



Cite this: *Chem. Sci.*, 2022, **13**, 13588

All publication charges for this article have been paid for by the Royal Society of Chemistry

Received 11th August 2022
Accepted 11th October 2022

DOI: 10.1039/d2sc04511g
rsc.li/chemical-science

1 Introduction

How molecules pack has vital ramifications for their application in areas such as optoelectronics,¹ catalysis,² and drug delivery.³ However, the lack of strong bonds has resulted in the long-standing challenge of how to design molecules to manipulate crystal formation for specific packings with desired properties.⁴ The prediction of solid-state packing is non-trivial and is more complex for molecular materials than their inorganic counterparts. Reticular design rules implemented in extended framework formation such as metal- and covalent-organic frameworks fail when applied to organic molecules due to their packing being determined by weak dispersion forces.

Currently, state-of-the-art computational efforts focus on crystal structure prediction (CSP) which, although successful,^{5–7} is computationally expensive and must be employed on a case-by-case basis. This results in difficulty formulating design rules between structural motifs, chemical functionality, and their packing behaviour. Moreover, weak intermolecular interactions result in a high number of possible low-energy polymorphs and, coupled with the high computational cost of these calculations,⁸

Department of Chemistry, Imperial College London, Molecular Sciences Research Hub, White City Campus, Wood Lane, London, W12 0BZ, UK. E-mail: e.wolpert@imperial.ac.uk; k.jelfs@imperial.ac.uk; Tel: +44 (0)20759 43438

† Electronic supplementary information (ESI) available: Additional computational details, results, figures, structure files and movies of simulations. See DOI: <https://doi.org/10.1039/d2sc04511g>

Coarse-grained modelling to predict the packing of porous organic cages†

Emma H. Wolpert * and Kim E. Jelfs *

How molecules pack has vital ramifications for their applications as functional molecular materials. Small changes in a molecule's functionality can lead to large, non-intuitive, changes in their global solid-state packing, resulting in difficulty in targeted design. Predicting the crystal structure of organic molecules from only their molecular structure is a well-known problem plaguing crystal engineering. Although relevant to the properties of many organic molecules, the packing behaviour of modular porous materials, such as porous organic cages (POCs), greatly impacts the properties of the material. We present a novel way of predicting the solid-state phase behaviour of POCs by using a simplistic model containing the dominant degrees of freedom driving crystalline phase formation. We employ coarse-grained simulations to systematically study how chemical functionality of pseudo-octahedral cages can be used to manipulate the solid-state phase formation of POCs. Our results support those of experimentally reported structures, showing that for cages which pack via their windows forming a porous network, only one phase is formed, whereas when cages pack via their windows and arenes, the phase behaviour is more complex. While presenting a lower computational cost route for predicting molecular crystal packing, coarse-grained models also allow for the development of design rules which we start to formulate through our results.

this restricts the ability to use high-throughput approaches.⁵ Additionally, it is even more computationally expensive to apply CSP to multicomponent systems,⁹ where it has been shown that combining different molecules can allow property tuning in a way that is not possible with extended framework materials.¹⁰ Therefore, a novel method of predicting the packing of molecular materials must be introduced in order to simultaneously improve our understanding of molecular crystals, reduce computational efforts, and elucidate design rules for targeted phase formation.

In this paper, we develop a novel way of predicting the molecular solid-state based on a simplistic, coarse-grained model containing the dominant degrees of freedom driving crystalline phase formation. We focus on determining the packing behaviour of a subset of porous organic cages (POCs) formed through a condensation reaction between a trialdehyde and diamine in a [4 + 6] cycloimination, creating discrete organic molecules which contain a permanent internal cavity (Fig. 1). Although the cages formed are pseudo-octahedral geometrically, the molecules have pseudo-tetrahedral symmetry due to the different chemical features of the cages, where each face of the octahedra is either a cavity, known as a window, or contains an arene (Fig. 1(c)). The porous nature of POCs leads to a variety of potential applications, including in catalysis,¹¹ sensing,¹² encapsulation¹³ and molecules separations.¹⁴ Despite significant interest in POCs, their utilisation in industrial applications is hindered by the lack of understanding

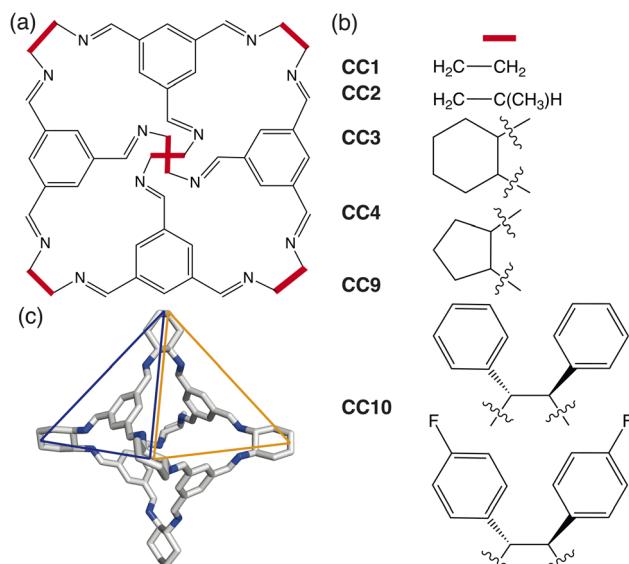


Fig. 1 (a) Basic structure of a POC formed from the condensation of a trialdehyde and diamine in a [4 + 6] cycloimination where the vertices have different functional groups depending on the diamine used. (b) Examples of the different chemical functionalities of common octahedral POCs reported within the literature. (c) An example of the 3D structure of the octahedral POC CC3 which has pseudo-tetrahedral symmetry due to the different chemical features of the octahedral cage where there is either an arene (blue) or a window (orange) at the centre of the octahedral face. The carbon/nitrogen atoms are shown in white/blue and the hydrogen atoms are omitted for clarity.

of how to control their assembly,¹⁵ both on a molecular level and in the solid-state. As techniques for predicting the molecular assembly of cage molecules become more refined,^{16,17} there still remains the question of how to predict and control the assembly in the solid-state.

As the internal cavity of the cages can only be accessed through their windows, changes in their solid-state phase behaviour can have a significant affect on the pore network in the material. Small changes in the chemical functionality of POCs can lead to non-intuitive changes in the global solid-state packing due to the fine balance between the weak dispersion forces (Fig. 2). These weak interactions are responsible for many of the advantages of POCs over extended frameworks such as their solution processability¹⁸ and the potential to control their solid-state assembly by manipulating the functionality of the cages or solvent used.^{19,20} However, facile prediction and control of the packing behaviour of novel cages is currently intractable as there remains a lack of understanding of how chemical functionality and solvent interplay in the formation of molecular solid-state materials. This complexity in predicting the packing behaviour of molecular crystals makes targeted design difficult.²¹

To implement the research presented in this paper, we use a general-purpose particle simulation toolkit, HOOMD-blue,²² to employ hard particle Monte Carlo (HPMC) simulations with directional interactions through “patches” (Fig. 3(a)). HPMC simulations as implemented within HOOMD-blue are often

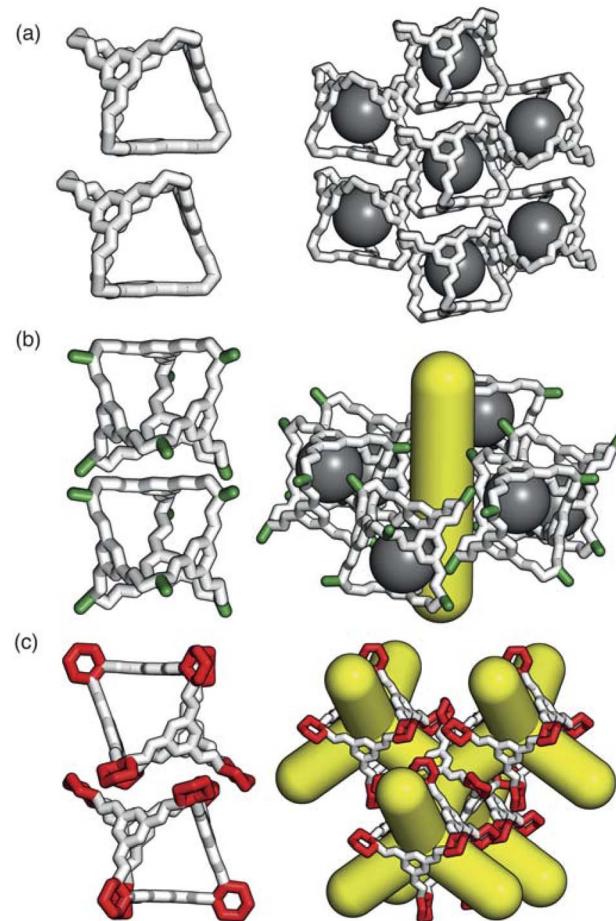


Fig. 2 The different functionalities of the vertices of the cages lead to vastly different packing arrangements, affecting the pore networks of the material as highlighted through (a) the 0D pore network formed by CC1 α' , (b) the 1D pore network formed by CC2 due to the methyl functionality (green), and (c) the 3D pore network formed by CC3 α due to the cyclohexyl groups (red). Isolated voids formed by window-to-arene packings are shown in grey and the pore networks are shown in yellow. Both the carbon and nitrogen atoms are shown in white and the hydrogen atoms are omitted for clarity.

used to investigate the self assembly of structures formed by packing non-interacting polyhedra.^{23,24} Through these simulations, Glotzer and coworkers have shown that geometry alone can direct complex structural formation through an “entropic bond” governed by the polyhedral shape. The entropic bond is a purely statistical phenomena which explains that densely packed structures form from non-interacting particles due to the system maximising the number of microstates available. While the studies focused on colloidal, non-interacting particles, this has strong implications for packing in molecular materials. The entropic bonds are weak,²⁵ on the order of a few $k_B T$,²⁶ which means enthalpic interactions dominate in most materials. However, the packing of organic molecules is dictated by weak intermolecular interactions, suggesting that the entropy gained from increasing the number of microstates available may influence the phase behaviour.



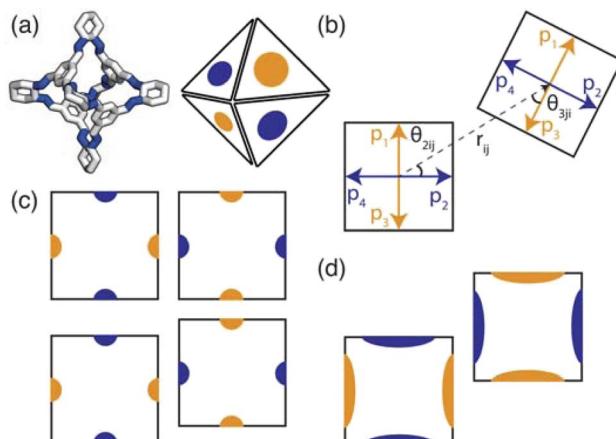


Fig. 3 (a) How the POC CC3 relates to the coarse-grained model where the orange and blue patches represent the different chemical nature of the POC's facets (*i.e.* window or arene face). Here the carbon/nitrogen atoms are shown in white/blue and the hydrogens atoms are omitted for clarity. (b–d) Two-dimensional representation of the patchy particle model used. (b) A representation of the geometry of the interaction between two particles for a window-to-arene simulation. Here each of the particles have four patches described by the patch vectors p_α and p_β where α and β refer to patch $\alpha(\beta)$ on atom $i(j)$. In the particular case shown in the figure, patch 2 on particle i interacts with patch 3 on particle j as they are the closest patches to the interparticle vector, r_{ij} . To relate the patches to our model we have coloured the patch vectors according to the types of patches. (c) Top: an example of the lowest energy orientation of an interaction with preferred alignment of opposing patches. Bottom: the maximum displacement between the two particles that is likely to occur with a small patch width (small σ_{ang}). (d) The maximum displacement between the two particles that is likely to occur with a large patch width (large σ_{ang}) leading to the formation of more ordered structures when σ_{ang} is small than for larger values of σ_{ang} .

Within HOOMD-blue, patches can be added to hard particles in order to mimic directional interactions,²⁷ which has been successfully applied to molecular materials as seen in the investigation of the self assembly of π -conjugated optoelectronic peptides.²⁸ The benefit of using HOOMD-blue is the ease in using user-defined potentials as required for the introduction of coarse-grained models. This was demonstrated by Mansbach *et al.* who tracked the effects of core and side chain interaction strength and sterics on the morphology and kinetics of assembly using a coarse-grained model.²⁸ These models allow for the formation of design rules by developing an understanding of the interactions and geometric building blocks required for targeted phase behaviour.²⁹

Here we use HPMC simulations in conjunction with patchy particle models to investigate the effect of the combination of building block geometry and directional interactions on the solid-state phase behaviour of POCs. The basic framework rigidity of POCs means that geometrically, the molecules can be considered as polyhedra. For the POCs considered in this paper (Fig. 1), that means we can represent the molecular cages as hard octahedra for the HPMC simulations. The directional interactions are introduced through patches on each of the facets, where the different colour patches represent the

reduction of symmetry of the octahedra due to the chemical makeup of the cages (*i.e.* the alternation of facets between windows or arenes) (Fig. 3(a)). Due to the simplicity of the model, this procedure drastically reduces computational efforts compared to CSP and facilitates the development of design rules required for targeted phase formation in the future. Although a simplistic model, we show that by manipulating the parameters of our coarse-grained model, we can reproduce the phase space spanned by the majority of octahedral POCs present in the Cambridge Structural Database.³⁰ Finally, our results highlight new phases which are unreported within the literature. While the focus of this paper is on POCs, many of the motifs studied here are seen in other areas of molecular chemistry such as metal–organic cages.^{31,32} Thus our results may have wider reaching applications to predict and rationalise the solid-state behaviour of a range of molecular materials.

2 Methods

2.1 Hard particle Monte Carlo simulations

The Monte Carlo simulations were performed using HPMC,³³ a plugin to the HOOMD-blue simulation toolkit.²² The Metropolis Monte Carlo algorithm in HPMC works by performing moves of either a translation or rotation of the particles. If the move leads to no overlap between any hard particles, it is accepted, whereas if an overlap between neighbouring particles occurs, then the move is rejected. Therefore the simulations are often deemed to be temperature independent as each move either leads to no change in energy, or an infinite increase.³³ However, in our case we applied interactions between neighbouring particles which caused the simulation results to be affected by temperature, modifying the acceptance rules based on the energetic penalty of a given move determined by the interaction potential.

2.2 Patchy particle potential

To determine the phase behaviour spanned by pseudo-octahedral cages, we implemented HPMC simulations on hard octahedra with different directional interactions introduced through patches on the polyhedra. The form of these interactions is based on the patchy particle potential described in ref. 34, which decorates the particle with a number of sticky patches, reducing its symmetry. As shown in Fig. 3(a), a patch was added to the centre of each facet (*i.e.* window or arene face), where the different colour patches represent the reduction of symmetry of the octahedra due to the chemical makeup of the cages (*i.e.* the alternation of facets between windows or arenes). To include the likely degrees of freedom that occur between POCs, the patches on the hard octahedra were introduced *via* an interaction term, V_{ij} , which contains three components; an attractive Lennard-Jones potential V_{LJ} , an angular modulation term V_{ang} , and a torsional modulation term V_{tor} :

$$V_{ij} = \begin{cases} V_{\text{LJ}} \times V_{\text{ang}} \times V_{\text{tor}} & r < r_{\text{cut}} \\ 0 & r \geq r_{\text{cut}} \end{cases} \quad (1)$$



$$V_{\text{LJ}} = 4J \left[\left(\frac{\sigma_{\text{LJ}}}{r} \right)^{12} - \left(\frac{\sigma_{\text{LJ}}}{r} \right)^6 \right] \quad (2)$$

$$V_{\text{ang}} = \exp \left(-\frac{\theta_{\alpha\beta}^2}{2\sigma_{\text{ang}}^2} \right) \exp \left(-\frac{\theta_{\beta\gamma}^2}{2\sigma_{\text{ang}}^2} \right) \quad (3)$$

$$V_{\text{tor}} = \exp \left(-\frac{1}{2\sigma_{\text{tor}}^2} \left[\min \left(\phi_{\alpha\beta} - \phi_{\alpha\beta}^{\text{offset}} \right) \right]^2 \right) \quad (4)$$

where α and β are patches on particles i and j respectively, Fig. 3(b). Throughout the simulations, σ_{LJ} , which is normally the measure of the diameter of the interacting particles, was set to 0.6 \AA . As each octahedra had unit length, this value is slightly larger than the minimum distance required for the two hard octahedra packed along the facets (0.5 \AA). This is to allow for a small gap between the hard octahedra, which is likely to exist between the analogous two neighbouring cages due to their van der Waals radii. The cutoff distance, r_{cut} , is set to $1.8\sigma_{\text{LJ}}$ so that the patches interact with nearest neighbour polyhedra only as r is a measure of the distance between the centres of the two neighbouring particles.

The angular term, V_{ang} , is a measure of how directly the patches on adjacent particles point at each other. Here, $\theta_{\alpha\beta}$ is the angle between the patch vector and the vector between the two neighbouring particles, \mathbf{r}_{ij} , the magnitude of which is given by r (Fig. 3(b)). σ_{ang} dictates the energetic penalty of particles deviating from being perfectly aligned, such that a larger σ_{ang} value relates to worse alignment of the two octahedral facets, Fig. 3(c and d). σ_{ang} is referred to as the width of the patch as changing σ_{ang} narrows or widens the Gaussian function that describes the patchy interaction, effectively changing the angular width of the patch where the particle interacts through. Through this work, we examine how changing σ_{ang} between 0.1 and 1.0 effects the solid-state phase behaviour of the interacting octahedral particles. We expect that this parameter encapsulates the effects of different substituents on the vertices as the substituents would likely change the directionality of the interaction due to steric effects, which is controlled by the patch width.

The torsional term, V_{tor} , describes the variation in the potential with rotation of the particle about the interparticle vector \mathbf{r}_{ij} . $\phi_{\alpha\beta}^{\text{offset}}$ is the preferred torsional angle between patches α and β , whereas $\phi_{\alpha\beta}$ is the actual torsional angle between patches α and β , such that there is an energy minimum where the two angles are the same. The preferred torsional angles used for each simulation are summarised in Table S3† and were chosen based on the likely lowest energy orientations of window-to-window and window-to-arene packings due to steric interactions, *i.e.* window-to-window prefers to be anti-aligned and window-to-arene prefers to be aligned. Similarly to σ_{ang} , σ_{tor} dictates the energetic penalty of the particles deviation from the perfect torsional angle. In our simulations, we found that the results did not vary with σ_{tor} , and as such, much like other studies on patchy particles,³⁵ we kept a constant ratio between σ_{tor} and σ_{ang} such that $\sigma_{\text{tor}} = 2\sigma_{\text{ang}}$. This is a physically reasonable approximation as, from chemical considerations, it is likely that σ_{tor} and σ_{ang} are coupled such that the less the

interaction cares about the alignment of the two molecules, the less likely it has a strong torsional preference *i.e.* the interaction overall will be less directional.

J is the measure of the interaction strength between the two octahedra. Varying σ_{ang} in our model causes the clusters to be formed at different temperatures for the same value of J . This is because a smaller value of σ_{ang} narrows the Gaussian function in eqn (3), resulting in a smaller value of V_{ang} . As the absolute energy scale is not important to the results of this study, for simplicity a value of J was chosen for each simulation such that the transition temperature T_t occurred when $k_{\text{B}}T_t \approx 1$. The interaction parameters used are summarised in Tables S1 and S2.† We note that for a given value of σ_{ang} , J is larger for the window-to-window simulations than for the window-to-arene simulations. This is because each cage only has four windows, resulting in a maximum of four stabilising window-to-window “interactions” for each cage, whereas in the window-to-arene simulations, there can be up to eight stabilising interactions as the cages can be stabilised through interactions through the arene as well. This means that J must be larger to result in a transition occurring at the same temperature.

2.3 Simulation details

The simulations were slowly cooled in the NVT ensemble over the transition point, $k_{\text{B}}T \approx 1$, starting at $k_{\text{B}}T = 1.35$, to ensure full equilibration at steps of $T_{i+1} = T_i \times 0.95$ where $k_{\text{B}} = 0.00831 \text{ kJ mol}^{-1} \text{ K}^{-1}$. Each simulation was run for 24 hours using 64 cores on ARCHER2 UK National Supercomputing Service, with the exception of the window-to-window simulations where $\sigma_{\text{ang}} = 0.4$ and 0.8 , which were run using Imperial College London’s Research Computing Service until $k_{\text{B}}T = 0.51$. During the simulations, at least nine temperatures were sampled with 10^8 timesteps—where one trial move is applied to a random number of particles in each cell during one timestep—taken at every temperature. The time taken to complete a timestep depends on the number of interactions being calculated, therefore once the polyhedra form a cluster, the time taken to complete each timestep increases. Consequently, due to the finite wall time, some simulations, particularly those at low σ_{ang} , have a fewer number of temperature points simulated. For these cases, once the simulations were terminated, the resulting configuration is considered to be representative of the phase behaviour at all lower temperatures. This assumption is corroborated by results from the window-to-window simulations where $\sigma_{\text{ang}} = 0.4$ and 0.8 , as both the ordered state at $\sigma_{\text{ang}} = 0.4$ and amorphous phase at $\sigma_{\text{ang}} = 0.8$ persisted down to $k_{\text{B}}T = 0.51$.

The simulations were performed on 512 particles of unit length in cubic boxes with a box length of 12 \AA and periodic boundary conditions. As the potentials are attractive and the systems were initialised at low density, the simulations formed compact clusters with no mechanical stresses and structure defects that are often seen in bulk simulations. To determine if there were any finite size effects, some simulations were repeated with 4096 particles and evidenced the same results.



2.4 Phase determination

For a structure to be considered a representative phase at any given simulation condition, we imposed a minimum cluster limit of 50 polyhedra before undergoing any further analysis. To ensure we were sampling from the bulk structure, the outer layer of the cluster is then removed and clusters that still contained over 50 polyhedra underwent subsequent analysis. This limit was chosen as the clusters were roughly spherical and as such, a cluster of at least 50 polyhedra would encapsulate the main structural features up to second-nearest-neighbours. In reality, most clusters had a lot more than 50 polyhedra and in fact only two of the configurations from the simulations had fewer than 100 polyhedra (Section S2†).

To determine the different structural phases formed, we calculated and compared the radial distribution functions (RDFs) of the centres of each octahedra within the clusters. Although this removed the orientational behaviour of the different phases, structures with similar RDFs were visually inspected to ensure that the orientational behaviour led to no meaningful differences between the structures. For structures at low σ_{ang} , we were able to determine the space group of the clusters by abstracting a unit cell and using FINDSYM,^{36,37} a program used to identify the space group of a crystal, on a coarse-grained version of the structure and converted the solved structure back onto the equivalent structure formed with the cages. An overview of this process is given in Section S4.†

For structures at higher σ_{ang} , as they are inherently more disordered due to the worse alignment of the cages (Fig. 3(c and d)), our procedure struggled to find a space group. In these cases, we were still able to abstract a unit cell and instead visually compared our structures to those found in the literature. An overview of which ones were solved using FINDSYM and which ones required visual inspection is given in Table S6† along with representative configurations of each visually solved phase (Fig. S3†). We then took the most prevalent ordered phases from the simulations and used them to colour the phase diagram by comparing the similarity of the RDF from the simulated data to the RDF of the solved phase using dynamic time warping, details of which are in Section S5.†

3 Results and discussion

Due to the two different types of patches separately representing windows and arenes, we ran simulations looking at the three possible packing scenarios; (i) when interactions between only one of the types of patches is favoured *i.e.* window-to-window or arene-to-arene (Fig. 4(a and b)), (ii) when interactions between patches of the same type are favoured *i.e.* window-to-window and arene-to-arene, and (iii) when interactions between patches of different types are favoured *i.e.* window-to-arene (Fig. 4(c)). For each of these conditions, we varied the temperature and patch width, σ_{ang} , of our model to fully explore the phase space spanned by patchy octahedra. Our results from these simulations are summarised below.

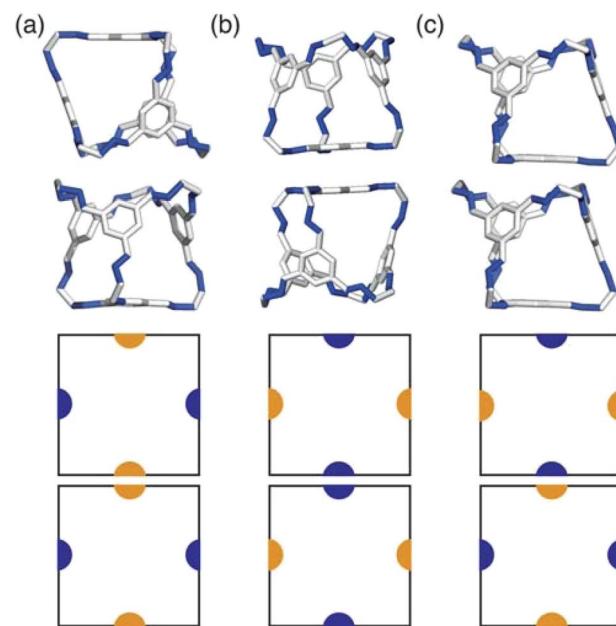


Fig. 4 Different packing types in the simulations: (a) window-to-window, (b) arene-to-arene, and (c) window-to-arene with the corresponding favourable patchy interactions shown below in 2D. The configurations are shown with the cage CC1 for simplicity, but in reality any of the cages could be substituted into these configurations.

3.1 Phase diagrams

3.1.1 Window-to-window simulations. The first simulations we performed were when only window-to-window (and by analogy arene-to-arene) packing is preferred. The low temperature RDFs only evidenced one ordered phase (Fig. S4 and S6†), a representative cluster of which is shown in Fig. 5(b) (see Fig. S7† for the analogous arene-to-arene configuration). To determine how the phase behaviour changes with temperature and also patch width, σ_{ang} , we compared the similarities of the RDFs at each temperature and σ_{ang} to the RDF of the ordered phase (Fig. S6(a)†), and coloured our phase diagram (Fig. 5(a)) such that the darker orange the region of the phase diagram, the more similar our simulated structure is to the ordered phase, as described in Section S5.† The phase diagram shows that the ordered structure is present in a large area of phase space, approximately $\sigma_{\text{ang}} = 0.1\text{--}0.6$. At higher values of σ_{ang} , an amorphous phase forms instead, an example of which is shown in Fig. S1(a).† When $\sigma_{\text{ang}} = 1.0$, although there is another phase that has an ordered RDF at low temperature (Fig. S4(j)†), on further inspection the orientation of the cages were disordered (Fig. S1(b)†). As such we deduced that this phase behaves more like a plastic crystal‡—where the structure has positional order but orientational disorder—a phase behaviour, unsurprisingly, unreported for porous organic cages, and as such was not considered an ordered phase. Therefore, these results indicate that cages that only interact through the same type of facets (window-to-window or arene-to-arene) produce a single ordered phase in the solid state.

3.1.2 Window-to-window and arene-to-arene simulations. Since the window-to-window simulations, and thus analogously



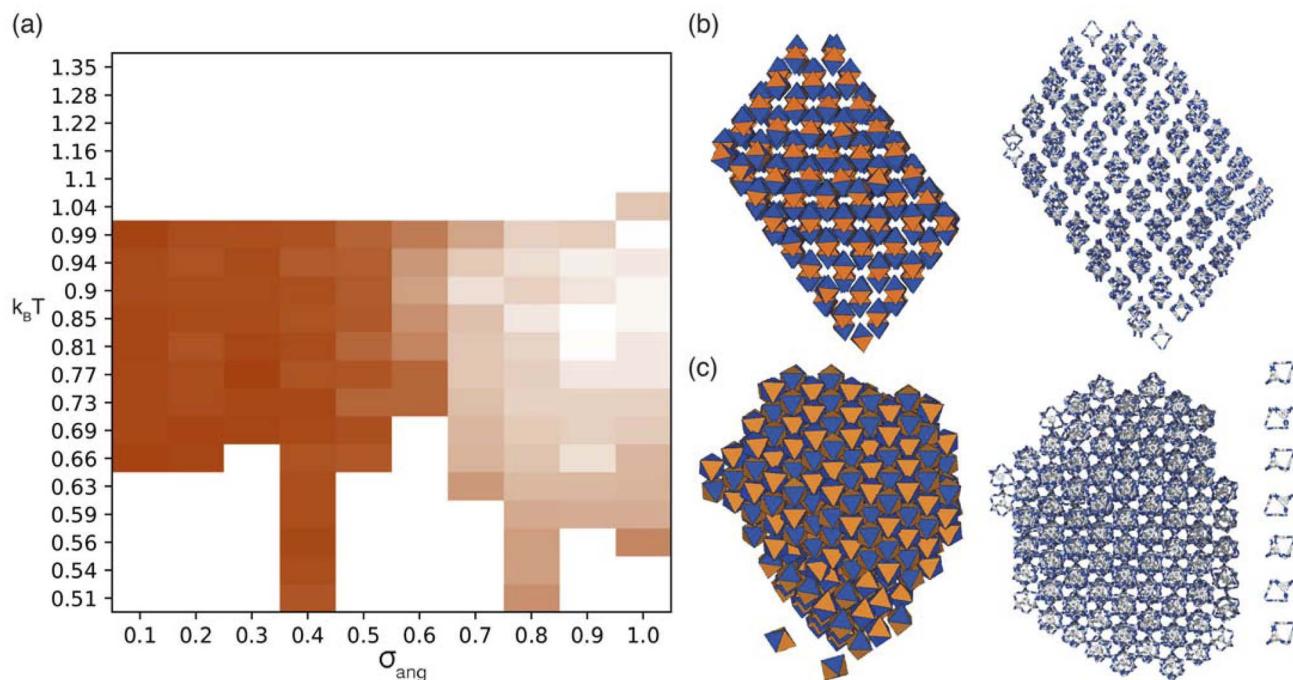


Fig. 5 (a) Phase diagram produced from simulations where patches of one type interact with each other where the darker orange phase is a measure of how similar the structure is to the ordered phase shown in (b). We note that uncoloured areas of the phase diagram at low temperatures are due to the simulations timing out (as explained in Section 2.3) and the phase behaviour at these temperatures is considered to be unchanged from the last simulated structure. Representative configurations of the ordered structures formed by simulations with (b) window-to-window packing and (c) window-to-window and arene-to-arene packing. Left is a representative configuration from the coarse-grained simulations, right is the mapping of the cluster onto the POC CC1 where the outer layer of the cluster has been removed. Here the different facets of the hard octahedra are coloured completely orange or blue to represent the window and arene functionality respectively. For (c) a representative chain of cages is shown on the right where there is alternating window-to-window and arene-to-arene packing.

the arene-to-arene simulations, only evidenced one ordered phase, to determine the phase behaviour where patches on both the windows and the arenes interact with the same patch type, we ran one targeted simulation where $\sigma_{\text{ang}} = 0.2$ for both the interactions representing window-to-window and arene-to-arene packing. This is equivalent to having interactions only between patches of the same type *e.g.* orange interacting with orange and blue with blue between neighbouring hard polyhedra (Fig. 3(a)). This simulation exhibited an ordered phase that contains alternating chains of window-to-window and arene-to-arene packing (Fig. 5(c)). To ensure we were not missing other potential phases by only sampling a narrow region of phase space, we sampled other combinations of σ_{ang} for both the interactions representing window-to-window and arene-to-arene packing between 0.1 and 0.6; these produced the same structure. The results from these simulations suggest that, similar to the window-to-window simulations, cages that interact between facets of the same type only form one ordered phase.

3.1.3 Window-to-arene simulations. For simulations where patches preferentially interact with those of the opposite type, *i.e.* window-to-arene packing, we found that when $\sigma_{\text{ang}} = 0.1$, the configurations did not order into clusters bigger than our limit of 50 polyhedra and as such we excluded these results. For simulations where $\sigma_{\text{ang}} \geq 0.2$, looking at the low temperature

RDFs, we found four distinct phases which changed as a function of increasing σ_{ang} (Fig. S5†). We chose the three most disparate phases (in terms of their RDF, Fig. S6(b-d)†) which became the red, green, and blue components of our phase diagram, Fig. 6(a). As with the window-to-window simulations, this phase diagram loses some information as we exclude the orientational behaviour of our structure due to the RDF data containing only the central positions of the cages, but again all phases were inspected visually to ensure we were not incorrectly characterising two orientationally different structures as the same phase. As expected, the phase behaviour at low temperatures of our phase diagram evidences the four different phases seen in the RDFs, shown through the different coloured sections at $\sigma_{\text{ang}} = 0.2\text{--}0.4$, 0.5, 0.6, and 0.7–1.0, Fig. 6(a). To ensure we were not missing any other phases due to the discrete values of σ_{ang} chosen, we ran simulations at intermediate values between $\sigma_{\text{ang}} = 0.4\text{--}0.7$ where the majority of the different phase behaviour lies, but this did not result in any new phases.

For the four different low temperature phases, the obvious question is how the phase behaviour differs with changing σ_{ang} . Although the RDFs at higher values of σ_{ang} suggest that the same phase persists from $\sigma_{\text{ang}} = 0.7\text{--}1.0$, on inspection of the structures, the cages orientations are disordered when $\sigma_{\text{ang}} \geq 0.8$ (Fig. S8(a)†), likely due to the large values of σ_{ang} and σ_{tor} leading to less strongly directional interactions towards the



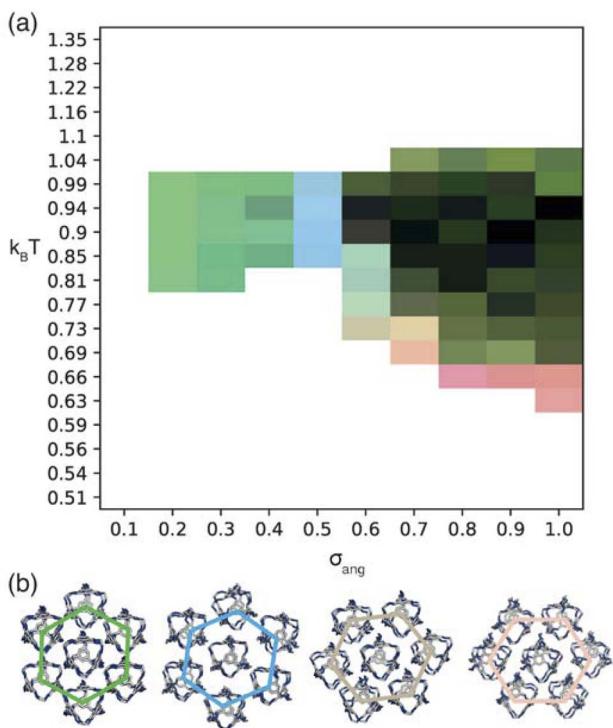


Fig. 6 (a) Phase diagram produced from simulations where patches of opposite types interact. Here the red, green, and blue components of the phase diagram are coloured based on the similarities of the phases to low temperature structures found at $\sigma_{\text{ang}} = 0.7$, 0.2–0.4, and 0.5 respectively as described in more detail in Section S5.[†] We note that uncoloured areas of the phase diagram at low temperatures are due to the simulations timing out (as explained in Section 2.3) and the phase behaviour at these temperatures is considered to be unchanged from the last simulated structure. (b) Representative configurations of the low temperature phases showing how that angle of rotation of the nearest shell of cages rotates as a function of σ_{ang} . From left to right the configurations represent the phases at $\sigma_{\text{ang}} = 0.2$ –0.4, 0.5, 0.6 and 0.7. The colour of the hexagons relate to the colour of the phase diagram where the phase is found.

facets. This means that although the phase behaviour looks similar, we are hesitant to call this the same phase and instead suggest that phases at higher σ_{ang} behave more like plastic crystals, much like in the case of high values of σ_{ang} for the window-to-window simulations.

For the four orientationally ordered phases between $\sigma_{\text{ang}} = 0.2$ –0.7, when looking down one of the high symmetry directions, each cage is surrounded by a hexagonal arrangement of cages. Interestingly, the angle of the hexagonal arrangement around the central cage compared to the orientation of the central cage changes as a function of σ_{ang} , Fig. 6(b). This suggests that the arrangement of the cages can be tuned almost linearly as a function of the patch width. To understand how this change in structure can affect the properties of the material, we examined the possible pore connectivity in each structure. We found that both the structures which form at small values of σ_{ang} may be able to host 1D pore channels due to their extrinsic porosity (Fig. S9[†]), whereas the two ordered structures at higher values of σ_{ang} can not. Therefore, understanding how

the structure can be tuned as a function of patch width is a valuable task, both from a crystal engineering point of view, and for future studies focused on finding novel cages with pore connectivity. However, even if cages form either of the two structures at small values of σ_{ang} , they might not exhibit porosity as if there were bulky groups on the vertices, the bulky groups could obstruct the connectivity by sitting within the pore channels.

On a molecular level, the manipulation of σ_{ang} may be achieved by changing the chemical functionality of the vertices or by using directional solvents. Bulkier groups on the vertices are likely to result in smaller values of σ_{ang} . Although bulkier groups might be thought to change the equilibrium distance between the cages (and thus σ_{LJ} in eqn (1)), results from force field calculations show that there is no change in the equilibrium bond distance for cages with vastly different steric behaviour *e.g.* CC1 and CC9 as shown in Section S9.[†] Instead, the reason why we expect that bulkier groups can be related to a decrease in σ_{ang} is that the bulkier the group, the greater the energetic penalty for neighbouring cages to be unaligned from a central position—a measure of the patch width—due to repulsive steric interactions. Similarly, solvents can be used to obtain more directional interactions, for example through hydrogen bonding between the pores, and therefore smaller values of σ_{ang} . We note that our simple model is unable to encapsulate all types of solvation effects, but instead use this as an example to highlight that some solvents may lead to more directional interactions along the $\langle 111 \rangle$ axes which could be used to navigate the phase behaviour seen here.

On changing temperature, the number of phase transitions seen in our simulations differs depending on the value of σ_{ang} . At low σ_{ang} (0.2–0.5), each phase undergoes one phase transition at $k_{\text{B}}T \approx 1$ to an ordered phase that persists down to low temperature. However, when $\sigma_{\text{ang}} \geq 0.6$, the structures undergo two phase transitions, one at $k_{\text{B}}T \approx 1$ and one at lower temperature, as evidenced by the different colours in the phase diagram (Fig. 6(a)). This first transition is to an amorphous cluster, a representative structure is included in Fig. S8(b),[†] and the second is to a positionally ordered phase at low temperature. Interestingly, the amorphous to ordered phase transition occurs for a wide range of σ_{ang} , spanning areas of phase space with two distinct ordered structures at low temperature—both when $\sigma_{\text{ang}} = 0.6$ and 0.7—as well as the plastic crystal phases where $\sigma_{\text{ang}} \geq 0.8$. The existence of two phases as a function of temperature suggests that the solid-state structure of cages which have a higher σ_{ang} can be controlled by temperature. Thus, cages that relate to this area of phase space may form both an amorphous and ordered structure based on the temperature of the synthesis, whereas molecules which exist at lower values of σ_{ang} may only evidence ordered structures no matter the synthesis temperature.

3.2 Comparison to known phases

Having determined the phase behaviour observed by changing parameters of our coarse-grained model, the question remains whether it is a suitable process for predicting the phase



behaviour of POCs. Given the simplicity of our model, we were particularly interested in whether we could recreate the packing behaviour spanned by enantiopure desolvated octahedral cages that do not contain any compositional disorder, of which there are five phases reported in the literature,⁴³ as summarised in Table 1 and Fig. 7. The specific focus on desolvated phases is a consequence of many of the functional properties of POCs relying on the interconnected pore network which are typically disrupted in solvated phases due to solvent molecules residing within the channels. So to determine if our process had predictive capabilities, we set out to establish whether these crystal structures were found within our simulated phase diagrams.

For cages where the dominant packing type is window-to-window, and a combination of window-to-window and arene-to-arene, only one solid-state structure is reported to exist (Fig. 7(a), window-to-window packing, and Fig. 7(b), window-to-window and arene-to-arene packing). This observation fits well with our simulations which also produce only one ordered structure from the two simulations, and on further analysis the structures produced in the simulations correspond to those seen experimentally (Fig. 8(a and b)).

When the dominant packing type is window-to-arene, three different crystal structures are reported to exist experimentally (Fig. 7(c–e)) which seemingly corroborates our window-to-arene results from the simulations that evidence a much richer phase diagram than for window-to-window. Yet, only two of the three experimentally known phases are found in our phase diagram, **CC9-R3** (at $\sigma_{\text{ang}} = 0.6$) (Fig. 8(c)) and **CC1 β'** (at $\sigma_{\text{ang}} = 0.7$) (Fig. 8(d)). The remaining phase reported experimentally does not exist within any of our phase diagrams. However, this is not a surprising result. Experimentally, this structure is formed with both cages **CC3** and **CC4** upon desolvation from another ordered structure. As such, it has been suggested that the crystal structure identified is formed due to a solvent templation effect resulting in a single-crystal-to-single-crystal transformation upon desolvation, something not taken into account by our model, and so it is not expected to be found in our simulations.^{44,46} This result highlights how solvent can be used to alter the polymorphic behaviour of the crystal and even change the preference of the dominant interaction between facets. In this case, the change in packing behaviour for **CC3** from **CC3 α** to **CC3 β** would likely decrease the porosity from the loss of the

three dimensional pore network, however, for **CC4**, *para*-xylene can direct the packing behaviour from window-to-arene to window-to-window as seen in the formation of **CC4 β** .⁴⁴

While our results show that we can predict most of the crystal structures known to be formed by ordered, enantiopure, desolvated octahedral cages, there still remains two ordered phases produced in our window-to-arene simulations that do not match up to any of the structures in Table 1. The first of these structures is a novel phase (when $\sigma_{\text{ang}} = 0.5$) (Fig. 8(e)) not reported within the literature, and the second is found in solvated forms of both **CC3** with Et_2O and CH_2Cl_2 and **CC4** with MeOH (Fig. 8(f)). Although we were not intending to find solvated structures, the occurrence of the solvated crystal in this region of phase space is logical as the structure occurs at low values of $\sigma_{\text{ang}} = 0.2\text{--}0.4$, which suggests the inclusion of the solvent increases the directionality, as it occurs at small patch widths. This is a logical result as the inclusion of solvents is likely to make the intermolecular interactions more directional, analogous to decreasing σ_{ang} . Moreover, this phase is the same as the solvated structure which produces **CC3 β** /**CC4 α** upon desolvation.^{44,46} We note that the discussion of whether the affects of solvent can be represented by our model depends on how the solvent affects the phases. When the solvent is included in the pores, it leads to a more directional interaction which is able to be captured by our model. However, when the solvent is removed from the pores, as in the case for **CC3 β** /**CC4 α** , the act of desolvation disrupts the directional interactions introduced by the solvent without allowing the phase to properly re-equilibrate, causing a phase transition which our model cannot capture.^{44,46} An overview of where each experimental phase is found in our simulations is given in Table 2.

We have shown that when window-to-window packing dominates, only one phase is formed, as seen experimentally, whereas window-to-arene packing results in a much richer phase space containing four phases, only three of which have been seen experimentally. The coarse-grained model laid out in eqn (1) is able to produce the majority of the structures formed by the enantiopure, ordered, desolvated octahedral cages. By changing the width of the patch, and thus the directionality of the interactions between cages, we are able to manipulate the structural phases formed by our model, mimicking the structural changes seen when changing the chemical functionality. Now, the natural question is, given a novel cage, can we predict

Table 1 Previously reported experimental crystal structures and their dominant packing motif(s) for ordered, enantiopure, desolvated octahedral cages

Structure	Dominant packing type	Space group
PUDXES ¹⁹ (CC3α)	Window-to-window	$F4_132$
OZECAY03 (ref. 44) (CC4β)	Window-to-window and arene-to-arene	$R32$
GANDUW ⁴⁵ (CC10)	Window-to-arene	$R3$
ELALAF ¹⁰ (CC1β')	Window-to-arene	$R3$
GANDAC ⁴⁵ (CC9-R3)	Window-to-arene	$R3$
PUDXES02 (ref. 44) (CC3β)	Window-to-arene	$P3$
OZECAY ⁴⁶ (CC4α)		



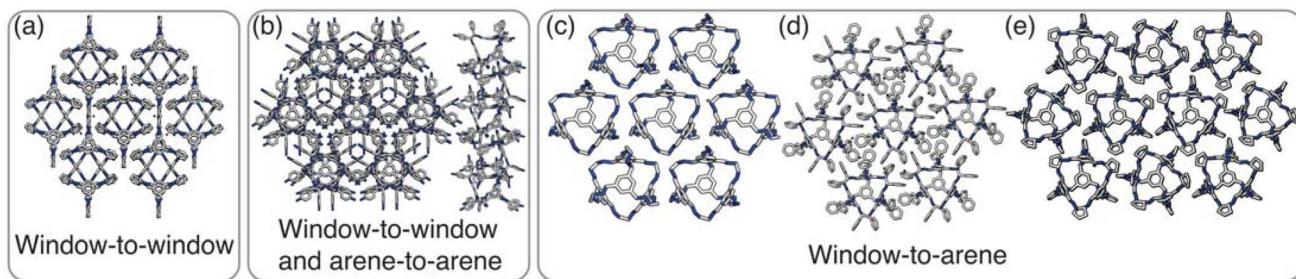


Fig. 7 The experimentally reported enantiopure, desolvated, octahedral cages that contain no occupational disorder. (a) $\text{CC3}\alpha$ which is also the same reported structure for $\text{CC4}\beta$, (b) CC10 , (c) $\text{CC1}\beta'$, (d) $\text{CC9-}R3$, (e) $\text{CC4}\alpha$ which is also the same reported structure for $\text{CC3}\beta$.

its likely solid-state structure using our simulations? And what design rules can we assimilate from our coarse-grained model?

3.3 Predictive capabilities and design rules

In order to gain insight into the most dominant packing types (*e.g.* window-to-arene) and likely value of σ_{ang} , a process must be created to relate the interactions between cages, as calculated through atomistic methods, to our phase space. Although density functional theory (DFT) dimer pair calculations examining packing motifs have proven to be deterministic of preferred packing behaviour,¹⁰ there is no obvious reason why one cage prefers one packing type over the other. This subtlety is what makes the crystal engineering of POCs so difficult. Moreover, establishing a robust relationship between DFT results and a continuous parameter such as σ_{ang} is a more complex task. We have started to address this problem by examining the energetics of two cages slipping around the central patch position (see Section S9†). The goal here is to relate the energies

produced from atomistic calculations to the parameters of our model, creating a relationship between each region of our phase space to a range of interaction energies. Therefore, by using this relationship for novel cages, we can use atomistic simulations of two cage molecules to extrapolate to solid-state, crystalline phase behaviour. However, for both the window-to-window and window-to-arene simulations there are only two desolvated cages that have been reported to form any of the predicted crystal structures: (i) $\text{CC9-}R3$ and $\text{CC1}\beta'$ in the window-to-arene simulations, (ii) $\text{CC3}\alpha$ and $\text{CC4}\beta$ in the window-to-window simulations. Moreover, the predicted crystal structures are found in a narrow, if not the same, region of phase space. This means that we are unable to relate the energies produced from atomistic calculations to the parameters of our model and thus crystal phase behaviour.

Nevertheless, for the cages which pack preferentially *via* their window and arene facets, the energetic differences produced by the two cages slipping around the central patch position are in

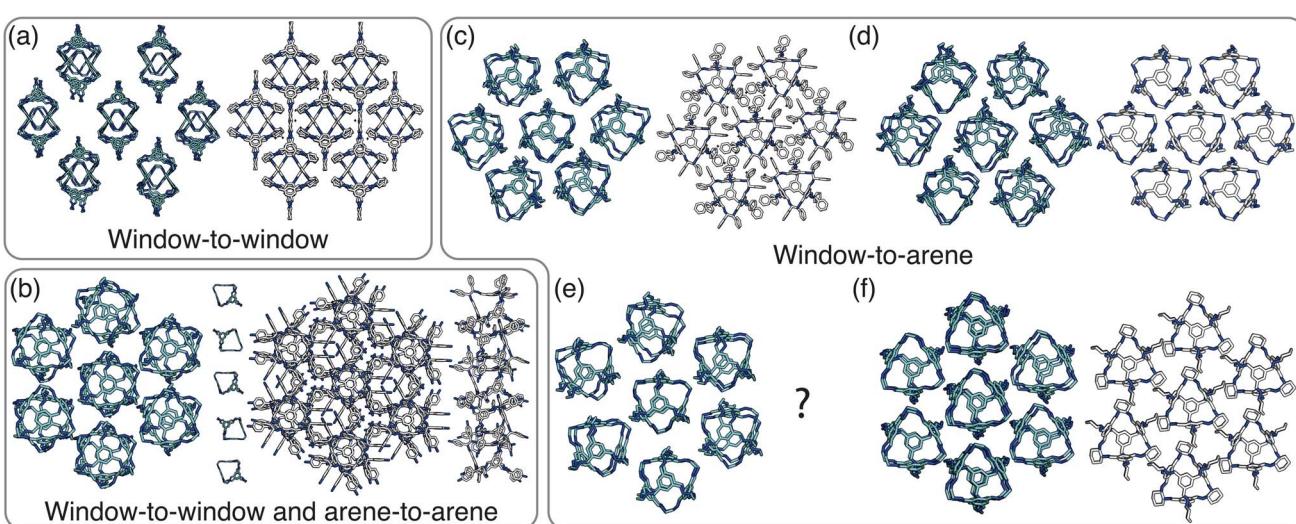


Fig. 8 Our simulated results compared to experimental structures. A representative cluster using the cage CC1 from each simulation is shown on the left in cyan next to its analogous experimental structure on the right. (a) The ordered structure produced by window-to-window simulations which matches up to $\text{CC3}\alpha/\text{CC4}\beta$. (b) The ordered structure produced by simulations with both window-to-window and arene-to-arene packing which matches up to the solid-state structure formed by CC10 . (c-f) The ordered structures formed in window-to-arene simulations when: (c) $\sigma_{\text{ang}} = 0.6$ which is related to the structure formed by CC9 , (d) $\sigma_{\text{ang}} = 0.7$ which has the same structure as $\text{CC1}\beta'$, (e) $\sigma_{\text{ang}} = 0.5$ which is unreported in the literature, and (f) $\sigma_{\text{ang}} = 0.2-0.4$, which is related to the solvated form of CC3 with Et_2O and CH_2Cl_2 and CC4 with MeOH .



Table 2 Match between the parameters of the coarse-grained model that produce different crystal structures and their related experimentally reported structures

Experimentally reported structure	Dominant packing type	σ_{ang}
PUDXES (CC3 α)	Window-to-window	0.1–0.6
OZECAY03 (CC4 β)		
GANDUW (CC10)	Window-to-window/arene-to-arene	0.1–0.6/0.1–0.6
NODVIN (CC3 with Et ₂ O and CH ₂ Cl ₂)	Window-to-arene	0.2–0.4
OZEBUR (CC4 with MeOH)		
N/A	Window-to-arene	0.5
GANDAC (CC9-R3)	Window-to-arene	0.6
ELALAF (CC1 β')	Window-to-arene	0.7
PUDXES02 (CC3 β)	Window-to-arene	N/A
OZECAY (CC4 α)		

agreement with where their crystal structures are found relative to one another in the phase diagram. CC9 has a higher energetic penalty associated with displacements from the central patch position than CC1, mirroring the relationship that the solid-state structure of CC9-R3 is found at a lower σ_{ang} than CC1 β' (Fig. S12 \dagger). These results indicate that similar calculations looking at the slipping of cages might be able to predict the packing behaviour of novel cages, particularly once a threshold is established between atomistic results and σ_{ang} . For this we will need to find molecular cages whose crystalline structures relate to areas of our phase space currently unobserved experimentally.

As for design rules, our results have started to establish a relationship between chemical functionality, solvent, and solid-state phase space as represented by our model. Cages synthesised with bulkier diamine vertices (*e.g.* CC3, CC9, and CC10) tend to sit at lower values of σ_{ang} than less bulky cages such as CC1 (Table 2). Additionally, the inclusion of solvent, as in the case of structures formed by CC3 and CC4, leads to low values of σ_{ang} . These trends can inform design rules, as they show that solvating groups can be used to introduce more directional interactions, increasing the number of polymorphs available for a given cage and thus opportunities for increased porosity, and that bulkier groups on the vertices lead to less rotation and slipping between neighbouring cages.

Finding more desolvated cages whose solid-state structures match those in our phase space, particularly for structures at low σ_{ang} , will lead to the development of a better relationship between results from atomistic calculations and the parameters of our coarse-grained model. This can lead to more concrete rules about the relative energies required to navigate phase space, allowing for easy prediction of the crystal structures formed by novel cages in a much less computationally expensive process than for CSP—only requiring dimer pair DFT calculations compared to many solid-state structure calculations. Through this we hope to further decipher the relationship between solvent, chemical functionality, and σ_{ang} to indicate design rules for tuning the solid-state phase behaviour of POCs.

On top of information on the structural behaviour of cages, our results may hold insights on the crystallisation process of POCs. The window-to-arene simulations show that, depending

on the temperature of the simulation, either an amorphous or ordered phase forms. Perhaps a similar trend could be found with synthesis temperatures, forming a link between our simulations and crystallisation effects such that our model can act as a guide for POCs crystallisation pathways. For example, our phase diagram suggests that crystallising CC9 or CC1 at high temperatures could lead to the formation of amorphous phases, whereas at low temperature an ordered crystalline structure will form. However for cages such as CC3, or any cages that pack window-to-window, the phase formed is independent of synthesis temperature as either an ordered phase forms at any temperature, as in CC3, or an amorphous phase is formed, as with cages that exist at higher σ_{ang} . This is a particularly useful result as it highlights routes for targeting amorphous cages which can exhibit enhanced gas selectivity⁴⁷ and increased porosity relative to their crystalline counterparts.^{47–49} Moreover, our results may provide insight into the polymorphic behaviour of POCs by giving an overview of how many phases exist given the dominant interaction type. For example, for window-to-window packing, the only known ordered structure is that of the 3D connected diamondoid structure found in CC3, which matches our window-to-window simulation results, whereas the window-to-arene simulations exhibit a much richer polymorphic phase space, as seen experimentally.

4 Conclusion

This work presents a coarse-grained model to examine the solid-state packing behaviour of POCs with dominant window-to-window and window-to-arene packing. Our results show that when cages interact most strongly through their windows then only one phase is seen. However, when cages interact most strongly between the windows and arenes, a much richer phase diagram is produced with four structurally distinct phases. These observations match up well to experimental results, where only one solid-state structure is reported when window-to-window packing dominates, but a range of phases are observed for window-to-arene packing.

Through these simulations, we have shown that our coarse-grained model is able to reproduce the majority of the packing behaviour seen by the enantiopure, ordered, desolvated



octahedral cages. These results are the first steps towards using computationally inexpensive coarse-grained simulations to predict the packing of POCs from their molecular structure, as well as developing design rules based on the chemical functionality of the vertices. Here we have shown how both solvent and the size of the functional group on the cages vertices can be used to direct crystal structure formation through changing the directionality of the interactions. Thus, both the addition of solvent and the inclusion of bulky functional groups can manipulate structural formation by reducing rotation and slipping between neighbouring cages, changing the packing behaviour. Additionally we have shown that synthesis temperature may be used to influence the crystallinity in cages, guiding efforts for amorphous phase formation which can lead to increased porosity.⁴⁸ As far as we are aware, this is the first study that looks to coarse-grain molecules for crystal structure prediction. We hope that the results highlighted here will be expanded upon to use coarse-grained modelling for molecules with a richer experimental phase space in order to map molecular interactions onto coarse-grained phase space using processes outlined in Section 3.3 for novel molecular structure prediction.

Although the focus of this paper has been on POCs, the results outlined here have ramifications in many areas of molecular chemistry. Motifs common in POCs are also widespread in other porous materials such as elemental carbon cages⁵⁰ and metal-organic cages.³¹ Therefore, solid-state structures predicted by methods outlined in this study may also relate to crystals formed by other molecular systems and the methodology outlined in this paper could be expanded to other areas of molecular chemistry to help predict the organic solid-state. Eventually, we aim to introduce our computationally inexpensive methodology into existing high-throughput workflows using robotic automation and computational modelling,¹⁵ creating a streamlined workflow for structural prediction of novel molecules. Applying this process to multi-component systems could help drive phase exploration at orders of magnitude currently intractable by current computational techniques.

Data availability

The structures of novel, amorphous, and plastic phases simulated here have been uploaded as part of the ESI.†

Author contributions

E. H. W. conceptualised and designed the project, carried out and analysed the simulations. K. E. J. supervised the project and acquired funding. E. H. W. wrote the manuscript and all authors contributed to the final version.

Conflicts of interest

There are no conflicts to declare.

Acknowledgements

K. E. J. acknowledges the Royal Society for a University Research Fellowship and Enhancement Award and the ERC through Agreement No. 758370 (ERC-StG-PE5-CoMMaD). This work used the ARCHER2 UK National Supercomputing Service (<https://www.archer2.ac.uk>) via our membership of the UK's HEC Materials Chemistry Consortium, which is funded by EPSRC (EP/R029431), and Imperial College London's Research Computing Service, <https://doi.org/10.14469/hpc/2232>. We thank Julia Dshemuchadse and Rachael Skye for helpful conversations about HOOMD-blue and Austin Mroz for commenting on drafts of the manuscript.

Notes and references

† Although plastic crystals are typically formed by spherical particles,^{24,38,39} particle orientations in the plastic crystal phase may not be entirely random and some orientations can be more likely to occur than others,^{38,40–42} similar to what we observe in these simulations.

- O. Ostroverkhova, *Chem. Rev.*, 2016, **116**, 13279–13412.
- A. I. Cooper, *ACS Cent. Sci.*, 2017, **3**, 544–553.
- S. Datta and D. J. W. Grant, *Nat. Rev. Drug Discovery*, 2004, **3**, 42–57.
- C. Zhao, L. Chen, Y. Che, Z. Pang, X. Wu, Y. Lu, H. Liu, G. M. Day and A. I. Cooper, *Nat. Commun.*, 2021, **12**, 817.
- E. O. Pyzer-Knapp, H. P. G. Thompson, F. Schiffmann, K. E. Jelfs, S. Y. Chong, M. A. Little, A. I. Cooper and G. M. Day, *Chem. Sci.*, 2014, **5**, 2235–2245.
- D. H. Case, J. E. Campbell, P. J. Bygrave and G. M. Day, *J. Chem. Theory Comput.*, 2016, **12**, 910–924.
- D. P. McMahon, A. Stephenson, S. Y. Chong, M. A. Little, J. T. A. Jones, A. I. Cooper and G. M. Day, *Faraday Discuss.*, 2018, **211**, 383–399.
- 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. 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 and C. R. Groom, *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.*, 2016, **72**, 439–459.

9 T. Hasell, S. Y. Chong, M. Schmidtmann, D. J. Adams and A. I. Cooper, *Angew. Chem., Int. Ed.*, 2012, **51**, 7154–7157.

10 J. T. A. Jones, T. Hasell, X. Wu, J. Bacsa, K. E. Jelfs, M. Schmidtmann, S. Y. Chong, D. J. Adams, A. Trewin, F. Schiffman, F. Cora, B. Slater, A. Steiner, G. M. Day and A. I. Cooper, *Nature*, 2011, **474**, 367–371.

11 A. G. Slater and A. I. Cooper, *Science*, 2015, **348**, 8075.

12 M. Brutschy, M. W. Schneider, M. Mastalerz and S. R. Waldvogel, *Adv. Mater.*, 2012, **24**, 6049–6052.

13 X. Yang, J.-K. Sun, M. Kitta, H. Pang and Q. Xu, *Nat. Catal.*, 2018, **1**, 214–220.

14 T. Hasell and A. I. Cooper, *Nat. Rev. Mater.*, 2016, **1**, 1–14.

15 R. L. Greenaway, V. Santolini, M. J. Bennison, B. M. Alston, C. J. Pugh, M. A. Little, M. Miklitz, E. G. B. Eden-Rump, R. Clowes, A. Shakil, H. J. Cuthbertson, H. Armstrong, M. E. Briggs, K. E. Jelfs and A. I. Cooper, *Nat. Commun.*, 2018, **9**, 2849.

16 S. Bennett, F. T. Szczypinski, L. Turcani, M. E. Briggs, R. L. Greenaway and K. E. Jelfs, *J. Chem. Inf. Model.*, 2021, **61**, 4342–4356.

17 R. L. Greenaway and K. E. Jelfs, *Adv. Mater.*, 2021, **33**, 2004831.

18 A. He, Z. Jiang, Y. Wu, H. Hussain, J. Rawle, M. E. Briggs, M. A. Little, A. G. Livingston and A. I. Cooper, *Nat. Mater.*, 2022, **21**, 463–470.

19 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 and A. I. Cooper, *Nat. Mater.*, 2009, **8**, 973–978.

20 T. Hasell, J. L. Culshaw, S. Y. Chong, M. Schmidtmann, M. A. Little, K. E. Jelfs, E. O. Pyzer-Knapp, H. Shepherd, D. J. Adams, G. M. Day and A. I. Cooper, *J. Am. Chem. Soc.*, 2014, **136**, 1438–1448.

21 G. R. Desiraju, *J. Am. Chem. Soc.*, 2013, **135**, 9952–9967.

22 J. A. Anderson, J. Glaser and S. C. Glotzer, *Comput. Mater. Sci.*, 2020, **173**, 109363.

23 A. T. Cadotte, J. Dshemuchadse, P. F. Damasceno, R. S. Newman and S. C. Glotzer, *Soft Matter*, 2016, **12**, 7073–7078.

24 P. F. Damasceno, M. Engel and S. C. Glotzer, *Science*, 2012, **337**, 453–457.

25 E. S. Harper, G. van Anders and S. C. Glotzer, *Proc. Natl. Acad. Sci. U. S. A.*, 2019, **116**, 16703.

26 G. van Anders, D. Klotsa, N. K. Ahmed, M. Engel and S. C. Glotzer, *Proc. Natl. Acad. Sci. U. S. A.*, 2014, **111**, E4812.

27 T. D. Nguyen, C. L. Phillips, J. A. Anderson and S. C. Glotzer, *Comput. Phys. Commun.*, 2011, **182**, 2307–2313.

28 R. A. Mansbach and A. L. Ferguson, *J. Phys. Chem. B*, 2018, **122**, 10219–10236.

29 A. W. Long and A. L. Ferguson, *Mol. Syst. Des. Eng.*, 2018, **3**, 49–65.

30 C. R. Groom, I. J. Bruno, M. P. Lightfoot and S. C. Ward, *Acta Crystallogr., Sect. B: Struct. Sci., Cryst. Eng. Mater.*, 2016, **72**, 171–179.

31 N. J. Young and B. P. Hay, *Chem. Commun.*, 2013, **49**, 1354–1379.

32 V. Santolini, M. Miklitz, E. Berardo and K. E. Jelfs, *Nanoscale*, 2017, **9**, 5280–5298.

33 J. A. Anderson, M. Eric Irrgang and S. C. Glotzer, *Comput. Phys. Commun.*, 2016, **204**, 21–30.

34 E. G. Noya, C. K. Wong, P. Llombart and J. P. K. Doye, *Nature*, 2021, **596**, 367–371.

35 A. W. Wilber, J. P. K. Doye, A. A. Louis and A. C. F. Lewis, *J. Chem. Phys.*, 2009, **131**, 175102.

36 H. T. Stokes and D. M. Hatch, *J. Appl. Crystallogr.*, 2005, **38**, 237–238.

37 H. T. Stokes, D. M. Hatch and B. J. Campbell, *FINDSYM, ISOTROPY Software Suite*, <https://iso.byu.edu>.

38 A. P. Gantapara, J. de Graaf, R. van Roij and M. Dijkstra, *J. Chem. Phys.*, 2015, **142**, 054904.

39 U. Agarwal and F. A. Escobedo, *Nat. Mater.*, 2011, **10**, 230–235.

40 A. K. Sharma, V. Thapar and F. A. Escobedo, *Soft Matter*, 2018, **14**, 1996–2005.

41 M. Burian, C. Karner, M. Yarema, W. Heiss, H. Amenitsch, C. Dellago and R. T. Lechner, *Adv. Mater.*, 2018, **30**, 1802078.

42 W. Shen, J. Antonaglia, J. A. Anderson, M. Engel, G. van Anders and S. C. Glotzer, *Soft Matter*, 2019, **15**, 2571–2579.

43 M. Bernabe, R. Pérez-Soto, I. Gómez García and M. Haranczyk, *Mol. Syst. Des. Eng.*, 2018, **3**, 942–950.

44 M. A. Little, S. Y. Chong, M. Schmidtmann, T. Hasell and A. I. Cooper, *Chem. Commun.*, 2014, **50**, 9465–9468.

45 M. J. Bojdys, M. E. Briggs, J. T. A. Jones, D. J. Adams, S. Y. Chong, M. Schmidtmann and A. I. Cooper, *J. Am. Chem. Soc.*, 2011, **133**, 16566–16571.

46 T. Mitra, X. Wu, R. Clowes, J. T. A. Jones, K. E. Jelfs, D. J. Adams, A. Trewin, J. Bacsa, A. Steiner and A. I. Cooper, *Chem.-Eur. J.*, 2011, **17**, 10235–10240.

47 S. Jiang, K. E. Jelfs, D. Holden, T. Hasell, S. Y. Chong, M. Haranczyk, A. Trewin and A. I. Cooper, *J. Am. Chem. Soc.*, 2013, **135**, 17818–17830.

48 J. D. Evans, D. M. Huang, M. R. Hill, C. J. Sumby, D. S. Sholl, A. W. Thornton and C. J. Doonan, *J. Phys. Chem. C*, 2015, **119**, 7746–7754.

49 T. Hasell, S. Y. Chong, K. E. Jelfs, D. J. Adams and A. I. Cooper, *J. Am. Chem. Soc.*, 2012, **134**, 588–598.

50 G. Maier, *Angew. Chem., Int. Ed. Engl.*, 1988, **27**, 309–332.

