

Energy–Structure–Function Maps: Cartography for Materials Discovery

Graeme M. Day* and Andrew I. Cooper*

Some of the most successful approaches to structural design in materials chemistry have exploited strong directional bonds, whose geometric reliability lends predictability to solid-state assembly. For example, metal–organic frameworks are an important design platform in materials chemistry. By contrast, the structure of molecular crystals is defined by a balance of weaker intermolecular forces, and small changes to the molecular building blocks can lead to large changes in crystal packing. Hence, empirical rules are inherently less reliable for engineering the structures of molecular solids. Energy–structure–function (ESF) maps are a new approach for the discovery of functional organic crystals. These maps fuse crystal-structure prediction with the computation of physical properties to allow researchers to choose the most promising molecule for a given application, prior to its synthesis. ESF maps were used recently to discover a highly porous molecular crystal that has a high methane deliverable capacity and the lowest density molecular crystal reported to date ($\rho = 0.41 \text{ g cm}^{-3}$, $S_{\text{ABET}} = 3425 \text{ m}^2 \text{ g}^{-1}$). Progress in this field is reviewed, with emphasis on the future opportunities and challenges for a design strategy based on computed ESF maps.

bonding motif that dominates the lattice energy of the framework: that is, directional coordination bonds in the case of PCPs and MOFs, or covalent bonds for COFs. The dominance of a particular arrangement of interactions allows families of structurally related or “isorecticular”^[4] frameworks to be prepared because the same bonding pattern can be assumed for multiple linkers. Hence, MOFs and COFs have become enormously popular platforms for a wide variety of materials applications, enabled by the relative ease with which one can explore the available “structure–function space.” The importance of this transcends specific materials functions, such as porosity: isorecticular assembly is a powerful, generalizable tool to assemble molecular subunits in the solid state.^[5]

The situation is different for molecular crystals, which lack extended networks of coordination or covalent bonds. Despite much progress in the field of crystal engineering,^[6] it is still rather challenging to program the assembly

1. Introduction

Extended frameworks such as porous coordination polymers (PCPs),^[1] metal–organic frameworks (MOFs),^[2] and covalent organic frameworks (COFs)^[3] have changed the face of materials chemistry. The crowning achievement has been the development of modular synthetic strategies where frameworks can be built to order because the linkers assemble in an intuitively predictable fashion. This predictability relies on a specific, directional

of molecules in crystalline solids. This is because the intermolecular forces that define molecular crystals—van der Waals and electrostatic interactions—are typically weaker and span a smaller energy range. Consequently, the lattice energies of molecular crystals are more evenly composed of various weak interactions and so the structure-directing forces are a more subtle balance than for PCPs, MOFs, and COFs. Strategies to make molecules crystallize more systematically, such as introducing strong hydrogen-bond donors and acceptors, can succeed but this may also introduce synthetic complexity. Moreover, the addition of such groups to create directional molecular building blocks, or “tectons,”^[7] can profoundly change the nature of the molecule—for example, by affecting its solubility or polarity—in ways that might not be aligned with the intended function. This can also be true for extended frameworks: in some PCPs and MOFs, the metal serves a specific chemical function, for example, as a catalyst, while in others, the metal is mostly there to hold the framework together—it serves no explicit chemical role.

The introduction of directional functional groups such as carboxylic acids into organic tectons exemplifies the strategy of purposeful crystal engineering, whether this be in a hydrogen-bonded molecular organic framework or in the organic linker of a MOF. Arguably, such intuitive design strategies will be defeated by complexity for molecular crystals, and even “well-behaved” isorecticular extended frameworks might carry an unacknowledged overhead in terms of incorporating directing

Prof. G. M. Day
Computational Systems Chemistry
School of Chemistry
University of Southampton
Southampton SO17 1BJ, UK
E-mail: G.M.Day@soton.ac.uk

Prof. A. I. Cooper
Department of Chemistry and Materials Innovation Factory
Leverhulme Centre for Functional Materials Design
51 Oxford Street, Liverpool L7 3NY, UK
E-mail: aicooper@liverpool.ac.uk

 The ORCID identification number(s) for the author(s) of this article can be found under <https://doi.org/10.1002/adma.201704944>.

© 2017 The Authors. Published by WILEY-VCH Verlag GmbH & Co. KGaA, Weinheim. 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.

DOI: 10.1002/adma.201704944

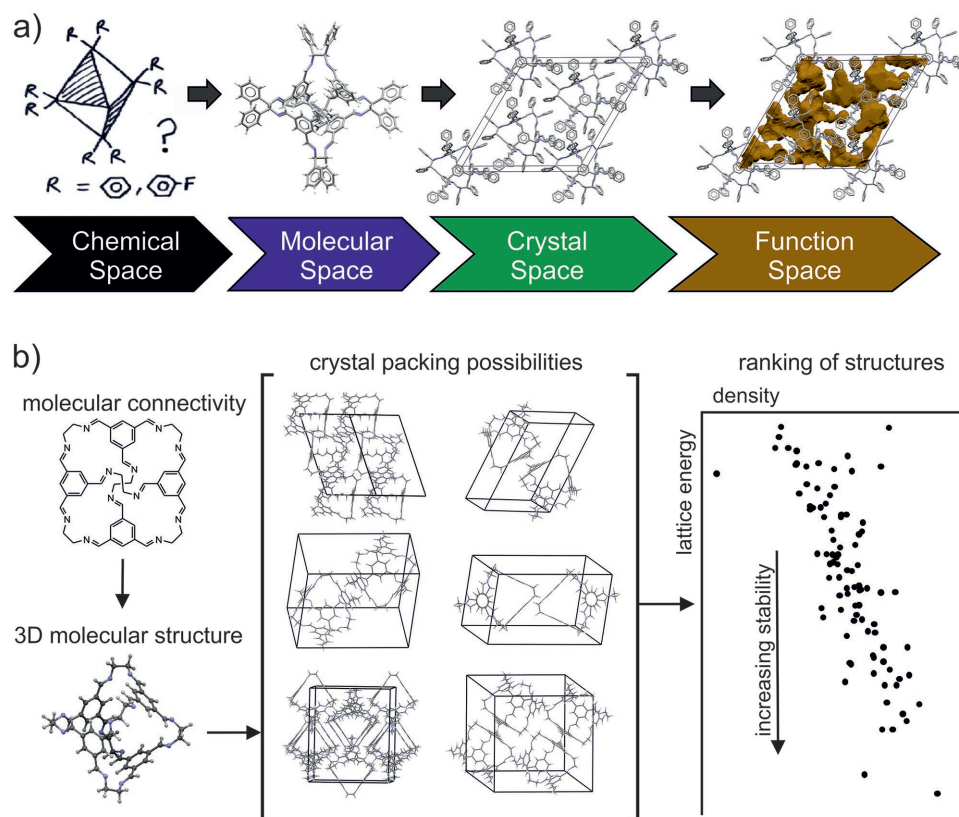


Figure 1. a) General scheme for the selection of functional materials using de novo structure and property prediction. Starting with a conceptual molecule chosen from “chemical space,” the molecular structure is predicted, followed by the crystal structures. The physical properties are then computed from the predicted crystal structures to describe the “function space.” b) Scheme for crystal-structure prediction, starting with a chemical diagram; each point on the density–lattice energy plot represents a local minimum structure on the lattice energy landscape.

functionality into materials to ensure that the components assemble correctly.

A different strategy is to *predict* molecular assembly, rather than coerce or systematize it: that is, to introduce only the basic molecular functionality that is desired for a given application (e.g., light absorption, acidity/basicity, a shape-persistent cavity, etc.) and then to identify such molecules that happen to assemble in an appropriate fashion—for example, to provide high charge carrier mobility or to render the material porous. This is a strategy of selection—or exploration—more than purposeful solid-state design, and it relates to ideas discussed by Jansen and Schön,^[8] and the concept that “humans are the explorers, and not the creators of chemical worlds.” The differences here are more than philosophical; such strategies might, ultimately, uncover nonintuitive materials with properties that are hard to access by more conventional crystal engineering.

There is an obvious problem: in lieu of a complete and predictive understanding of molecular assembly, how do we select the appropriate molecular building blocks? One experimental approach is high-throughput screening, but even with sophisticated robotics, there is usually a property measurement bottleneck that thwarts brute-force screening, especially for materials with hard-to-measure functions, such as magnetism, porosity, or charge carrier mobility, which are not amenable to automated solution-based assays. An alternative approach is the computational prediction of molecular crystal structures and

their associated physical properties (Figure 1a). De novo computational strategies are applicable to both hypothetical chemical species and known molecules. In principle, the screening can therefore be brought forward in the discovery process, in advance of molecular synthesis, which is a particular advantage for complex molecular building blocks that might require months of synthetic work to access even milligram quantities for preliminary testing.

Computational methods for crystal structure prediction (CSP) involve a global exploration of the multidimensional lattice energy surface for stable energy minima, and an assessment of the relative stabilities of the resulting structures. An exhaustive search of crystal packing possibilities using sufficiently accurate lattice energy calculations is a demanding computational exercise. The computational cost of the methods and unreliability of the off-the-shelf force fields used in early implementations of CSP^[9] have limited the widespread use of these methods. The application of CSP to organic molecules has largely focused on the anticipation of polymorphism in pharmaceutical materials,^[10] where the desired output is a complete, detailed understanding of the possible crystal structures. In this application area, the use of solid-state density functional theory (DFT) for assessing relative stabilities has gained popularity;^[11] the high cost of DFT can be justified when the computational effort is focused on the crystal packing of one target molecule. Applications of CSP to functional materials have

been less common,^[12] at least for molecular crystals, aside from a relatively small number of studies for small-molecule organic semiconductors,^[13] pigments,^[14] fluorescent molecules,^[15] and explosives.^[16] However, the development of efficient, accurate force fields, coupled with highly parallelized structure searching methods^[17] and rapid improvements in computational hardware mean that it is now possible to apply CSP in screening applications for the discovery of functional organic materials (Figure 1b) and to couple this with property predictions. Recently, this has led to materials with unprecedented properties, such as the lowest density molecular solid.^[18]

2. Crystal Structure Prediction for Functional Organic Materials

2.1. Porous Organic Cages

One important insight from applications of CSP to molecular organic crystals is the close energetic spacing of structurally disparate crystal structures of the same molecule, which results from the weak nature of intermolecular interactions. A consequence of small energy differences between different packing motifs is that small changes to the molecular structure easily shake up their energetic ordering. Thus, even closely related molecules can crystallize quite differently to produce materials with divergent physical properties.

As an example from the field of porous molecular crystals,^[19] the organic cage molecule **CC1** is nonporous because the cage cavities are disconnected in the crystal structure (Figure 2a), but a close analogue, **CC2**, has 1D pores between the cages. A second analogue, **CC3**, has a 3D interconnected diamondoid pore structure that runs through the cages themselves.^[20] As such, the solid-state properties of the three cages shown in Figure 2a cannot be anticipated intuitively from the structure of the cage molecules alone, and these structurally related cages do not form isostructural crystals or isorecticular pore networks. Application of CSP to such molecules, however, can predict the most stable crystalline forms correctly based on knowledge of only the molecular formulae (Figure 2b).^[21] CSP calculations also explain the marked preference for certain organic cages to crystallize as heterochiral crystals, rather than in homochiral forms, as manifested by the large calculated lattice energy gaps for the (**CC3-R**, **CC3-S**) racemate relative to enantiomerically pure **CC3-R** (32.1 kJ mol⁻¹ for **A** vs **B**) and the (**CC1-S**, **CC3-R**) quasiracemate versus homochiral cage cocrystal (18.8 kJ mol⁻¹ for structure **E** vs structure **F**). The heterochiral preference is mostly a result of favorable intermolecular interactions between cage windows of opposite chirality. However, this chiral pairing rule—which holds quite broadly for small [4+6] imine cages derived from triformylbenzene^[22]—breaks down for a larger analogue, **CC5**, even though that cage is produced using the same chemistry and it has the same molecular symmetry. Racemic **CC5** crystallizes as a chiral conglomerate in all solvents tested, rather than as a crystalline racemate. This illustrates again the nonintuitive nature of molecular assembly, but the effect is rationalized by CSP: there is a 32.5 kJ mol⁻¹ energy gain in forming the chiral conglomerate for **CC5**, as opposed to the racemate (Figure 2b; **C** vs **D**).^[21] These large energy gaps

are atypical: for other, less directional cage molecules, such as **CC1**, the calculated lattice energy differences between structures were smaller^[23] and in-line with the landscapes of most organic molecules that have been studied by CSP.^[24] Hence, CSP methods provide a means to distinguish between molecules such as **CC3** with deep-lying, unique global minima—that is, “robust tectons”—and molecules such as **CC1** that are more likely to exhibit polymorphism.^[25]

To reduce computational expense, these CSP calculations were performed on the neat cage structures, without considering the possible effects of enclathrated crystallization solvent on relative stabilities. However, the lattice energy differences between the candidate structures are sufficiently large to allow predictions that agree with experiment. Also, the cages were treated as rigid in these simulations. In other cases, inclusion of molecular flexibility in the calculations is required to capture the experimental crystallization behavior: this can succeed,^[23] albeit with increased computational cost. CSP approaches can also be applied to porous cages with different geometries, such as ditopic tubular imine cages^[26] with the same window geometry as **CC3**. Again, the energy landscapes revealed using CSP rationalize the preservation of the chiral window pairing in these systems, suggesting that chiral window–window interactions between cages are strong and directional enough to be compared with directional coordination bonds in PCPs and MOFs.

For these organic cages, the main property of interest—porosity—is prefabricated into the cage molecule itself. As a result, a high percentage of the hypothetical structures on the CSP energy landscapes in Figure 2b are predicted to be porous. In fact, the lowest energy predicted structures for both chiral **CC3-R** and racemic (**CC3-R**, **CC3-S**) are also the *densest* predicted structures—they minimize the free volume in the crystal—and indeed experiments show that amorphous **CC3** is significantly more porous than its most stable crystalline forms.^[22a] Hence, these energy landscapes offer an insight into possible design strategies for the future: one way to increase porosity in organic cage crystals is to make the cage molecules larger^[27] (e.g., surface areas of **CC3-R** and **CC5-R** are ≈400 and ≈1300 m² g⁻¹, respectively),^[21] while another strategy is to identify cages that pack less effectively^[28]—although this might also reduce the large energy gaps illustrated in Figure 2b, promoting polymorphism.

2.2. Energy–Structure–Function (ESF) Maps

CSP is a valuable tool for predicting the crystal structures of porous organic cages and, hence, their likely porous properties. However, the predicted energy landscapes in Figure 2b convey no direct information about the likely physical properties of the various structures. Each energy landscape can contain hundreds, sometimes even thousands, of structures within the accessible energy window and each structure—that is, each point on the landscape—is associated with a different set of physical properties. Given the size and potential complexity of these landscapes, it is challenging to assess the fitness of a given molecule for a target application based on the energy landscape alone. To address this, we have developed ESF maps,

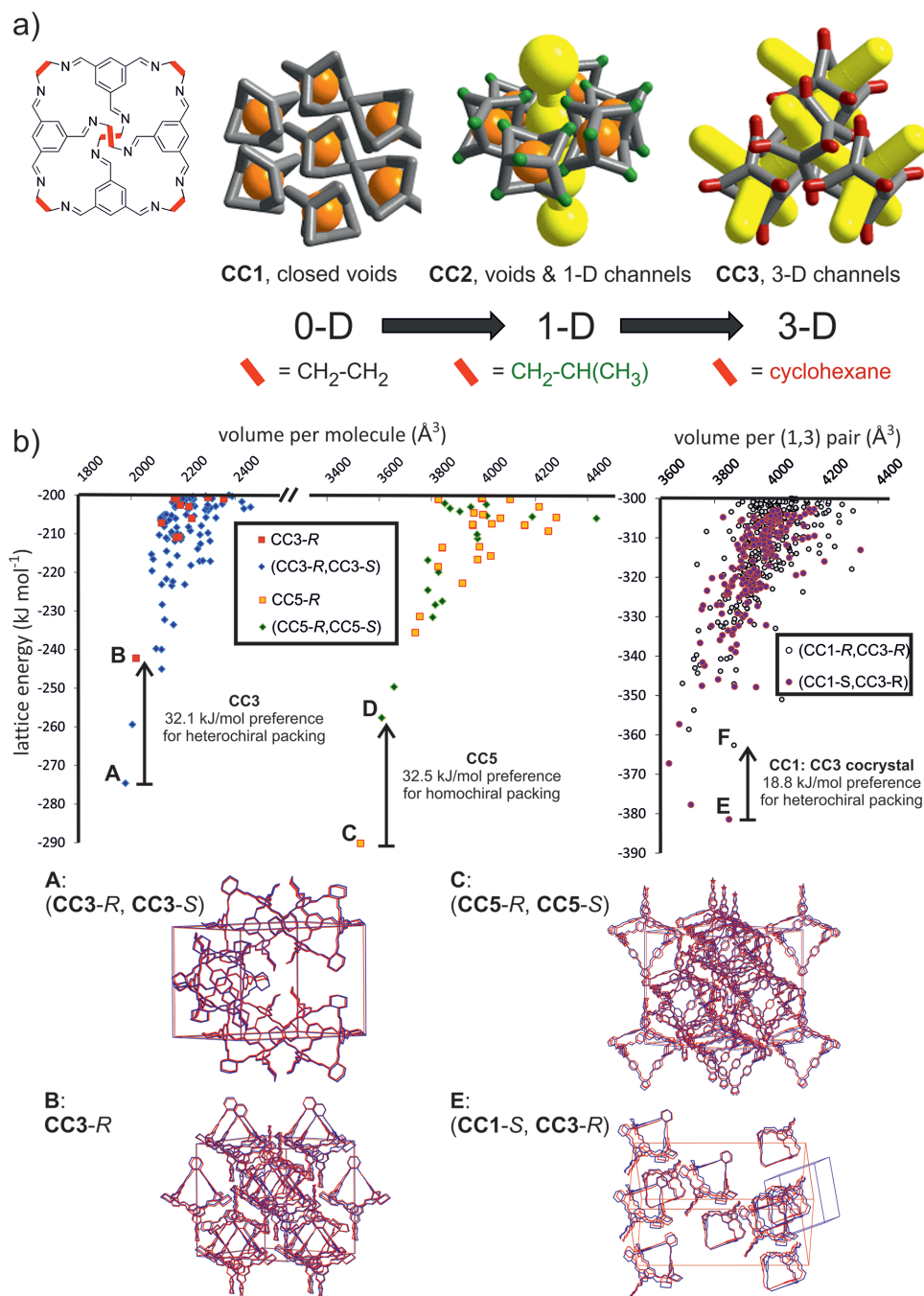


Figure 2. a) A fundamental problem in molecular crystal engineering is that even quite similar molecules can assemble differently to give crystals with divergent physical properties. Here, three structurally related [4+6] organic cages crystallize to give a nonporous material (CC1), a material with 1D pores (CC2), and a material with 3D pores (CC3).^[20] b) Crystal-structure prediction explains the crystal packing for porous cages. The experimental single-crystal structures (red) are predicted accurately (blue) using only the molecular diagram as the input.^[21] The chiral crystallization preferences of these cages are explained by the large calculated energy gaps between structures A and B, C and D, and E and F.

where calculated properties of interest are superimposed onto the lattice energy landscape, represented here via a color-coding of the points (Figure 3).^[18,29] We first exemplified this for extrinsically porous molecules, such as the hydrogen-bonded frameworks developed by Mastalerz.^[30] Lattice energy landscapes for such molecules can be surprisingly complex: for example, a trigonal benzimidazole, T2, was found to have at least

four stable porous polymorphs,^[18] all of which were located on the low-energy boundary,^[31] or “leading edge” of the calculated energy–structure landscape (Figure 3a). Three of these polymorphs (T2- β , T2- γ , and T2- δ) were previously unknown, illustrating the potential for CSP to reveal hidden detail for molecules that have already been studied. The T2- γ polymorph (Figure 3b) is particularly interesting: it has a density of just

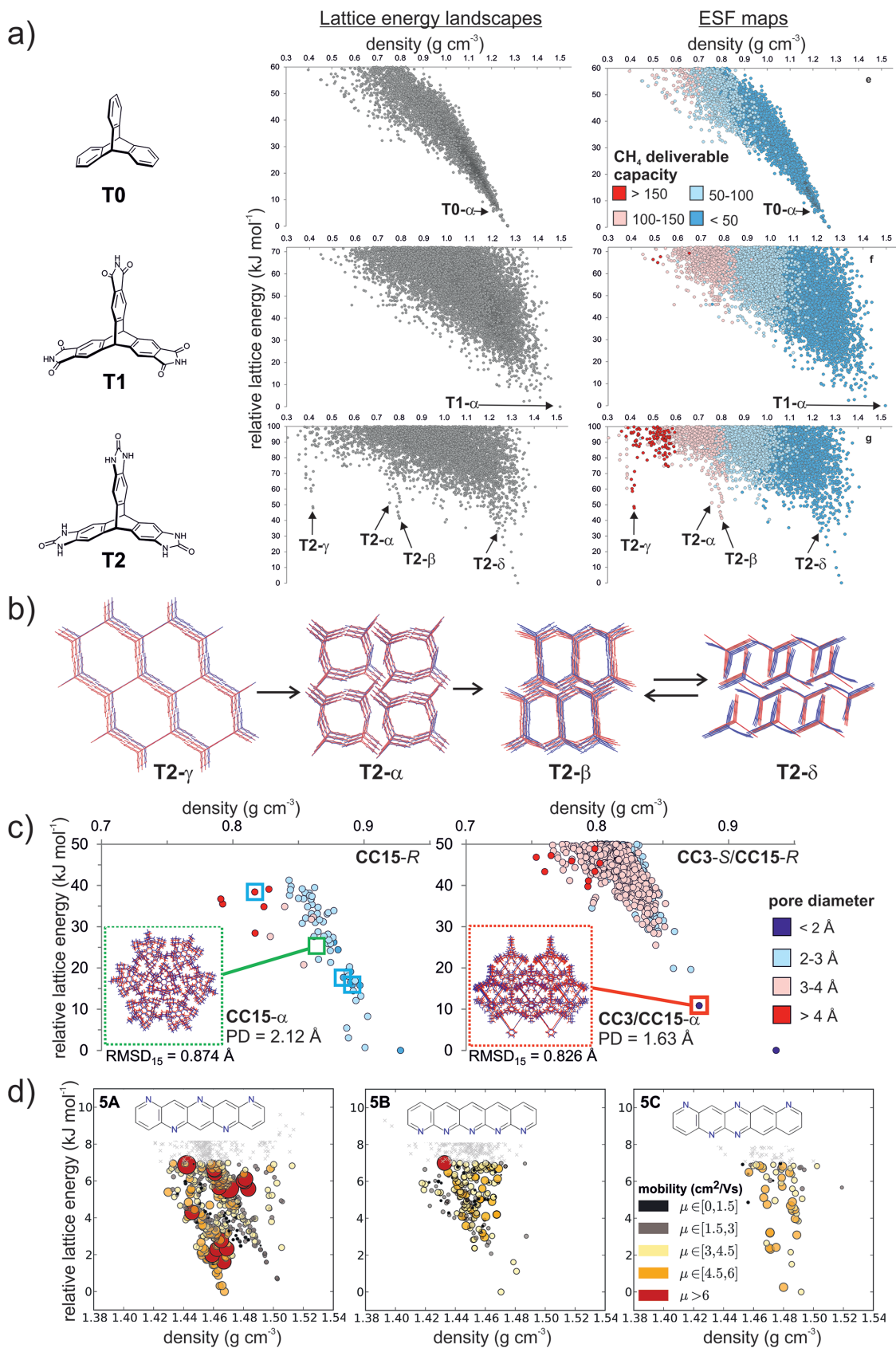


Figure 3. a) Energy–structure–function (ESF) maps allow the discovery of new porous crystals.^[18] T2 has at least four porous polymorphs (T2-α–T2-δ), three of which are found in “spikes” on the lattice energy landscape. No such stable regions are found for T0 or T1. b) Four stable polymorphs of T2

0.412 g cm⁻³ and an experimental surface area of 3425 m² g⁻¹, close to the a priori predicted values of 0.417 g cm⁻³ and 3230 m² g⁻¹. At the time of writing, **T2-γ** is the lowest density molecular crystal reported in the Cambridge Structural Database.

These ESF maps can also suggest, at a glance, which molecule in a candidate set might be the most appropriate for a given application. For example, to be of practical interest for methane storage, a porous solid must deliver at least 150 v STP/v of methane—and preferably more than 180 v STP/v—over a practicable storage and release cycle. ESF maps predict that **T2** is the only molecule in a set of eight geometrically diverse candidates that has potentially viable structures (e.g., **T2-γ**, Figure 3a) that might be interesting for methane storage. Methane sorption experiments for **T2-γ** validated these predictions: this polymorph has a saturation methane storage capacity of 437.4 v STP/v at 115 K, in agreement with simulations, and a deliverable capacity calculated from sorption isotherms at higher temperatures that match the ESF map prediction of 159 v STP/v.

It may be noted that the four observed polymorphs for **T2** all have calculated lattice energies that are substantially higher than a denser form, **T2-0**, which is the global minimum structure on the predicted energy landscape (Figure 3a). This latter polymorph has never been observed to crystallize from solution. The preferential formation of the higher-energy polymorphs **T2-α**–**T2-δ** by solution crystallization can be explained by solvent stabilization calculations,^[18] which showed that the solvated forms of these polymorphs are more stable than **T2-0**.

This CSP approach was subsequently applied to a hypothetical, extended analogue of **T2**, **T2E**, which was predicted to crystallize isostructurally to **T2-γ** but with 2.83 nm hexagonal mesopores. Based on this, **T2E** was synthesized and the predicted hexagonal form, **T2E-α**, was crystallized from a mixture of DMF and acetone. If it can be desolvated without loss of crystallinity, **T2E-α** is computed to have an unprecedentedly low density of 0.303 g cm⁻³ and a high gravimetric H₂ delivery capacity of 14.5 wt% at 100 bar and 77 K.

It is also possible to predict more complex, second-order properties for these hydrogen-bonded molecular frameworks—for example, we produced “energy–structure–selectivity maps” that can be used to identify porous materials with the highest gas selectivity.^[18] ESF maps have also been applied to porous organic cages, in this case mapping out the narrowest pore diameter to investigate the effects of adding methyl groups to the cage windows.^[25] These computational predictions are revealing: in cage **CC15-R**, the addition of methyl groups disrupts the window-to-window packing mode, and this is apparent in the computed structure-energy landscapes (Figure 3c, left). By contrast, in a cocrystal of **CC15-R**, (**CC3-S**, **CC15-R**), the window-to-window packing is preserved, as

highlighted by the prediction of a well-separated, unique global minimum structure (Figure 3c, right) in which the methyl groups narrow the pore diameter, as intended. This shows how ESF maps can be used in tandem with more intuitive design strategies: methyl groups in **CC15-R** were included purposefully to narrow the pore size in the molecule, but this had the unintended consequence of disrupting the window-to-window packing, leading to a crystal where the pore size was not, in fact, narrowed as we had planned. Both the effect on structure and the resulting effect on function are made clear by color-coded ESF maps (Figure 3c).

2.3. Beyond Porous Solids

ESF maps are a general concept that can be applied to any physical property that can be computed from crystal structures. As such, this approach can be extended beyond porous solids. For example, in a pharmaceutical materials context, the mechanical properties evaluated for low-energy predicted structures of an active pharmaceutical ingredient can predict their compressibility and tabletability.^[32] In the field of organic electronics, ESF maps were calculated for electron mobility in azapentacene molecular crystals, and a Boltzmann-like averaging function was proposed to select the most promising candidate from a set of hypothetical molecules (Figure 3d).^[13a] The study of substituted pentacenes using ESF maps also demonstrates how this approach can highlight structure–property relationships, in this case quantifying the differences in average charge carrier mobilities between herringbone, stacked, and sheet-like packings. Unlike trends drawn from experimental measurements on crystals of families of related molecules, here the structure–property relationships can be observed across a large set of crystal packings of the *same* molecule, eliminating differences due to intrinsic molecular properties and isolating the effect of crystal packing. We expect that ESF maps will play a role in uncovering new structure–property relationships as their use is extended to more properties. There is a bright future for the selection of molecular organic crystals for a range of applications, provided that the function of interest—or a descriptor that relates to this—is computable with sufficient reliability and on an affordable timescale.

3. Outlook

As computational tools and hardware develop, crystal structure prediction is set to have a disruptive impact on the field of crystal engineering, which is currently based on simple, intuitive concepts of molecular assembly. Today, the dominant design strategy for both extended frameworks and molecular

were isolated in the laboratory, with good agreement between prediction (red) and experiment (blue). The superior methane deliverable capacity for **T2-γ** (159 v STP/v), suggested by ESF maps in (a), was also confirmed experimentally. c) ESF maps for organic cages reveal that methylation of the cage disrupts window-to-window packing for **CC15-R**, leading to loss of the unique global-minimum packing mode (left), while a (**CC3-S**, **CC15-R**) cocrystal preserves the window-to-window packing mode, leading to a material with a very narrow pore diameter (PD = 1.63 Å), as per the color-coded ESF map (right).^[25] d) ESF maps of electron mobility for three azapentacene molecules (**5A**–**5C**).^[13a] While the landscape for **5A** (left) has multiple low-energy structures with high calculated electron mobility, **5C** (right) was proposed as the favored molecule based on a Boltzmann-like probabilistic analysis.

crystals is purposeful crystal engineering via structural analogues, achieved by exploiting directional covalent, coordinative, or noncovalent bonding. Indeed, these design tools have become so powerful and appealing that the design “pull” sometimes exceeds the technology “push”: MOFs, in particular, have been evaluated for an enormous range of applications; in some cases, the elegant modularity of the chemistry, more than the inherent nature of these solids, seems to have driven this. As argued above, a different strategy is the selection of materials based on computational prediction of structure and function—in essence, to purposefully design only the core, target functionality into the molecular building block, and to let self-assembly do the rest. In the long view, this is an attractive approach that could lead to simple, pared-down functional materials, shorn of any superfluous structure-directing functions. There are several challenges, though, that must be tackled first for such approaches to become more widespread.

The first question raised is usually speed, or computational expense. At least for some classes of molecules, timescales for computation of structure and properties can already compete with experimental timescales. For example, computation of ESF maps for the largest hydrogen-bonded system that we studied, **T2E**, required around one month using national high-performance computing facilities,^[33] whereas to work out a synthesis for **T2E** and to grow diffractable crystals, initially on the scale of a few milligrams, involved more than three months of laboratory effort. Computational selection strategies are therefore particularly attractive for large, synthetically complex molecular materials. The high cost involved with CSP for **T2E** was partly due to the molecular size, but also to the need to exhaustively search for crystal structures within a wide (100 kJ mol^{−1}) energy window, in order to capture all potentially porous crystal structures that could be stabilized by solvent inclusion. CSP can be completed on much shorter timescales for smaller molecules, and applications where all relevant structures will be found within the usual energy range for polymorphism (≈7–10 kJ mol^{−1}). For example, CSP for substituted pentacenes for organic electronics^[13a] were completed in under a day for each molecule using modest computational resources; as such, higher throughput computational screening can be envisioned for some applications, where many molecular variations can be assessed computationally on relatively short timescales.

Applications of ESF maps more generally face some challenges: for example, **T2E** is large but it is also rigid, and the cost of CSP for more flexible molecules increases dramatically; flexible molecules often do not adopt their lowest energy conformation in their crystal structures,^[34] meaning that the search for stable crystal structures might need to explore broad, strongly coupled inter- and intramolecular phase spaces. For some ESF maps, the computational expense will be defined more by the physical properties to be calculated. Properties such as pore volume are computationally very cheap, even for maps containing thousands of predicted structures, while other properties, such as gas selectivity and many electronic properties, are much more expensive. In this regard, porous molecular solids have both advantages and disadvantages: surface area, pore volume, and pore topology are cheap to calculate, but it is also necessary to consider a broad range of densities and, hence, a

larger number of structures (see, e.g., Figure 3a). Also, solvent stabilization, which is currently expensive to compute, plays a particularly important role for porous crystals.

Another challenge is the *de novo* prediction of composition. In all the examples given here, the composition of the material was fixed, either to contain a single molecular entity or two molecules in a binary 1:1 cocrystal. In principle, CSP can be applied to an organic cocrystal system with multiple fixed stoichiometries of the components, thus allowing the comparison of lattice energies for different compositions^[35] using a convex hull construction. However, this is currently too expensive for routine application. This problem has been tackled for relatively simple inorganic materials using variable-composition algorithms for predicting structures^[36] and more complex systems by using “extended modules” to predict new layered materials of varying compositions^[37]—but these methods are not directly transferrable to organic crystals where the molecule, with all its independent degrees of freedom is, *de facto*, the module. This might not, in fact, be such a limitation because our modern understanding of organic synthesis allows complex retrosynthetic analysis. That is, it might be simpler to synthesize a single molecule containing two different functionalities than it is to engineer the cocrystallization of two separate functional molecules, thus raising the twin questions of (i) will the molecules cocrystallize at all, and (ii) if so, in what stoichiometry? In this respect, organic synthesis might be the preferred route to controlling composition, particularly if we can access ESF maps to persuade us that these synthetic labors will yield the desired solid-state function.

A further, difficult question is where to start: that is, how do we choose the molecular building blocks prior to computation? In each of the systems presented here, the building blocks were selected with the intention of encoding the main physical property into the molecule itself. For example, shape-persistent organic cages have intrinsic molecular porosity: as a result, many packing modes lead to porous crystalline solids, even if the nature of that porosity depends strongly on the crystal packing, which is not intuitively predictable (Figures 1a and 3c). Likewise, molecules such as **T2** were chosen to have both structural rigidity and directional hydrogen bonding; that is, they were also “designed” to be porous although, in fact, the first polymorph of **T2** to be discovered^[30] (**T2-α**) is by no means the most intuitive, nor the most stable, and it was impossible to foresee without computation that there would be at least three other stable porous polymorphs for **T2**, one of which is almost half as dense (Figure 3a,b). As a third example, pentacene and azapentacenes were investigated by CSP because they are known candidates for organic electronics.^[13a] This is due to the extended π -system and low intramolecular reorganization energy—that is, the energetic cost associated with accepting an electron or hole. However, the charge carrier mobilities in these materials are strongly affected by their crystal packing (Figure 3d), and hence this property cannot be predicted from the molecules in isolation. More broadly, this molecular design strategy could be extended into many other materials: for example, by selecting molecules that contain chromophores for light absorption, or that give rise to magnetic properties. This approach uses human design input to select the building blocks. One can also envisage strategies where that selection

step, too, becomes driven by computation. For instance, the core molecular property of interest—let us say light absorption at a given wavelength—could be selected by an algorithmic or high-throughput^[38] search prior to the CSP step to produce an initial candidate set of molecular building blocks. More ambitiously, by looping the four steps shown in Figure 1a and selecting for predicted crystals that meet the target property, we have a potential basis for the autonomous design of functional molecular crystals.

Strategies for autonomous design would demand, among other things, a substantial increase in the speed at which we can calculate—or perhaps learn and extrapolate—the ESF maps. This also raises questions about how we interpret the ESF maps. In most of the examples given here, the maps were read by humans in an empirical fashion; for example, we observed low energy “spikes” (sets of low-lying structures on the lattice energy landscape) for T2 and targeted structures within these spikes (Figure 3a). In the future, and certainly for more autonomous, looped strategies, these ESF maps will need to be interpreted by computers. The recent azapentacene study (Figure 3d) gives one indication of how this might be approached in a Boltzmann-like, probabilistic fashion.^[13a] Such developments are important because many functional molecules will not give rise to large lattice energy gaps. For porous cages^[21] (Figure 2b), these gaps were adventitious, while for T2,^[18] low-energy regions on the crystal structure landscape were programmed in by the choice of a strong directional hydrogen bonding functionality—although it should be noted that this intuitive crystal engineering strategy does not succeed for the structural analogue, T1 (Figure 3a). Thinking more generally, it might be advantageous to dispense with structure directing groups or, as was the case with porous cages,^[21] to discover molecules where such energy gaps are intrinsic to the molecule. In the absence of strong structure directing groups, we will need to think in terms of more probabilistic design,^[13a] since we will be working with complex maps that may not lend themselves to simple visual interpretations.

The integration of CSP with property predictions to create ESF maps raises a wide range of opportunities for future work. The associated challenges span several research fields including computational chemistry, informatics, retrosynthetic analysis, and areas of computer science such as algorithms and machine learning. In the shorter term, we highlight three areas of materials chemistry where ESF maps might pay early dividends. First, the area of hydrogen-bonded organic frameworks, or HOFs, has grown rapidly in recent years,^[39] and there are doubtless many systems that remain to be discovered. Perhaps many of these, like T2, will be susceptible to polymorphism. We propose that ESF maps could have a valuable role to play in the selection of HOF building blocks, and that this approach might be quite general: for example, the experimentally observed porous form of a hydrogen-bonded macrocycle was shown recently to be located on the “leading edge” of the CSP landscape for that molecule.^[40] An important target here is the development of methods to calculate or to approximate solvent stabilization across energy–structure landscapes with much lower computational expense. Second, the area of porous organic cages is also burgeoning,^[41] and the use of ESF maps and other computational approaches^[42] to choose cages for

specific applications is a promising direction. Here, front-end methods to select target cages that will be synthesizable and shape-persistent are a necessary precursor.^[43] Third, materials for organic electronics gain their function from a combination of their molecular properties and their extended, solid-state properties, and this is a natural area for materials selection using ESF maps;^[13a] also, these molecules are frequently rigid in nature, which makes them good targets for CSP.

Acknowledgements

The authors thank the many colleagues that contributed to the work discussed here. The authors acknowledge the European Research Council under the European Union's Seventh Framework Programme (FP/2007-2013)/ERC through Grant Agreement Nos. 321156 (ERC-AG-PE5-ROBOT) and 307358 (ERC-stG-2012-ANGLE), EPSRC (Grant Nos. EP/N004884/1, EP/K018396/1, and EP/K018132/1), and the Leverhulme Trust (Leverhulme Research Centre for Functional Materials Design) for funding. Finally, The authors thank Prof. Susumu Kitagawa for inspiration, advice, and support, and specifically for acting as an External Advisor for our EPSRC Programme Grant (EP/H000925/1, 2009–2015).

Conflict of Interest

The authors declare no conflict of interest.

Keywords

crystal structure prediction, hydrogen-bonded organic frameworks, porous molecular solids, porous organic cages

Received: August 29, 2017

Revised: September 20, 2017

Published online:

- [1] a) M. Kondo, T. Yoshitomi, H. Matsuzaka, S. Kitagawa, K. Seki, *Angew. Chem., Int. Ed.* **1997**, 36, 1725; b) S. Kitagawa, *Angew. Chem., Int. Ed.* **2015**, 54, 10686.
- [2] a) M. Eddaoudi, J. Kim, N. Rosi, D. Vodak, J. Wachter, M. O’Keeffe, O. M. Yaghi, *Science* **2002**, 295, 469; b) A. K. Cheetham, G. Férey, T. Loiseau, *Angew. Chem., Int. Ed.* **1999**, 38, 3268; c) C. H. Hendon, A. J. Rieth, M. D. Korzynski, M. Dinca, *ACS Cent. Sci.* **2017**, 3, 554.
- [3] a) A. P. Côté, A. I. Benin, N. W. Ockwig, M. O’Keeffe, A. J. Matzger, O. M. Yaghi, *Science* **2005**, 310, 1166; b) R. P. Bisbey, W. R. Dichtel, *ACS Cent. Sci.* **2017**, 3, 533.
- [4] a) D. J. L. Tranchemontagne, Z. Ni, M. O’Keeffe, O. M. Yaghi, *Angew. Chem., Int. Ed.* **2008**, 47, 5136; b) O. M. Yaghi, M. O’Keeffe, N. W. Ockwig, H. K. Chae, M. Eddaoudi, J. Kim, *Nature* **2003**, 423, 705.
- [5] a) C. S. Diercks, O. M. Yaghi, *Science* **2017**, 355; b) L. Öhrström, *ACS Cent. Sci.* **2017**, 3, 528.
- [6] a) G. R. Desiraju, *Angew. Chem., Int. Ed.* **1995**, 34, 2311; b) B. Moulton, M. J. Zaworotko, *Chem. Rev.* **2001**, 101, 1629; c) M. Simard, D. Su, J. D. Wuest, *J. Am. Chem. Soc.* **1991**, 113, 4696.
- [7] P. Brunet, M. Simard, J. D. Wuest, *J. Am. Chem. Soc.* **1997**, 119, 2737.
- [8] M. Jansen, J. C. Schon, *Angew. Chem., Int. Ed.* **2006**, 45, 3406.
- [9] a) J. P. M. Lommerse, W. D. S. Motherwell, H. L. Ammon, J. D. Dunitz, A. Gavezzotti, D. W. M. Hofmann, F. J. J. Leusen,

- W. T. M. Mooij, S. L. Price, B. Schweizer, M. U. Schmidt, B. P. van Eijck, P. Verwer, D. E. Williams, *Acta Crystallogr., Sect. B* **2000**, 56, 697; b) G. M. Day, W. D. S. Motherwell, H. L. Ammon, S. X. M. Boerrigter, R. G. Della Valle, E. Venuti, A. Dzyabchenko, J. D. Dunitz, B. Schweizer, B. P. van Eijck, P. Erk, J. C. Facelli, V. E. Bazterra, M. B. Ferraro, D. W. M. Hofmann, F. J. J. Leusen, C. Liang, C. C. Pantelides, P. G. Karamertzanis, S. L. Price, T. C. Lewis, H. Nowell, A. Torrisi, H. A. Scheraga, Y. A. Arnautova, M. U. Schmidt, P. Verwer, *Acta Crystallogr., Sect. B* **2005**, 61, 511.
- [10] a) A. V. Kazantsev, P. G. Karamertzanis, C. S. Adjiman, C. C. Pantelides, S. L. Price, P. T. A. Galek, G. M. Day, A. J. Cruz-Cabeza, *Int. J. Pharm.* **2011**, 418, 168; b) M. A. Neumann, J. van de Streek, F. P. A. Fabbiani, P. Hidber, O. Grassmann, *Nat. Commun.* **2015**, 6, 7793; c) S. L. Price, S. M. Reutzel-Edens, *Drug Discovery Today* **2016**, 21, 912.
- [11] a) M. A. Neumann, F. J. J. Leusen, J. Kendrick, *Angew. Chem., Int. Ed.* **2008**, 47, 2427; b) A. M. Reilly, R. I. Cooper, C. S. Adjiman, S. Bhattacharya, A. D. Boese, J. G. Brandenburg, P. J. Bygrave, R. Bylsma, J. E. Campbell, R. Car, D. H. Case, R. Chadha, J. C. Cole, K. Cosburn, H. M. Cuppen, F. Curtis, G. M. Day, R. A. DiStasio Jr., A. Dzyabchenko, B. P. van Eijck, D. M. Elking, J. A. van den Ende, J. C. Facelli, M. B. Ferraro, L. Fusti-Molnar, C.-A. Gatsiou, T. S. Gee, R. de Gelder, L. M. Ghiringhelli, H. Goto, S. Grimme, R. Guo, D. W. M. Hofmann, J. Hoja, R. K. Hylton, L. Iuzzolino, W. Jankiewicz, D. T. de Jong, J. Kendrick, N. J. de Klerk, H.-Y. Ko, L. N. Kuleshova, X. Li, S. Lohani, F. J. J. Leusen, A. M. Lund, J. Lv, Y. Ma, N. Marom, A. E. Masunov, P. McCabe, D. P. McMahon, H. Meekes, M. P. Metz, A. J. Misquitta, S. Mohamed, B. Monserrat, R. J. Needs, M. A. Neumann, J. Nyman, S. Obata, H. Oberhofer, A. R. Oganov, A. M. Orendt, G. I. Pagola, C. C. Pantelides, C. J. Pickard, R. Podeszwa, L. S. Price, S. L. Price, A. Pulido, M. G. Read, K. Reuter, E. Schneider, C. Schober, G. P. Shields, P. Singh, I. J. Sugden, K. Szalewicz, C. R. Taylor, A. Tkatchenko, M. E. Tuckerman, F. Vacarro, M. Vasileiadis, A. Vazquez-Mayagoitia, L. Vogt, Y. Wang, R. E. Watson, G. A. de Wijs, J. Yang, Q. Zhu, C. R. Groom, *Acta Crystallogr., Sect. B* **2016**, 72, 439.
- [12] J. D. Evans, K. E. Jelfs, G. M. Day, C. J. Doonan, *Chem. Soc. Rev.* **2017**, 46, 3286.
- [13] a) J. E. Campbell, J. Yang, G. M. Day, *J. Mater. Chem. C* **2017**, 5, 7574; b) R. G. D. Valle, E. Venuti, A. Brillante, A. Girlando, *J. Chem. Phys.* **2003**, 118, 807; c) Y. Yang, B. Rice, X. Shi, J. R. Brandt, R. Correa da Costa, G. J. Hedley, D.-M. Smilgies, J. M. Frost, I. D. W. Samuel, A. Otero de la Roza, E. R. Johnson, K. E. Jelfs, J. Nelson, A. J. Campbell, M. J. Fuchter, *ACS Nano* **2017**, 11, 8329.
- [14] E. F. Paulus, F. J. J. Leusen, M. U. Schmidt, *CrystEngComm* **2007**, 9, 131.
- [15] M. Arhangelskis, M. D. Eddleston, D. G. Reid, G. M. Day, D.-K. Bučar, A. J. Morris, W. Jones, *Chem. Eur. J.* **2016**, 22, 10065.
- [16] R. Podeszwa, B. M. Rice, K. Szalewicz, *PCCP* **2009**, 11, 5512.
- [17] a) D. H. Case, J. E. Campbell, P. J. Bygrave, G. M. Day, *J. Chem. Theory Comput.* **2016**, 12, 910; b) M. Vasileiadis, A. V. Kazantsev, P. G. Karamertzanis, C. S. Adjiman, C. C. Pantelides, *Acta Crystallogr., Sect. B* **2012**, 68, 677.
- [18] A. Pulido, L. J. 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.
- [19] a) P. Sozzani, A. Comotti, R. Simonutti, T. Meersmann, J. W. Logan, A. Pines, *Angew. Chem., Int. Ed.* **2000**, 39, 2695; b) J. L. Atwood, L. J. Barbour, A. Jerga, B. L. Schottel, *Science* **2002**, 298, 1000; c) A. Avellaneda, P. Valente, A. Burgun, J. D. Evans, A. W. Markwell-Heys, D. Rankine, D. J. Nielsen, M. R. Hill, C. J. Sumby, C. J. Doonan, *Angew. Chem., Int. Ed.* **2013**, 52, 3746; d) G. Zhang, M. Mastalerz, *Chem. Soc. Rev.* **2014**, 43, 1934; e) M. E. Briggs, A. I. Cooper, *Chem. Mater.* **2017**, 29, 149; f) A. I. Cooper, *ACS Cent. Sci.* **2017**, 3, 544.
- [20] 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.
- [21] J. T. A. Jones, T. Hasell, X. Wu, J. Bacsa, K. E. Jelfs, M. Schmidtman, 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.
- [22] a) T. Hasell, S. Y. Chong, K. E. Jelfs, D. J. Adams, A. I. Cooper, *J. Am. Chem. Soc.* **2012**, 134, 588; b) T. Hasell, S. Y. Chong, M. Schmidtman, D. J. Adams, A. I. Cooper, *Angew. Chem., Int. Ed.* **2012**, 51, 7154; c) T. Hasell, J. L. Culshaw, S. Y. Chong, M. Schmidtman, 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; d) T. Hasell, M. A. Little, S. Y. Chong, M. Schmidtman, M. E. Briggs, V. Santolini, K. E. Jelfs, A. I. Cooper, *Nanoscale* **2017**, 9, 6783; e) S. Tothadi, M. A. Little, T. Hasell, M. E. Briggs, S. Y. Chong, M. Liu, A. I. Cooper, *CrystEngComm* **2017**, 19, 4933.
- [23] E. O. Pyzer-Knapp, H. P. G. Thompson, F. Schiffmann, K. E. Jelfs, S. Y. Chong, M. A. Little, A. I. Cooper, G. M. Day, *Chem. Sci.* **2014**, 5, 2235.
- [24] a) G. M. Day, J. Chisholm, N. Shan, W. D. S. Motherwell, W. Jones, *Cryst. Growth Des.* **2004**, 4, 1327; b) G. M. Day, *Crystallogr. Rev.* **2011**, 17, 3.
- [25] A. G. Slater, P. S. Reiss, A. Pulido, M. A. Little, D. L. Holden, L. J. 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.
- [26] 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.
- [27] G. Zhang, O. Presly, F. White, I. M. Oppel, M. Mastalerz, *Angew. Chem., Int. Ed.* **2014**, 53, 1516.
- [28] M. J. Bojdis, M. E. Briggs, J. T. A. Jones, D. J. Adams, S. Y. Chong, M. Schmidtman, A. I. Cooper, *J. Am. Chem. Soc.* **2011**, 133, 16566.
- [29] G. J. O. Beran, *Nat. Mater.* **2017**, 16, 602.
- [30] M. Mastalerz, I. M. Oppel, *Angew. Chem., Int. Ed.* **2012**, 51, 5252.
- [31] A. J. Cruz-Cabeza, G. M. Day, W. Jones, *Chem. Eur. J.* **2009**, 15, 13033.
- [32] a) G. M. Day, S. L. Price, M. Leslie, *Cryst. Growth Des.* **2001**, 1, 13; b) T. Beyer, G. M. Day, S. L. Price, *J. Am. Chem. Soc.* **2001**, 123, 5086.
- [33] *Crystal structure prediction on T2E required ≈400 000 CPU hours of computing, which was performed in about three weeks (real time) using the UK's national High Performance Computing facility, ARCHER. Property calculations were performed on smaller computers with times ranging from 1 d (pore size and dimensionality analysis) to 6 weeks (methane capacity predictions).*
- [34] H. P. G. Thompson, G. M. Day, *Chem. Sci.* **2014**, 5, 3173.
- [35] a) A. J. Cruz-Cabeza, G. M. Day, W. Jones, *Chem. Eur. J.* **2008**, 14, 8830; b) A. J. Cruz-Cabeza, S. Karki, L. Fabian, T. Friscic, G. M. Day, W. Jones, *Chem. Commun.* **2010**, 46, 2224.
- [36] a) W. W. Zhang, A. R. Oganov, A. F. Goncharov, Q. Zhu, S. E. Boulfelfel, A. O. Lyakhov, E. Stavrou, M. Somayazulu, V. B. Prakapenka, Z. Konopkova, *Science* **2013**, 342, 1502; b) W. W. Zhang, A. R. Oganov, Q. Zhu, S. S. Lobanov, E. Stavrou, A. F. Goncharov, *Sci. Rep.* **2016**, 6, 6.
- [37] a) M. S. Dyer, C. Collins, D. Hodgeman, P. A. Chater, A. Demont, S. Romani, R. Sayers, M. F. Thomas, J. B. Claridge, G. R. Darling, M. J. Rosseinsky, *Science* **2013**, 340, 847; b) 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.

- [38] a) A. N. Sokolov, S. Atahan-Evrenk, R. Mondal, H. B. Akkerman, R. S. Sánchez-Carrera, S. Granados-Focil, J. Schrier, S. C. B. Mannsfeld, A. P. Zoombelt, Z. Bao, A. Aspuru-Guzik, *Nat. Commun.* **2011**, 2, 437; b) J. Hachmann, R. Olivares-Amaya, S. Atahan-Evrenk, C. Amador-Bedolla, R. S. Sánchez-Carrera, A. Gold-Parker, L. Vogt, A. M. Brockway, A. Aspuru-Guzik, *J. Phys. Chem. Lett.* **2011**, 2, 2241; c) E. O. Pyzer-Knapp, C. Suh, R. Gómez-Bombarelli, J. Aguilera-Iparraguirre, A. Aspuru-Guzik, *Annu. Rev. Mater. Res.* **2015**, 45, 195.
- [39] a) W. B. Yang, A. Greenaway, X. A. Lin, R. Matsuda, A. J. Blake, C. Wilson, W. Lewis, P. Hubberstey, S. Kitagawa, N. R. Champness, M. Schroder, *J. Am. Chem. Soc.* **2010**, 132, 14457; b) X. Z. Luo, X. J. Jia, J. H. Deng, J. L. Zhong, H. J. Liu, K. J. Wang, D. C. Zhong, *J. Am. Chem. Soc.* **2013**, 135, 11684; c) P. Li, Y. B. He, J. Guang, L. H. Weng, J. C. G. Zhao, S. C. Xiang, B. L. Chen, *J. Am. Chem. Soc.* **2014**, 136, 547; d) P. Li, Y. B. He, H. D. Arman, R. Krishna, H. L. Wang, L. H. Weng, B. L. Chen, *Chem. Commun.* **2014**, 50, 13081; e) H. L. Wang, B. Li, H. Wu, T. L. Hu, Z. Z. Yao, W. Zhou, S. C. Xiang, B. L. Chen, *J. Am. Chem. Soc.* **2015**, 137, 9963; f) W. Yang, B. Li, H. L. Wang, O. Alduhaish, K. Alfooty, M. A. Zayed, P. Li, H. D. Arman, B. L. Chen, *Cryst. Growth Des.* **2015**, 15, 2000; g) P. Li, Y. B. He, Y. F. Zhao, L. H. Weng, H. L. Wang, R. Krishna, H. Wu, W. Zhou, M. O'Keeffe, Y. Han, B. L. Chen, *Angew. Chem., Int. Ed.* **2015**, 54, 574; h) W. Yan, X. Yu, T. Yan, D. Wu, E. Ning, Y. Qi, Y.-F. Han, Q. Li, *Chem. Commun.* **2017**, 53, 3677.
- [40] I. Hisaki, S. Nakagawa, N. Ikenaka, Y. Imamura, M. Katouda, M. Tashiro, H. Tsuchida, T. Ogoshi, H. Sato, N. Tohnai, M. Miyata, *J. Am. Chem. Soc.* **2016**, 138, 6617.
- [41] a) M. Mastalerz, *Chem. Eur. J.* **2012**, 18, 10082; b) M. Xiong, H. M. Ding, B. J. Li, T. L. Zhou, C. Wang, *Curr. Org. Chem.* **2014**, 18, 1965; c) J. D. Evans, C. J. Sumby, C. J. Doonan, *Chem. Lett.* **2015**, 44, 582; d) M. I. Hashim, C. W. Hsu, H. T. M. Le, O. S. Miljanic, *Synlett* **2016**, 27, 1907; e) T. Hasell, A. I. Cooper, *Nat. Rev. Mater.* **2016**, 1, 14.
- [42] a) K. E. Jelfs, A. I. Cooper, *Curr. Opin. Solid State Mater. Sci.* **2013**, 17, 19; b) 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; c) J. S. Camp, D. S. Sholl, *J. Phys. Chem. C* **2015**; d) J. D. Evans, D. M. Huang, M. Haranczyk, A. W. Thornton, C. J. Sumby, C. J. Doonan, *CrystEngComm* **2016**; e) M. Liu, L. Chen, S. Lewis, S. Y. Chong, M. A. Little, T. Hasell, I. M. Aldous, C. M. Brown, M. W. Smith, C. A. Morrison, L. J. Hardwick, A. I. Cooper, *Nat. Commun.* **2016**, 7, 12750; f) X. Kong, J. Jiang, *PCCP* **2017**, 19, 18178; g) M. Miklitz, S. Jiang, R. Clowes, M. E. Briggs, A. I. Cooper, K. E. Jelfs, *J. Phys. Chem. C* **2017**, 121, 15211; h) F. Zhang, Y. D. He, J. S. Huang, B. G. Sumpter, R. Qiao, *J. Phys. Chem. C* **2017**, 121, 12426.
- [43] V. Santolini, M. Miklitz, E. Berardo, K. E. Jelfs, *Nanoscale* **2017**, 9, 5280.