

# Functional materials discovery using energy–structure–function maps

Angeles Pulido<sup>1</sup>, Linjiang Chen<sup>2</sup>, Tomasz Kaczorowski<sup>2</sup>, Daniel Holden<sup>2</sup>, Marc A. Little<sup>2</sup>, Samantha Y. Chong<sup>2</sup>, Benjamin J. Slater<sup>2</sup>, David P. McMahon<sup>1</sup>, Baltasar Bonillo<sup>2</sup>, Chloe J. Stackhouse<sup>2</sup>, Andrew Stephenson<sup>2</sup>, Christopher M. Kane<sup>2</sup>, Rob Clowes<sup>2</sup>, Tom Hasell<sup>2</sup>, Andrew I. Cooper<sup>2</sup> & Graeme M. Day<sup>1</sup>

Molecular crystals cannot be designed in the same manner as macroscopic objects, because they do not assemble according to simple, intuitive rules. Their structures result from the balance of many weak interactions, rather than from the strong and predictable bonding patterns found in metal–organic frameworks and covalent organic frameworks. Hence, design strategies that assume a topology or other structural blueprint will often fail. Here we combine computational crystal structure prediction and property prediction to build energy–structure–function maps that describe the possible structures and properties that are available to a candidate molecule. Using these maps, we identify a highly porous solid, which has the lowest density reported for a molecular crystal so far. Both the structure of the crystal and its physical properties, such as methane storage capacity and guest–molecule selectivity, are predicted using the molecular structure as the only input. More generally, energy–structure–function maps could be used to guide the experimental discovery of materials with any target function that can be calculated from predicted crystal structures, such as electronic structure or mechanical properties.

Predictive calculations of material structure and properties have successfully been applied to a range of materials, including zeolites<sup>1</sup>, new allotropes of common elements<sup>2</sup>, cathode materials for batteries<sup>3</sup>, redox-active frameworks<sup>4</sup>, organic photovoltaics<sup>5</sup>, metal oxides<sup>6</sup> and porous solids<sup>7</sup>. However, it is a key challenge for computational materials research to identify new materials that are not only hypothetical. This requires us to compute both the property of interest and the stability of the material with respect to alternative atomic configurations. Computational prediction of stability and function has great potential for the discovery of materials with arresting properties<sup>8,9</sup>, but it is difficult in practice because of the computational expense of exploring vast structural landscapes, coupled with the need for accurate lattice energies and reliable property predictions. Inexpensive calculations have been used to enumerate large libraries of metal–organic frameworks (MOFs) and to predict their gas adsorption properties<sup>7,10</sup>. However, these methods are based on assumed framework topologies and they do not tell us about the relative energies of the hypothetical structures or which structures, if any, can be synthesized. To predict, select and then synthesize new functional materials, we need a simple, understandable description of the probable structure–function space, rather than the potential space, which will always be extremely large<sup>11</sup>.

This design challenge is particularly acute for molecular crystals, whose complex structural landscapes are defined by competing, weak structure-determining interactions. As a result, small changes to molecular structure can cause profound changes in crystal packing and polymorphism is commonplace in molecular crystals<sup>12</sup>. Hence, it is difficult to apply structure–function relationships learned from one system to a new molecule, which might pack in a totally different way, and so there are few molecular analogues of isoreticular MOFs<sup>13</sup> or covalent organic frameworks<sup>14</sup>, the lattice energies of which are dominated by a specific bonding pattern across a broad range of building blocks. Therefore, the *a priori* design of functional molecular crystals requires a predictive strategy that does not rely on intuitive bonding rules or assumed topologies. Crystal structure prediction (CSP) methods<sup>15,16</sup>

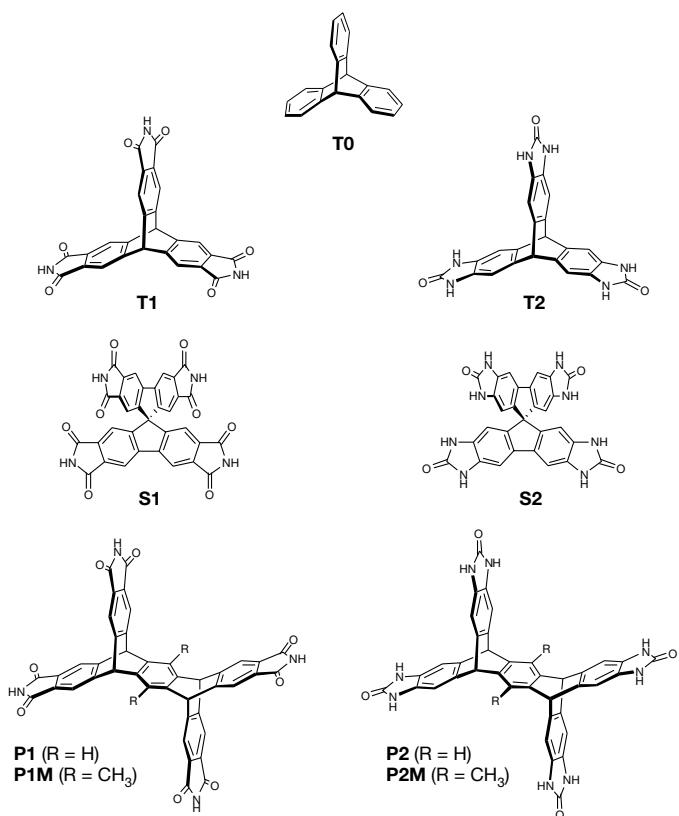
have been developed to determine the stable crystalline arrangements that are available to a molecule. Each predicted structure encodes a set of physical properties: this ensemble of structures and their properties defines an energy–structure–function (ESF) map, which represents the possible material properties associated with the molecule. The likelihood of each structure being stable and accessible to experiment relates to its predicted lattice energy. We illustrate this function mapping approach for porous organic molecular crystals<sup>17</sup>, which are rare because molecules tend to pack densely<sup>18</sup>. Previously, we predicted the crystal structures of organic cages<sup>19,20</sup>, for which most of the possible crystal packings are porous because of the intrinsically porous molecular structure. Without such built-in porosity, the lowest-energy crystal structures for molecules are, with a few exceptions<sup>21</sup>, close-packed and non-porous. Here, we use ESF maps to guide us to molecular materials with remarkable porosity levels and high predicted gas selectivities, avoiding any assumptions or intuitive guesses about the crystal packing. We show that ESF maps can reveal a rich landscape of undiscovered structures and properties for known molecules<sup>22</sup>, and can predict the properties of hypothetical molecules for specific target applications before these molecules are realized in the laboratory.

## Energy landscapes predict porous phases

Directional intermolecular interactions<sup>22–24</sup> and geometries that hinder close packing<sup>25</sup> are both known to promote porosity. We therefore studied a series of awkwardly shaped molecules with different hydrogen-bonding functionalities (Fig. 1) with the aim of locating stable, porous crystals.

Benzimidazolone **T2** was synthesized previously<sup>22</sup>, and a crystal structure with a low density ( $0.755 \text{ g cm}^{-3}$ ) and an experimental Brunauer–Emmett–Teller surface area ( $\text{SA}_{\text{BET}}$ ) of  $2,796 \text{ m}^2 \text{ g}^{-1}$  was reported. **T1** is the imide analogue of **T2**, and its only known crystal structure is a solvate that exhibits strong hydrogen bonding to the solvent<sup>26</sup>. We include triptycene, **T0**, for comparison, as a non-hydrogen-bonding analogue. The spiro-linked tetrahedral analogues, **S1** and

<sup>1</sup>Computational Systems Chemistry, School of Chemistry, University of Southampton, Southampton, UK. <sup>2</sup>Department of Chemistry, University of Liverpool, Liverpool, UK.  
§These authors jointly supervised this work.



**Figure 1 | Candidate building blocks for porous solids.** The molecules are based on triptycene (**T0**, **T1** and **T2**), spiro-biphenyl (**S1** and **S2**) or pentiptycene (**P1**, **P2**, **P1M** and **P2M**) cores. 1, imide series; 2, benzimidazolone series.

**S2**, and the expanded, tie-shaped pentiptycene analogues, **P1** and **P2**, are all unknown. It is not obvious which of these molecules should give rise to stable porous phases: except for **T0**, they all have potential to form intermolecular hydrogen bonds and, in principle, **P1** and **P2** (or their methylated analogues **P1M** and **P2M**, Fig. 1) might form more open frameworks than does **T2** because of their longer organic struts, by analogy with isoreticular framework design strategies<sup>13,14</sup>.

Unbiased searches of the lattice energy surface<sup>27</sup> were used to predict the possible crystal structures of these molecules (Fig. 2a–d for **T0**, **T1**, **T2** and **P2**, and Supplementary Figs 1 and 2 for **S1**, **S2**, **P1**, **P1M** and **P2M**). Analysis of the lower edge of the energy–density representation of the crystal structure landscape is important in the search for stable porous structures. This ‘leading edge’ comprises the structures with the lowest energy possible at any given density, and stable porous molecular frameworks have previously been identified in this region<sup>28</sup>.

The landscapes for **T0** and **T1** (Fig. 2a, b) are typical for organic molecules: the lowest-energy structures are densely packed and the leading edge of the energy–density distribution decreases nearly monotonically, owing to the energetic cost of void space in a solid. The known, non-porous crystal structure of triptycene is accurately reproduced by one of the lowest-energy calculated structures (**T0- $\alpha$** , Fig. 2a, Supplementary Fig. 3). Intermolecular hydrogen bonding in imide **T1** broadens the density distribution of the predicted structures (Fig. 2b), but again the lattice energy decreases monotonically and most structures are non-porous.

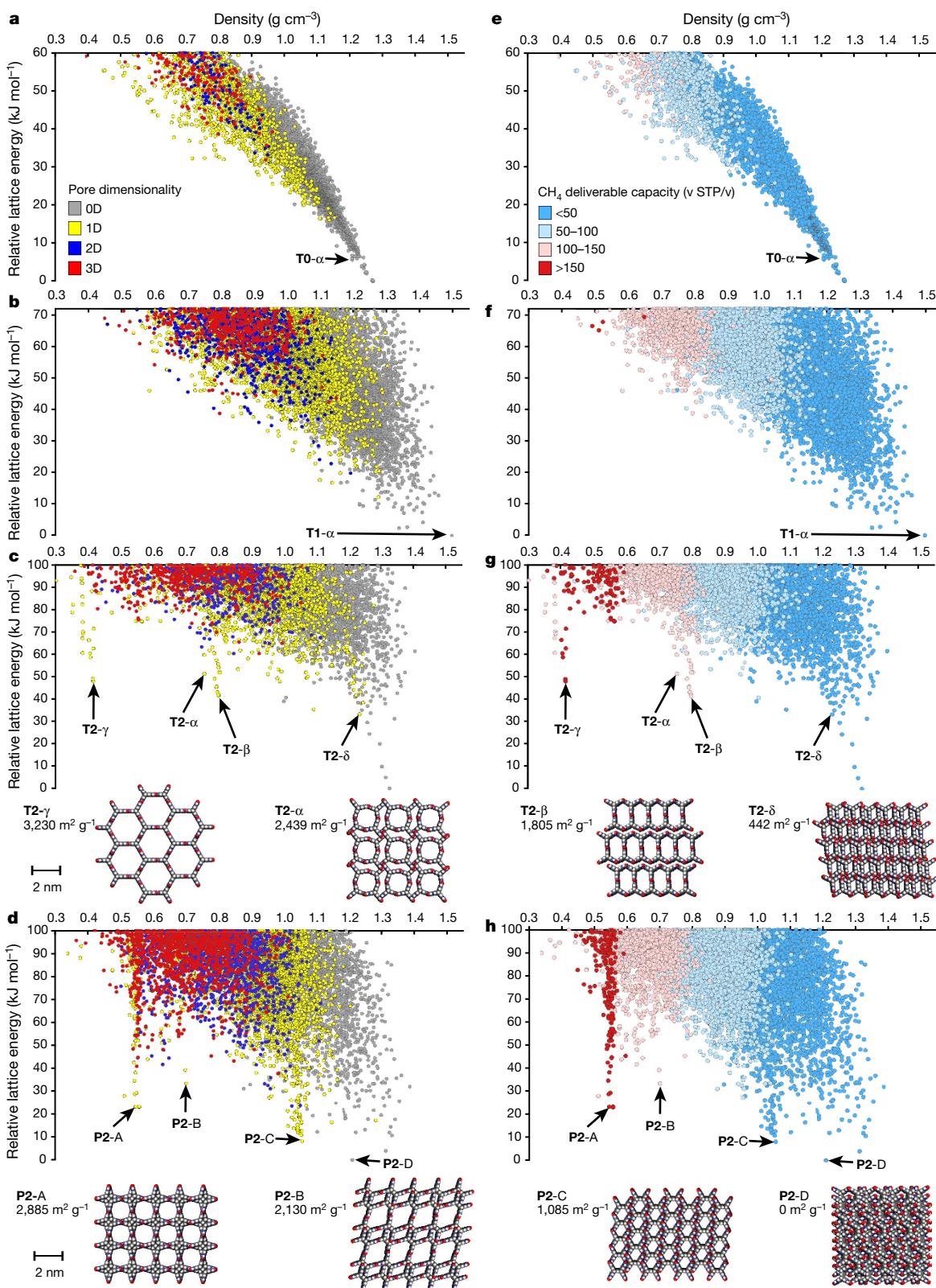
The energy–density distribution of structures for **T2** (Fig. 2c) is markedly different. Multiple low-density structures are predicted with energies substantially below the bulk of the landscape, indicating unusual stability for their respective densities. In particular, two low-energy ‘spikes’ are apparent at densities of about  $0.8 \text{ g cm}^{-3}$  and  $0.4 \text{ g cm}^{-3}$ ,

both containing structures with one-dimensional pores. Their energetic separation from the bulk of the landscape suggests a sizeable energy barrier to transformation to denser structures. Structures within these spikes feature hydrogen-bonded networks (Supplementary Fig. 4) with two-dimensional rings propagating along a third direction to form one-dimensional pore channels. Decomposition of the lattice energy into its physical contributions shows that the structures on the leading edge of the landscape are those with the most stabilizing electrostatic interactions (Supplementary Figs 5–7). The low-energy spikes on the **T2** landscape result from optimization of the strongly directional electrostatic interactions via oriented, polar hydrogen bonds. These spikes are an immediate indication that **T2** might be a better choice for porosity generation than its imide analogue **T1**, even without analysis of the various predicted structures. As discussed later, all structures highlighted for **T2** in Fig. 2c (**T2- $\alpha$** –**T2- $\delta$** ) are stabilized by solvent inclusion in their voids, such that they are more stable than the predicted global minimum.

The previously reported<sup>22</sup> crystal structure for **T2** is found in the  $0.8 \text{ g cm}^{-3}$  spike (**T2- $\alpha$** , Fig. 2c) and its calculated lattice energy is  $51 \text{ kJ mol}^{-1}$  above the predicted global energy minimum. Because **T2- $\alpha$**  is a known structure—a ‘landmark’ on the energy map—other structures with comparable energies might also be experimentally accessible. The predicted energy minimum in the  $0.8 \text{ g cm}^{-3}$  spike corresponds to a new porous structure, **T2- $\beta$** , which contains hydrogen-bonded chains along the pore channels as well as extensive  $\pi$ - $\pi$  stacking (Fig. 2) that is not found in **T2- $\alpha$** . We also predict structures in the  $0.4 \text{ g cm}^{-3}$  spike with lattice energies that are comparable with those of **T2- $\alpha$**  and **T2- $\beta$** , despite having exceptionally low packing fractions of approximately 0.2 (compared with 0.37 and 0.39 for **T2- $\alpha$**  and **T2- $\beta$** , respectively). The minimum-energy structure in this spike, **T2- $\gamma$** , is again dominated by hydrogen-bonded chains, but here the hydrogen bonds are linear, optimizing the directional intermolecular electrostatic interactions (Supplementary Fig. 6) and resulting in hexagonal pore channels (pore diameter,  $1.99 \text{ nm}$ ; Fig. 2). **T2- $\gamma$**  has a high calculated surface area of  $3,230 \text{ m}^2 \text{ g}^{-1}$  and an exceptionally low predicted density of  $0.417 \text{ g cm}^{-3}$ . The densest one-dimensional porous structure on the energy landscape, **T2- $\delta$** , is also highlighted in Fig. 2c. **T2- $\delta$**  is predicted to be the most stable of the possible porous structures for **T2**, with a calculated lattice energy that is lower than the lattice energies of **T2- $\alpha$** , **T2- $\beta$**  and **T2- $\gamma$** . The previously reported structure<sup>22</sup> for **T2**, **T2- $\alpha$** , has the highest calculated lattice energy of the four porous forms highlighted here, suggesting that **T2- $\alpha$**  might not be the only porous polymorph of this molecule.

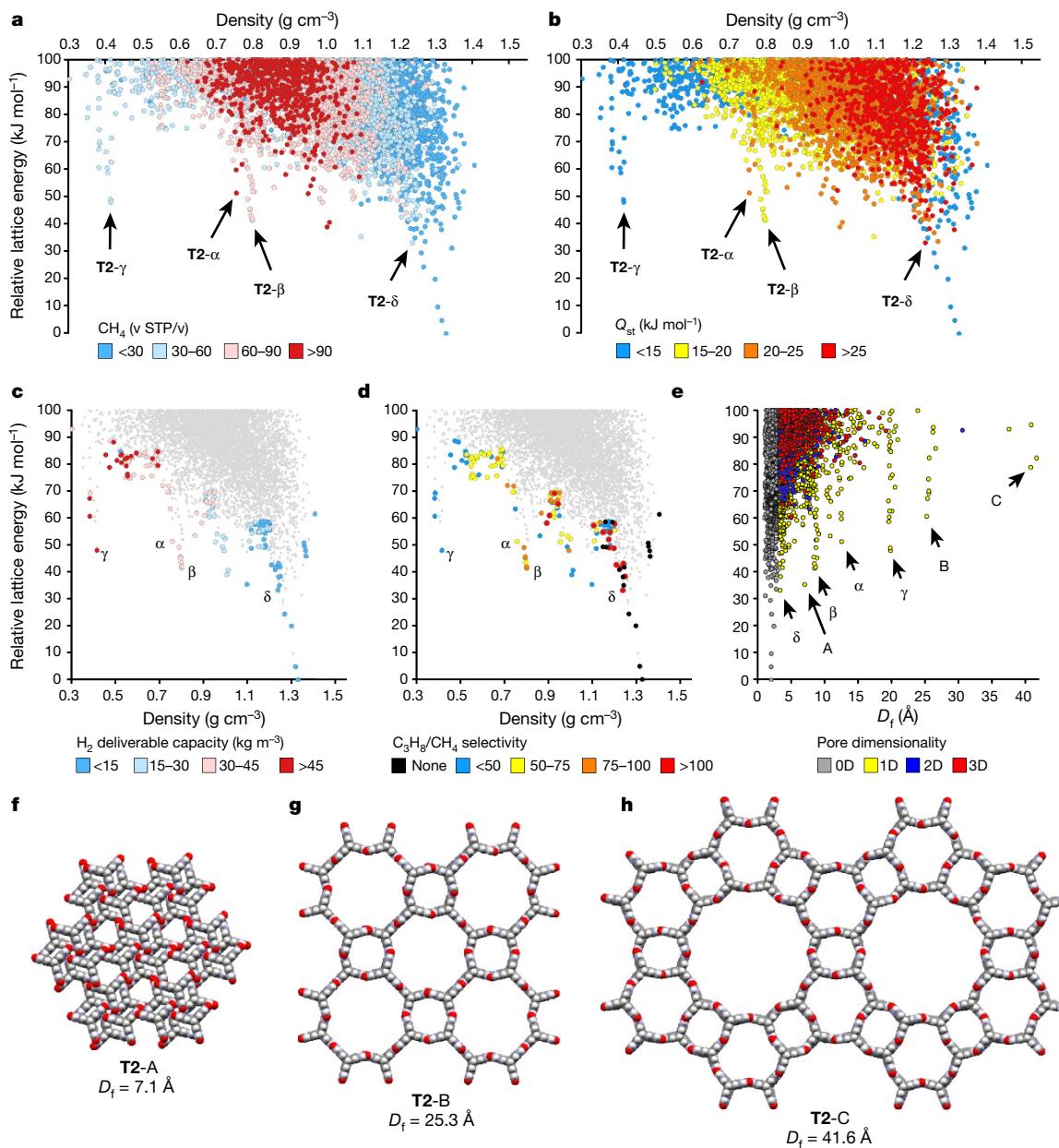
The spikes in the energy surface for **T2** stem from its shape and symmetry, which allows a strong hydrogen-bonding network to form along the channels, frustrating interpenetration. By contrast, the spiro-linked benzimidazolone, **S2**, which is pseudo-tetrahedral, shows no unusually stable low-density regions (Supplementary Fig. 1). **P2** also shows pronounced spikes extending below the bulk of its energy landscape (Fig. 2d), and we highlight three porous structures from its leading edge, **P2-A**, **P2-B** and **P2-C** (Fig. 2d), each of which exhibits extensive intermolecular hydrogen bonding and a high predicted surface area. As for **T2**, the most stable predicted porous structures for **P2** have one-dimensional pore channels. By contrast, the dimethyl analogue of **P2**, **P2M**, shows less pronounced spikes in its energy surface (Supplementary Fig. 1), and some of the most stable predicted porous structures have two- or three-dimensional pore topologies. This sensitivity to small chemical changes highlights the non-intuitive relationship between molecular structure and the energy–structure landscape of a crystal.

There were no pronounced spikes in the lattice energy surfaces for the pentiptycene imides **P1** and **P1M**, or the spiro-linked imide **S1** (Supplementary Fig. 1), suggesting that imides will be less effective than benzimidazolone at generating porosity, for both triptycene and pentiptycene frameworks.



**Figure 2 |** From structure prediction to energy-structure-function (ESF) maps. **a–d**, Crystal structure prediction (CSP) energy–density plots for T0 (a), T1 (b), T2 (c) and P2 (d), with each point corresponding to a computed crystal structure. The symbols are colour coded by the dimensionality of the pore channels, assessed using a CH<sub>4</sub> probe radius, 1.7 Å. **e–h**, ESF maps showing the calculated methane deliverable

capacities for T0 (e), T1 (f), T2 (g) and P2 (h), projected onto the energy–density plot. The symbols are colour coded by deliverable capacity (in units of v STP/v; 65–5.8 bar, 298 K). T2 (c, g) and P2 (d, h) structures selected from the leading edge of the distributions are also shown. Grey, white, blue and red atoms represent carbon, hydrogen, nitrogen and oxygen, respectively.



**Figure 3 |** ESF maps for T2. **a**, Volumetric methane capacity at 5.8 bar and 298 K. **b**, Isosteric heat of adsorption ( $Q_{\text{st}}$ ) for methane. **c**, Calculated  $\text{H}_2$  deliverable capacity (assuming gas storage at 100 bar and 77 K and gas delivery at 5 bar and 160 K) for structures selected from the leading edge of the ESF map; T2- $\gamma$  has the highest deliverable capacity. **d**, Simulated propane/methane ( $\text{C}_3\text{H}_8/\text{CH}_4$ ) selectivity (1 bar and 298 K); T2- $\beta$  and

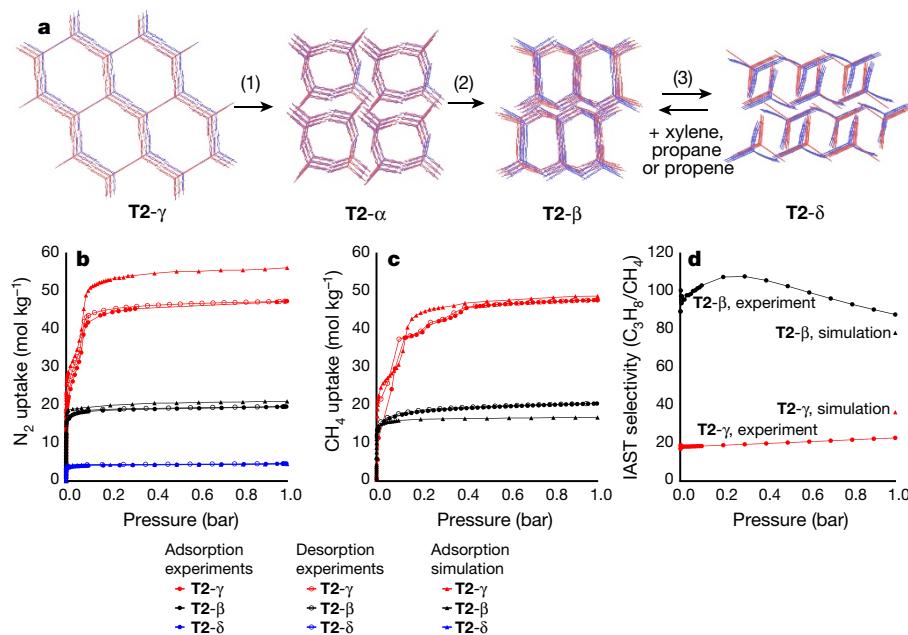
T2- $\delta$  have the highest selectivities. **e**, Relative lattice energy versus  $D_f$ , the largest free sphere, which relates to pore size; symbols are coloured by pore dimensionality. **f–h**, Selected hypothetical structures from the leading edge: T2-A (**f**), T2-B (**g**) and T2-C (**h**) correspond to labels A, B and C in **e**.

### Energy–structure–function maps

Pore topologies and computed surface areas (Fig. 2a–d) are inexpensive descriptors for porosity, but for practical applications we are rarely interested in those parameters *per se*. However, we can project any computable property onto these energy landscapes to guide the selection of materials for specific purposes. For example, methane storage<sup>29,30</sup> is an important goal for natural-gas-powered vehicles, and we therefore calculated the methane deliverable capacity<sup>7</sup> (298 K, 65–5.8 bar) for each structure in the predicted ensembles to create ESF maps (Fig. 2e–h, Supplementary Fig. 2). Unsurprisingly, the T0 map shows no structures with high methane capacities (Fig. 2e). There are also very few structures predicted for T1 with reasonable deliverable capacities (>150 volumes of methane at standard temperature and pressure per volume of solid, v STP/v), and their relative lattice energies are high (Fig. 2f). By contrast, the T2 map shows multiple

structures with capacities that exceed 150 v STP/v (Fig. 2g). More importantly, the minimum-energy structure in the  $0.4 \text{ g cm}^{-3}$  spike, T2- $\gamma$ , has one of the highest predicted methane capacities in the T2 structure ensemble (159 v STP/v). The deliverable capacity for T2- $\alpha$  is calculated to be 109 v STP/v—32% lower than for T2- $\gamma$ , and closer to the capacity for T2- $\beta$  (115 v STP/v). We can therefore predict, *a priori*, that T2 has stronger potential for methane storage than is suggested by its known structure, T2- $\alpha$ .

P2 also has low-energy predicted structures with calculated methane capacities of more than 150 v STP/v (Fig. 2h). Hence, ESF maps suggest, at a glance, that T2 and P2 are the most promising candidates here for methane storage. These maps also suggest that the more synthetically elaborate, hypothetical P2 offers no advantage over T2 for methane storage because both T2- $\gamma$  and P2-A have similar predicted capacities of 159 v STP/v and 153 v STP/v, respectively.



**Figure 4 | Predicted and experimental structures and gas adsorption isotherms for polymorphs of T2.** **a**, Overlays of predicted (red) and experimental (blue) structures for T2- $\gamma$ , T2- $\alpha$ , T2- $\beta$  and T2- $\delta$ , ordered by increasing predicted density. The transformation conditions for interconverting these polymorphs were as follows: (1) loss of solvent at room temperature, heating at 340 K or mechanical grinding at room temperature; (2) heating at 358–383 K; (3) direct removal of DMSO and then acetone from DMSO/acetone solvate. T2- $\delta$  transforms back to a solvate of T2- $\beta$  upon exposure to xylene, propane or propene. All four phases can be isolated as stable solvent-free frameworks in the

laboratory. **b, c**, Predicted and experimental gas adsorption isotherms for T2- $\gamma$  (red), T2- $\beta$  (black) and T2- $\delta$  (blue). **b**, Nitrogen isotherms (77 K); **c**, methane isotherms (115 K); filled circles, adsorption experiments; unfilled circles, desorption experiments; filled triangles, adsorption simulations. All simulations were performed using the CSP structures. **d**, Pressure-dependent IAST selectivity of propane over methane ( $C_3H_8/CH_4$ ) determined for equimolar mixtures, using experimental isotherms at 298 K (Supplementary Fig. 19). The ESF selectivity predictions are marked as triangles and coloured accordingly for T2- $\beta$  (black) and T2- $\gamma$  (red).

The ESF map for methane adsorption in T2 at 5.8 bar (the depletion pressure) shows a different pattern (Fig. 3a); smaller-pore structures such as T2- $\alpha$  and T2- $\beta$  adsorb more methane than does T2- $\gamma$ , as reflected in the isosteric heats of adsorption (Fig. 3b). This reduced adsorption in T2- $\gamma$  at the depletion pressure is the main reason for its superior predicted methane deliverable capacity (Supplementary Fig. 8). T2- $\gamma$  is also predicted to have the highest hydrogen deliverable capacity ( $47.1\text{ kg m}^{-3}$  or 11.3 wt%) assuming gas storage at 100 bar and 77 K and gas delivery at 5 bar and 160 K (Fig. 3c)<sup>31</sup>.

Small pores and relatively strong adsorption are often beneficial for applications such as hydrocarbon separation. Calculations for propane and methane adsorption in structures on the leading edge of the T2 energy landscape (Supplementary Figs 9, 10) suggest that the hypothetical small-pore polymorph, T2- $\beta$ , might have a good balance of adsorption capacity and selectivity for separating these two industrially important gases at 298 K and 1 bar (Fig. 3d; ratio of predicted propane and predicted methane adsorption,  $C_3H_8/CH_4 = 78.1$ ). There are also denser porous structures on the leading edge of this map, such as T2- $\delta$ , that are predicted to have even higher propane/methane selectivity ( $C_3H_8/CH_4 = 119.7$ ). By contrast, T2- $\alpha$  and T2- $\gamma$  have good predicted adsorption capacities for hydrocarbon gases, but poorer selectivities. We therefore predict that T2- $\beta$  and T2- $\delta$  (or other leading-edge structures with similar densities to T2- $\delta$ ; see red points in Fig. 3d) have good potential for propane/methane separation. Likewise, T2- $\delta$  is predicted to have high selectivity for *ortho*-xylene over the *meta*- and *para*-xylene isomers (Supplementary Fig. 11). T2- $\alpha$ , T2- $\beta$  and T2- $\gamma$  are predicted to be weakly *ortho*-xylene selective, but much less so than T2- $\delta$ .

'Spikes' also emerge when the lattice energy for T2 is plotted against calculated pore size ( $D_f$ ), rather than density (Fig. 3e). This is because there are related structures with common packing motifs, such as hydrogen-bonded channels with specific pore diameters (Fig. 3f–h and Supplementary Figs 12, 13). These channels are reminiscent of inorganic zeolites, and it is conceivable that some of the higher-energy

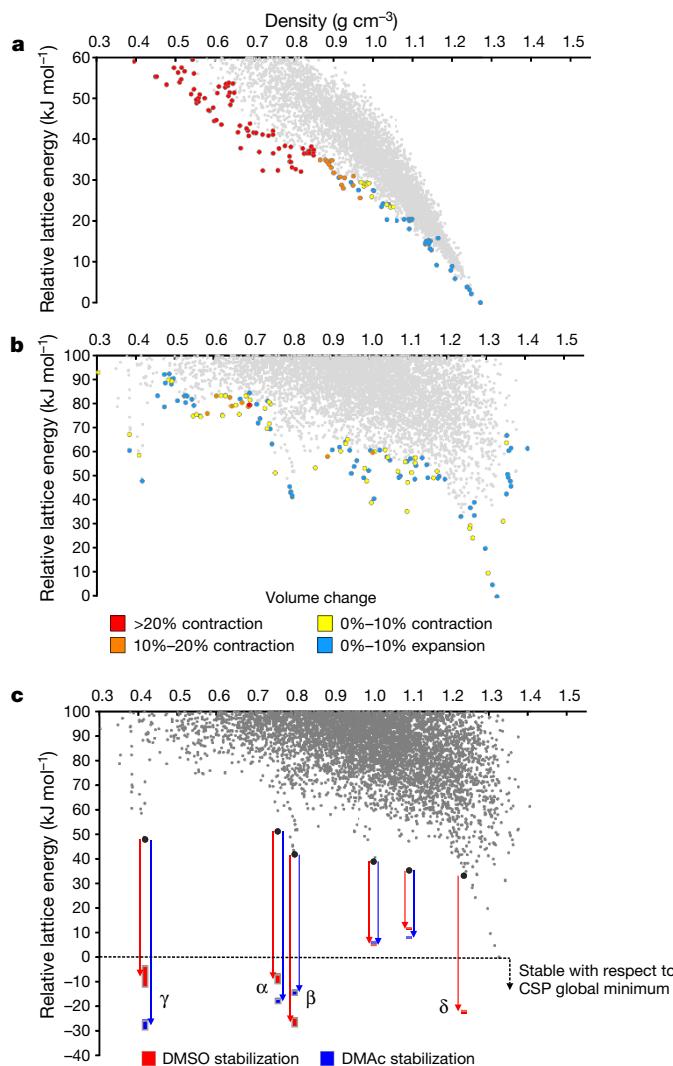
forms, such as T2-B or T2-C, might be accessed by using structure-directing agents<sup>1</sup>. Again, ESF maps are revealing, even without analysis of the individual structures: comparison with other candidate molecules (Supplementary Fig. 14) shows that T2 and P2 are the only molecules here that might form a mesoporous solid with pores of 2 nm or larger. Thinking beyond porous solids, ESF maps could be prepared in which lattice energy is plotted against any other calculable parameter, for example, to identify molecules that have low-energy structures with a suitable electronic bandgap for solar energy harvesting.

### Crystallization verifies ESF map predictions

Computation suggests that T2 has a rich, unrevealed chemistry; crystallization screens were therefore performed for T2 and its imide analogue T1, as a comparison.

Despite extensive studies across multiple solvents and desolvation protocols, we were unable to isolate a stable, unsolvated porous crystalline phase of T1 (Supplementary Figs 15, 16). However, T1 crystals grown by sublimation revealed a previously unreported dense structure that was accurately predicted by the global-energy-minimum structure on the T1 energy landscape (T1- $\alpha$ , Supplementary Fig. 16). Efforts to grow equivalent solvent-free crystals of T2 failed, owing to decomposition before sublimation. However, slow diffusion of acetone into a saturated dimethylacetamide (DMAc) solution of T2 led to crystallization of a solvated form of the predicted hexagonal T2- $\gamma$ , isolated initially as T2-(DMAc)<sub>7.79</sub>. The T2- $\gamma$  polymorph was also found to crystallize from other solvent combinations, such as DMSO/acetone and N-methyl-2-pyrrolidone/acetone.

The T2-(DMAc)<sub>7.79</sub> material was desolvated by first exchanging DMAc for acetone and then pentane, followed by overnight evacuation. At 240 K, no residual solvent was found in the one-dimensional pores and a density of  $0.412\text{ g cm}^{-3}$  was determined by crystallography. Single-crystal data were recorded up to 500 K, demonstrating good thermal stability for desolvated T2- $\gamma$ . There is excellent agreement



**Figure 5 | Crystal structure stability and solvent stabilization.** **a, b,** Volume change during molecular dynamics calculations at 300 K for leading-edge structures of **T0** (**a**) and **T2** (**b**). Large (>10%) contraction corresponds to collapse of porosity present in the temperature-free predicted structures. **c,** Calculated stability of DMSO- and DMAc-solvated structures of **T2- $\alpha$** , **T2- $\beta$** , **T2- $\gamma$**  and **T2- $\delta$** , and the DMSO- and DMAc-solvates of two predicted porous structures of intermediate density. The solid bars give energy ranges for each fully solvated structure. **T2- $\delta$**  is unable to accommodate DMAc in the simulations. All energies are shown relative to the CSP global-energy-minimum structure.

between the predicted structure and the experimental desolvated structure for **T2- $\gamma$**  (Fig. 4a). The structure of the previously reported polymorph<sup>22</sup>, **T2- $\alpha$** , is also accurately reproduced by one of the predicted structures (Fig. 4a).

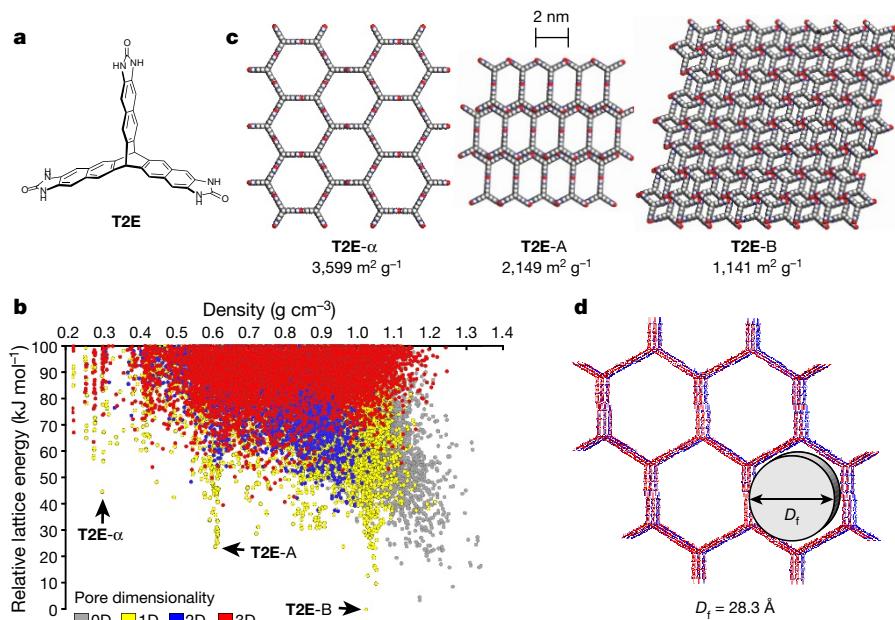
A type-IV nitrogen ( $N_2$ ) adsorption isotherm was obtained for fully desolvated **T2- $\gamma$**  with a sharp step at  $P/P_0 = 0.06$  (where  $P$  is the adsorbate pressure and  $P_0 = 1$  bar is the final pressure), in agreement with grand canonical Monte Carlo (GCMC) adsorption simulations (Fig. 4b). This sharp step is attributable to nitrogen condensation in the uniform hexagonal pores<sup>32</sup> (Supplementary Fig. 17). The experimental surface area  $S_{\text{BET}}$  was estimated to be  $3,425 \text{ m}^2 \text{ g}^{-1}$  (1 s.d. = 3%) in good agreement with our *a priori* prediction ( $3,230 \text{ m}^2 \text{ g}^{-1}$ ). **T2- $\gamma$**  has the lowest density of any molecular solid reported in the Cambridge Structural Database so far<sup>21,33</sup>. It is also the largest pore size observed for an extrinsically porous molecule (1.99 nm), rather than a cage<sup>33</sup>. The promise of **T2- $\gamma$**  for methane adsorption, suggested by ESF maps (Fig. 2g, Supplementary Fig. 18), was also confirmed by experiment.

The saturation methane capacity at 115 K for **T2- $\gamma$**  was found to be  $47.4 \text{ mol kg}^{-1}$  (437.4 v STP/v), both by experiment and by simulations (Fig. 4c).

**T2- $\gamma$**  is stabilized by solvent exchange with pentane, but the DMAc/acetone solvate of **T2- $\gamma$**  taken directly from the mother liquor of the crystallization transforms to a solvate of **T2- $\alpha$**  under light grinding or when left to stand at room temperature. Likewise, heating the **T2- $\gamma$**  solvate isolated from a DMSO/acetone mixture for 60 min at 340 K transformed it to a **T2- $\alpha$**  solvate. Further heating of this **T2- $\alpha$**  solvate (30 min, 358 K) caused another transformation to a third phase, which was identified by comparison of CSP-derived and experimental powder X-ray diffraction (PXRD) patterns as the predicted **T2- $\beta$**  polymorph. **T2- $\beta$**  could also be isolated in desolvated form by heating **T2- $\alpha$**  in DMSO/acetone to 383 K for 3 h before exchanging the solvent in the pores with acetone and then *n*-pentane before evacuation. If the acetone-solvent-exchanged material was evacuated directly, without first exchanging for *n*-pentane, then the predicted unsolvated **T2- $\delta$**  phase could also be isolated. A single-crystal structure of the **T2- $\delta$**  acetone solvate revealed a close size match between the one-dimensional pores and the ordered acetone guests, which interact with the **T2** molecules. The absence of such strong solvent-framework interactions rationalizes why the *n*-pentane solvate of **T2- $\beta$**  does not transform to **T2- $\delta$**  during evacuation of the pores. Hence, by using different crystallization solvents and desolvation protocols, we can access in the laboratory at least four of the predicted structures on the leading edge of the energy landscape for **T2**: the known **T2- $\alpha$**  form, plus three new polymorphs, **T2- $\beta$** , **T2- $\gamma$**  and **T2- $\delta$** .

There is good agreement between experimental nitrogen adsorption isotherms and adsorption isotherms predicted from the CSP-derived structures for all three of the new polymorphs, **T2- $\beta$** , **T2- $\gamma$**  and **T2- $\delta$**  (Fig. 4b). Likewise, the methane adsorption isotherms for **T2- $\beta$**  and **T2- $\gamma$**  can be predicted *a priori* (Fig. 4c). Our prediction that **T2- $\beta$**  would have better propane/methane selectivity than **T2- $\gamma$**  (Fig. 3d) was also realized by experiments that showed enhanced ideal propane/methane selectivity for **T2- $\beta$** , as calculated using ideal adsorbed solution theory (IAST) (Fig. 4d, Supplementary Fig. 19). However, studies on the denser **T2- $\delta$**  polymorph revealed some limitations of these *a priori* predictions. **T2- $\delta$**  was predicted to have superior hydrocarbon selectivity compared to **T2- $\beta$**  (Fig. 3d, e), but experiments show that **T2- $\delta$**  transforms back to **T2- $\beta$**  upon exposure to either *ortho*-xylene or propane (Fig. 4a, Supplementary Figs 20–22). By contrast, neither nitrogen nor methane adsorption causes the expansion of **T2- $\delta$**  to the **T2- $\beta$**  phase.

To better understand the phase changes between the various **T2** polymorphs, and to explore the thermal stability of the predicted, solvent-free structures of **T0**, **T1** and **T2**, we used molecular dynamics simulations to probe the dynamical stability to thermal motion at 300 K (Fig. 5a, b, Supplementary Figs 23, 24). Most of the leading-edge structures for **T2** were found to be stable; in particular, **T2- $\delta$** , **T2- $\beta$**  and **T2- $\gamma$**  showed only small structural fluctuations about the CSP structure during 500-ps simulations (Fig. 5b, Supplementary Fig. 24). By contrast, molecular dynamics simulations for **T2- $\alpha$**  show a partial transformation to the **T2- $\beta$**  structure, in line with experimental observations (Fig. 4a). The phase transformations observed for solvates in our crystallization experiments suggest that the presence of polar solvent in the pores of **T2- $\gamma$**  and **T2- $\alpha$**  lowers the barrier to interconversion to denser forms. In the case of **T2- $\delta$** , solvation by propane or by *ortho*-xylene (but not  $N_2$  or  $CH_4$ ) can cause expansion to the less dense form, **T2- $\beta$** . In contrast to **T2**, almost all low-density ( $\rho < 1 \text{ g cm}^{-3}$ ) predicted structures for **T0** were unstable during the molecular dynamics simulations, rapidly collapsing to denser structures at room temperature (Fig. 5a). **T1**, which did not yield stable porous structures by experiment, showed intermediate behaviour, with many low-density structures collapsing during molecular dynamics, but some remaining stable at 300 K (Supplementary Fig. 23). These results show that molecular dynamics simulations for leading-edge structures can provide a



**Figure 6 | Predicted and experimental structures and properties for T2E.** **a**, Extended benzimidazolone analogue of T2, T2E. **b**, CSP energy-density plot. **c**, Selected structures for T2E, drawn from the leading edge

qualitative indication of the likely experimental stability of interesting, low-density forms.

We also probed the influence of solvent on the T2 energy landscape to understand why porous polymorphs are formed instead of the dense and non-porous predicted global-minimum structure. Solvent stabilization calculations were performed on the four observed polymorphs, T2- $\alpha$ , T2- $\beta$ , T2- $\gamma$  and T2- $\delta$ , plus two other leading-edge structures, using Monte Carlo and lattice-energy-minimization methods (Fig. 5c, Supplementary Figs 25, 26). Simulations with both DMSO and DMAc show that the four observed polymorphs are energetically favoured over the global-minimum predicted structure when the channels are fully solvated, explaining why porous networks crystallize from solution in preference to a dense, non-porous phase.

### ESF maps predict an ultra-low-density solid

The existence of T2- $\gamma$  and the electrostatic stabilization that we calculated for this material (Supplementary Figs 5, 6) suggests the possibility of designing molecular crystals with even higher porosity levels. The energy landscape for the extended pentiptycene molecule, P2, did not predict the existence of any competitive structures with densities lower than that of T2- $\gamma$ . For example, P2-A (Fig. 2d) has a higher predicted density ( $0.559 \text{ g cm}^{-3}$ ) than T2- $\gamma$ . We therefore calculated the crystal-energy landscape for a hypothetical extended form of T2, T2E (Fig. 6a). This landscape again contains distinct low-energy spikes, as for T2 and P2, and a leading-edge structure was identified (T2E- $\alpha$ , Fig. 6b) that has a predicted density of just  $0.303 \text{ g cm}^{-3}$  and hexagonal pore channels with diameters of  $2.83 \text{ nm}$ . T2E- $\alpha$  is isostructural to T2- $\gamma$  (Fig. 6c) and it is located  $44.7 \text{ kJ mol}^{-1}$  above the predicted global-minimum structure (T2E-B). Comparison with T2- $\gamma$ , which lies  $47.8 \text{ kJ mol}^{-1}$  above its predicted global minimum, suggested that T2E- $\alpha$  might also be stabilized by solvent inclusion. We therefore devised a synthetic route to T2E (Supplementary Information, Supplementary Fig. 27) and grew crystals of this material from DMF/acetone and DMSO/CHCl<sub>3</sub>, yielding the predicted T2E- $\alpha$  phase as a solvate (Fig. 6d). Preliminary studies suggest that T2E- $\alpha$  can be prepared as a bulk, phase-pure solvate (Supplementary Fig. 27) with  $2.8\text{-nm}$  hexagonal pore channels. This is the largest pore size observed for any molecular organic crystal so far, including intrinsically mesoporous organic cages<sup>33</sup>. We predict that T2E- $\alpha$  will have remarkable properties, if scaled up and successfully

of the energy–density landscape: T2E- $\alpha$ , T2E-A and the global-minimum structure, T2E-B. **d**, Overlay of predicted (red) and experimental (blue) structures for T2E- $\alpha$ . ESF maps for T2E are shown in Supplementary Fig. 28.

desolvated in the bulk; for example, it should have high volumetric and gravimetric deliverable capacities for hydrogen ( $43.8 \text{ kg m}^{-3}$  and  $14.5 \text{ wt\%}$ , respectively) assuming gas storage at  $100 \text{ bar}$  and  $77 \text{ K}$  and gas delivery at  $5 \text{ bar}$  and  $160 \text{ K}$ . Unlike all of the other molecules studied here, T2E has a predicted global-minimum structure, T2E-B (Fig. 6b, c), that is predicted to be highly porous ( $1,141 \text{ m}^2 \text{ g}^{-1}$ ). We therefore anticipate that T2E might be resistant to loss of porosity, so long as it remains crystalline, which is not obvious from the molecular structure in isolation.

### Outlook

In summary, we have demonstrated the power of ESF maps for the computationally led discovery of new materials, in this case guiding us to three new porous polymorphs of a known molecule, T2, each of which has different physical properties, plus an ultra-low-density form of a new molecule, T2E- $\alpha$ . It is challenging for computers to generate abstract chemical design hypotheses or rules of thumb that might be intuitive to chemists, such as the principle that an equal number of hydrogen-bond donors and acceptors makes imidazolones better than imides for open-framework generation. ESF maps therefore open the door to more autonomous computer-aided materials design schemes in the future, for example, in which a computational algorithm evolves a set of candidate molecules, perhaps chosen initially using human knowledge, and selects for energetically favourable candidates with superior function that emerge from the resulting dynamic molecular library. This mapping approach is not limited to porous materials, but could find wider use in the discovery of functional solids, particularly as we expand our ability to calculate other physical properties from structure.

**Online Content** Methods, along with any additional Extended Data display items and Source Data, are available in the online version of the paper; references unique to these sections appear only in the online paper.

Received 3 June 2016; accepted 20 January 2017.

Published online 22 March 2017.

- Lewis, D. W., Wilcock, D. J., Catlow, C. R. A., Thomas, J. M. & Hutchings, G. J. *De novo* design of structure-directing agents for the synthesis of microporous solids. *Nature* **382**, 604–606 (1996).
- Oganov, A. R. et al. Ionic high-pressure form of elemental boron. *Nature* **457**, 863–867 (2009).

3. Ceder, G. *et al.* Identification of cathode materials for lithium batteries guided by first-principles calculations. *Nature* **392**, 694–696 (1998).
4. Marleny Rodriguez-Albelo, L. *et al.* Zeolitic polyoxometalate-based metal–organic frameworks (Z-POMOFs): computational evaluation of hypothetical polymorphs and the successful targeted synthesis of the redox-active Z-POMOF1. *J. Am. Chem. Soc.* **131**, 16078–16087 (2009).
5. Hachmann, J. *et al.* The Harvard Clean Energy Project: large-scale computational screening and design of organic photovoltaics on the World Community Grid. *J. Phys. Chem. Lett.* **2**, 2241–2251 (2011).
6. Dyer, M. S. *et al.* Computationally assisted identification of functional inorganic materials. *Science* **340**, 847–852 (2013).
7. Simon, C. M. *et al.* The materials genome in action: identifying the performance limits for methane storage. *Energy Environ. Sci.* **8**, 1190–1199 (2015).
8. Ma, Y. *et al.* Transparent dense sodium. *Nature* **458**, 182–185 (2009).
9. Sharma, V. *et al.* Rational design of all organic polymer dielectrics. *Nat. Commun.* **5**, 4845 (2014).
10. Wilmer, C. E. *et al.* Large-scale screening of hypothetical metal–organic frameworks. *Nat. Chem.* **4**, 83–89 (2012).
11. Jansen, M. & Schön, J. C. “Design” in chemical synthesis—an illusion? *Angew. Chem. Int. Ed.* **45**, 3406–3412 (2006).
12. Cruz-Cabeza, A. J., Reutzel-Edens, S. M. & Bernstein, J. Facts and fictions about polymorphism. *Chem. Soc. Rev.* **44**, 8619–8635 (2015).
13. Eddaoudi, M. *et al.* Systematic design of pore size and functionality in isoreticular MOFs and their application in methane storage. *Science* **295**, 469–472 (2002).
14. Côté, A. P. *et al.* Porous, crystalline, covalent organic frameworks. *Science* **310**, 1166–1170 (2005).
15. Woodley, S. M. & Catlow, R. Crystal structure prediction from first principles. *Nat. Mater.* **7**, 937–946 (2008).
16. Price, S. L. Predicting crystal structures of organic compounds. *Chem. Soc. Rev.* **43**, 2098–2111 (2014).
17. Mastalerz, M. Permanent porous materials from discrete organic molecules—towards ultra-high surface areas. *Chem. Eur. J.* **18**, 10082–10091 (2012).
18. Kitagorodskii, A. I. The principle of close packing and the condition of thermodynamic stability of organic crystals. *Acta Crystallogr.* **18**, 585–590 (1965).
19. Jones, J. T. A. *et al.* Modular and predictable assembly of porous organic molecular crystals. *Nature* **474**, 367–371 (2011).
20. Pyzer-Knapp, E. O. *et al.* Predicted crystal energy landscapes of porous organic cages. *Chem. Sci.* **5**, 2235–2245 (2014).
21. Evans, J. D. *et al.* Computational identification of organic porous molecular crystals. *CrystEngComm* **18**, 4133–4141 (2016).
22. Mastalerz, M. & Oppel, I. M. Rational construction of an extrinsic porous molecular crystal with an extraordinary high specific surface area. *Angew. Chem. Int. Ed.* **51**, 5252–5255 (2012).
23. Wang, H. *et al.* A flexible microporous hydrogen-bonded organic framework for gas sorption and separation. *J. Am. Chem. Soc.* **137**, 9963–9970 (2015).
24. Simard, M., Su, D. & Wuest, J. D. Use of hydrogen bonds to control molecular aggregation. Self-assembly of three-dimensional networks with large chambers. *J. Am. Chem. Soc.* **113**, 4696–4698 (1991).
25. Kohl, B., Rominger, F. & Mastalerz, M. Crystal structures of a molecule designed not to pack tightly. *Chem. Eur. J.* **21**, 17308–17313 (2015).
26. Zonta, C., De Lucchi, O., Linden, A. & Lutz, M. Synthesis and structure of  $D_{3h}$ -symmetric triptycene trimaleimide. *Molecules* **15**, 226–232 (2010).
27. Case, D. H., Campbell, J. E., Bygrave, P. J. & Day, G. M. Convergence properties of crystal structure prediction by quasi-random sampling. *J. Chem. Theory Comput.* **12**, 910–924 (2016).
28. Cruz-Cabeza, A. J., Day, G. M. & Jones, W. Predicting inclusion behaviour and framework structures in organic crystals. *Chem. Eur. J.* **15**, 13033–13040 (2009).
29. Gándara, F., Furukawa, H., Lee, S. & Yaghi, O. M. High methane storage capacity in aluminum metal–organic frameworks. *J. Am. Chem. Soc.* **136**, 5271–5274 (2014).
30. Mason, J. A. *et al.* Methane storage in flexible metal–organic frameworks with intrinsic thermal management. *Nature* **527**, 357–361 (2015).
31. Gómez-Gualdrón, D. A. *et al.* Evaluating topologically diverse metal–organic frameworks for cryo-adsorbed hydrogen storage. *Energy Environ. Sci.* **9**, 3279–3289 (2016).
32. Côté, A. P., El-Kaderi, H. M., Furukawa, H., Hunt, J. R. & Yaghi, O. M. Reticular synthesis of microporous and mesoporous 2D covalent organic frameworks. *J. Am. Chem. Soc.* **129**, 12914–12915 (2007).
33. Zhang, G., Presly, O., White, F., Oppel, I. M. & Mastalerz, M. A permanent mesoporous organic cage with an exceptionally high surface area. *Angew. Chem. Int. Ed.* **53**, 1516–1520 (2014).

**Supplementary Information** is available in the online version of the paper.

**Acknowledgements** We acknowledge the European Research Council under the European Union’s Seventh Framework Programme (FP/2007–2013)/ERC through grant agreement numbers 321156 (ERC-AG-PE5-ROBOT) and 307358 (ERC-stG-2012-ANGLE), and EPSRC (grants EP/N004884/1, EP/K018396/1 and EP/K018132/1) for funding. This work made use of the facilities of N8 HPC Centre of Excellence, provided and funded by the N8 consortium and EPSRC (grant number EP/K000225/1). T.H. thanks the Royal Society for a University Research Fellowship. We thank Diamond Light Source for access to beamlines I19 (MT8728) and I11 (EE12336). We thank the Advanced Light Source, supported by the Director, Office of Science, Office of Basic Energy Sciences, of the US Department of Energy under contract number DE-AC02-05CH11231, and S. J. Teat and K. J. Gagnon for their assistance. We acknowledge the ARCHER UK National Supercomputing Service via the UK’s HEC Materials Chemistry Consortium membership and a Programme Grant, which are funded by EPSRC (grants EP/L000202 and EP/N004884), and use of the IRIDIS High Performance Computing Facility at the University of Southampton.

**Author Contributions** A.P. performed the CSPs. T.K. synthesized **T2** and isolated **T2**- $\gamma$ . L.C. carried out the methane and hydrogen capacity simulations, the  $Q_{st}$  calculations, the hydrocarbon separation simulations, the adsorption isotherm simulations and the IAST calculations. D.H. analysed the pore geometries for the predicted structures, produced most of the ESF maps and wrote scripts to analyse the data. S.Y.C. and M.A.L. isolated **T2**- $\beta$  and **T2**- $\delta$  and carried out the PXRD experiments. S.Y.C. analysed the hydrogen-bonding patterns for **T2**. B.J.S. synthesized **T1**, **T2** and **T2E**. B.B. developed the initial synthetic route to **T2E**. D.P.M. carried out the molecular dynamics stability and solvent stabilization calculations. M.A.L., C.J.S. and A.S. collected single-crystal X-ray diffraction data and solved the structures; M.A.L., S.Y.C., A.S. and B.J.S. performed experiments on the stability of various phases. M.A.L. isolated **T2E**- $\alpha$ , solved its crystal structure and led the single-crystal diffraction work. T.K., C.M.K. and B.J.S. carried out crystallization and sublimation studies for **T1** and **T2**. T.K., R.C., M.A.L., A.S. and T.H. collected and interpreted the gas sorption isotherms. A.I.C. and G.M.D. conceived the project and the concept of ESF maps, and led the writing of the manuscript with contributions from all co-authors.

**Author Information** Reprints and permissions information is available at [www.nature.com/reprints](http://www.nature.com/reprints). The authors declare no competing financial interests. Readers are welcome to comment on the online version of the paper. Correspondence and requests for materials should be addressed to G.M.D. (G.M.Day@soton.ac.uk).

**Reviewer Information** *Nature* thanks R. Catlow and the other anonymous reviewer(s) for their contribution to the peer review of this work.

## METHODS

**Crystal structure prediction (CSP).** CSP was performed using a quasi-random sampling procedure, as implemented in the Global Lattice Energy Explorer software<sup>27</sup>. Molecules were first sketched in ChemDraw, followed by an initial molecular-geometry optimization with the COMPASS force field, as implemented in the Materials Studio software package<sup>34</sup>. Force-field-optimized molecular geometries were further refined by re-optimization using density functional theory (DFT) with the M06-2X exchange-correlation functional and 6-31G\*\* basis set. Molecular DFT calculations were performed with the Gaussian09 software<sup>35</sup>. These molecular geometries were held rigid throughout crystal structure generation and lattice energy minimization.

Trial crystal structures were generated with one molecule in the asymmetric unit in each of the 89 space groups in which non-polymeric  $Z' = 1$  organic molecular crystal structures are reported in the Cambridge Structure Database ( $Z'$  is number of molecules in the asymmetric unit):  $P_{2_1}/c$ ,  $P_{2_1}2_{1_1}$ ,  $\bar{P}_1$ ,  $P_{2_1}$ ,  $Pbcn$ ,  $C_2/c$ ,  $Pna_2$ ,  $C_c$ ,  $Pca_2$ ,  $C_2$ ,  $P_1$ ,  $Pbcn$ ,  $P_c$ ,  $P_{2_1}2_{1_2}$ ,  $P_{4_3}2_{1_2}$ ,  $P_{4_1}$ ,  $P_{3_2}$ ,  $Fdd2$ ,  $Pccn$ ,  $P_2/c$ ,  $P_6_1$ ,  $I_{4_1}/a$ ,  $R\bar{3}$ ,  $C22_2$ ,  $P_4/n$ ,  $P_{3_2}2_1$ ,  $Aba2$ ,  $P\bar{3}_1c$ ,  $Iba2$ ,  $R_3$ ,  $\bar{I}4$ ,  $Pnma$ ,  $R3c$ ,  $I_{4_1}cd$ ,  $P_{2_1}/m$ ,  $I_4$ ,  $Pnna$ ,  $P_4bc$ ,  $P_3$ ,  $I_{4_1}$ ,  $P_4$ ,  $Pmc_2$ ,  $Pbam$ ,  $P_6_3$ ,  $Pnn2$ ,  $Pnc2$ ,  $C_2/m$ ,  $P_4/n$ ,  $Ibca$ ,  $P_6_22$ ,  $Fddd$ ,  $Pcc$ ,  $P_2$ ,  $P_6_2$ ,  $P_{31}c$ ,  $P_3$ ,  $Ccc_2$ ,  $Pba2$ ,  $P_{4_2}2_{1_2}$ ,  $P_{4_1}2_2$ ,  $\bar{P}4$ ,  $I22_2$ ,  $I\bar{4}_2d$ ,  $P4nc$ ,  $Ccc_2$ ,  $Pm$ ,  $Pnna$ ,  $Cmc_2$ ,  $Pcc_2$ ,  $P_6_22$ ,  $P_6_4$ ,  $P\bar{3}_1c$ ,  $P222$ ,  $P_4/nc$ ,  $\bar{P}4n2$ ,  $I_{4_1}22$ ,  $Pbcm$ ,  $Pmn_2$ ,  $F222$ ,  $C22_2$ ,  $P_{2_1}3$ ,  $I2_3$ ,  $P22_2$ ,  $R\bar{3}c$ ,  $P_{3_2}12$ ,  $P_{3_1}12$ ,  $\bar{P}4b_2$ ,  $P\bar{4}_2c$  and  $P_2/m$ .

The generation of crystal structures involves a low-discrepancy sampling of all structural variables within each space group: unit cell lengths and angles, and molecular positions and orientations within the asymmetric unit. Space-group symmetry was then applied and a geometric test was performed for overlap between molecules. Molecular clashes were removed by lattice expansion (the SAT-expand method in ref. 27). All of the accepted trial structures were lattice-energy-minimized and the search was run until a total of 5,000 lattice energy minimizations had been performed in each space group (445,000 energy minimizations per molecule). Lattice energy calculations were performed with an anisotropic atom–atom potential using DMACRYS<sup>36</sup>. Electrostatic interactions were modelled using an atomic multipole description of the molecular charge distribution (up to hexadecapole on all atoms) from the B3LYP/6-31G\*\*-calculated charge density using a distributed multipole analysis. Atom–atom repulsion and dispersion interactions were modelled using a revised Williams intermolecular potential<sup>37</sup>. Charge–charge, charge–dipole and dipole–dipole interactions were calculated using Ewald summation; all other intermolecular interactions were summed to a 30-Å cut-off between molecular centres-of-mass. For a given molecule, the relative lattice energy of a given predicted crystal structure was evaluated as the difference between its calculated lattice energy and the lattice energy of the global minimum on the energy–density landscape of that given molecule. Duplicate structures were removed from the set using COMPACK<sup>38</sup>, on the basis of matching interatomic distances within 30-molecule clusters. Unless stated otherwise, best geometric overlays between predicted and experimentally determined single crystal structures are given by the lowest root-mean-square deviation of atomic positions in a 15-molecule cluster (r.m.s.d.<sub>15</sub>).

The packing coefficient, PC, of a given predicted crystal structure was evaluated as the ratio between the gas-phase molecular volume,  $V_g$ , and the crystal volume per molecule,  $V_c$ :

$$PC = \frac{V_g}{V_c}$$

Connolly molecular volumes at DFT equilibrium geometries were used as the gas-phase molecular volume,  $V_g$ , and calculated using Materials Studio<sup>34</sup>. Crystal volume per molecule,  $V_c$ , was evaluated as the unit cell volume divided by the number of molecules per unit cell. Calculated T0, T1, T2, T2E, S1, S2, P1, P2, P1M and P2M gas-phase molecular volumes,  $V_g$ , are 244.47, 369.09, 344.46, 479.91, 474.40, 442.78, 568.05, 544.45, 607.28 and 586.52 Å<sup>3</sup>, respectively.

Throughout, Greek letters  $\alpha$ ,  $\beta$ ,  $\gamma$ , and so on, are used to refer to predicted crystal structures corresponding to experimentally observed polymorphs, whereas as-yet unobserved predicted structures are labelled A, B, C, and so on.

**Energy–structure–function (ESF) maps.** Each structure in each CSP dataset was assigned a unique identification code. To generate the ESF maps, these structures then underwent a series of property calculations and the results were tabulated. For properties such as methane capacity, calculations were performed for all CSP-derived structures for a given molecule. For computationally more expensive property calculations, a ‘leading edge’ of low-energy structures across the computed density range was first selected.

**Supercell generation.** For each of the CSP structures used to generate the ESF maps, a supercell representation of the crystal structure was adopted in the geometric property calculations and adsorption/separation simulations. A supercell was

obtained by replicating the unit cell in three dimensions so that the minimum of the projections of its crystallographic  $a$ ,  $b$  and  $c$  cell vectors onto the Cartesian  $X$ ,  $Y$  and  $Z$  axes was greater than 24.0 Å.

**Geometric analysis.** Full topological analysis was performed on the CSP structures using the void analysis tool Zeo++ (ref. 39). This calculates the geometrical parameters of the pore space, if any, within each structure on the basis of the Voronoi decomposition. The outputs from this analysis include the pore dimensionality (0D, 1D, 2D or 3D), the largest inclusion sphere ( $D_i$ ) and the largest free sphere ( $D_f$ ). A probe radius of 1.70 Å was used in all calculations to represent methane. Monte Carlo sampling was also performed to calculate accessible surfaces (for example, Fig. 2) and pore volumes.

**Methane and hydrogen capacity calculations.** Methane adsorption was predicted for each structure at a temperature of 298 K and pressures of 5.8 bar, 65 bar and 100 bar. Hydrogen uptakes were predicted for each structure at 100 bar/77 K and 5 bar/160 K. All of the adsorption predictions were performed using grand-canonical Monte Carlo (GCMC) simulations involving a 50,000-cycle equilibration period and a 50,000-cycle production run. In hydrogen adsorption simulations, quantum diffraction effects of hydrogen at low temperatures were accounted for using the so-called Feynman–Hibbs quantum effective potentials<sup>40</sup>.

**Heats of adsorption.** The isosteric heat of adsorption ( $Q_{st}$ ) for methane at zero coverage (infinite dilution) in each adsorbent at 298 K was obtained from energy differences computed in the canonical (NVT) ensemble, via the Widom test particle method:

$$Q_{st} = RT - (\langle U_1 \rangle_1 - \langle U_0 \rangle_0 - \langle U_g \rangle)$$

where  $R$  is the gas constant,  $U_N$  is the total energy of the host with  $N \in \{0, 1\}$  guest molecules present,  $\langle \dots \rangle_X$  is the ensemble average at constant volume  $V$ , temperature  $T$ , and with  $X$  guest molecules, and  $\langle U_g \rangle$  is the average energy of an isolated guest molecule *in vacuo* at the same temperature.

**Leading-edge collation.** Because molecular dynamics and separation simulations are computationally expensive, it was necessary to limit the number of structures that were analysed. This was achieved by creating the leading edge for each CSP dataset. A script was written to separate all of the structures into bins of width 0.05 g cm<sup>-3</sup>, and any structures up to 10 kJ mol<sup>-1</sup> above the lowest-energy structure in each bin were extracted and tabulated to create the leading edge. Hydrogen-bond analysis, hydrocarbon separation simulations and thermal stability simulations were carried out for a subset of T2 leading-edge structures that were determined to be distinct on the basis of their simulated powder diffraction profiles.

**Hydrocarbon separation simulations.** Configurational-bias Monte Carlo (CBMC) simulations were performed for leading-edge structures for T2 to predict the separation performance of structures in the predicted ensemble for three hydrocarbon separations: propane/methane, xylene isomers and propane/propane. Simulations were conducted for equimolar mixtures at a total pressure of 1 bar (298 K). The CBMC simulations performed included a 300,000-cycle equilibration period and a 300,000-cycle production run.

**Hydrogen-bonding analysis.** Hydrogen bonds were identified using PLATON<sup>41</sup>. The frequency of hydrogen-bond patterns was determined using the ‘Motif search’ functionality in Mercury<sup>42</sup>.

**Approximate computational costs.** Approximately three weeks (in real time) and 400,000 CPU hours, with over 200 CPUs devoted uniquely to these calculations, were needed to perform CSP for T2E—the largest molecule investigated here—resulting in the final set of structures and associated lattice energies. This timescale is much shorter than the time required to develop a synthesis for T2E and to isolate diffractable single crystals (around three months). The computational cost of CSP for these molecules scales approximately as the square of the number of atoms in the molecule, so was correspondingly smaller for the other molecules studied.

Estimates, in real time, for various property calculations, using 256 CPUs are:

- Geometry-based analysis: 1 day, 6,000 crystal structures
- Zero-coverage isosteric heats of adsorption: 1 week, 10,000 crystal structures
- Methane capacity: 6 weeks, 10,000 crystal structures
- Hydrogen capacity: 2 weeks, 150 crystal structures
- Propane/methane or propane/propane separation: 3 weeks, 150 crystal structures
- Xylene separation: 4 weeks, 150 crystal structures

**Isotherm simulations.** Adsorption isotherms were simulated using the GCMC method with the RASPA code<sup>43</sup>. All systems were modelled using force fields, and the corresponding CSP structures for the experimental T2 phases were used.

**Thermal stability simulations.** Molecular dynamics simulations were performed in DL\_POLY4.07 (ref. 44) for structures selected from the low-energy leading edge of the CSP energy landscapes for T0, T1 and T2, and on select structures for P2M and T2E. A force field was generated, based on equilibrium bond lengths and angles derived using Gaussian09 (M06-2X/6-31G\*\*), with OPLS2005 atom

types, bond, angle and torsional terms, assigned using DL\_FIELD<sup>45</sup>. Constraints on bond lengths were applied to all bonds to conserve their equilibrium values. MULFIT<sup>46,47</sup> was used to fit atomic charges to the molecular electrostatic potential generated by the B3LYP/6-31G\*\*-distributed multipole analysis atomic multipoles. van der Waals interactions were modelled using the revised Williams intermolecular potential<sup>37</sup> with a long-range interaction cut-off of 15 Å and an Ewald precision of  $1 \times 10^{-6}$ . Each structure was expanded to a supercell such that the minimum cell dimension was greater than 65 Å and then relaxed using 10,000 0.001-ps time-steps in a zero-temperature minimization in the NVT ensemble. This relaxation used the Berendsen thermostat (with a relaxation time of 0.1 ps) and the Velocity Verlet integrator. The relaxed cell was used as the input for an extended 500-ps simulation in the NPT ensemble using the Berendsen thermostat and barostat (with relaxation times of 0.1 ps and 10.0 ps, respectively) at a temperature of 300 K and a pressure of 1 atm, with a time-step of 0.0025 ps. Additional high-temperature molecular dynamics simulations on T2- $\alpha$ , T2- $\beta$ , T2- $\gamma$  and T2- $\delta$  were performed using a temperature of 500 K. For P2M, CHELPG-derived point-charges from a Gaussian09 B3LYP/6-31G\*\* calculation were used in place of MULFIT-generated charges.

**Solvent stabilization calculations.** Monte Carlo simulations were performed to insert acetone, DMSO and DMAc into selected porous structures using towhee-7.1.0<sup>48</sup> in the NVT ensemble, using a temperature of 5,000 K to enhance sampling of solvent positions. The resulting structures were lattice-energy-minimized. A force field was generated, based on DFT equilibrium bond lengths and angles, with UFF-derived force constants, dihedral terms and improper torsions, and with CHELPG-derived point-charges. Atom–atom repulsion and dispersion interactions were modelled using the W99 intermolecular potential with the addition of isotropic parameters for sulfur. Simulations of solvent insertion into selected predicted structures were performed in fixed molecule:N ratios (where N is the number of solvent molecules per molecule of the host structure, with values in the range 1–14), with the solvent positions initiated from a template structure placed on a cubic grid. Each simulation was run for 20,000 Monte Carlo cycles with sampling performed every 10 cycles. An interaction cut-off of half the minimum cell dimension was used. A selection of Monte Carlo moves were allowed (intra-box deletion/insertion, partial configuration bias regrowth, single-atom translations, centre-of-mass rotations and translations) using CBMC. The resulting configurations were processed and the gas-phase-optimized molecular geometry overlaid back onto the simulation structures. Rigid molecule lattice energy minimization was attempted on all valid structures using DMACRYS<sup>36</sup>, with electrostatic and atom–atom repulsion and dispersion interactions modelled as described in Methods section ‘Crystal structure prediction (CSP)’. Each trial structure was subject to a three-step optimization procedure. In the first stage, electrostatic interactions were modelled using point charges derived from a B3LYP/6-311G\*\* calculation. In the subsequent steps, electrostatics were modelling using an atomic multipole description up to the hexadecapole on all atoms. Structures in which the host framework became distorted were removed. The distortion was monitored by comparing the host frameworks to the starting solvent-free crystal structure using the COMPACK algorithm<sup>38</sup> with an overlay based on 200-molecule clusters; only structures with r.m.s.d.<sub>200</sub> < 0.8 Å were retained.

Comparing the energies of solvated crystal structures with the low-energy, non-solvated predicted structures requires a correction for the energetic cost of removing solvent molecules from their pure phase. For this, Monte Carlo simulations were performed on solvent boxes containing 50 molecules, in the NVT ensemble at 300 K. The simulation cell lengths were taken such that the density matched the experimental density of each solvent. Simulations consisted of 100,000 Monte Carlo cycles, with Monte Carlo centre-of-mass translations and rotations using a cut-off of 5.5 Å. Configurations from the last 50,000 steps were sampled every 50 Monte Carlo cycles and subject to constant-volume lattice energy minimizations using DMACRYS, with parameters as previously described and an interaction cut-off of 15 Å.

The stability of the solvated porous crystal structures, relative to the dense, global-minimum CSP structure and pure solvent,  $E_{\text{rel,solvated}}$ , is calculated as

$$E_{\text{rel,solvated}} = E_{\text{latt(host+solvent)}} + N \left( -E_{\text{solvent}} + \frac{3}{2}RT \right)$$

where  $E_{\text{latt(host+solvent)}}$  is the calculated lattice energy of the solvated CSP structure containing N solvent molecules per host molecule, and  $E_{\text{solvent}}$  is the calculated energy of the pure solvent. The 3RT/2 term (calculated at  $T = 300$  K) is an equipartition estimate of the loss in internal energy of solvent molecules, assuming free rotation in the liquid phase and loss of solvent molecule rotational freedom in the crystal.

For all solvents (acetone, DMAc and DMSO), there are porous structures (T2- $\alpha$ , T2- $\beta$  and T2- $\gamma$ ) that have lower calculated energies than the lowest-energy unsolvated T2 crystal structure. The presence of solvent sufficiently stabilizes these structures such that they are energetically more favourable than the densely packed phase, T2-D, when fully solvated. In all cases, the porous structures (T2- $\alpha$ , T2- $\beta$  and T2- $\gamma$ ) are the most favoured forms. The more densely packed porous polymorphs (T2- $\delta$ , T2-A and T2-E) show negligible solvent stabilization, with the exception of T2- $\delta$  with DMSO, which displays a similar extent of stabilization to the more porous structures. The voids of T2- $\delta$  are too small to accommodate DMAc. For T2- $\gamma$ , the maximum in the solvent stabilization energy is seen at a ratio of 8 DMAc per T2 molecule, in good agreement with the experimentally determined ratio of 7.79.

Full details of all computational methods are given in Supplementary Information.

**Synthesis and measurements.** Full details of synthesis, crystallization and sublimation methods, solvent exchange protocols, X-ray structure determination, and gas sorption analysis are given in Supplementary Information. T1 (ref. 26) and T2 (ref. 22) were synthesized as described previously. T2E was synthesized by new a route that is described in Supplementary Information.

**Data availability.** The computational data that support the findings of this study, including all of the predicted crystal structures and properties, are available at <http://eprints.soton.ac.uk/404749/>. The experimentally determined crystal structures, including structure factors, have been deposited as CIFs with the Cambridge Crystallographic Data Centre (entries 1478355–1478365), available free of charge at [http://www.ccdc.cam.ac.uk/data\\_request/cif](http://www.ccdc.cam.ac.uk/data_request/cif).

34. Materials Studio v.6.1.200, <http://accelrys.com/products/collaborative-science/biovia-materials-studio/> (Accelrys Software Inc., 2012).
35. Frisch, M. J. *et al.* Gaussian 09 Revision D.01 (Gaussian Inc., 2013).
36. Price, S. L. *et al.* Modelling organic crystal structures using distributed multipole and polarizability-based model intermolecular potentials. *Phys. Chem. Chem. Phys.* **12**, 8478–8490 (2010).
37. Pyzer-Knapp, E. O., Thompson, H. P. G. & Day, G. M. An optimized intermolecular force field for hydrogen-bonded organic molecular crystals using atomic multipole electrostatics. *Acta Crystallogr. B* **72**, 477–487 (2016).
38. Chisholm, J. A. & Motherwell, S. COMPACK: a program for identifying crystal structure similarity using distances. *J. Appl. Cryst.* **38**, 228–231 (2005).
39. Willems, T. F., Rycroft, C., Kazi, M., Meza, J. C. & Haranczyk, M. Algorithms and tools for high-throughput geometry-based analysis of crystalline porous materials. *Micropor. Mesopor. Mater.* **149**, 134–141 (2012).
40. Sesé, L. M. Feynman-Hibbs potentials and path integrals for quantum Lennard-Jones systems: theory and Monte Carlo simulations. *Mol. Phys.* **85**, 931–947 (1995).
41. Spek, A. L. Structure validation in chemical crystallography. *Acta Crystallogr. D* **65**, 148–155 (2009).
42. Macrae, C. F. *et al.* Mercury CSD 2.0 – new features for the visualization and investigation of crystal structures. *J. Appl. Cryst.* **41**, 466–470 (2008).
43. Dubbeldam, D., Calero, S., Ellis, D. E. & Snurr, R. Q. RASPA: molecular simulation software for adsorption and diffusion in flexible nanoporous materials. *Mol. Simul.* **42**, 81–101 (2016).
44. Todorov, I. T., Smith, W., Trachenko, K. & Dove, M. T. DL\_POLY\_3: new dimensions in molecular dynamics simulations via massive parallelism. *J. Mater. Chem.* **16**, 1911–1918 (2006).
45. Yong, C. W. Descriptions and implementations of DL\_F notation: a natural chemical expression system of atom types for molecular simulations. *J. Chem. Inf. Model.* **56**, 1405–1409 (2016).
46. Winn, P. J., Ferenczy, G. G. & Reynolds, C. A. Towards improved force fields. 1. Multipole-derived atomic charges. *J. Phys. Chem. A* **101**, 5437–5445 (1997).
47. Ferenczy, G. G., Winn, P. J. & Reynolds, C. A. Towards improved force fields. 2. Effective distributed multipoles. *J. Phys. Chem. A* **101**, 5446–5455 (1997).
48. Martin, M. G. MCCCS Towhee: a tool for Monte Carlo molecular simulation. *Mol. Simul.* **39**, 1212–1222 (2013).