational characterisation of the known experimental hits can then be used to identify the most promising candidates for experimental scale up and full characterisation. This study therefore revealed the most promising routes for integrating experimental and computational studies, so as to exert the power of each technique to give the largest benefit.

A priori prediction has also successfully been applied for an entire process from molecular precursors of a multi-component system through to crystal structure prediction (Figure 4c), which was then experimentally realised.[47] First, a multi-component system using dynamic combinatorial chemistry, where there are many potential outcomes from mixing three different components together, was predicted and then synthesised. A single socially self-sorted assembly was formed, rather than self-sorting into separate two-component cages or scrambling to form a statistical mixture (Figure 1). Unusually, this social self-

© 2020 The Authors. Published by Wiley-VCH GmbH

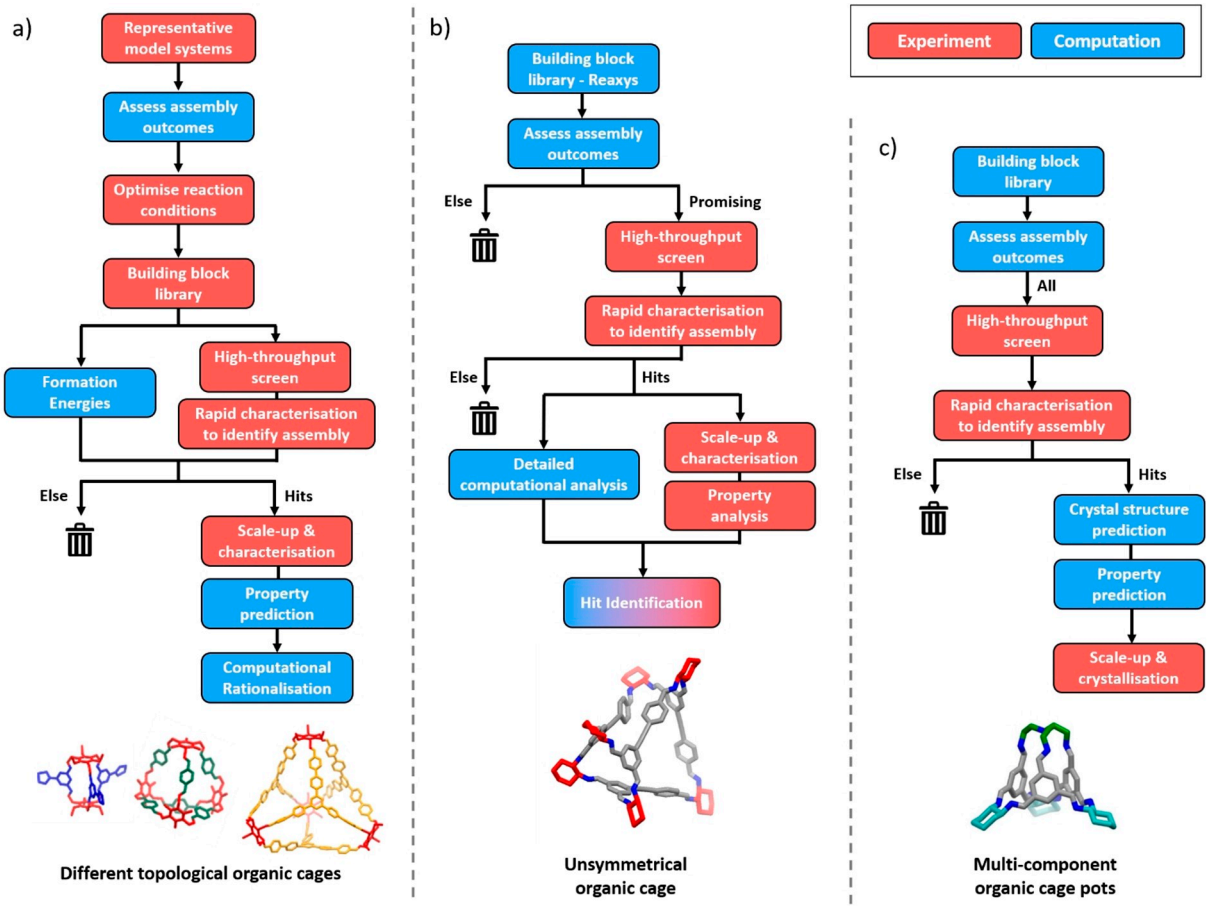


Figure 4. Hybrid workflows fusing high-throughput automated synthesis with computation for the: (a) accelerated design and discovery of different topological organic cages; (b) targeted synthesis of a computationally inspired and rationalised unsymmetrical organic cage; (c) realisation of a computationally predicted multi-component self-sorted structure from precursors through to crystallisation. Experimental stages are shown in red and computational stages in blue.

sorting did not exploit the use of orthogonal reactivity, nor did it use an additional component of the same topicity, both of which are common strategies to access these types of materials: it is purely the result of energetic preferences, as predicted in our discovery workflow. This allowed us to form a "cage pot", a new topology within the existing family of porous organic imine cages. After realising our molecular prediction of the organic cage pots by synthesis (as confirmed by mass spectrometry and NMR), we then conducted crystal structure prediction to determine the low-energy packings to be expected if the system was crystallised. We found that no single mode of crystal packing was dominant, and further identified that chiral recognition is preferred in these systems. We could also identify preferred crystal packing motifs, such as the fact that window-to-window packings were more preferred in racemic structures than homochiral ones. These preferences were then realised in experimental single-crystal X-ray diffraction structures. Overall, we believe this to be the first study to have accomplished this complete prediction and realisation workflow, from starting materials to the crystal, and that this strategy will be broadly applicable within the field of supramolecular chemistry and we expect to see increasing usage in other groups over time, particularly since the computational timescales for the predictions are now becoming much more tractable.

Experimental Realisation and A Posteriori Rationalisation

During our initial hybrid workflow comparing the predicted formation energies with the experimental synthesis of binary organic cages,[39] recrystallisation studies of the successfully formed organic cages were carried out. During these studies, a unique new cage topology – a doubly-bridged, triply-interlocked cage catenane – was discovered, formed by the re-equilibration of the originally formed **Tri^{4}Di^6** cage (Figure 5). *A posteriori* computational modelling was used to rationalise the formation of these unique structures, with the bridged catenane being found to be more stable than the parent **Tri^{4}Di^6** cage which was initially formed in solution.

© 2020 The Authors. Published by Wiley-VCH GmbH

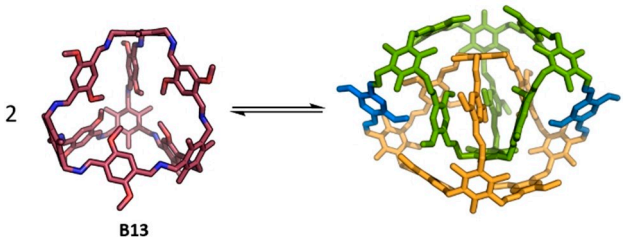


Figure 5. Re-equilibration of a **Tri^{4}Di^6** organic cage (**B13**), discovered during high-throughput screening, to an unexpected [8 + 12] covalently bridged, triply interlocked cage catenane.[39] Reproduced (adapted) from reference [39] with permission from Springer Nature.

A Priori Prediction and A Posteriori Rationalisation

Building on our initial report that provided proof-of-concept that fusing computational screening with robotic synthesis accelerated the discovery of supramolecular organic cages,[39] and in an attempt to reduce bias during precursor selection, we carried out *a priori* prediction by using the supramolecular toolkit (*stk*) to automate computational screening for promising organic cage precursors from a building block library of ~10,000 possibilities from the Reaxys database. The assembled combinations were screened for shape-persistence, before combinations that we would not have designed intuitively were targeted in a high-throughput experimental screen. This led to the discovery of the first example of a porous organic cage that lacked any symmetry elements (Figure 4b).[72] However, experimental characterisation alone could not identify the formed cage isomer. Therefore, we returned to computation to narrow down the 162 possible structural isomers of this cage to the most plausible candidate. Using a combination of DFT calculations to determine the relative stabilities of the isomers, alongside characterisation of the cage void and window dimensions, and NMR shifts for the most likely hypothetical isomers, alongside the detailed experimental analyses, the most likely isomer was proposed. Finally, an amorphous solid-state model was generated to rationalise the experimentally observed surface area which can be attributed to inefficient packing between the organic cages, an insight that is not possible by experiment alone. This demonstrates how computation can be used to guide experimental efforts but can also be used to aid in characterisation and rationalise observed behaviour.

Summary and Outlook

Overall, advances in computational modelling allow the prediction of the most likely cage topology that will be formed experimentally,[26,80] and indicates whether the cage will be shape-persistent prior to synthesis.[37,38] However, it is still difficult to predict the synthetic outcome prior to carrying out the reaction. For example, it is not yet possible to predict unambiguously the reaction outcome with the potential for the formation of oligomeric species, different cage species, and/or polymer, or indicate what conditions are required. The use of an increased number of components makes this problem even more complex as it can lead to a larger range of different outcomes, including social self-sorting into a single mixed assembly, 'scrambling' to form statistical mixtures of assemblies, or self-sorting into separate discrete species.[27] In addition, if computational prediction is relied on solely, it will miss the serendipitous discovery of new materials.

The use of high-throughput automation can rapidly accelerate the screening and discovery process, and whilst this capability can allow a large chemical space to be screened, this approach is inefficient and is still not widely implemented. Arguably a number of factors influence the uptake of robotic automation in supramolecular and organic materials discovery, including the high-cost of the majority of state-of-the-art synthetic platforms, the need for specific synthesis conditions for each different materials class, the difficulty in characterising and analysing the large amount of resulting data from a screen, the bottleneck of product isolation and purification, and the challenge of developing equally high-throughput methods for property screening. Therefore, the advancement of low-cost automation and automated data analysis would be transformative in this area.

Computational modelling can be used to narrow the potential search space and direct experimental efforts. By combining computation with experiment, the possibility of computational suggestions not being synthetically viable is reduced, although there is no guarantee that the material can be synthetically realised even if it is theoretically viable if the appropriate conditions aren't used. High-throughput automation can further streamline the screening process by allowing optimisation and substrate screens to be carried out rapidly, and allows for the generation of large databases of data that could be fed back into computational modelling to close the design loop.

In summary, the combination of both high-throughput computational modelling and automated robotic screening can greatly accelerate the discovery process, as exemplified by these recent advances in applying this to the area of supramolecular synthesis and dynamic covalent chemistry. In future work, it would be interesting to quantitatively determine to what extent a combination of automation and computation in a combined hybrid workflow is more powerful than using either one separately, for example, by attempting different approaches for the same problem in a blind parallel investigation. Such an approach has not yet been applied to supramolecular organic cages, but for a specific target application, a range of different strategies could be attempted and compared. This could include (i) computational modelling followed by conventional synthesis of the most promising targets, (ii) computational modelling followed by automated synthesis of a series of cages selected by either a human or an algorithm, and (iii) automated synthesis based on human intuition and precursor selection. A similar strategy has been used to investigate the formation and crystallisation of polyoxometalate clusters.[81] By comparing different exploration methods involving either a combination of: (i) automation with

© 2020 The Authors. Published by Wiley-VCH GmbH

a machine learning algorithm, (ii) automation with human experimental intuition, or (iii) automation with random experiment selection, it was found that the algorithm was able to search more crystallisation space and improve the crystallisation prediction accuracy than both a random search and human experimenters, highlighting the benefits of combining the two approaches.

Acknowledgements

R.L.G. and K.E.J. thank the Royal Society for University Research Fellowships. K.E.J. acknowledges the ERC for funding through grant agreement number 758370 (ERC-StG-PE5-CoMMaD).

Conflict of Interest

The authors declare no conflict of interest.

Keywords: computational modelling · high-throughput automation · organic cages · self-sorting · supramolecular assembly

[1] Z. Liu, S. K. M. Nalluri, J. F. Stoddart, *Chem. Soc. Rev.* **2017**, *46*, 2459–2478.
[2] G. Zhang, M. Mastalerz, *Chem. Soc. Rev.* **2014**, *43*, 1934–1947.
[3] T. Hasell, A. I. Cooper, *Nat. Rev. Mater.* **2016**, *1*, 16053.
[4] F. Beuerle, B. Gole, *Angew. Chem. Int. Ed.* **2018**, *57*, 4850–4878; *Angew. Chem.* **2018**, *130*, 4942–4972.
[5] S. D. P. Fielden, D. A. Leigh, S. L. Woltering, *Angew. Chem. Int. Ed.* **2017**, *56*, 11166–11194; *Angew. Chem.* **2017**, *129*, 11318–11347.
[6] G. Gil-Ramirez, D. A. Leigh, A. J. Stephens, *Angew. Chem. Int. Ed.* **2015**, *54*, 6110–6150; *Angew. Chem.* **2015**, *127*, 6208–6249.
[7] D. Fujita, Y. Ueda, S. Sato, N. Mizuno, T. Kumasaka, M. Fujita, *Nature* **2016**, *540*, 563–566.
[8] J. J. Danon, A. Krüger, D. A. Leigh, J.-F. Lemonnier, A. J. Stephens, I. J. Vitorica-Yrezabal, S. L. Woltering, *Science* **2017**, *355*, 159–162.
[9] G. Zhang, O. Presly, F. White, I. M. Oppel, M. Mastalerz, *Angew. Chem. Int. Ed.* **2014**, *53*, 5126–5130; *Angew. Chem.* **2014**, *126*, 5226–5230.
[10] J. D. Evans, K. E. Jelfs, G. M. Day, C. J. Doonan, *Chem. Soc. Rev.* **2017**, *46*, 3286–3301.
[11] G. Fraux, F. X. Coudert, *Chem. Commun.* **2017**, *53*, 7211–7221.
[12] S. Curtarolo, G. L. W. Hart, M. B. Nardelli, N. Mingo, S. Sanvito, O. Levy, *Nat. Mater.* **2013**, *12*, 191–201.
[13] R. Potyrailo, K. Rajan, K. Stoewe, I. Takeuchi, B. Chisholm, H. Lam, *ACS Comb. Sci.* **2011**, *13*, 579–633.
[14] B. Roth, S. Savagatrup, N. V. de los Santos, O. Hagemann, J. E. Carlé, M. Helgesen, F. Livi, E. Bundgaard, R. R. Søndergaard, F. C. Krebs, D. J. Lipomi, *Chem. Mater.* **2016**, *28*, 2363–2373.
[15] F. Blanc, S. Y. Chong, T. O. McDonald, D. J. Adams, S. Pawsey, M. A. Caporini, A. I. Cooper, *J. Am. Chem. Soc.* **2013**, *135*, 15290–15293.
[16] R. J. Kearsey, B. M. Alston, M. E. Briggs, R. L. Greenaway, A. I. Cooper, *Chem. Sci.* **2019**, *10*, 9454–9465.
[17] E. Berardo, R. L. Greenaway, M. Miklitz, A. I. Cooper, K. E. Jelfs, *Mol. Syst. Des. Eng.* **2020**, *5*, 186–196.
[18] E. Berardo, L. Turcani, M. Miklitz, K. E. Jelfs, *Chem. Sci.* **2018**, *9*, 8513–8527.
[19] S. Er, C. Suh, M. P. Marshak, A. Aspuru-Guzik, *Chem. Sci.* **2015**, *6*, 885–893.
[20] R. Banerjee, A. Phan, B. Wang, C. Knobler, H. Furukawa, M. O'Keeffe, O. M. Yaghi, *Science* **2008**, *319*, 939–943.
[21] M. L. Kelty, W. Morris, A. T. Gallagher, J. S. Anderson, K. A. Brown, C. A. Mirkin, T. D. Harris, *Chem. Commun.* **2016**, *52*, 7854–7857.
[22] S. Li, Y. G. Chung, R. Q. Snurr, *Langmuir* **2016**, *32*, 10368–10376.
[23] P. Wollmann, M. Leistner, U. Stoeck, R. Grünker, K. Gedrich, N. Klein, O. Throl, W. Grählert, I. Senkovska, F. Dreisbach, S. Kaskel, *Chem. Commun.* **2011**, *47*, 5151.
[24] E. Braun, A. F. Zurhelle, W. Thijssen, S. K. Schnell, L.-C. Lin, J. Kim, J. A. Thompson, B. Smit, *Mol. Syst. Des. Eng.* **2016**, *1*, 175–188.
[25] T. Tozawa, J. T. A. Jones, S. I. Swamy, S. Jiang, D. J. Adams, S. Shakespeare, R. Clowes, D. Bradshaw, T. Hasell, S. Y. Chong, C. Tang, S. Thompson, J. Parker, A. Trewin, J. Bacsa, A. M. Z. Slawin, A. Steiner, A. I. Cooper, *Nat. Mater.* **2009**, *8*, 973–978.
[26] V. Santolini, M. Miklitz, E. Berardo, K. E. Jelfs, *Nanoscale* **2017**, *9*, 5280–5298.
[27] K. Ono, N. Iwasawa, *Chem. Eur. J.* **2018**, *24*, 17856–17868.
[28] T. Hasell, J. L. Culshaw, S. Y. Chong, M. Schmidtmann, M. A. Little, K. E. Jelfs, E. O. Pyzer-Knapp, H. Shepherd, D. J. Adams, G. M. Day, A. I. Cooper, *J. Am. Chem. Soc.* **2014**, *136*, 1438–1448.
[29] A. Kewley, A. Stephenson, L. Chen, M. E. Briggs, T. Hasell, A. I. Cooper, *Chem. Mater.* **2015**, *27*, 3207–3210.
[30] L. Chen, P. S. Reiss, S. Y. Chong, D. Holden, K. E. Jelfs, T. Hasell, M. A. Little, A. Kewley, M. E. Briggs, A. Stephenson, K. M. Thomas, J. A. Armstrong, J. Bell, J. Busto, R. Noel, J. Liu, D. M. Strachan, P. K. Thallapally, A. I. Cooper, *Nat. Mater.* **2014**, *13*, 954–960.
[31] Q. Song, S. Jiang, T. Hasell, M. Liu, S. Sun, A. K. Cheetham, E. Sivaniah, A. I. Cooper, *Adv. Mater.* **2016**, *28*, 2629–2637.
[32] K. Acharyya, P. S. Mukherjee, *Chem. Commun.* **2014**, *50*, 15788–15791.
[33] N. Giri, M. G. Del Popolo, G. Melaugh, R. L. Greenaway, K. Ratzke, T. Koschine, L. Pison, M. F. Gomes, A. I. Cooper, S. L. James, *Nature* **2015**, *527*, 216–220.
[34] B. Teng, M. A. Little, T. Hasell, S. Y. Chong, K. E. Jelfs, R. Clowes, M. E. Briggs, A. I. Cooper, *Cryst. Growth Des.* **2019**, *19*, 3647–3651.
[35] L. Turcani, E. Berardo, K. E. Jelfs, *J. Comput. Chem.* **2018**, *39*, 1931–1942.
[36] "stk," can be found under https://github.com/JelfsMaterialsGroup/stk.
[37] K. E. Jelfs, X. Wu, M. Schmidtmann, J. T. A. Jones, J. E. Warren, D. J. Adams, A. I. Cooper, *Angew. Chem. Int. Ed.* **2011**, *50*, 10653–10656; *Angew. Chem.* **2011**, *123*, 10841–10844.
[38] V. Santolini, G. A. Tribello, K. E. Jelfs, *Chem. Commun.* **2015**, *51*, 15542–15545.
[39] R. L. Greenaway, V. Santolini, M. J. Bennison, B. M. Alston, C. Stackhouse, M. A. Little, M. Miklitz, E. G. B. Eden, R. Clowes, A. Shakil, H. J. Cuthbertson, H. Armstrong, M. E. Briggs, K. E. Jelfs, A. I. Cooper, *Nat. Commun.* **2018**, *9*, 2849.
[40] R. L. Greenaway, V. Santolini, F. T. Szczypiński, M. J. Bennison, M. A. Little, A. Marsh, K. E. Jelfs, A. I. Cooper, *Chem. Eur. J.* **2020**, *26*, 3718–3722.
[41] A. G. Slater, M. A. Little, A. Pulido, S. Y. Chong, D. Holden, L. Chen, C. Morgan, X. Wu, G. Cheng, R. Clowes, M. E. Briggs, T. Hasell, K. E. Jelfs, G. M. Day, A. I. Cooper, *Nat. Chem.* **2017**, *9*, 17–25.
[42] J. T. A. Jones, T. Hasell, X. Wu, J. Bacsa, K. E. Jelfs, M. Schmidtmann, S. Y. Chong, D. J. Adams, A. Trewin, F. Schiffman, F. Cora, B. Slater, A. Steiner, G. M. Day, A. I. Cooper, *Nature* **2011**, *474*, 367–371.
[43] A. G. Slater, P. S. Reiss, A. Pulido, M. A. Little, D. L. Holden, L. Chen, S. Y. Chong, B. M. Alston, R. Clowes, M. Haranczyk, M. E. Briggs, T. Hasell, G. M. Day, A. I. Cooper, *ACS Cent. Sci.* **2017**, *3*, 734–742.
[44] A. Pulido, L. Chen, T. Kaczorowski, D. Holden, M. A. Little, S. Y. Chong, B. J. Slater, D. P. McMahon, B. Bonillo, C. J. Stackhouse, A. Stephenson, C. M. Kane, R. Clowes, T. Hasell, A. I. Cooper, G. M. Day, *Nature* **2017**, *543*, 657–664.
[45] M. Bernabei, R. Pérez-Soto, I. Gómez García, M. Haranczyk, *Mol. Syst. Des. Eng.* **2018**, *3*, 942–950.
[46] P. Cui, D. P. McMahon, P. R. Spackman, B. M. Alston, M. A. Little, G. M. Day, A. I. Cooper, *Chem. Sci.* **2019**, *10*, 9988–9997.
[47] R. L. Greenaway, V. Santolini, A. Pulido, M. A. Little, B. M. Alston, M. E. Briggs, G. M. Day, A. I. Cooper, K. E. Jelfs, *Angew. Chem. Int. Ed.* **2019**, *58*, 16275–16281.
[48] E. Jackson, M. Miklitz, Q. Song, G. A. Tribello, K. E. Jelfs, *J. Phys. Chem. C* **2019**, *123*, 21011–21021.
[49] J. D. Evans, K. E. Jelfs, G. M. Day, C. J. Doonan, *Chem. Soc. Rev.* **2017**, *46*, 3286–3301.
[50] M. Miklitz, S. Jiang, R. Clowes, M. E. Briggs, A. I. Cooper, K. E. Jelfs, *J. Phys. Chem. C* **2017**, *121*, 15211–15222.
[51] M. Miklitz, K. E. Jelfs, *J. Chem. Inf. Model.* **2018**, *58*, 2387–2391.
[52] "pywindow," can be found under https://github.com/JelfsMaterialsGroup/pywindow.
[53] D. Holden, K. E. Jelfs, A. Trewin, D. J. Willock, M. Haranczyk, A. I. Cooper, *J. Phys. Chem. C* **2014**, *118*, 12734–12743.

© 2020 The Authors. Published by Wiley-VCH GmbH

[54] M. Miklitz, L. Turcani, R. L. Greenaway, K. E. Jelfs, *Commun. Chem.* **2020**, *3*, 10.
[55] L. Turcani, R. L. Greenaway, K. E. Jelfs, *Chem. Mater.* **2019**, *31*, 714–727.
[56] R. B. Merrifield, *Science* **1965**, *150*, 178–185.
[57] M. Caruthers, *Science* **1985**, *230*, 281–285.
[58] O. J. Plante, *Science* **2001**, *291*, 1523–1527.
[59] D. W. Robbins, J. F. Hartwig, *Science* **2011**, *333*, 1423–1427.
[60] V. Sans, L. Cronin, *Chem. Soc. Rev.* **2016**, *45*, 2032–2043.
[61] J. Li, S. G. Ballmer, E. P. Gillis, S. Fujii, M. J. Schmidt, A. M. E. Palazzolo, J. W. Lehmann, G. F. Morehouse, M. D. Burke, *Science* **2015**, *347*, 1221–1226.
[62] A.-C. Bédard, A. Adamo, K. C. Aroh, M. G. Russell, A. A. Bedermann, J. Torosian, B. Yue, K. F. Jensen, T. F. Jamison, *Science* **2018**, *361*, 1220–1225.
[63] A. McNally, C. K. Prier, D. W. C. MacMillan, *Science* **2011**, *334*, 1114–1117.
[64] S. Steiner, J. Wolf, S. Glatzel, A. Andreou, J. M. Granda, G. Keenan, T. Hinkley, G. Aragon-Camarasa, P. J. Kitson, D. Angelone, L. Cronin, *Science* **2019**, *363*, eaav2211.
[65] P. H. Seeberger, *Chem. Soc. Rev.* **2008**, *37*, 19–28.
[66] C. A. Shukla, A. A. Kulkarni, *Beilstein J. Org. Chem.* **2017**, *13*, 960–987.
[67] M. E. Briggs, A. I. Cooper, *Chem. Mater.* **2017**, *29*, 149–157.
[68] M. E. Briggs, A. G. Slater, N. Lunt, S. Jiang, M. A. Little, R. L. Greenaway, T. Hasell, C. Battilocchio, S. V. Ley, A. I. Cooper, *Chem. Commun.* **2015**, *51*, 17390–17393.
[69] J. Lucero, C. Osuna, J. M. Crawford, M. A. Carreon, *CrystEngComm* **2019**, *21*, 4534–4537.
[70] B. Içli, N. Christinat, J. Tönnemann, C. Schüttler, R. Scopelliti, K. Severin, *J. Am. Chem. Soc.* **2009**, *131*, 3154–3155.
[71] B. D. Egleston, M. C. Brand, F. Greenwell, M. E. Briggs, S. L. James, A. I. Cooper, D. E. Crawford, R. L. Greenaway, *Chem. Sci.* **2020**, *11*, 6582–6589.
[72] E. Berardo, R. L. Greenaway, L. Turcani, B. M. Alston, M. J. Bennison, M. Miklitz, R. Clowes, M. E. Briggs, A. I. Cooper, K. E. Jelfs, *Nanoscale* **2018**, *10*, 22381–22388.
[73] S. Jiang, J. T. A. Jones, T. Hasell, C. E. Blythe, D. J. Adams, A. Trewin, A. I. Cooper, *Nat. Commun.* **2011**, *2*, 207.
[74] Y. Bai, L. Wilbraham, B. J. Slater, M. A. Zwijnenburg, R. S. Sprick, A. I. Cooper, *J. Am. Chem. Soc.* **2019**, *141*, 9063–9071.
[75] J. Gregoire, H. S. Stein, *Chem. Sci.* **2019**, *10*, 9640–9649.
[76] Q. Yan, J. Yu, S. K. Suram, L. Zhou, A. Shinde, P. F. Newhouse, W. Chen, G. Li, K. A. Persson, J. M. Gregoire, et al., *Proc. Mont. Acad. Sci.* **2017**, *114*, 3040–3043.
[77] C. Collins, M. S. Dyer, M. J. Pitcher, G. F. S. Whitehead, M. Zanella, P. Mandal, J. B. Claridge, G. R. Darling, M. J. Rosseinsky, *Nature* **2017**, *546*, 280–284.
[78] A. Ludwig, *npj Comput. Mater.* **2019**, *5*, 70.
[79] L. Porwol, D. J. Kowalski, A. Henson, D. Long, N. L. Bell, L. Cronin, *Angew. Chem. Int. Ed.* **2020**, *59*, 11256–11261.
[80] G. Zhu, Y. Liu, L. Flores, Z. R. Lee, C. W. Jones, D. A. Dixon, D. S. Sholl, R. P. Lively, *Chem. Mater.* **2017**, *30*, 262–272.
[81] V. Duros, J. Grizou, W. Xuan, Z. Hosni, D.-L. Long, H. N. Miras, L. Cronin, *Angew. Chem. Int. Ed.* **2017**, *56*, 10815–10820; *Angew. Chem.* **2017**, *129*, 10955–10960.

Manuscript received: June 5, 2020
Revised manuscript received: July 27, 2020
Accepted manuscript online: July 28, 2020

© 2020 The Authors. Published by Wiley-VCH GmbH