Journal Name

ARTICLE TYPE

Cite this: DOI: 10.1039/xxxxxxxxxx

Optimising Minimal Building Blocks for Addressable Self-Assembly

Jim Madge and Mark A. Miller

Received Date
Accepted Date

DOI: 10.1039/xxxxxxxxxx

www.rsc.org/journalname

Addressable structures are characterised by the set of unique components from which they are built and by the specific location that each component occupies. For an addressable structure to self-assemble, its constituent building blocks must be encoded with sufficient information to define their positions with respect to each other and to enable them to navigate to those positions. DNA, with its vast scope for encoding specific interactions, has been successfully used to synthesise addressable systems of several hundred components. In this work we examine the complementary question of the minimal requirements for building blocks to undergo addressable self-assembly driven by a controlled temperature quench. Our testbed is an idealised model of cubic particles patterned with attractive interactions. We introduce a scheme for optimising the interactions using a variant of basin-hopping and a negative design principle. The designed building blocks are tested dynamically in simple target structures to establish how their complexity affects the limits of reliable self-assembly.

Introduction

Advances in nanoparticle synthesis and fabrication mean that we are beginning to see bespoke building blocks that are designed from the ground up for self-assembly.[1,2] In particular, there has been progress in synthetic methods for patterning the surfaces of nanoparticles with regions of different chemical identity. Examples include single-patch Janus particles[3], functionalised clusters with geometries that resemble hybrid orbitals in organic chemistry,[4,5] and striped triblock spheres.[6] These patchy particles provide the opportunity to control not only the valency and directionality of interactions, but also the underlying chemical or physical origin of the interactions.[5,7–11]

Such building blocks are being used to realise both discrete and periodic self-assembling structures of increasing complexity, in which the geometry and connectivity are specified in detail. In many cases, such targets can be achieved using many copies of one or a just a few different building blocks. However, to self-assemble functioning nanodevices, it is necessary to exert an even greater level of control by creating structures from a large number of different components. In the limiting case, every component is unique and has a specified position and orientation in the target. Such a structure is termed “fully addressable”.

The new challenges introduced by addressable self-assembly are being tackled experimentally, theoretically and computationally.[12–16] The correct behaviour of particles requires highly specific interactions, so that building blocks bind strongly only with their neighbours in the target structure. Hence, there are stringent energetic requirements. The uniqueness of the building blocks and their interactions has the advantage of helping to avoid incorrect binding or aggregation during self-assembly, but it also incurs an entropic cost because distinct components cannot be interchanged in the target structure. This constraint contrasts with targets built from multiple copies of the same particle, where the statistical weight of the target and the number of paths towards it are greatly increased by permutations of equivalent building blocks. A fully addressable structure relies on the energetic interactions between neighbouring particles in the target being sufficiently strong to overcome the entropic cost of assembling a structure where no permutations are possible.[17]

Because of its highly specific interactions, single-stranded DNA (ss-DNA) is widely used to mediate self-assembly. Grafting ss-DNA onto colloidal particles encodes them with information (in the form of the nucleotide sequence) that determines which other particles (those with the complementary sequence) they will bind to. DNA-grafted colloids have been used to assemble clusters and crystals[5,18–21] and even to create partially addressable structures.[16] In the most impressive experimental examples of addressable assembly, ss-DNA itself constitutes the building blocks without being grafted to a colloid. These “DNA bricks”[12,22] are folded ss-DNA strands that interact addressably with their neighbours in a three-dimensional, cubic canvas. By excluding selected bricks from a mixture, it is possible in effect to carve out regions of the canvas, leaving a desired structure. Excluding different com-

Department of Chemistry, Durham University, South Road, Durham DH1 3LE, United Kingdom. E-mail: m.a.miller@durham.ac.uk

binations of bricks allows a single canvas to be used to assemble any distinct discrete cluster that fits within the dimensions of the canvas. Hence, there is no need to design new building blocks for different targets.

DNA bricks[12,22] and tiles[23] have shown that addressable assembly with several hundred different components is possible. The success of these schemes relies on the possibility of generating as many distinct nucleotide sequences as needed to encode the interactions of each building block and thereby to determine its position in the target structure. DNA is unique in its scope for implementing specific interactions by systematic generation of nucleotide sequences. This raises the question of how difficult it would be to mediate addressable self-assembly by other types of interactions. In cases where the range of exclusive pairwise interactions is more restricted than with DNA, it would be desirable for building blocks to be as simple as possible to ensure reliable self-assembly of a given target. Similarly, starting from building blocks of a given complexity, we might ask what would be the most complex target that could be constructed.

The complexity of building blocks may be qualitatively described in terms of the amount of information encoded by their interactions. For instance, colloidal particles grafted isotropically with identical polymers, or particles with a single deposited metallic patch are examples of designs with low complexity. However, particles of this type are not able to self-assemble into arbitrary structures. In contrast, colloidal particles grafted with several ss-DNA sequences in well defined directions[16] are of greater complexity because they embody more information about specific binding partners and anisotropic interactions.

A complexity metric has been computed for certain idealised models of self-assembly.[24,25] In that work the particles consist of square tiles, the edges of which may be given an interaction label, with rules to dictate which labels interact with each other. An assembly kit can then be defined as a set of particles that assemble into a given target, or "polyomino". In cases where the symmetry of a target allows, kits containing repeated particles or symmetric particles significantly reduced the complexity of kits. Of course, the success of any self-assembling system also depends on external conditions and on the level of performance required with respect to yield or robustness against perturbations.

Theoretical and computational work has provided important insight into the special properties of addressable assembly.[13,15,26–28] The starting point for such studies is usually an idealised model where the exclusivity of interactions between neighbouring building blocks is built into the potential. This approach is equivalent to defining an interaction "alphabet" whose size is equal to the number of pairwise contacts in the target structure, akin to the case of DNA bricks. In this paper, we start to tackle the question of whether sufficient specificity of interactions can be achieved with a smaller alphabet, and how to maximise the performance of a set of building blocks that deploy a given alphabet. The aim is to design building blocks of minimal complexity that are capable of self-assembling into a given target.

Our test model consists of hard cubic building blocks patterned with attractive patches.[29] Patches are drawn from an alphabet of different types, and the pairwise interaction between two patches is determined by their types. We will test how the number of patches on the faces of the building blocks and the size of the available alphabet affect the quality and limitations of self-assembly. The goal is to implement the interactions between all exclusive binding partners using an alphabet that is far smaller than the number of bonds in a particular target.

Complementarity of intended binding partners is readily arranged by constraining the patterns of patches on corresponding faces to be mirror images. However, for the binding of such pairs to be exclusive, any accidental interactions with other faces must be minimised. Our optimisation of building-block interactions will therefore be a form of negative design,[30] where the desired structure is promoted by actively seeking to destabilise competing structures. In other forms of self-assembly, such as engineered proteins complexes,[31,32] the emphasis is on positive design by maximising the stability of the intended target. In addition to the fact that the ground state of our designs is fixed by construction, negative design is needed to fulfill our aim of making building blocks that are as simple as possible. As we shall see, accidental interactions are more likely to occur by chance in a system where the building blocks are encoded with less information. Hence, it is necessary to minimise competition from unwanted structures in the design process, in contrast to cases where the building blocks are encoded with more information.[12]

A successful self-assembling system would have a number of desirable attributes, including efficiency (speed of assembly), high yield, and robustness with respect to changes in external conditions. The task of design procedures is to optimise some metric that quantifies such criteria. One flexible class of optimisation method is genetic algorithms. In the context of self-assembly, a population of competing designs would be maintained. Their respective performance (fitness) would be measured and used to propagate features of the best designs to a new generation of designs. Genetic algorithms have been successfully applied to the design of self-assembling systems,[33] but they rely on an efficient method for evaluating the fitness since many iterations (generations) may be required to produce a satisfactory design.

Miskin *et al.* have proposed another general approach to design, which exploits information from all configurations visited during a simulation to guide design parameters towards their optimal values in a rational way.[34] This elegant method nevertheless also requires iterative simulations of the full system. To avoid repeated simulations, it is sometimes possible to optimise a system on the fly, such as in the design of isotropic potentials for the self-assembly of special periodic structures.[35]

In some cases, it is possible to use energetic arguments to design interactions for self-assembly without feedback from full simulation or experiment. Notable successes of this sort lie in protein folding[36–38] and assembly,[31,32,39] where the design process searches for sequences that have a unique low-energy state (the target structure) that is an outlier with respect to the main distribution.[40,41] Our method for designing interactions for addressable self-assembly takes a related approach where we make binding partners as distinct as possible from each other within set constraints of complexity. This approach has the advantage of requiring simulations only to test performance, and not for the

design process itself.

In Section 2, we summarise the model and simulation method for the cubic building blocks.[29] In Section 3 we then introduce the design principles and global optimisation procedure for obtaining the interaction interfaces. Designs for addressable dimers, a compact octamer, and a more sparsely connected cluster are presented and tested in Section 4. Section 5 provides a concluding discussion.

Model and Simulation Method

Patchy Building Blocks

In this work we use our previously developed model for patchy building blocks.[29] The particles consist of a hard cubic core with edge length d, the faces of which can be patterned arbitrarily with any number of attractive patches. Patches are drawn from an alphabet of types whose interactions are defined by a matrix of pairwise combinations. The cores of the building blocks bear a resemblance to cubic nanoparticles,[42–44] the phase behaviour and assembly of which has attracted much interest in recent years both in experiment and in theory.[45–47] Although it is not currently possible to pattern the surface interactions of such cubes in detail, we note that synthetic routes to patchy colloids are advancing rapidly.[5,7,8,11]

Target structures for self-assembly can be built up by arranging the building blocks with their faces in contact. The faces of the cubes are therefore the interaction interfaces[48] that must drive self-assembly and the task will be to design the pattern of interactions on the interfaces so as to optimise the efficiency and reliability of the self-assembly process.

Overlaps between the hard cores of the particles are forbidden and are detected in the simulations by treating the particles as oriented bounding boxes.[45,49,50] The pairwise interaction between patches is represented by a Morse potential with an angular attenuation. The interaction between two patches i and j depends on the distance r_{ij} between them according to

$$V_{ij}^{\mathrm{M}}(r_{ij}) = \varepsilon_{ij}\left[e^{-2\alpha(r_{ij}-d)} - 2e^{-\alpha(r_{ij}-d)}\right], \qquad (1)$$

where α is a parameter that controls the range of the potential. We have chosen $\alpha = 6d^{-1}$, which produces a curvature of the potential at its minimum identical to that of the Lennard-Jones potential. ε_{ij} determines the depth of the potential at its minimum, and in the following is always chosen from $\{\varepsilon, 0\}$, thereby simply switching the interaction between a given pair of patches on or off. The Morse site for each patch is embedded in the particles to a depth of $d/2$ in a direction perpendicular to the face on which the patch acts (fig. 1). The distance dependence of the exponent in eqn. 1 therefore ensures that the minimum of the interaction between two patches occurs when two particles sit with faces in contact and with the patches perfectly overlapping at the interface.

The Morse potential is truncated at a distance of $2d$. To avoid a discontinuity at the cut-off, the potential is shifted by $V_{ij}^{\mathrm{M}}(2d)$. The potential is then rescaled to restore the value of the potential at the minimum to ε_{ij}.

The angular attenuation of the potential is a Gaussian of the form

$$V^{\mathrm{ang}}(\hat{\mathbf{r}}_{ij}, \hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j) = \exp\left(-\frac{\theta_i^2 + \theta_j^2}{2\sigma^2}\right), \qquad (2)$$

where $\hat{\mathbf{r}}_{ij}$ is the unit vector from patch i to patch j. $\theta_i = \cos^{-1}(\hat{\mathbf{r}}_{ij} \cdot \hat{\mathbf{u}}_i)$ and $\theta_j = \cos^{-1}(\hat{\mathbf{r}}_{ji} \cdot \hat{\mathbf{u}}_j)$ are the angles between the patch directions and $\hat{\mathbf{r}}_{ij}$, respectively. $\hat{\mathbf{u}}_i$ and $\hat{\mathbf{u}}_j$ are the vectors normal to the faces upon which patches i and j sit (fig. 1). σ determines how quickly the potential decays with any deviation from perfect alignment of the normal vectors and in this work is set to $\sigma = 0.2$. The embedded interaction sites with angular attenuation are similar to a previous model for patchy spherical particles.[51–53]

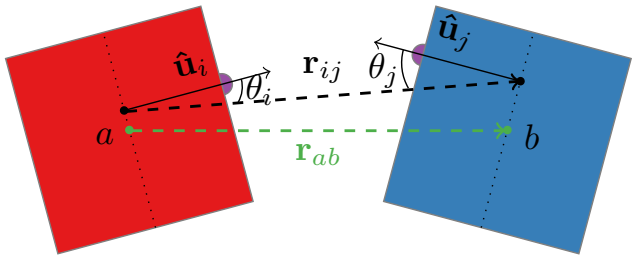

Fig. 1 Schematic representation of the interaction between patches i and j on two cubes a and b, showing the definition of the angles θ_i and θ_j.

The overall form of the potential is therefore

$$V_{ij}^{\mathrm{patch}}(\mathbf{r}_{ij}, \hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j) = \left[\frac{V_{ij}^{\mathrm{M}}(r_{ij}) - V_{ij}^{\mathrm{M}}(2d)}{\varepsilon_{ij} - V_{ij}^{\mathrm{M}}(2d)}\right] \times \Theta(2d - r_{ij}) V^{\mathrm{ang}}(\hat{\mathbf{r}}_{ij}, \hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j), \qquad (3)$$

where $\mathbf{r}_{ij} = r_{ij}\hat{\mathbf{r}}_{ij}$ and Θ is the Heaviside step function, which enforces the cut-off at $r_{ij} = 2d$. The total interaction energy between any two building blocks is given by

$$V_{ab}^{\mathrm{cube}}(\mathbf{r}_{ab}, \Omega_a, \Omega_b) = \sum_{i\in a}\sum_{j\in b} V_{ij}^{\mathrm{patch}}(\mathbf{r}_{ij}, \hat{\mathbf{u}}_i, \hat{\mathbf{u}}_j)\Delta_{ij}(\hat{\mathbf{r}}_{ij}, \Omega_a, \Omega_b), \qquad (4)$$

where Ω_a represents the orientation of particle a and $\mathbf{r}_{ab}$ is the vector from the centre of particle a to the centre of particle b. $\Delta_{ij} = 1$ if the faces on which patches i and j sit are the closest pair of most aligned faces of the two cubes, and 0 otherwise. Δ_{ij} therefore acts as an angular truncation of the potential. The strength of the interaction is typically negligible at the point of truncation (less than $10^{-6}\varepsilon_{ij}$).

Dynamical Monte Carlo algorithm

To simulate the dynamics of self-assembly, we use a special hybrid Monte Carlo scheme.[29] Trial moves are proposed and accepted so as to capture both cooperative internal motion of clusters and Stokes–Einstein-like diffusion of whole clusters. The full details of the algorithm have been described previously.[29] Here we summarise its key features.

Trial moves attempt either to alter the internal structure of a cluster of building blocks or to move the cluster rigidly to produce diffusion. The two types of move are chosen with equal

probability. Diffusion moves act on isolated clusters—groups of building blocks that are mutually connected by paths of non-zero patch–patch interactions. Half of the diffusion moves are translational, and the other half rotational, with step sizes selected to account for the size-dependence of cluster diffusion. To satisfy detailed balance, a bulk move must be rejected if it would bring two clusters within interaction range, since that would alter the decomposition of the system into isolated clusters.

The internal relaxation of aggregates of particles is handled using symmetrised Virtual Move Monte Carlo (VMMC).[54–56] This algorithm begins by picking a seed particle and proposing a translational or rotational move, and then recruits appropriate particles according to the energy changes that would be incurred. This algorithm implicitly takes into account the gradient of the energy without the need to evaluate derivatives, and is highly successful in generating moves to build aggregates from smaller fragments. In order to ensure that the VMMC moves do not interfere with the bulk diffusion moves, a VMMC move must be rejected if it proposes a move of an entire cluster that also remains isolated after the move.

The bulk and internal cluster moves are made consistent by matching the step size for single particles; the distribution of VMMC step sizes is equal to that for the diffusive moves of an isolated monomer so that there is a smooth hand-over between the two types of move. Displacement steps are chosen from a Gaussian distribution of width $0.2d$, which results in a satisfactory acceptance of virtual move Monte Carlo (VMMC) moves across a wide range of temperatures.

Reduced units

Throughout this work we will refer to reduced temperatures T^* defined by

$$T^* = k_B T/\varepsilon_{\text{face}}\,, \tag{5}$$

where $\varepsilon_{\text{face}}$ is the optimal interaction energy between a face and its partner in the target structure. This pairwise "native" interaction between building blocks is a convenient reference for our purposes since, for all simulation results presented here, it is simply equal to ε multiplied by the number of patches per face, and all faces have the same number of patches in a given simulation.

The control over diffusion rates of particles provided by the dynamical Monte Carlo (MC) algorithm means that the number of MC moves can be interpreted as a time scale.[29] Because of the fixed width of the step-size distribution, the diffusion rates, measured with respect to cycles of MC steps, of aggregates of a given size are independent of temperature. However Stokes–Einstein diffusion constants are directly proportional to temperature. Therefore, to compare simulations conducted at different temperatures, we must map the number of MC steps to a relative (reduced) time according to

$$t^* = s/T^*\,, \tag{6}$$

where s is the number of MC cycles completed.

Optimisation of Interactions

An ideal design scheme would optimise the interactions between building blocks to maximise the actual yield of the target structure, the rate of its formation or some combination of the two. However, measuring these properties would require simulations of self-assembly to be used in the design process itself either iteratively or on the fly. To avoid the computational expense and statistical uncertainty associated with such approaches, we instead appeal to energetic arguments. Here, an analogy may be drawn with protein design, where efficient folding can be promoted by maximising the energy gap between the target native conformation and the ensemble of misfolded states, as rationalised in lattice proteins by the random energy model.[40,41]

In the case of proteins, an energy gap can be engineered by minimising the energy of the target structure with respect to the sequence of amino acids from the alphabet of 20 available residues while maintaining a sufficient level of sequence heterogeneity,[36–38] leading to a minimally frustrated energy landscape.[57] In contrast, for the present case of rigid, unlinked building blocks, the interaction energy of two given binding partners can be fixed in advance by making the pattern of patches mirror images and choosing complementary patch types. An energy gap must therefore be achieved by maximising the interaction energy of building blocks that are not intended to bind, and of binding partners that have incorrect mutual orientations.

We therefore seek to design a library of patchy interfaces that are as distinct and energetically incompatible with each other as possible, apart from the interaction of each interface with its single intended binding partner. Since the patches define rigid patterns, the task is a geometric one for a given number of patches. If two interfaces possess a pair of complementary patches, those interfaces can always be brought into a geometry that allows the two patches to interact. However, further interaction between two interfaces that are not intended to bind can be limited by making it geometrically impossible (or unlikely) for two or more complementary pairs of patches to be aligned simultaneously. Hence, to a first approximation, we should ensure that the distance between a given pair of patch types is as different as possible on all the independent interfaces where that pair occurs. This maximisation of the difference needs to be satisfied for all pairs of types simultaneously.

It is possible to envisage a more comprehensive comparison of faces. For example, the overall interaction between two interfaces could be measured by integration over all mutual positions and orientations (akin to the second virial coefficient), or partition functions for small clusters could be built up.[58] However, we will see that a first-order measure of dissimilarity based on simple distances has advantages for the optimisation process. We also note that some variation in the strength of bonds within a structure can be advantageous for dynamical aspects of self-assembly.[59] Our optimisation of the collective interactions will naturally produce slight variations in the free energy of binding of different partners in the set due to differences in rigidity that arise from the distances between patches. However, we will not attempt to optimise the width of the distribution of binding free energies during

the design process.

Let $\left(x_\alpha^{(i)}, y_\alpha^{(i)}\right)$ be the coordinates of patch α on interface i, $1 \leq \alpha \leq n_\mathrm{p}$ where n_p is the number of patches per face and has the same value for all faces. $1 \leq i \leq n_\mathrm{f}$ where n_f is the number of faces being designed. Each $x_\alpha^{(i)}$ or $y_\alpha^{(i)}$ may take values in the range $[-d/2, d/2]$. Each face i also has a complementary face to which it is intended to bind. This complementary face is simply the mirror image of i, as this is the design for which all patches will perfectly align. The mirror faces to not contribute to n_f.

The separation between a pair of patches α and β on face i is

$$r_{\alpha\beta}^{(i)} = \left[\left(x_\alpha^{(i)} - x_\beta^{(i)}\right)^2 + \left(y_\alpha^{(i)} - y_\beta^{(i)}\right)^2\right]^{1/2}. \tag{7}$$

The difference between two patch separations is then

$$\Delta_{\alpha\beta\gamma\delta}^{(ij)} = r_{\alpha\beta}^{(i)} - r_{\gamma\delta}^{(j)}. \tag{8}$$

There are $\frac{1}{2} n_\mathrm{p}\left(n_\mathrm{p} - 1\right)$ pairs of patches per face. Hence, the total number of separations is

$$n_\mathrm{s} = \tfrac{1}{2} n_\mathrm{f} n_\mathrm{p}\left(n_\mathrm{p} - 1\right), \tag{9}$$

and the number of unique differences between separations is

$$n_\Delta = \tfrac{1}{2} n_\mathrm{s}\left(n_\mathrm{s} - 1\right). \tag{10}$$

To make the separations as distinct as possible, we minimise the sum of the reciprocal squares of the differences in eqn. 8. The square ensures that the function is invariant with respect to the order of patch labels. The objective function for minimisation is

$$\chi = \sum_{j>i}^{n_\mathrm{f}} \sum_{\beta>\alpha}^{n_\mathrm{p}} \sum_{\delta>\gamma}^{n_\mathrm{p}} \left(\frac{d}{\Delta_{\alpha\beta\gamma\delta}^{(ij)}}\right)^2 + \sum_{i=1}^{n_\mathrm{f}} \sum_{\alpha=1}^{n_\mathrm{p}} \sum_{\beta>\alpha}^{n_\mathrm{p}} \sum_{\gamma\geq\alpha}^{n_\mathrm{p}} \sum_{\delta>\beta}^{n_\mathrm{p}} \left(\frac{d}{\Delta_{\alpha\beta\gamma\delta}^{(ii)}}\right)^2 + \sum_{i=1}^{n_\mathrm{f}} \sum_{\beta>\alpha}^{n_\mathrm{p}} \left(\frac{2r_0}{r_{\alpha\beta}^{(i)}}\right)^N + \sum_{i=1}^{n_\mathrm{f}} \sum_{\alpha=1}^{n_\mathrm{p}} \left[\left(\frac{x_\alpha^{(i)}}{d/2 - r_0}\right)^M + \left(\frac{y_\alpha^{(i)}}{d/2 - r_0}\right)^M\right], \tag{11}$$

where r_0 is the nominal radius of patches, and M and N are large, positive, even integers. In the following work, $r_0 = 0.1d$, and $M = N = 100$. The terms in eqn. 11 describe the following contributions.

1. Comparison of patch separations where the two pairs of patches belong to different interfaces.

2. Comparison of patch separations where the patches belong to the same interface. The conditions $\gamma \geq \alpha$ and $\delta > \beta$ are required to prevent double counting and to exclude terms like $\Delta_{\alpha\beta\alpha\beta}^{(ii)}$ which equal zero.

3. A penalty to prevent patches on the same interface from coinciding.

4. A penalty to prevent patches from being placed closer than their own radius from the edge of the interface.

Hence, by minimising eqn. 11 for given values of n_f and n_p we can obtain optimally distinct interfaces subject to physically reasonable constraints.

Global minima of the objective function were located using a basin-hopping[60] procedure in the space of patch positions. Basin-hopping consists of a MC simulation with local minimisation between steps, and is therefore a MC simulation where all states are local minima. Starting from a set of n_f faces, each with n_p patches randomly distributed within the bounds of the face, patches are selected at random and displaced in both the x and y dimensions by a uniformly distributed random step in the range $[-d/4, d/4]$. χ for this perturbed configuration is then optimised by conjugate gradient minimisation to a tolerance of 10^{-6}. The local minimisation employs the derivatives of the objective function, which are provided in Appendix A.

The probability of a move being accepted is given by a Metropolis-like criterion

$$W_\mathrm{acc} = \min\left\{1, \exp\left(-\beta\Delta\chi\right)\right\}, \tag{12}$$

where $\Delta\chi$ is the difference between the score (χ) of the new minimised configuration and the current configuration. β is a fictitious reciprocal temperature, chosen to allow efficient exploration of the χ landscape, and here is set to $\beta = 10$. Basin-hopping procedures of 5×10^4 steps were run for various values of n_f and n_p. For each run, the configuration with the lowest value of χ was taken to be the global minimum and recorded. The running lowest value of χ rapidly decreased at the start of a simulation, generally converging after around 10^3 steps, after which any decreases in score were generally small.

The preceding equations assume that all pairs of patch distances must be made distinct to suppress non-native interactions. However, if not all patches interact with each other then separations of different patch combinations do not need to be made distinct. For instance, consider the case of two patch types, A and B and their binding partners, A′ and B′, on the complementary interface. If there are no A–B interactions then there is no need to make A–A distances distinct from A–B or B–B distances because alignment of different patch types would not lead to a lowering of energy and stabilisation of unwanted structures. In such cases, only contributions corresponding to matching pairs of types are included in the first two terms of eqn. 11. Nevertheless, it would be unphysical to place two patches too close to each other, even if they are of different types. Hence, the groups of different patches do influence each other's positions through the third term of eqn. 11, which must include all pairs.

Results

Interface Design

The interface optimisation algorithm successfully produces designs for a wide range of values of n_f and n_p, a selection of which are presented in fig. 2 for the case where all patches interact equally with each other. We may confirm that differences between pairwise patch distances are as distinct as possible by plotting the distribution of $r_{\alpha\beta}^{(i)}$ values, as in fig. 3. For a given number n_f of faces, the values of $r_{\alpha\beta}^{(i)}$ are approximately equally

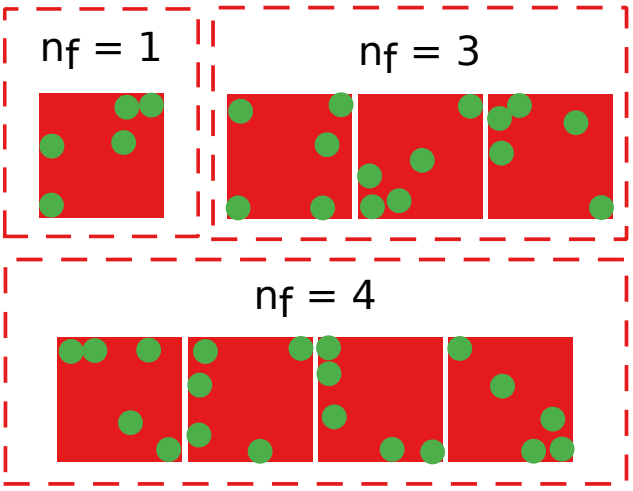

Fig. 2 Example optimised patch patterns with n_f faces in $\{1,3,4\}$ and $n_p = 5$ patches per face for the case where all patches interact equally.

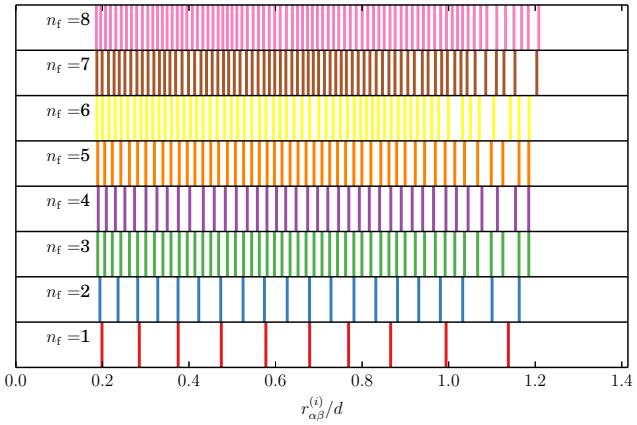

Fig. 3 Differences in pairwise distances for optimised designs with $1 \leq n_f \leq 8$ interfaces and $n_p = 5$ patches per face for identical patch interactions.

distributed over the available range. The minimum and maximum are achieved in all cases by having two patches adjacent to each other on the same face, and two patches in opposite corners on the same face, respectively.

It is important to note that, when initiated from different random starting points, the global minimisation scheme described in section 3 reaches final optimal values of χ that are very similar, but that correspond to significantly different patterns of patches on the designed interfaces. In other words, there is no single decisive global minimum of χ, but rather a large number of comparably good solutions. In our simulation tests of self-assembly, we will therefore average over a set of optimal designs for given values of n_f and n_p.

As well as the case of identical patches, we will investigate designs with a small alphabet of different patches, denoted by different colours. Each patch type has a complementary type, which is used on the matching faces of the binding partners. For designs displaying, for example, three colours of patch, the overall alphabet size is six, once the binding partners are taken into account. The interaction between complementary types has uniform strength ε and all other combinations do not interact. As discussed in section 3, the introduction of different patch types releases some of the constraints on the optimisation procedure, allowing the same pairwise distance to be used more than once in a design if it arises between different combinations of patch types. Two examples of designs with five patch colours are shown in fig. 4.

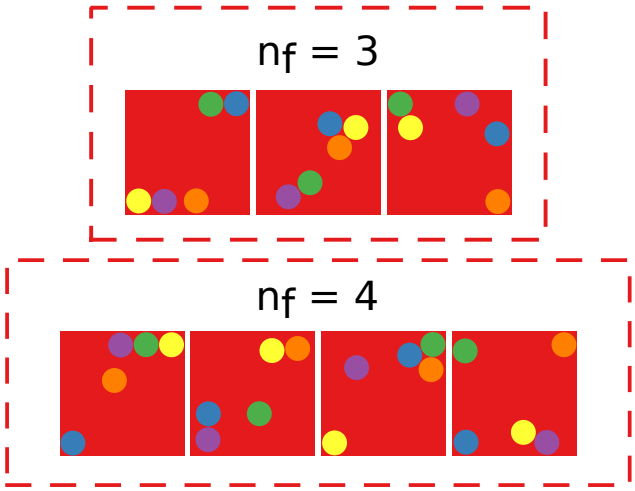

Fig. 4 Optimised patch designs for $n_f = 3$ and 4 interfaces using $n_p = 5$ patches of different types.

The plot of pairwise distances in fig. 5 demonstrates that relevant pairwise distances are indeed well separated by the design procedure. The distribution of separations for each combination of patch types in a design involving $n_p = 3$ patches per face is shown. The three patches are of different types, giving three pairwise combinations of types. Comparing pairs of types within a given design, we tend to see duplicate distances, showing that pairwise distances are being reused where possible to spread matching pairs as widely as possible. Nevertheless, separations of different pairs of types in fig. 5 are usually not perfectly mirrored. This does not indicate the failure of the optimisation procedure, but rather demonstrates how the restrictions imposed on our system (terms to prevent patches from overlapping or moving outside the bounds of an interface) cause some frustration and may mean that a perfectly distributed set of distances is not geometrically possible.

Assembly simulations

We have tested the optimal designs obtained in section 4.1 on three target structures. In the first case, each cube has only one patterned face and therefore forms a simple dimer when it binds to its complementary building block. The dimers are nevertheless unique, thereby providing the simplest possible test of whether the designed interfaces can drive addressable assembly, while also being efficient to simulate. We then examine two contrasting cluster targets, which will be introduced in section 4.2.3.

As discussed in section 4.1, when running the optimisation algorithm from different starting points, distinct final designs with similar scores are obtained. To ensure that the particular behaviour of any one design does not produce misleading results,

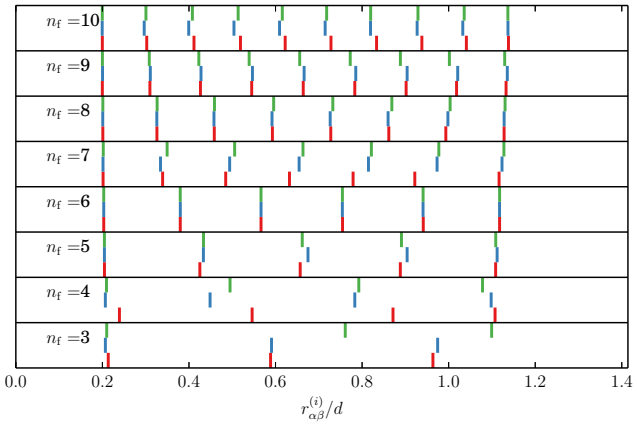

Fig. 5 Distribution of pairwise distances for a set of optimised designs with $3 \leq n_f \leq 10$ faces and $n_p = 3$ patches of different types per face. For each value of n_f there are three distributions, one for each of the three pairs of patch types.

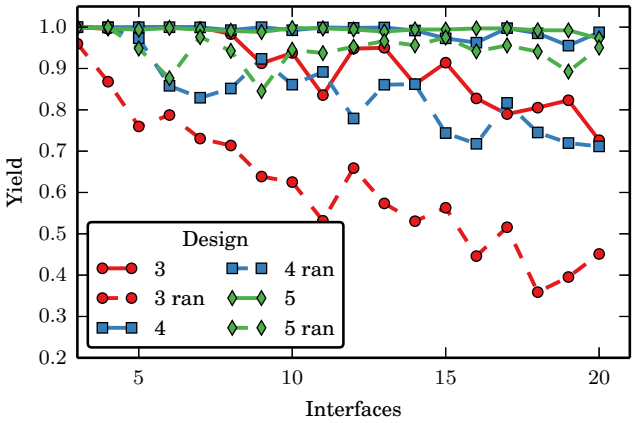

Fig. 6 Yield of addressable dimers as a function of the number of interfaces designed. Yields for designed interfaces with $n_p = 3,4$ and 5 patches per face are shown alongside those for randomly generated interfaces.

all results for given n_f and n_p in the following work are averaged over five optimised designs. For each design, a further ensemble average over five dynamic runs is taken, leading to a total of 25 trajectories per measurement.

In all the simulations, only one copy of each unique building block is present. Our motivation for this simplification is to avoid the independent question of competition between partially formed fragments of multiple copies of the target. This important phenomenon will be addressed in a separate study.[61] However, even in the absence of fragment competition, it is important to appreciate that the restriction to a small number of copies (which is common in simulation studies of self-assembly) affects self-assembly because not all relevant density fluctuations are captured. Such artefacts, and corrections under certain circumstances, have been discussed in detail by Ouldridge *et al.*[62] In the present study, the cross-interactions between building blocks make absolute corrections impractical to apply, but we note that our main concern in this article is comparison of different designs under equivalent conditions.

Identical patches

It was quickly established that designs using a single type of patch were capable of assembling only targets with a very small number of independent interfaces. The restriction to identical patches does not allow sufficiently distinct interfaces to be designed because unintended interactions abound. Even if patches are highly directional, so that they only interact when perfectly aligned, any two patches on different interfaces may bind with energy $-\varepsilon$, and the number of non-native possibilities of this sort is proportional to both n_f and n_p. Furthermore, even pairs of patches on complementary faces can always be perfectly but incorrectly aligned by rotating one building block by 180°. Hence, although it is tempting to add more patches to make the interface patterns more distinct, the number of incorrectly oriented pairs, each with binding energy -2ε grows as $n_p(n_p-1)$. In the more practical case, where patches have a finite angular width, pairs of building blocks that are incorrectly bound in this way are likely to be further stabilised by additional, accidental partial interactions.

From now on, we will therefore consider designs with an alphabet larger than one. The size of the interaction alphabet for all simulations will be twice n_p, meaning every patch per face is a unique type and only interacts with a complementary type on its partner face. The intermediate case, where the alphabet size lies between 1 and n_p, as well as the case where the patch composition of faces is allowed to vary between interfaces, are briefly discussed in the Supplementary Material.

Dimers

As an initial assessment of the limits of the designed interfaces, we have tested systems of addressable dimers with $n_p = 3,4$ and 5 patches per face for an increasing number n_f of distinct interfaces. Starting from disordered configurations with each species of building block present at a number density of $0.0005d^{-3}$, assembly was simulated for 10^7 MC cycles before averaging the yield of correctly assembled dimers over 10^5 cycles. These simulations were conducted at $T^* = 0.0267$, which is low enough for the target to remain effectively irreversibly bonded once correctly formed.

Final yields for $3 \leq n_f \leq 20$ interfaces are plotted in fig. 6. The simplest designs, featuring three patches per interface, display high yields for small numbers of interfaces ($n_f \leq 8$). However, beyond this point the capacity to encode sufficiently distinct patterns is reached, with an associated fall in yield for larger values of n_f.

Although the yields for $n_p = 3$ may appear to remain quite high, dropping only to about 0.8, any interface then has a one-in-five chance of binding incorrectly. For a target structure with N independent links, we would therefore have only a 0.8^N chance of all links forming correctly at once. Hence, performance at this level may not be of practical use for some targets.

fig. 6 demonstrates how additional information in the form of a modest increase in the number of patches n_p and patch types aids in creating unique designs. Yields for both $n_p = 4$ and $n_p = 5$ remain very high for the entire range of n_f covered, with $n_p = 4$ yields showing some tendency to fall towards the upper region of n_f covered. This can be understood by considering how additional

patches allow the design procedure more degrees of freedom with which to explore the scoring function. Furthermore, with more patches, the difference between the native energy and unwanted interactions can be enlarged.

Also plotted in fig. 6 are yields for interface patterns generated by placing patches randomly, subject to the constraints that patches are not within a nominal patch diameter ($2r_0$) of each other or within r_0 of the edge of the interface. Each interface still has an exact mirror image for its binding partner, giving the target structure approximately the same energy as with the designed interfaces. Similarly to the designed case, each point for random patterns in fig. 6 was obtained from 25 simulations using 5 random patterns.

The yields for $n_\mathrm{p} = 3$ and 4 random patches demonstrate how the optimisation procedure produces superior designs to random patterns. This confirms both that design is necessary, and that our proposition of maximising the differences between pairwise distances between patches is an effective method for generating distinct interfaces. Yields for random patterns with $n_\mathrm{p} = 3$ and 4 clearly perform worse than their equivalent designed interactions, rapidly falling to unacceptably low yields for use in cluster assembly.

The yield for random patterns with $n_\mathrm{p} = 5$ remains relatively competitive across the range of n_f simulated, although perhaps not high enough for more complex applications. Enough information has been added to the faces to make even random faces sufficiently distinct to assemble correctly up to the largest value of n_f simulated here, which does not begin to probe the breakdown self-assembly with such particles.

The noise in fig. 6 has two sources. First, there is a small variation in performance for a given design amongst its five independent simulations, which start from different disordered configurations. Second, there are intrinsic differences in the performance of different designs. When the mean yield is high (above 0.9), the standard deviation about that mean for a given design is small (around 0.01). The standard deviation between designs at a given combination of n_f and n_p is somewhat larger (typically in the range $[0.02, 0.05]$). This variation acts as a reminder that χ is not a perfect predictor of performance. For the random designs, where the yield tends to be poorer, the standard deviation between designs is much larger (in the range $[0.1, 0.2]$). Hence, random designs are highly unpredictable; it is possible to be fortunate or unfortunate in the susceptibility to accidental interactions.

We may further test the robustness of the designed interfaces by comparing them with versions where the optimised patch positions are randomly perturbed from their idealised positions. To measure this effect, we took our optimal designs for $n_\mathrm{p} = 3$ at $n_\mathrm{f} = 5, 13$ and 20 and, for each of the five designs in each case, created twenty perturbed designs by displacing the patches about their original positions on a Gaussian distribution with standard deviations in the range $[0.05d, 0.5d]$. The corresponding displacements were also applied to the partner faces. Each of the perturbed designs was then simulated in the manner described above and the final yield of dimers determined. These yields are shown in fig. 7. The simplest case of $n_\mathrm{f} = 5$ interfaces withstands perturbation up to about a width of 0.3. The cases of $n_\mathrm{f} = 13$ and 20, which are already difficult for the perfectly optimised three-patch designs, immediately suffer when perturbed. This vulnerability of the designs for higher n_f is a further measure of the limitations of building blocks with this particular number of patches and indicates where additional degrees of freedom (in the form of more patches) are needed to produce a robust energy gap between the target structure and its competitors.

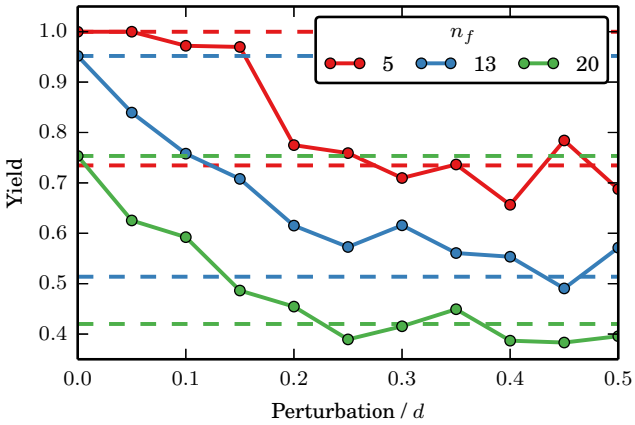

Fig. 7 Yields for a series of perturbed designs for addressable dimers with $n_\mathrm{p} = 3$ patches per face and $n_\mathrm{f} = 5, 13$ and 20 independent interfaces. The upper and lower dashed lined for each value of n_f represent the yield for the unperturbed designs, and fully random patterns, respectively.

It is desirable for self-assembly to be insensitive to the details of the external conditions, such as temperature. We have measured the yield of addressable dimers as a function of the temperature of the quench from the initial disordered state. fig. 8 shows the cases of $n_\mathrm{f} = 3, 10$ and 20 interfaces designed with $n_\mathrm{p} = 3, 4$ and 5 patches. A number of characteristics of the assembly can be extracted from fig. 8. Firstly, for all values of n_p and n_f, randomly patterned interfaces universally have a lower peak yield with respect to temperature and have a narrower range of near-peak performance, *i.e.,* they are less robust with respect to deviation from the optimal temperature. The poorer performance is on the low-temperature side, where unintended interactions act as kinetic traps that can permanently stall assembly. On the high-temperature side of the yield curves, close to where the target is not thermodynamically stable, the random patterns give similar yields to designed interactions. However, yields at these temperatures are suboptimal, and so this cannot be taken to mean that the random designs are competitive.

Secondly, we also see that, for a given number of patches, designs with an increasing number of interfaces become less resilient to changes in temperature. This effect reflects a limit on the complexity of independent designs for a given n_p. The high-temperature edge of the yield curves are similar across the different values of n_f because the temperature is scaled to (approximately) the energy of correctly bound interfaces in each case. However, the low-temperature edge falls sooner for greater values of n_f as more numerous kinetic traps start to hinder assembly.

Thirdly, fig. 8 demonstrates that increasing the information content of the designs (through n_p) can improve the reliability

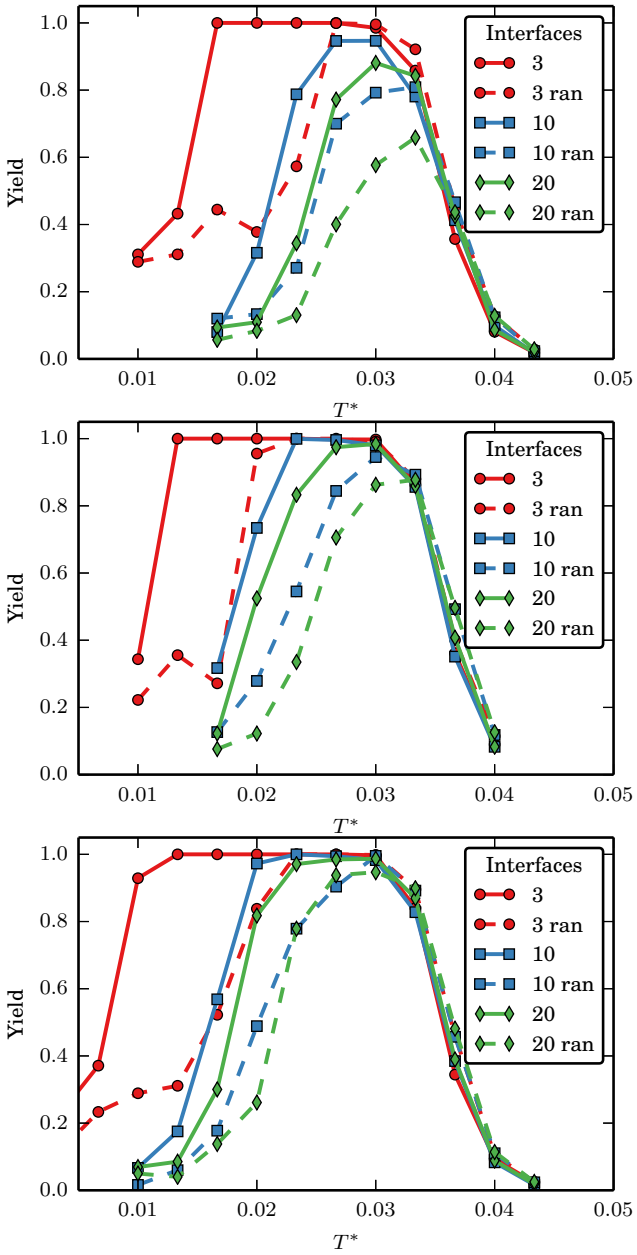

Fig. 8 Yields of addressable dimers as a function of temperature for $n_p = 3$ (top), 4 (middle) and 5 (bottom) patches per face. For each value of n_p three values of the number n_f of interfaces are plotted (3, 10 and 20) for both optimally designed and randomly generated interfaces.

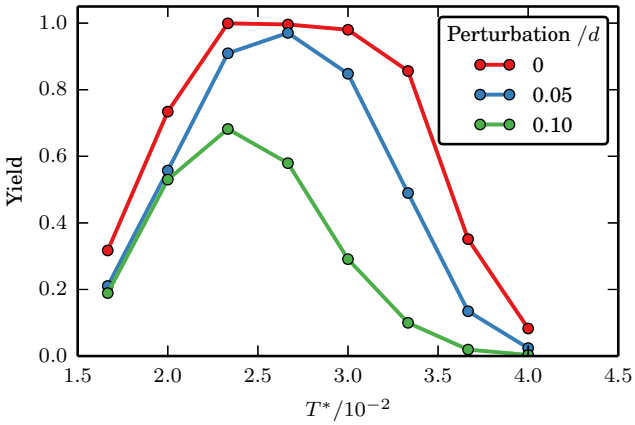

Fig. 9 Yield of dimers as a function of temperature for optimised designs and for designs with binding partners independently perturbed, for $n_f = 10$ and $n_p = 4$.

of assembly across a range of temperatures. Comparing yields for the same n_f on the panels for different n_p, we observe the regions of high yield broaden somewhat. This is a composite effect of the increase of native binding energy compared to unwanted interactions, and also the extra complexity in design allowing for a larger number of unique interfaces to be designed. In all these comparisons, we have quenched directly to a given assembly temperature. We note that yields might be further increased by optimising a programme of temperature change such as gradual annealing, or a two-step protocol.[14]

Before leaving the case of addressable dimers, we consider the implications of any imprecision in the placement of patches during the synthesis of an optimised design. In fig. 7, patch positions on binding partners were perturbed identically to test the underlying design. We now introduce independent perturbations of patch positions on binding partners. The mismatch between partners now raises the ground-state energy, shifting optimal assembly to lower temperatures as shown in fig. 9 for the case of $n_f = 10$ interfaces with $n_p = 4$ patches per face. For small perturbations (a Gaussian distribution of width $0.5d$), peak yield does not suffer, but the range of temperature for good assembly is narrowed. For larger perturbations, the peak yield inevitably drops. One important consequence of imprecision in synthesis is therefore that the temperature for maximum performance must be adjusted to account for the imprecision. Working at the optimal temperature of the ideal designs may result in unnecessary underperformance of the realised building blocks.

Cluster Assembly

Having tested binding of the designed interfaces independently in the context of dimers, we now turn to self-assembly of addressable clusters. The two test clusters are illustrated in fig. 10 and were chosen for their contrasting connectivity. The first is a compact cube of eight building blocks, which we have previously studied in depth using potentials where the exclusivity of binding was imposed rather than designed.[29] This target is maximally connected, with all adjacent faces in the target binding to each other. The other cluster target resembles the Vitruvian man,

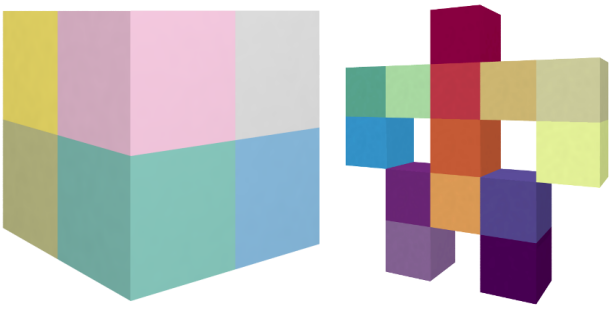

Fig. 10 The two discrete clusters used to test assembly of addressable building blocks. The octamer cube (left) contains eight particles and $n_f = 12$ interfaces, while the Vitruvian man (right) contains fourteen particles and $n_f = 13$ interfaces.

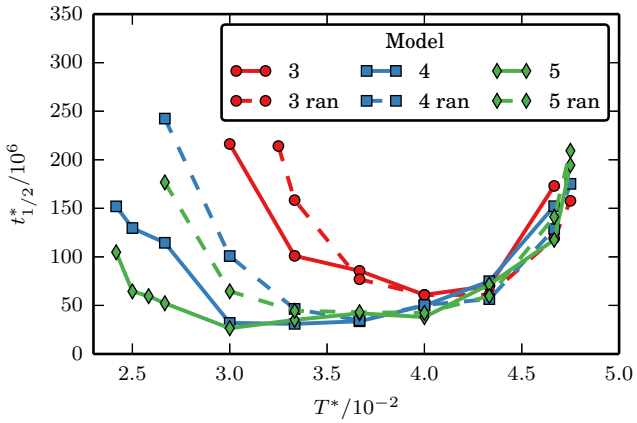

Fig. 11 Average reduced time $t_{1/2}$ for 50% yield for the octamer target as a function of temperature for designs with $n_p = 3, 4$ and 5 patches per face with both optimal and random designs.

a partially addressable version of which was first assembled experimentally.[16] This target is minimally connected and requires all bonds to be present for the structure to be intact.

The octamer target requires 12 pairs of complementary interfaces, while the Vitruvian man requires 13. As for the addressable dimer tests, each simulation includes only one copy of each building block, so that any competition between partially formed fragments is excluded. Also as before, all results are ensemble averages are taken over five runs for each of five interface designs. The overall number density of building blocks for the cluster targets is $0.05d^{-3}$.

As a test of both the efficiency and the sensitivity of self-assembly, fig. 11 shows the time $t_{1/2}$ taken to achieve an average of 50% yield of the octamer target as a function of temperature, using eqn. 6 to compare times at different temperatures. Interfaces with three different numbers n_p of patches were simulated with both designed and random patterns of patches. The plots have a characteristic U shape. The rise in $t_{1/2}$ at high temperatures is due to the thermodynamic instability of the target, which causes the time for assembly to diverge. At low temperature, $t_{1/2}$ increases due to the time taken to escape from the kinetic traps of incorrectly bound building blocks. Efficient assembly takes place between these extremes.

fig. 11 shows that designs with $n_p = 4$ and 5 patches per face lead to significantly more efficient assembly than the simplest case of $n_p = 3$, with an optimal assembly time that is about half that of the latter. There is also a progressive improvement in the reliability of assembly as n_p is increased, reflected by the considerably broader region of rapid assembly. This broadening arises from extension of successful assembly to lower temperatures, due to the reduced competition from incorrectly bound structures when there are more patches per face. At high temperatures, thermodynamic instability develops in unison for all cases as expected from the definition of reduced temperature in eqn. 5.

Building blocks with random patterns of patches perform less robustly at low temperatures, showing the importance of making the interfaces between unintended binding partners as incompatible as possible by the design process. One mechanism for kinetic trapping in self-assembly of the octamer at low temperature starts when two building blocks join correctly, thereby creating two interfaces of exposed patches with dimensions $2d \times d$ for the structure to continue growing. However, it is then also possible for a third building block to bind to this larger interface by straddling the boundary between the first two particles. Inter-patch distances created by placing two interfaces adjacent to each other are not included in the optimisation algorithm and can generate kinetic traps for both designed and random interfaces. The three-patch system suffers the most from such cross-interface alignments because the accidental coincidence of even just two patches represents a larger fraction of ε_{face} compared to systems with more patches.

Partially counterbalancing the vulnerability of extended exposed interfaces is the fact that the octamer contains many closed bonding loops. The smallest such fragment consists of four particles in a plane, where correct placement of the fourth particle results in two favourable interfaces forming at the same time. Since even the randomly patterned interfaces have complementary binding partners by construction, the additional energetic return on completion of a bonding loop helps to make the random interfaces partly viable.

The corresponding plot of times for 50% assembly of the Vitruvian man target is shown in fig. 12. Unlike the octamer target, all interfaces must bind simultaneously and independently, without reinforcement of closed loops in the bonding topology. Here, randomly patterned interfaces perform very poorly. The random designs for both $n_p = 3$ and 4 failed to achieve 50% yield at any temperature simulated and so are absent from the plot. Random patterns with $n_p = 5$ are capable of driving assembly, but only do so efficiently over a narrow range of temperatures, comparable to that for optimised designs of the simpler building blocks with $n_p = 3$. In contrast, the optimised designs with $n_p = 4$ and 5 both exhibit efficient assembly over a wide range of temperature. Unlike for the octamer target, assembly of the Vitruvian man never involves intermediate steps with an enlarged exposed interface, and this helps to extend successful assembly to lower reduced temperatures.

The simulation snapshots in fig. 13 demonstrate how the limited information in designs with $n_p = 3$ struggles to define enough unique interfaces. The figure shows a correctly assembled cluster,

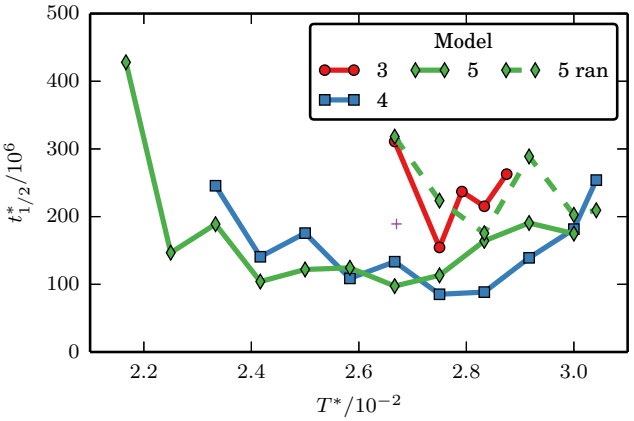

Fig. 12 Time to reach 50% yield for the Vitruvian man cluster as a function of temperature. Displayed are the results for $n_p = 3, 4$ and 5 and also $n_p = 5$ with randomly places patches. The randomly patterned faces for $n_p = 3$ and 4 did not achieve 50% yield within the simulation time at any temperature. The purple plus sign denotes the time and temperature of the dimer tests from fig. 6.

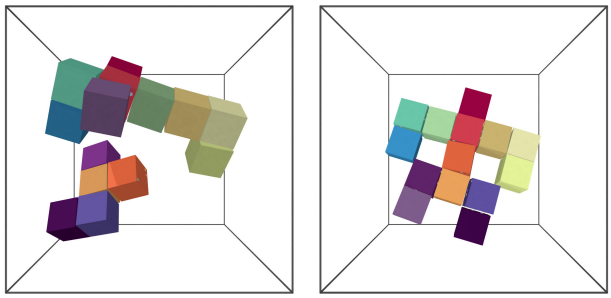

Fig. 13 Simulation snapshots showing an incorrectly assembled (left) and correctly assembled (right) Vitruvian man. The failed example is taken from a simulation of an optimised patch design with $n_p = 3$ patches per face, whilst the successful cluster has $n_p = 5$. Both simulations were conducted at $T^* = 0.0267$.

using an optimal design with $n_p = 5$, and a system with incorrectly assembled clusters obtained from a design with $n_p = 3$. Each particle type is given a unique colour, which is shared between the panels. The binding errors in the failed structure can therefore be seen by comparing the colours.

For the Vitruvian man target, where only 13 independent interfaces are required, $n_p = 4$ seems to be the best compromise between complexity of the building blocks and the reliability of assembly, since the additional flexibility provided by a fifth patch produces a diminishing return in robustness of assembly. We expect from the dimer results in fig. 6 that $n_p = 5$ could encode addressable assembly for loopless structures with at least 20 independent interfaces and that this extra capacity is somewhat redundant for the Vitruvian man. A rough calculation corroborates this conclusion as follows. The average yield in fig. 6 can be interpreted as the probability that any given dimer has correctly formed at the specified temperature and after the allowed time. The temperature and time corresponding to fig. 6 are denoted by a plus symbol on fig. 12. Raising the yield of dimers for $n_p = 4$ and $n_p = 5$ to the power of 13 to find the probability that all 13 interfaces needed for the Vitruvian man are simultaneously present, we obtain values well above 0.5. Hence, $t_{1/2}$ for the Vitruvian man lies below the time allowed in the tests of dimer assembly when it is assembled with 4 or 5 patches per face. In contrast, raising the yield of dimers for $n_p = 3$ to the power of 13 gives a value of about 0.5, indicating that such designs should be marginal in their ability to produce a 50% yield of the Vitruvian man under those conditions. From fig. 12 we see that $t_{1/2}$ is indeed steeply rising and passes above the plus symbol for the dimers.

Conclusion

In this study we have focused on the design of a simple but general class of building block for self-assembly, consisting of a rigid core and fixed anisotropic interactions. Moreover, the flat faces of the cubic particles have allowed us to consider the pattern of energetic interactions independently from particular steric effects. We have adopted a key principle from protein design, where the target (native) state is generally expected to have a significantly lower energy than the ensemble of competing configurations.[36,40] Binding partners in our self-assembling system are exactly complementary by construction, so the task has been to minimise undesired interactions.

Our scheme for this optimisation has avoided an iterative dependence on measured performance by attempting to maximise the geometrical incompatibility of unintended binding partners. Comparison of such "negatively" optimised designs with random patterns of complementary interactions (fig. 6) has shown that this approach makes addressable self-assembly with an increasing number of components possible and, in some cases (fig. 12), can dramatically transform the viability of the process. Nevertheless, there is scope for further refinement of the design process. For example, for a given target structure, it is conceivable to include inter-patch distances that straddle pairs of correctly bound building blocks in the design process. This additional consideration should help to avoid one of the mechanisms for kinetic trapping that we have observed.

In our simple model, the number of patches per designed interface is a measure of the complexity of the building blocks and of the amount of information encoded into them. Increasing the number of patches and the size of the patch alphabet generally provides (i) greater scope for binding specificity of multiple independent pairs, (ii) decreased sensitivity of self-assembly to the choice of temperature, (iii) greater robustness with respect to any imprecision in patch placement, and (iv) more rapid completion of the target. However, these advantages come at the cost of a potentially harder synthesis, and a compromise may be necessary in practice. Hence, investigations of this sort can provide valuable information on the minimal requirements for successful assembly for a given target.

The contrasting cluster targets of a highly connected octameric cube and the loopless Vitruvian man cluster demonstrate that the yield in addressable self-assembly can depend sensitively on the structure of the target. The presence of bonding loops can provide an energetic reinforcement of correct binding, permitting less optimised interactions to succeed. However, loops can also encourage competition between independently initiated fragments.[29]

This important effect has deliberately been excluded from the present simulations, which include only one copy of each distinct building block, but will be addressed in a separate investigation.[61]

Naturally, fixed interactions are only one way of driving self-assembly and encoding addressability. In addition, many polymer-based and biological self-assembly processes exploit internal degrees of freedom in the assembling units, as exemplified by the spectrum of mechanisms displayed in protein binding.[63] The additional flexibility provides further control over the free energy profile for assembly by balancing enthalpic interactions and internal entropy.[1,17] Another important attribute of building blocks is their shape,[2] which has been demonstrated to be an effective tool for driving self-assembly in work on lock-and-key colloids.[64] The importance of the steric fit between building blocks is also evident in protein-based self-assembly.[31,32] Like the optimisation of energetic interactions at a binding interface, the design of complementary shapes is a way of increasing the complexity of building blocks in order to encode them with the necessary information to drive assembly. The combination and optimisation of different properties of building blocks opens up the prospects for powerful control of addressable self-assembly.

Objective function derivatives

To minimise eqn. 11 we need the derivatives of χ with respect to the x and y coordinates of each patch on each face. Application of the chain rule leads to the following terms for the derivative with respect to the x coordinate of patch ε on face k:

$$\begin{aligned}\frac{\partial\chi}{\partial x_\varepsilon^{(k)}} = -2d^2 \sum_{\substack{j=1\\ j\neq k}}^{n_\mathrm{f}} \sum_{\substack{\beta=1\\ \beta\neq\varepsilon}}^{n_\mathrm{p}} \sum_{\gamma=1}^{n_\mathrm{p}} \sum_{\delta<\gamma}^{n_\mathrm{p}} \left(\Delta_{\varepsilon\beta\gamma\delta}^{(kj)}\right)^{-3} \frac{\partial \Delta_{\varepsilon\beta\gamma\delta}^{(kj)}}{\partial x_\varepsilon^{(k)}} \\ - 2d^2 \sum_{\substack{\beta=1\\ \beta\neq\varepsilon}}^{n_\mathrm{p}} \sum_{\substack{\gamma=1\\ \gamma\neq\varepsilon}}^{n_\mathrm{p}} \sum_{\substack{\delta>\gamma\\ \delta\neq\varepsilon}}^{n_\mathrm{p}} \left(\Delta_{\varepsilon\beta\gamma\delta}^{(kk)}\right)^{-3} \frac{\partial \Delta_{\varepsilon\beta\gamma\delta}^{(kk)}}{\partial x_\varepsilon^{(k)}} \\ - 2d^2 \sum_{\substack{\beta=1\\ \beta\neq\varepsilon}}^{n_\mathrm{p}} \sum_{\substack{\delta>\beta\\ \delta\neq\varepsilon}}^{n_\mathrm{p}} \left(\Delta_{\varepsilon\beta\varepsilon\delta}^{(kk)}\right)^{-3} \frac{\partial \Delta_{\varepsilon\beta\varepsilon\delta}^{(kk)}}{\partial x_\varepsilon^{(k)}} \\ - \frac{N}{2r_0} \sum_{\substack{\beta=1\\ \beta\neq\varepsilon}}^{n_\mathrm{p}} \left(\frac{2r_0}{r_{\varepsilon\beta}^{(k)}}\right)^{N+1} \frac{\partial r_{\varepsilon\beta}^{(k)}}{\partial x_\varepsilon^{(k)}} + M\left(\frac{1}{d/2-r_0}\right)^M \left(x_\varepsilon^{(k)}\right)^{M-1}. \end{aligned} \quad (13)$$

The derivatives with respect to a y coordinate have the same form, replacing each occurrence of $x_\varepsilon^{(k)}$ with $y_\varepsilon^{(k)}$.

Three of the derivatives in eqn. 13 simplify to the same result. In both $\Delta_{\varepsilon\beta\gamma\delta}^{(kj)}$ and $\Delta_{\varepsilon\beta\gamma\delta}^{(kk)}$ (which appear in the first and second terms of eqn. 13) only one difference of distances depends on $x_k^{(\varepsilon)}$, and so we may simplify these expressions:

$$\frac{\partial \Delta_{\varepsilon\beta\gamma\delta}^{(kj)}}{\partial x_\varepsilon^{(k)}} = \frac{\partial \Delta_{\varepsilon\beta\gamma\delta}^{(kk)}}{\partial x_\varepsilon^{(k)}} = \frac{\partial}{\partial x_\varepsilon^{(k)}} r_{\varepsilon\beta}^{(k)}. \quad (14)$$

This derivative also appears in the fourth term of eqn. 13 and evaluates to

$$\frac{\partial r_{\varepsilon\beta}^{(k)}}{\partial x_\varepsilon^{(k)}} = \frac{1}{r_{\varepsilon\beta}^{(k)}}\left(x_\varepsilon^{(k)} - x_\beta^{(k)}\right). \quad (15)$$

The remaining derivative (the third term in eqn. 13) features $\Delta_{\varepsilon\beta\varepsilon\delta}^{(kk)}$, in which both distances depend on $x_\varepsilon^{(k)}$, and so has two contributions:

$$\frac{\partial \Delta_{\varepsilon\beta\varepsilon\delta}^{(kk)}}{\partial x_\varepsilon^{(k)}} = \frac{1}{r_{\varepsilon\beta}^{(k)}}\left(x_\varepsilon^{(k)} - x_\beta^{(k)}\right) - \frac{1}{r_{\varepsilon\delta}^{(k)}}\left(x_\varepsilon^{(k)} - x_\delta^{(k)}\right). \quad (16)$$

References

1 L. Cademartiri and K. J. M. Bishop, *Nat. Mater.*, 2015, **14**, 2–9.
2 S. Sacanna, D. J. Pine and G.-R. Yi, *Soft Matter*, 2013, **9**, 8096–8106.
3 A. Walther and A. H. E. Müller, *Chem. Rev.*, 2013, **113**, 5194–5261.
4 G. Zhang, D. Wang and H. Möhwald, *Angew. Chem., Int. Ed.*, 2005, **44**, 7767–7770.
5 Y. Wang, Y. Wang, D. R. Breed, V. N. Manoharan, L. Feng, A. D. Hollingsworth, M. Weck and D. J. Pine, *Nature*, 2012, **491**, 51–55.
6 Q. Chen, S. C. Bae and S. Granick, *Nature*, 2011, **469**, 381–384.
7 A. B. Pawar and I. Kretzschmar, *Macromol. Rapid Commun.*, 2010, **31**, 150–168.
8 E. Bianchi, R. Blaak and C. N. Likos, *Phys. Chem. Chem. Phys.*, 2011, **13**, 6397–6410.
9 G.-R. Yi, D. J. Pine and S. Sacanna, *J. Phys.: Condens. Matter*, 2013, **25**, 193101.
10 M. Huang, C.-H. Hsu, J. Wang, S. Mei, X. Dong, Y. Li, M. Li, H. Liu, W. Zhang, T. Aida, W.-B. Zhang, K. Yue and S. Z. D. Cheng, *Science*, 2015, **348**, 424–428.
11 E. Bianchi, B. Capone, I. Coluzza, L. Rovigatti and P. D. J. van Oostrum, *Phys. Chem. Chem. Phys.*, 2017, **19**, 19847–19868.
12 Y. Ke, L. L. Ong, W. M. Shih and P. Yin, *Science*, 2012, **338**, 1177–1183.
13 Z. Zeravcic, V. N. Manoharan and M. P. Brenner, *PNAS*, 2014, **111**, 15918–15923.
14 W. M. Jacobs and D. Frenkel, *J. Am. Chem. Soc.*, 2016, **138**, 2457–2467.
15 A. Reinhardt and D. Frenkel, *Soft Matter*, 2016, **12**, 6253–6260.
16 W. Liu, J. Halverson, Y. Tian, A. V. Tkachenko and O. Gang, *Nat. Chem.*, 2016, **8**, 867–873.
17 D. Frenkel, *Nat. Mater.*, 2015, **14**, 9–12.
18 M. R. Jones, N. C. Seeman and C. A. Mirkin, *Science*, 2015, **347**, 1260901.
19 R. J. Macfarlane, M. N. O'Brien, S. H. Petrosko and C. A. Mirkin, *Angew. Chem., Int. Ed.*, 2013, **52**, 5688–5698.
20 C. Knorowski and A. Travesset, *J. Am. Chem. Soc.*, 2014, **136**, 653–659.
21 F. Lu, K. G. Yager, Y. Zhang, H. Xin and O. Gang, *Nat. Com-*

mun., 2015, **6**, 6912.
22 Y. Ke, L. L. Ong, W. Sun, J. Song, M. Dong, W. M. Shih and P. Yin, *Nat. Chem.*, 2014, **6**, 994–1002.
23 B. Wei, M. Dai and P. Yin, *Nature*, 2012, **485**, 623–626.
24 S. E. Ahnert, I. G. Johnston, T. M. A. Fink, J. P. K. Doye and A. A. Louis, *Phys. Rev. E*, 2010, **82**, 026117.
25 I. G. Johnston, S. E. Ahnert, J. P. K. Doye and A. A. Louis, *Phys. Rev. E*, 2011, **83**, 066105.
26 A. Reinhardt and D. Frenkel, *Phys. Rev. Lett.*, 2014, **112**, 238103.
27 L. O. Hedges, R. V. Mannige and S. Whitelam, *Soft Matter*, 2014, **10**, 6404–6416.
28 D. J. Wales, *J. Chem. Phys.*, 2017, **146**, 054406.
29 J. Madge and M. A. Miller, *J. Chem. Phys.*, 2015, **143**, 044905.
30 J. S. Richardson and D. C. Richardson, *Trends Biochem. Sci.*, 1989, **14**, 304–309.
31 N. P. King, W. Sheffler, M. R. Sawaya, B. S. Vollmar, J. P. Sumida, I. André, T. Gonen, T. O. Yeates and D. Baker, *Science*, 2012, **336**, 1171–1174.
32 N. P. King, J. B. Bale, W. Sheffler, D. E. McNamara, S. Gonen, T. Gonen, T. O. Yeates and D. Baker, *Nature*, 2014, **510**, 103–108.
33 B. Srinivasan, T. Vo, Y. Zhang, O. Gang, S. Kumar and V. Venkatasubramanian, *Proc. Nat. Acad. Sci. USA*, 2013, **110**, 18431–18435.
34 M. Z. Miskin, G. Khaira, J. J. d. Pablo and H. M. Jaeger, *Proc. Nat. Acad. Sci. USA*, 2016, **113**, 34–39.
35 B. A. Lindquist, R. B. Jadrich and T. M. Truskett, *J. Chem. Phys.*, 2016, **145**, 111101.
36 E. I. Shakhnovich, *Phys. Rev. Lett.*, 1994, **72**, 3907–3910.
37 I. Coluzza, H. G. Muller and D. Frenkel, *Phys. Rev. E*, 2003, **68**, 046703.
38 I. Coluzza and D. Frenkel, *Phys. Rev. E*, 2004, **70**, 051917.
39 F. DiMaio, A. Leaver-Fay, P. Bradley, D. Baker and I. André, *PLOS ONE*, 2011, **6**, e20450.
40 V. S. Pande, A. Y. Grosberg and T. Tanaka, *Biophys. J.*, 1997, **73**, 3192–3210.
41 I. Coluzza, *J. Phys. Cond. Matt.*, 2017, **29**, 143001.
42 Y. Sun and Y. Xia, *Science*, 2002, **298**, 2176–2179.
43 J. Zhang, A. Kumbhar, J. He, N. C. Das, K. Yang, J.-Q. Wang, H. Wang, K. L. Stokes and J. Fang, *J. Am. Chem. Soc.*, 2008, **130**, 15203–15209.
44 L. Rossi, S. Sacanna, W. T. M. Irvine, P. M. Chaikin, D. J. Pine and A. P. Philipse, *Soft Matter*, 2011, **7**, 4139–4142.
45 F. Smallenburg, L. Filion, M. Marechal and M. Dijkstra, *Proc. Nat. Acad. Sci. USA*, 2012, **109**, 17886–17890.
46 H. R. Vutukuri, F. Smallenburg, S. Badaire, A. Imhof, M. Dijkstra and A. van Blaaderen, *Soft Matter*, 2014, **10**, 9110–9119.
47 W. Yang, Y. Yu, L. Wang, C. Yang and H. Li, *Nanoscale*, 2015, **7**, 2877–2882.
48 R. P. Sear, *J. Chem. Phys.*, 2004, **120**, 998–1005.
49 S. Gottschalk, M. C. Lin and D. Manocha, *Proceedings of ACM Siggraph*, 1996, pp. 171–180.
50 B. S. John and F. A. Escobedo, *J. Phys. Chem. B*, 2005, **109**, 23008–23015.
51 J. P. K. Doye, A. A. Louis, I.-C. Lin, L. R. Allen, E. G. Noya, A. W. Wilber, H. C. Kok and R. Lyus, *Phys. Chem. Chem. Phys.*, 2007, **9**, 2197–2205.
52 G. Villar, A. W. Wilber, A. J. Williamson, P. Thiara, J. P. K. Doye, A. A. Louis, M. N. Jochum, A. C. F. Lewis and E. D. Levy, *Phys. Rev. Lett.*, 2009, **102**, 118106.
53 A. W. Wilber, J. P. K. Doye and A. A. Louis, *J. Chem. Phys.*, 2009, **131**, 175101.
54 S. Whitelam and P. L. Geissler, *J. Chem. Phys.*, 2007, **127**, 154101.
55 S. Whitelam and P. L. Geissler, *J. Chem. Phys.*, 2008, **128**, 219901.
56 S. Whitelam, E. H. Feng, M. F. Hagan and P. L. Geissler, *Soft Matter*, 2009, **5**, 1251–1262.
57 J. D. Bryngelson and P. G. Wolynes, *Proc. Nat. Acad. Sci. USA*, 1987, **84**, 7524–7528.
58 E. Jankowski and S. C. Glotzer, *J. Phys. Chem. B*, 2011, **115**, 14321–14326.
59 W. M. Jacobs, A. Reinhardt and D. Frenkel, *PNAS*, 2015, **112**, 6313–6318.
60 D. J. Wales and J. P. K. Doye, *J. Phys. Chem. A*, 1997, **101**, 5111–5116.
61 J. Madge, D. P. Bourne and M. A. Miller, *in preparation*.
62 T. E. Ouldridge, A. A. Louis and J. P. K. Doye, *J. Phys.: Condens. Matter*, 2010, **22**, 104102.
63 D. D. Boehr and P. E. Wright, *Science*, 2008, **320**, 1429–1430.
64 S. Sacanna, W. T. M. Irvine, P. M. Chaikin and D. J. Pine, *Nature*, 2010, **464**, 575–578.