Article

Self-assembly of emulsion droplets through programmable folding

Angus McMullen[1,4], Maitane Muñoz Basagoiti[2,4], Zorana Zeravcic[2✉] & Jasna Brujic[1,3✉]

In the realm of particle self-assembly, it is possible to reliably construct nearly arbitrary structures if all the pieces are distinct[1–3], but systems with fewer flavours of building blocks have so far been limited to the assembly of exotic crystals[4–6]. Here we introduce a minimal model system of colloidal droplet chains[7], with programmable DNA interactions that guide their downhill folding into specific geometries. Droplets are observed in real space and time, unravelling the rules of folding. Combining experiments, simulations and theory, we show that controlling the order in which interactions are switched on directs folding into unique structures, which we call colloidal foldamers[8]. The simplest alternating sequences (ABAB…) of up to 13 droplets yield 11 foldamers in two dimensions and one in three dimensions. Optimizing the droplet sequence and adding an extra flavour uniquely encodes more than half of the 619 possible two-dimensional geometries. Foldamers consisting of at least 13 droplets exhibit open structures with holes, offering porous design. Numerical simulations show that foldamers can further interact to make complex supracolloidal architectures, such as dimers, ribbons and mosaics. Our results are independent of the dynamics and therefore apply to polymeric materials with hierarchical interactions on all length scales, from organic molecules all the way to Rubik's Snakes. This toolbox enables the encoding of large-scale design into sequences of short polymers, placing folding at the forefront of materials self-assembly.

Self-assembly of materials currently requires a toolbox of building blocks with a given shape and a multitude of interaction flavours and strengths to ensure a unique product[9–11]. By contrast, achieving self-assembly of an arbitrary structure with high yield using a limited palette of flavours remains a key challenge. We therefore turn to the biological concept of self-assembly by the folding of linear chains, analogous to protein and RNA folding, and adapt it to materials science[12–14].

Our system consists of two flavours of colloidal droplets, labelled blue (A) and yellow (B), functionalized with complementary DNA strands (Methods). These droplets irreversibly bind with valence two to form the backbone of an alternating colloidomer[7,15], as depicted in Fig. 1a,b. The droplets are dispersed in an aqueous ferrofluid and we apply an intermittent magnetic field to accelerate the chaining process, giving rise to an exponential distribution of chain lengths, as shown in Extended Data Fig. 1. These chains are thermal and freely jointed because DNA diffuses on the surface even after the droplets are bound.

To mediate folding, each droplet flavour is in addition functionalized with DNA strands that act as weaker secondary interactions. Droplets have the advantage that they freely rearrange after binding, facilitating folding[16]. If all interactions are all simultaneously switched on, one obtains a mixture of folded geometries as the final product[17,18]. The number of possible geometries is singular for chains shorter than hexamers, but then grows exponentially with chain length. For example, an octamer can fold into nine distinct geometries, four of which are shown in Fig. 1b. By choosing DNA strands with distinct binding energies and therefore different melting temperatures[10] (Methods), we establish a hierarchy of bonds that are switched on as the temperature is lowered, as shown in Fig. 1c,d. Because the melting transition is sharp, working a few degrees below it ensures irreversible bond formation and downhill folding. For example, the decamer chain in Fig. 1d folds into the crown in a stepwise manner. First, the blue–blue palindrome interaction forms a pentamer core at high temperature, followed by the sequential locking in of yellow–blue and yellow–yellow bonds at progressively lower temperatures. Other protocols with a different sequence of secondary interactions are mediated by the same DNA strands, but grafted on droplets in different combinations (Methods).

Design of the folding landscape

Along the folding process, each new bond that forms causes the chain to adopt a different configuration. Those configurations that have the same contact matrix, ignoring chirality, are here defined to belong to a given state. All possible states between the linear chain and the final geometries map out an energy landscape that can be represented in a tree form[19]. In the folding tree in Fig. 2a, each row shows states with the same number of secondary bonds, that is, the same potential energy. Two states are connected in the tree if one can topologically transform into the other by making or breaking a single bond. Designing folding protocols, or the order of secondary droplet interactions, enables us to funnel the landscape to one final folded state.

[1]Center for Soft Matter Research, New York University, New York, NY, USA. [2]Gulliver UMR CNRS 7083, ESPCI Paris, Université PSL, Paris, France. [3]PMMH UMR CNRS 7636, ESPCI Paris, Université PSL, Paris, France. [4]These authors contributed equally: Angus McMullen, Maitane Muñoz Basagoiti. ✉e-mail: zorana.zeravcic@espci.fr; jb2929@nyu.edu
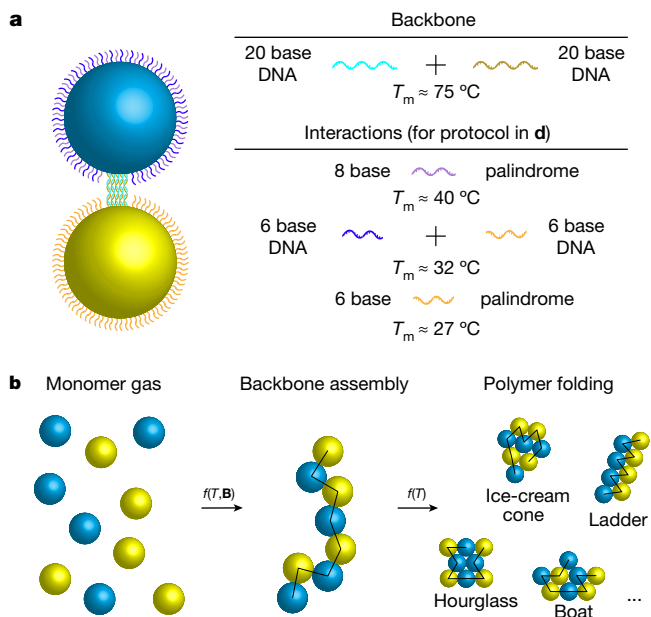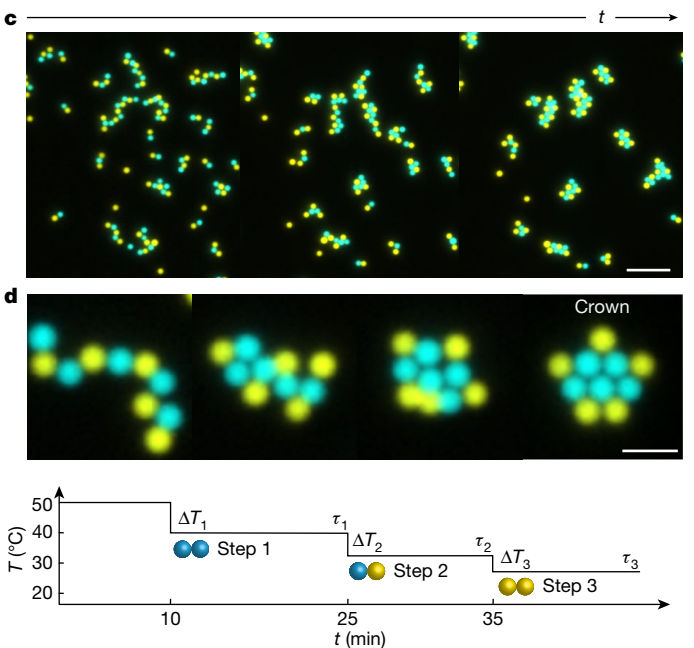
Fig. 1 | Colloidomer design and folding. a, Two flavours of droplets, A (blue) and B (yellow), are functionalized with complementary backbone strands of DNA to make alternating chains. They also carry weaker DNA interactions that mediate folding. The blue flavour carries two additional types of DNA, whereas the yellow flavour carries only one DNA strand that provides two distinct interactions. This 9 base pair (bp) DNA strand carries a consecutive complement to the 6 bp strand on the blue particle, activated at a melting temperature (T_m) around 32 °C, and 6 intermittent bp that are palindromes activated at T_m around 27 °C to mediate the yellow–yellow interaction (Methods). **b**, An emulsion first assembles into colloidomers using a magnetic field (**B**), after which a temperature (T) protocol triggers folding into diverse geometries. **c**, Fluorescent images show colloidomers of different lengths that undergo folding over time. Scale bar, 20 μm. **d**, A temperature protocol gives rise to stepwise folding, each step with a duration τ, of a decamer chain into the crown foldamer. Scale bar, 5 μm.

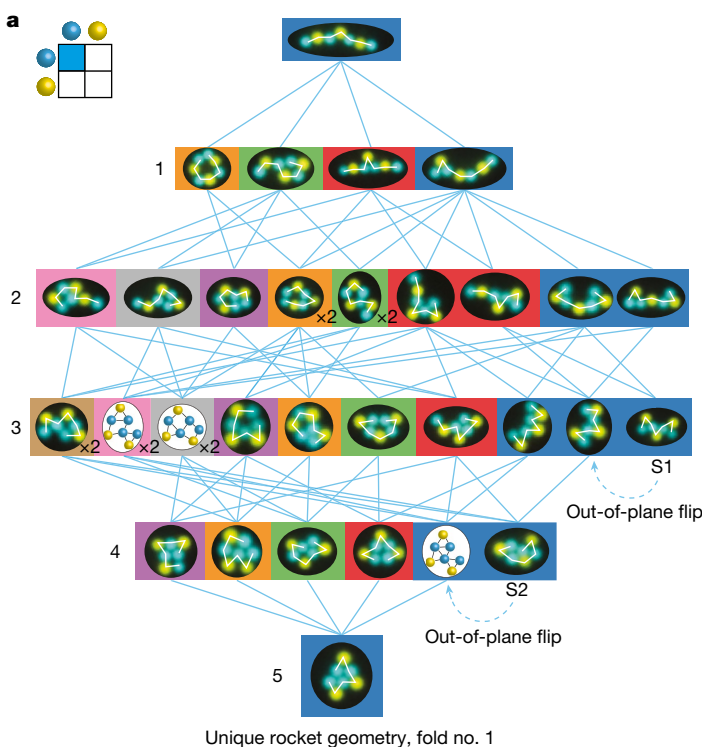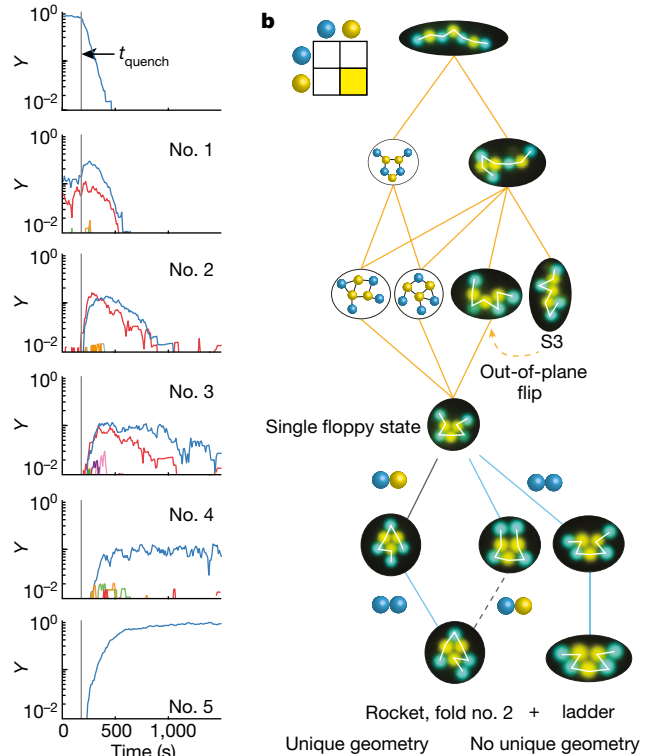**Fig. 2 | Folding pathways of a heptamer. a**, All folding pathways of a four-blue, three-yellow droplet heptamer result in a rocket foldamer when only the blue–blue interaction is turned on. Experimental images of states are superimposed with the theoretical tree, in order of frequency, to show the diversity of observed pathways. Each image contains an example of a backbone arrangement overlaid in white. The number of secondary bonds acquired is shown at each level of the tree. The plots on the right show the time evolution after the temperature quench t_{quench} of the yield Y of each colour-coordinated state. **b**, When the yellow–yellow interaction is switched on first, the same polymer folds into a single floppy state. Further interactions fold it into a rocket with a different fold, but reversing their order leads to a mixture of the rocket and the ladder. Note that switching on the yellow–blue bond last (dashed line) would require a different DNA strand design to that shown in Fig. 1a.
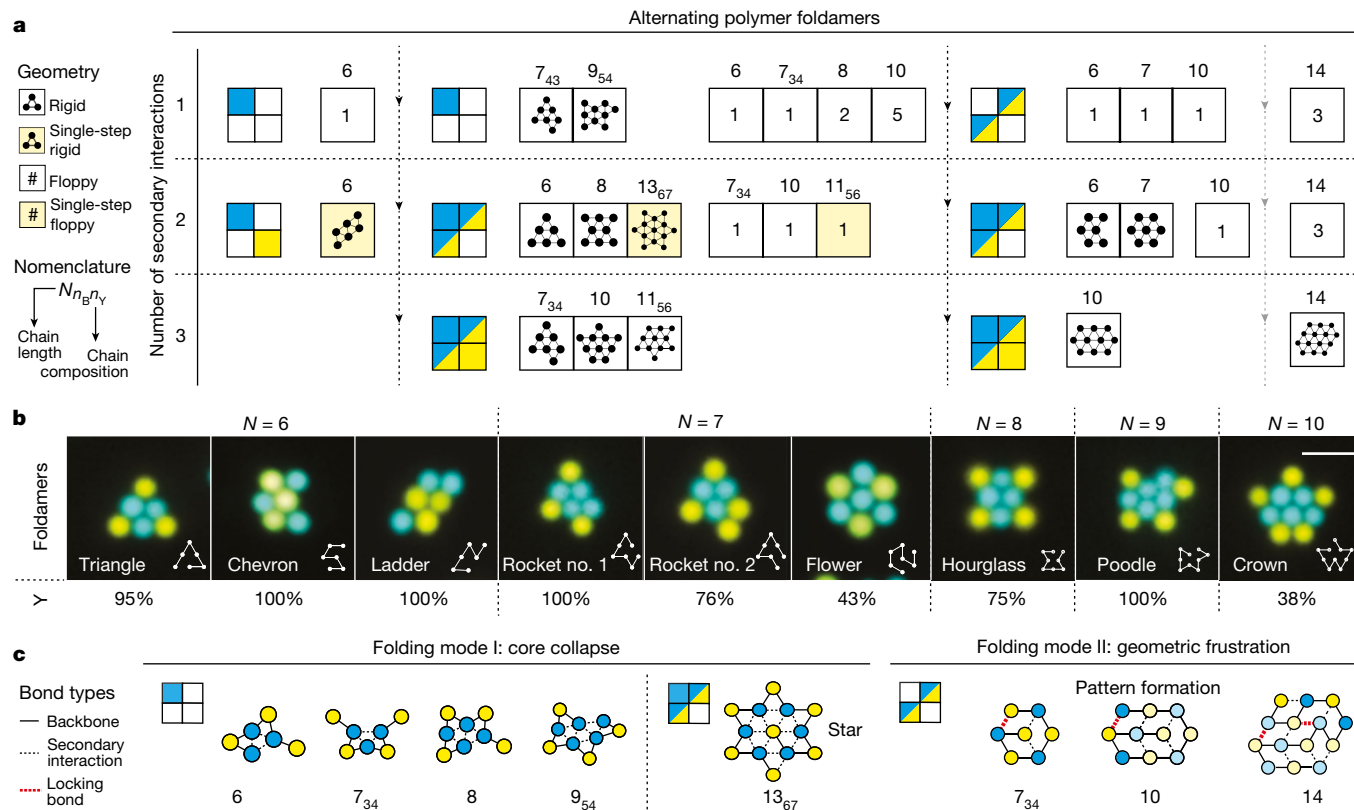
Fig. 3 | Predicted and experimentally realized foldamers. a, Alternating polymers of length $N = 6$–14 (subscripts indicate the number of blue and yellow droplets) can be successfully folded by distinct protocols (columns) with a maximum of three interactions (rows). Foldamers shaded in yellow require only one step, which can switch on one or more interactions. At the end of each step, foldamers are shown on the left and the number of floppy geometries on the right, in order of increasing chain length. **b**, Experimental results show fluorescent images of predicted foldamers up to decamers, as well as their relative folding yields Y. Scale bar, 5 μm. Experimental number of observations: for $N = 6$, [triangle, chevron, ladder] = (19, 86, 67); for $N = 7$, [rocket no. 1, rocket no. 2, flower] = (175, 25, 7); for $N = 8$, [hourglass] = 8; for $N = 9$, [poodle] = 24; for $N = 10$, [crown] = 8. **c**, Modes of folding: core collapse (left) and geometric frustration (right).

The example of an alternating heptamer chain in Fig. 2a shows that switching on only the blue–blue interaction yields a rocket foldamer as the final state. This tree was constructed theoretically and then populated by images of states that were observed along experimental folding pathways (Methods). The notable overlap between experiment and theory indicates that the experiments are sampling all the available states. Tracking $n = 255$ folding heptamers enables us to plot the evolution of the yield Y of the most popular states in each level of the tree in the side panels. Long-lived states correspond to local minima (states S1 and S2 in the tree) that are theoretical dead-ends, but are overcome in experiments because our system is quasi-two-dimensional and rare out-of-plane rearrangements are possible. As a result, all pathways lead to the rocket foldamer out of the four possible heptamer geometries on a timescale of approximately 20 min.

Because the heptamer comprises four blue and three yellow droplets, switching on the yellow–yellow interaction funnels the landscape into a much simpler tree, as shown in Fig. 2b. Here the final state is a unique floppy state that needs additional interactions to become rigid. Subsequently turning on the blue–blue interaction yields two new floppy states, one of which closes into a rigid ladder, whereas the other requires the remaining blue–yellow interaction to fold into the rocket shape. This particular protocol yields a mixture of the ladder and the rocket and does not qualify as a successful protocol. On the other hand, reversing the order of the last two steps leads only to the rocket foldamer, but with a different colour arrangement, or fold, to the one obtained from a single blue–blue interaction in Fig. 2a. This feature demonstrates the robustness of geometry to the protocol.

Foldamer search algorithm

In search of foldamers, we sweep all protocols for folding alternating sequences. The construction of folding trees becomes computationally expensive as the chain length grows, so we devise an alternative strategy for a systematic search (Extended Data Fig. 2), which enables us to reach chains with $N = 15$ droplets. We start by enumerating only the rigid states[17,20] and we map out all the possible backbone arrangements therein (Methods). Superimposing the alternating sequence on the backbones, we add secondary bonds between neighbouring droplets according to a specific interaction matrix. The resulting states are then classified as local or global minima. Keeping track of the minima each time an interaction is added, we determine if a colloidomer eventually folds into a unique geometry for a given sequence of interactions steps. The algorithm relies on the assumption that interactions are irreversible and that all bonds form, which requires a long enough waiting time at each temperature step in the experiment. This strategy is general for any linear polymer that can freely rearrange during folding by hierarchical interactions.
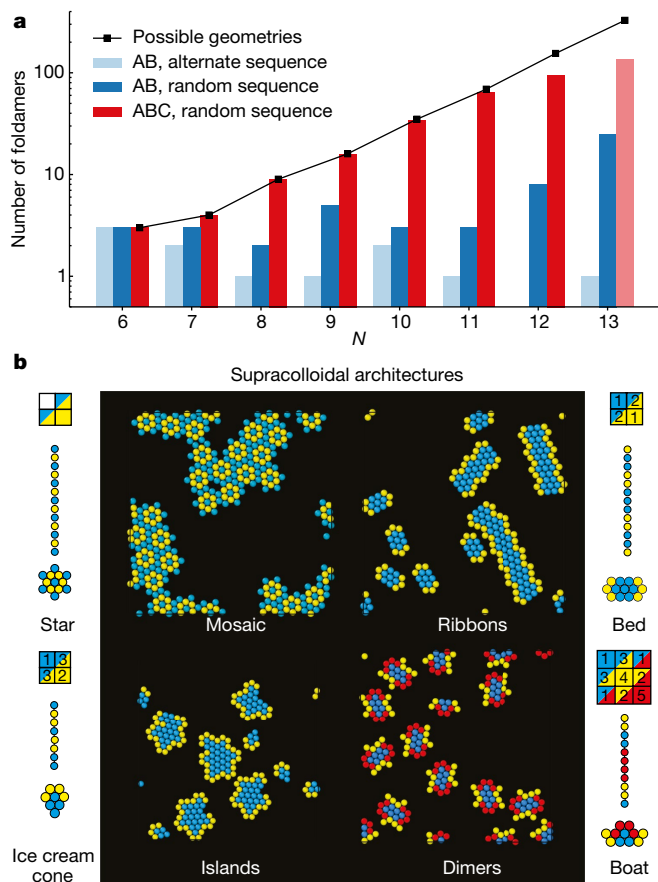
Fig. 4 | Colloidomer sequence design and supracolloidal assembly.
a, Exponential growth of the number of possible rigid geometries as a function of chain length N (black line). Numbers of foldamers encoded by an alternating AB sequence (light blue), any AB sequence (dark blue) and any ABC sequence (red) via all available protocols are shown as bars (the $N = 13$ bar is a lower bound). **b**, Simulated examples of supracolloidal self-assemblies using specific interaction protocol, shown with the droplet sequence that gives that foldamer. Numbers indicate the order in which interactions are activated (Extended Data Fig. 4b).

Alternating sequence foldamers

Our theory systematically identifies successful protocols that yield a total of 11 foldamer geometries for chains up to 13 droplets long, as shown in Fig. 3a. Following those protocols, experiments capture most of the predicted foldamers, as shown in Fig. 3b and Supplementary Videos 1–7. High relative yields, defined as the proportion of rigid structures that reach the correct geometry (Methods), are achieved in all but the flower and crown foldamers, owing to floppy dead-ends they encounter on timescales beyond the experimental window. This may explain why single-step quenches have perfect yields of rigid structures, whereas multiple quenches on average have lower yields. The incorrect structures arising from local minima can be suppressed by optimizing the bond strength, as shown in the simulations in Extended Data Figs. 3 and 4a.

Our foldamers demonstrate that the simplest alternating sequence encodes all the possible geometries of the hexamer: the ladder, the chevron and the triangle, as shown in Fig. 3b. Among longer foldamers, only the heptamer flower and the decamer bed correspond to the ground state of a folded homocolloidomer, whereas the rest are unlikely geometries in equilibrium[21]. For example, the octamer hourglass geometry has the highest free energy, that is, the smallest yield among the nine possible geometries because of its high symmetry

number[17,18]. Therefore, our foldamers correspond to kinetic states that are accessible on the basis of geometric considerations alone. Another example is the nonamer poodle, which is the longest chain that can be folded with a single interaction. By contrast, the decamer folds into the crown through a many-to-one transition, as an example of a funnel-like landscape[22].

Colloidomer folding mechanisms

More generally, alternating colloidomers follow two mechanisms to reach the foldamer state: core collapse and geometric frustration, as illustrated in Fig. 3c. The most common mechanism is the core collapse, which first forms a rigid core and then locks in the remaining droplets on the outside. Up to decamers, the cores consist of a maximum of five identical droplets in unique geometries. Beyond this length, foldamers comprise multiflavoured cores formed upon turning on two interactions simultaneously, as seen in the star foldamer.

The second mechanism of geometric frustration initially engages an interaction that traps the droplets by certain locking bonds into positions in which they are surrounded by neighbours with which they cannot form secondary bonds. Turning on other interactions adds the remaining bonds without changing the geometry. The Russian doll architecture of these foldamers as a function of N allows us to successfully predict the $N = 14$ foldamer following the same protocol, as shown in Fig. 3c.

From sequence to supracolloidal design

Next, we theoretically investigate how increasing complexity[23] improves the number and variety of possible foldamers. We run our search algorithm across all possible droplet flavour sequences, while preserving the number of each flavour in the chain. This process uncovers winning protocols, increasing the total number of foldamers by roughly an order of magnitude, particularly in longer chains, as shown in Fig. 4a (dark blue). Note that chains with at least $N = 13$ droplets are able to encode foldamers with stable holes (Extended Data Fig. 5), which can serve as precise sieves and offer porous design. In addition, the introduction of a third flavour while designing in both sequence and protocol spaces identifies more than a half of all possible geometries up to tridecamers, giving in total 310 foldamers (red). Whereas two flavours code for all three geometries in hexamers, three letters encode all geometries up to decamers, putting a bound on what can be achieved with a small number of flavours as a function of N (refs. [24,25]). To achieve these sequence-specific foldamers experimentally would require sequential droplet polymerization, as previously demonstrated in ref. [26].

With this lexicon of foldamers as building blocks, simulations show that they self-assemble by additional supracolloidal interactions into higher order architectures[27], as shown in the simulated examples in Fig. 4b. For instance, an interaction between blue droplets assembles star foldamers into a complex mosaic. Foldamers with polarized flavours self-assemble into ribbons or islands, whereas three flavours facilitate the assembly of unique dimers. All these examples could be experimentally realized if the chains were segregated by length (Methods), diluted and the DNA strands were subsequently activated for further assembly (Methods).

Our minimal model system exhibits many of the phenomena nominally associated with protein folding. Foldamers consisting of droplets with two or three flavours have the properties of uniqueness, robustness and kinetic accessibility in a funnel landscape[22,28]. The core collapse folding mechanism resembles the hydrophobic collapse in proteins[29], whereas that of geometric frustration has been proposed as a design principle in the assembly of peptides[30]. On the supracolloidal scale, foldamer assembly mimics the polymerization of fibrils[31], the formation of protein-based micelles[32] or protein dimerization[33]. These similarities

Article

occur even though our system is strictly out-of-equilibrium, highlighting the importance of geometry in guiding assembly.

Colloidal self-assembly has the advantage that the monomers are easily visualized under a microscope, giving access to the underlying rules that govern successful folding by dissecting the respective roles of sequence design, minimal number of flavours, hierarchy of interactions and topological constraints. This type of structural design influences function. Once folded, emulsions are readily polymerized to make solid two-dimensional patterns on the scale of the wavelength of near infrared light, enabling one to tune their optical properties.

Moreover, sequential secondary interactions can be programmed to fold into three-dimensional foldamers (Supplementary Videos 9 and 10). Using smaller droplets allows them to explore the available phase space in three dimensions more efficiently. An alternating hexamer uniquely gives a polytetrahedron following a three-step protocol, which we experimentally demonstrate with 100% yield ($n = 5$). Self-assembly of geometric clusters paves an alternative path towards materials with photonic band gaps, such as the colloidal diamond[6]. Instead of using droplets, one can imagine folding molecular polymers designed with hydrophobic and polar moieties[8], or building macroscopic beads-on-a-string models with specific interactions, facilitated by an external drive[12]. This new paradigm of hierarchical folding as a precursor for large-scale self-assembly offers design rules for biomimetic materials with tunable functionalities[34].

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-022-05198-8.

1. Ong, L. L. Programmable self-assembly of three-dimensional nanostructures from 10,000 unique components. *Nature* **552**, 72–77 (2017).
2. Zeravcic, Z., Manoharan, V. N. & Brenner, M. P. Size limits of self-assembled colloidal structures made using specific interactions. *Proc. Natl Acad. Sci. USA* **111**, 15918–15923 (2014).
3. Jacobs, W. M., Reinhardt, A. & Frenkel, D. Rational design of self-assembly pathways for complex multicomponent structures. *Proc. Natl Acad. Sci. USA* **112**, 6313–6318 (2015).
4. Macfarlane, R. J. Nanoparticle superlattice engineering with DNA. *Science* **334**, 204–208 (2011).
5. Wang, Y., Jenkins, I. C., McGinley, J. T., Sinno, T. & Crocker, J. C. Colloidal crystals with diamond symmetry at optical lengthscales. *Nat. Commun.* **8**, 14173 (2017).
6. He, M. Colloidal diamond. *Nature* **585**, 524–529 (2020).
7. McMullen, A., Holmes-Cerfon, M., Sciortino, F., Grosberg, A. Y. & Brujic, J. Freely jointed polymers made of droplets. *Phys. Rev. Lett.* **121**, 138002 (2018).
8. Hill, D. J., Mio, M. J., Prince, R. B., Hughes, T. S. & Moore, J. S. A field guide to foldamers. *Chem. Rev.* **101**, 3893–4011 (2001).
9. Rogers, W. B., Shih, W. M. & Manoharan, V. N. Using dna to program the self-assembly of colloidal nanoparticles and microparticles. *Nat. Rev. Mater.* **1**, 16008 (2016).
10. Seeman, N. C. & Sleiman, H. F. DNA nanotechnology. *Nat. Rev. Mater.* **3**, 17068 (2017).
11. Talapin, D. V., Engel, M. & Braun, P. V. Functional materials and devices by self-assembly. *MRS Bull.* **45**, 799–806 (2020).
12. Reches, M., Snyder, P. W. & Whitesides, G. M. Folding of electrostatically charged beads-on-a-string as an experimental realization of a theoretical model in polymer science. *Proc. Natl Acad. Sci. USA* **106**, 17644–17649 (2009).
13. Coulais, C., Sabbadini, A., Vink, F. & van Hecke, M. Multi-step self-guided pathways for shape-changing metamaterials. *Nature* **561**, 512–515 (2018).
14. Dodd, P. M., Damasceno, P. F. & Glotzer, S. C. Universal folding pathways of polyhedron nets. *Proc. Natl Acad. Sci. USA* **115**, 6690–6696 (2018).
15. McMullen, A., Hilgenfeldt, S. & Brujic, J. Dna self-organization controls valence in programmable colloid design. *Proc. Natl Acad. Sci. USA* **118**, e2112604118 (2021).
16. Zhang, Y. Multivalent, multiflavored droplets by design. *Proc. Natl Acad. Sci. USA* **115**, 9086–9091 (2018).
17. Meng, G., Arkus, N., Brenner, M. P. & Manoharan, V. N. The free-energy landscape of clusters of attractive hard spheres. *Science* **327**, 560–563 (2010).
18. Perry, R. W., Holmes-Cerfon, M. C., Brenner, M. P. & Manoharan, V. N. Two-dimensional clusters of colloidal spheres: ground states, excited states, and structural rearrangements. *Phys. Rev. Lett.* **114**, 228301 (2015).
19. Zeravcic, Z. & Brenner, M. P. Spontaneous emergence of catalytic cycles with colloidal spheres. *Proc. Natl Acad. Sci. USA* **114**, 4342–4347 (2017).
20. Holmes-Cerfon, M. C. Enumerating rigid sphere packings. *SIAM Rev.* **58**, 229–244 (2016).
21. Trubiano, A. & Holmes-Cerfon, M. Thermodynamic stability versus kinetic accessibility: Pareto fronts for programmable self-assembly. *Soft Matter* **17**, 6797–6807 (2021).
22. Bryngelson, J. D., Onuchic, J. N., Socci, N. D. & Wolynes, P. G. Funnels, pathways, and the energy landscape of protein folding: a synthesis. *Proteins: Struct. Funct. Genet* **21**, 167–195 (1995).
23. Du, R. & Grosberg, A. Y. Models of protein interactions: how to choose one. *Fold. Des.* **3**, 203–211 (1998).
24. Guarnera, E., Pellarin, R. & Caflisch, A. How does a simplified-sequence protein fold? *Biophys. J.* **97**, 1737–1746 (2009).
25. Fink, T. M. & Ball, R. C. How many conformations can a protein remember? *Phys. Rev. Lett.* **87**, 198103 (2001).
26. Zhang, Y. Sequential self-assembly of DNA functionalized droplets. *Nat. Commun.* **8**, 21 (2017).
27. Oh, J. S., Lee, S., Glotzer, S. C., Yi, G.-R. & Pine, D. J. Colloidal fibers and rings by cooperative assembly. *Nat. Commun.* **10**, 3936 (2019).
28. Anfinsen, C. B. Principles that govern the folding of protein chains. *Science* **181**, 223–230 (1973).
29. Gutin, A., Abkevich, V. & Shakhnovich, E. Is burst hydrophobic collapse necessary for protein folding? *Biochemistry* **34**, 3066–3076 (1995).
30. Jiang, T., Magnotti, E. L. & Conticello, V. P. Geometrical frustration as a potential design principle for peptide-based assemblies. *Interface Focus* **7**, 20160141 (2017).
31. Mohapatra, L., Goode, B. L. & Kondev, J. Antenna mechanism of length control of actin cables. *PLoS Comp. Biol.* **11**, e1004160 (2015).
32. Friedrich, P. *Supramolecular Enzyme Organization: Quaternary Structure and Beyond* (Elsevier, 1984).
33. Marianayagam, N. J., Sunde, M. & Matthews, J. M. The power of two: protein dimerization in biology. *Trends Biochem. Sci.* **29**, 618–625 (2004).
34. Zeravcic, Z., Manoharan, V. N. & Brenner, M. P. Colloquium: toward living matter with colloidal particles. *Rev. Mod. Phys.* **89**, 031001 (2017).

Methods

Droplet synthesis

Monodisperse polydimethylsiloxane droplets were synthesized according to a protocol modified from that outlined in refs. [7,15,26]. An equal volume of dimethoxydimethysilane (Sigma Aldrich) and (3,3,3-trifluoropropyl)methyldimethoxysilane (Gelest) was mixed together with DI water at approximately 2% v/v. The monomers were prehydrolysed by vortexing for 60 min. Ammonia was added at 1% v/v, and the droplets were left to grow for 24 h. The droplets were then dialysed against 5 mM sodium dodecyl sulfate (SDS, Sigma Aldrich) to remove the remaining ammonia and reaction byproducts. We then incubated the droplets in 1% volume of (3-glycidoxypropyl) methyldiethoxysilane (Gelest) with 10 mM sodium azide and 5 mM SDS. This embedded reactive azide groups inside the droplets, such that they can be fluorescently labelled at a later stage. This synthesis produced monodisperse oil droplets that were denser than water with a low gravitational height, forming a quasi-two-dimensional system.

DNA sequences and their interactions

The following is a complete list of DNA sequences used in this work, listed with their modifications from 5′ to 3′. The strands which formed the interactions were as follows:

A: azide Cy3A GCA TTA CTT TCC GTC CCG AGA GAC CTA ACT GAC ACG CTT CCC ATC GCT A GA GTT CAC AAG AGT TCA CAA

B: azide Cy5 A GCA TTA CTT TCC GTC CCG AGA GAC CTA ACT GAC ACG CTT CCC ATC GCT A TT GTG AAC TCT TGT GAA CTC

C: azide AG CAT TAC TTT CCG TCC CGA GAG ACC TAA CTG ACA CGC TTC CCA TCG CTA TTT TTA GTC

D: azide AG CAT TAC TTT CCG TCC CGA GAG ACC TAA CTG ACA CGC TTC CCA TCG CTA TTT GAC TAA

P: azide AG CAT TAC TTT CCG TCC CGA GAG ACC TAA CTG ACA CGC TTC CCA TCG CTA TTT ATC GAT

CS: TAG CGA TGG GAA GCG TGT CAG TTA GGT CTC TCG GGA CGG AAA GTA ATG CT azide

The strongest DNA interaction is the 20 bp hybridization of A and B strands providing the colloidomer (blue–yellow) backbone. In typical experimental conditions, the backbone melts at around 75 °C.

The remaining strands provide a hierarchy of secondary interactions strengths to mediate sequential folding:

(1) The strongest secondary interaction is realized by the P strand through palindromic self-interaction. In typical experimental conditions, it melts between 40 °C and 45 °C. This strand facilitates homophilic blue–blue or yellow–yellow interactions.

(2) A weaker secondary interaction is mediated by the complementary interaction of C and D strands, which melt between 30 °C and 35 °C. This interaction facilitates secondary yellow–blue bonds.

(3) The weakest interaction is provided by the D strand, by a weak palindromic self-interaction. In typical experimental conditions, it melts around 27 °C. This strand facilitates homophilic blue–blue or yellow–yellow interactions.

In Fig. 3b, we show foldamers obtained via three protocols, each of which uses a different combination of the DNA interactions coating the droplets. Protocol I uses interactions 1 and 3 (giving the ladder foldamer). Protocol II uses interactions 1, 2 and 3 (giving the triangle, rocket, hourglass, poodle and crown foldamers). Protocol III uses interactions 2 and 3 (giving the chevron and flower foldamers).

DNA-labelling of emulsion droplets

Before labelling with DNA, emulsion droplets were diluted into 1 mM SDS at a volume fraction of approximately 6%. DNA strands with sticky ends were reacted with a DBCO terminated pegylated lipid (DPSE-PEG-DBCO, Avanti Polar Lipids), and then annealed with a complementary spacer strand as described in refs. [7,15]. Droplets were incubated with backbone DNA at 200 nM concentrations with a volume fraction of 0.6% with 50 mM NaCl, 10 mM Tris pH 8 and 1 mM EDTA. After 30 min, secondary interaction DNA was added, bringing the total concentration to 5–25 μM. The droplets were then incubated for 2 h before being diluted by a factor of two with a buffer containing 50 mM NaCl, 10 mM Tris pH 8, 0.1% w/v Triton 165 and Cyanine 3 DBCO (or Cyanine 5 DBCO, both from Lumiprobe). The droplets were incubated for a further 30 min before being washed several times in 50 mM NaCl to remove all unreacted dye.

Colloidomer formation

Droplet polymerization was accelerated by dispersing the droplets in an aqueous ferrofluid (EMG 707, FerroTec) and aligning them with a magnetic field. The ferrofluid was washed several times into 0.3% F68 pluronic surfactant by centrifugation to remove the proprietary surfactant in the ferrofluid. Two sets of droplets were prepared with complementary backbone DNAs and secondary DNA strands of choice. The two droplet types were mixed at a 1:1 ratio along with a 1:3 dilution of the F68 ferrofluid buffer, 200 mM NaCl and 20 mM EDTA pH 8. The sample was added to a custom flow chamber made from a hexamethyldisilazane (Sigma Aldrich) treated glass slide and coverslip and parafilm. The flow cell was sealed with ultraviolet glue.

The sample was then heated up to 75 °C to break all bonds in the system, and then cooled down to just above the melting temperature of the strongest secondary interaction, typically 50 °C. The sample was then put through a repeated cycle of alignment with rare earth magnets and relaxation to grow the chains. Typically, this produced a mixed sample of monomers, linear chains and branched chains. The density of droplets was optimized such that they would grow sizable polymer chains, but that the chains would not aggregate on the timescale of the folding experiments. The colloidomers were allowed to relax in the absence of a magnetic field before the folding data were taken. Data were taken using a Nikon TI Eclipse with a ×20 objective using either single- or double-channel fluorescence imaging.

Temperature protocols and waiting times

The temperature was adjusted using a custom-made heating cell composed of an indium tin oxide coated glass slide (SPI) connected to a Thorlabs TC200 resistive heater with a thermocouple for feedback. The temperature protocol was programmed through custom software. For a given temperature protocol, first a sample of droplet polymers with the desired set of interactions was made. A manual sweep of the temperature was performed to determine where each interaction takes place, as the melting temperatures can change from sample to sample. The first temperature step lasting 10 min was programmed to be above the melting temperature of all interactions to identify the unfolded colloidomers.

Subsequently, there can be one, two or three additional steps depending on how many interactions are to be turned on. If there is more than one interaction that is turned on, the waiting step for the first interaction is the longest. For the data in Fig. 3c, the waiting time at the first step was 20 min (except for the $N = 6$ triangle, which had a waiting time of 30 min), whereas that for the second and third steps was typically 5–10 min. In principle, longer waiting times enable the resolution of local minima and lead to better yields. In practice, however, longer waiting times increase the chance that colloidomers aggregate during folding, which can be avoided in dilute samples.

Video analysis

Folding videos were analysed using a custom MATLAB data analysis software. All particles were identified and located using thresholding. These particles were then tracked through the whole video using custom software modelled after that in ref. [35]. Polymers were identified using the same metrics as in ref. [7] from the first 10 min of every recording, which was always above the melting temperature of the strongest secondary interaction. An $N \times N \times t$ (where N is the number of monomers in the polymer and t is the time) connectivity matrix was then calculated for each polymer using the particle locations and diameters. The

contact matrix was median filtered over t to remove transient interactions. Each contact matrix was then matched to a polymer configuration theoretically computed, allowing us to track the polymer configuration over time. Selections of data were vetted by hand afterwards to ensure the integrity of the data. Polymers that aggregated or that folded into three-dimensional structures were discarded.

In Fig. 2a, the plotted yields as a function of time of a given configuration are normalized by the total number of identified configurations having the same number of bonds, that is, ones within the same row of the folding tree. If a colloidomer is lost at a given time, that is, leaves the observational window, aggregates with another one or enters an unidentifiable configuration, it is removed from the analysis pool. For Fig. 3c, the yield is defined as the fraction of polymers of length N that fold to completion into the target rigid structure over the fraction of polymers of length N that fold to completion into any rigid structure of the same size. A chain-by-chain analysis reveals that the typical fraction of chains that successfully complete folding is on average 65%, ranging from 50 to 70% for chain lengths $N = 6\text{–}11$ droplets. In this work, the quoted folding yields consider only these chains. To increase the fraction of viable chains, our method could be improved with a larger density mismatch between the droplets and the aqueous phase to ensure two-dimensional folding, while using a sample cell with individual wells for each chain.

Possible extensions to colloidomer folding

To experimentally realize supracolloidal self-assembly, such as the ones shown in Fig. 4b, several further steps need to be taken. The emulsion polymerization protocol yields an exponential distribution of chain lengths shown in Extended Data Fig. 1. Therefore, our samples first need to be segregated by chain length. This could be achieved using the glycerol-based density gradient centrifugation method[36,37]. This method has been used to separate clusters of solid colloids with different size. It can now be extended to colloidomers consisting of emulsion droplets, as they are robust against centrifugation (see the washing steps of the current synthesis), and are not destabilized by glycerol, as shown by the refractive-index matching experiments in ref. [38].

To avoid chain aggregation, secondary interactions would be implemented using linker-mediated assembly[39–41]. The desired single-chain-length sample would then be diluted to low volume fraction to avoid aggregation during folding. The linker strands would then be added to implement the appropriate folding protocol. Temperature quenches can then be followed to create a uniform sample of foldamers for supracolloidal assembly. Once folded, the unused interactions can lead to supracolloidal architectures, such as in the case shown in Fig. 4b for which activation of the unused blue–blue bond in the star foldamer leads to the mosaic assembly. In other cases, specific binding between foldamers can be activated using strand displacement reactions[42] or triggered with linker-mediated interactions[41].

Enumerating two-dimensional geometries

We define as a geometry any colloidomer cluster in which deformations cost energy, that is, a deformation requires the breaking of a secondary bond. Geometries are therefore rigid clusters. To enumerate two-dimensional geometries for a system of size N, we start by selecting all possible sets of N neighbouring points on an $N \times N$ triangular lattice. We form bonds between points located at a unit distance and test the rigidity of the resulting geometries by analysing the normal modes of the dynamical matrix. We describe the ensemble of N_R geometries for a chain of length N by a set of planar graphs $\{G_{i,N}(V, E)\}$, with index $i \in (1, N_R)$, and where the vertices (V) are the droplets in the chain and the edges (E) are the DNA-mediated bonds. Edges may be of two types: backbone bonds and secondary bonds. Each graph is characterized by a contact matrix, which describes the bonds between droplets, and a distance matrix, which contains the distances between each droplet pair in a geometry. The first size with more than one geometry is $N = 6$ (ref. [18]). At $N \geq 13$ the first geometries with stable holes in the bulk appear.

Foldamer search algorithm

We develop a computationally efficient search algorithm to systematically scan protocol and sequence spaces and find foldamers of a given length N. The algorithm requires as input the ensemble of all backbone configurations within the geometries N_R for a chain of length N, that is, the set of Hamiltonian paths $\{H_{1,1}, \ldots, H_{p_1,1}, \ldots, H_{1,q}, \ldots, H_{p_q,q}\}$, for all $q \in (1, N_R)$, where p_q is the number of paths in the qth geometry. The total number of Hamiltonian paths grows exponentially and it does not depend on the sequence or the interaction matrix. Thus, they are computed only once per N, significantly reducing the computation time. The structure of the algorithm is shown in the Extended Data Fig. 1. For a given protocol and sequence, the algorithm can be summarized as follows:

Input. Map the sequence onto Hamiltonian paths.
(1) Form bonds. Apply the first interaction of the protocol. A bond will be formed between two vertices if they are in neighbouring lattice points and the interaction is allowed.
(2) Are there geometries?
  (i) Yes. If the classification flags geometries, the algorithm stops. If there is a single geometry, a foldamer is reported. We choose to report a solution even if there are competing floppy states with the same or more bonds as the foldamer geometry (this becomes possible when $N \geq 7$).
  (ii) No. A foldamer is not selected.
(3) Select global minima. This is analogous to selecting floppy states with the largest number of bonds. Note that this also implies that local minima in the first interaction tree are not considered (here we assume strict downhill folding).
(4) Continue the protocol of adding interactions. Update the interaction matrix according to the protocol.
(5) Form new bonds. Repeat the bond-making process iterating over the states from step 3.
(6) Classify states. We classify states into global and local minima, and transient states. Global minima are states of a tree that cannot acquire additional bonds either because they reached a rigid state or because spatially accessible neighbours do not have flavours with attractive interactions. Local minima are floppy states for which the topology prevents further formation of bonds. All other states are classified as transient states.
(7) Is the protocol over?
  (i) Yes. Analyse the resulting geometries. If a single geometry is found, a foldamer is reported.
  (ii) No. Repeat steps 4–7 until the protocol ends.

Simulation details

We perform Dissipative Particle Dynamics (DPD)[43] simulations using an in-house code. Our unit of length is the particle diameter $\sigma = 1$ and we assume all particles have the same mass $m = 1$. Energy is measured in units of $k_B T$, where k_B is the Boltzmann constant, and we fix the temperature of the system at $k_B T = 1$. When folding a colloidomer of length N, we set the simulation box size L to $L/\sigma = (N + 2)$. For the self-assembly of supracolloidal architectures, we choose $L/\sigma = 30$. In both cases we use periodic boundary conditions. We use a multiple-timestep simulation scheme to integrate the equations of motion with a timestep $dt_s = 10^{-2}$ to resolve the dynamics of the solvent and a timestep $dt_c = 10^{-4}$ for the dynamics of the colloids. DNA-mediated interactions are modelled by a short-range, isotropic interaction potential[44]

$$U(r) = \varepsilon \alpha(r_i, \sigma) \left[\left(\frac{\sigma}{r} \right)^2 - 1 \right] \left[\left(\frac{r_i}{r} \right)^2 - 1 \right]^2, \tag{1}$$

where r is the distance between two interacting particles, $r_i = 1.05\sigma$ is the interaction range, ε is the strength of the interaction and α is a

parameter that sets the minimum of the potential $U(r_{min}) = \varepsilon$ (see ref. [44] for further details). Primary bonds are made irreversible by setting $\varepsilon_P = 40k_BT$. To simulate secondary interactions, we gradually increase ε until it reaches ε_S, once the corresponding interaction is turned on. The increase is done over the course of 200 simulation steps to ensure downhill folding while preventing poor potential sampling.

Data availability

The data that support the findings of this study are available from the corresponding author upon request. The data includes experimental as well as computational datasets, Matlab scripts for experimental video analysis and Python scripts for computational dataset analysis.

Code availability

The custom computer codes to build folding trees, to identify foldamers and the Dissipative Particle Dynamics (DPD) code to simulate folding of colloidomers are available from the corresponding author upon request. We have provided pseudo-code for the enumeration code in Extended Data Figure 2. DPD is also available in open-source packages such as HOOMD and LAMMPS.

35. Guerra, R. E., Kelleher, C. P., Hollingsworth, A. D. & Chaikin, P. M. Freezing on a sphere. *Nature* **554**, 346–350 (2018).
36. Jerri, H. A., Sheehan, W. P., Snyder, C. E. & Velegol, D. Prolonging density gradient stability. *Langmuir* **26**, 4725–4731 (2009).
37. Wang, Y. Colloids with valence and specific directional bonding. *Nature* **491**, 51–55 (2012).
38. Pontani, L.-L., Jorjadze, I., Viasnoff, V. & Brujic, J. Biomimetic emulsions reveal the effect of mechanical forces on cell–cell adhesion. *Proc. Natl Acad. Sci. USA* **109**, 9839–9844 (2012).
39. Zhang, D. Y. & Seelig, G. Dynamic DNA nanotechnology using strand-displacement reactions. *Nat. Chem.* **3**, 103–113 (2011).
40. Mognetti, B. M., Leunissen, M. & Frenkel, D. Controlling the temperature sensitivity of dna-mediated colloidal interactions through competing linkages. *Soft Matter* **8**, 2213–2221 (2012).
41. Lowensohn, J., Oyarzún, B., Paliza, G. N., Mognetti, B. M. & Rogers, W. B. Linker-mediated phase behavior of dna-coated colloids. *Phys. Rev. X* **9**, 041054 (2019).
42. Srinivas, N. On the biophysics and kinetics of toehold-mediated DNA strand displacement. *Nucl. Acids Res.* **41**, 10641–10658 (2013).
43. Groot, R. D. & Warren, P. B. Dissipative particle dynamics: bridging the gap between atomistic and mesoscopic simulation. *J. Chem. Phys.* **107**, 4423–4435 (1997).
44. Wang, X., Ramírez-Hinestrosa, S., Dobnikar, J. & Frenkel, D. The Lennard-Jones potential: when (not) to use it. *Phys. Chem. Chem. Phys.* **22**, 10624–10633 (2020).

Author contributions A.M. designed the materials, synthesized the droplet system, performed the folding experiments and developed data analysis tools to extract time-dependent yields of foldamers. M.M.B. and Z.Z. constructed the theoretical model for generating folding trees and rigid states. M.M.B. developed the algorithm for enumerating foldamers, wrote the molecular dynamics codes and performed numerical simulations. J.B. and Z.Z. conceived the study and supervised the research. The manuscript was written by J.B. and Z.Z. together with A.M and M.M.B.

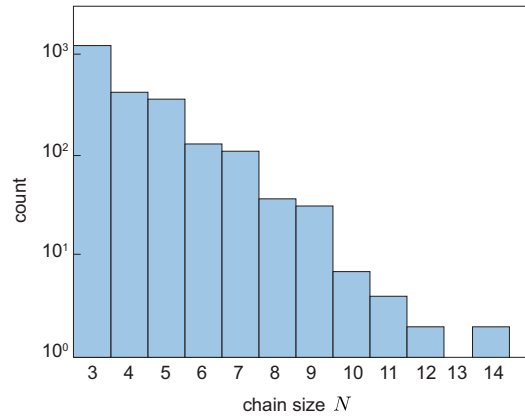**Competing interests** The authors declare no competing interests.

Extended Data Fig. 1 | Chain length distribution. This panel shows a distribution of chain lengths from a typical chaining experiment, on a semi-log scale. Observed chains in the range $N = 3$–14 are exponentially distributed.
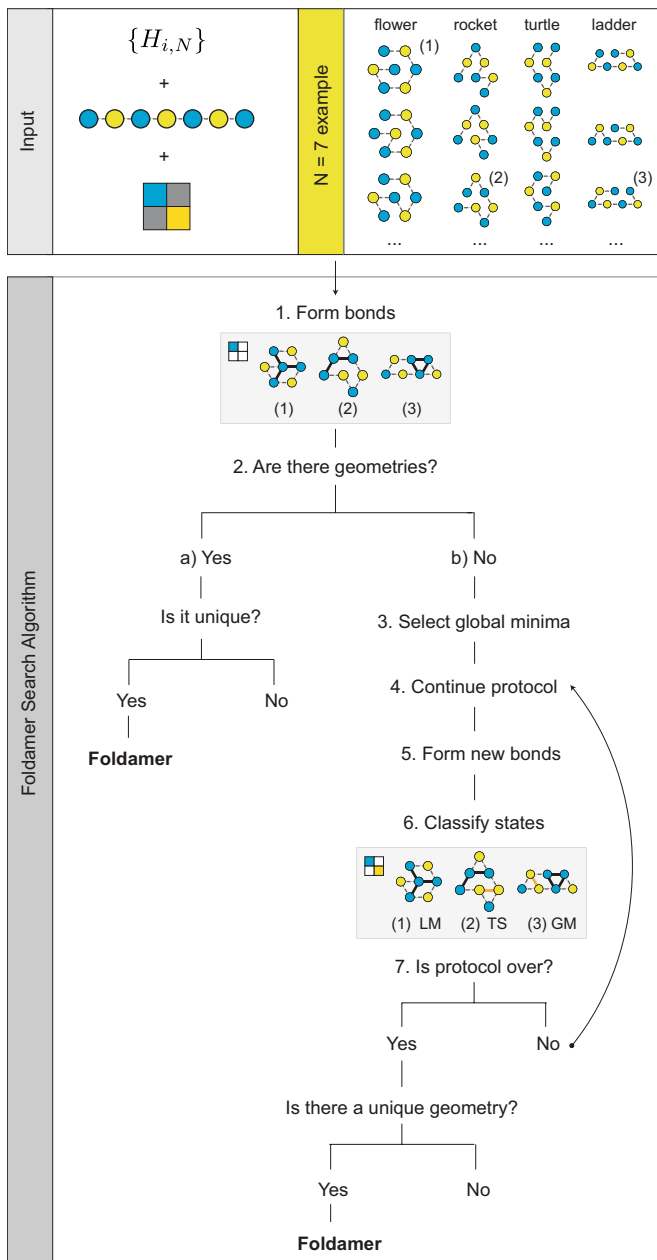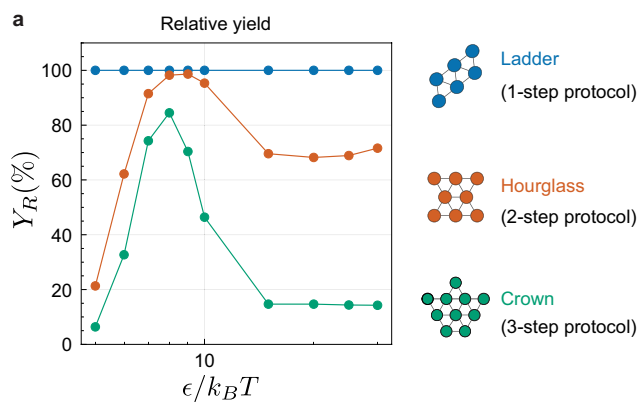
Extended Data Fig. 2 | Flowchart of the foldamer search algorithm. The top panel shows the ingredients required to run the algorithm: the ensemble of Hamiltonian paths $\{H_{i,N}\}$, a sequence, and a protocol. The input is a set of colored Hamiltonian paths embedded on the geometries, as shown for $N = 7$ and an alternate $ABABABA$ sequence. The bottom panel outlines the main steps of the algorithm. In the figure, LM stands for Local Minimum, TS for Transient State and GM for Global Minimum.

	Single quench				Double quench				Triple quench			
Protocol	1	1	1	1	1 → 2	1 → 2	1 → 2	1 → 2	1 → 2 → 3	1 → 2 → 3		1 → 2 → 3
N	6	7_{43}	9_{54}	13_{67}	6	8	6	7	7_{34}	10	11_{56}	10
Geometry												
Y_{Exp} (%)	100	100	100	ND	95	75	100	43	75	38	ND	ND
$Y_{SimDownhill}$ (%)	100	100	100	100	64	72	100	76	73	14	4	53
$Y_{SimStepThermal}$ (%)	100	100	100	100	88	98	100	40	86	85	23	20

Extended Data Fig. 3 | Foldamer yields for an alternating *ABAB* sequence with length N = 6–13. From left to right, we show the results for single, two, and three-step protocols. All yields are given as relative yields, in which the number of foldamers is normalized by the total number of rigid structures observed at the end of the corresponding protocol. The experimental number of observations is n_6[ladder, triangle, chevron] = (67, 19, 86), n_7[rocket#1, rocket#2, flower] = (175, 25, 7), n_8[hourglass] = 8, n_9[poodle] = 24 and n_{10}[crown] = 8. 'ND' stands for 'No Data'. Simulation yields $Y_{SimDownhill}$ and $Y_{SimStepThermal}$ come from two different protocols, the pure downhill and the step-thermalized quench, respectively, and result from averaging over > 2000 different initial conditions (sampling error is negligible). The significant difference between the two simulation protocols arises because the finite unbinding probability in the step-thermalized case is optimized to allow the colloidomer to escape kinetic traps, i.e., local minima, and fold correctly (see Supplementary Video 8 and the Extended Data Fig. 4(a)), whereas this is not possible in the downhill case. Exceptions are the N = 7 flower and N = 10 bed, foldamers that undergo geometric frustration, as the irreversibility of the locking bonds in the downhill protocol actually improves the yield. The two simulation methods give the range of yields one can access by folding strictly in 2D. Experiments undergo a finite temperature quench, which is better mimicked by the thermalized simulation protocol in all cases. Note that all but one (triangle) experimental yield fall within the range predicted by simulations. This experimental yield exceeds that of the optimized simulation, owing to the fact that local minima can be escaped by rare 3D excursions only possible in experiments.

a

Relative yield

Ladder (1-step protocol)	
Hourglass (2-step protocol)	
Crown (3-step protocol)	

b

	2 colors		3 colors
	1 — 2 — 3	1 — 2	1 — 2 — 3 — 4 — 5
Protocol			
N	8	13_{76}	11_{344}
Geometry			
$Y_{SimDownhill}$ (%)	83	11	9
$Y_{SimStepThermal}$ (%)	97	58	51

Extended Data Fig. 4 | Optimized foldamer yields for alternating and designed sequences. (a) Relative foldamer yield Y_R(%) as function of bond strength ϵ/k_BT for three foldamer structures: the ladder (blue), hourglass (orange) and crown (green). These structures exemplify foldamers resulting from 1-step, 2-step and 3-step protocols, respectively. The results were obtained by waiting $\tau = 10^3$ time units (t. u.) with the first interaction on, and setting $\tau = 10^2$ t. u. for subsequent interactions. We note that ϵ/k_BT refers only to the binding strength during the first step in the protocol; subsequent steps in the protocol are treated as pure downhill folding. (b) Relative yields of the three foldamers in Fig. 4(b). As discussed in Extended Data Fig. 3, $Y_{SimDownhill}$ corresponds to pure downhill folding, where the chain is likely to get trapped in a local minimum along the pathway. $Y_{SimStepThermal}$ yields are the result of optimized binding strengths ϵ and quench lengths τ. Results in the table were obtained by setting $\tau = 10^3$ t.u., and choosing $\epsilon/k_BT = 8$ for the $N = 8$ and $N = 13$ chains, and $\epsilon/k_BT = 7$ for the $N = 11$ chain.
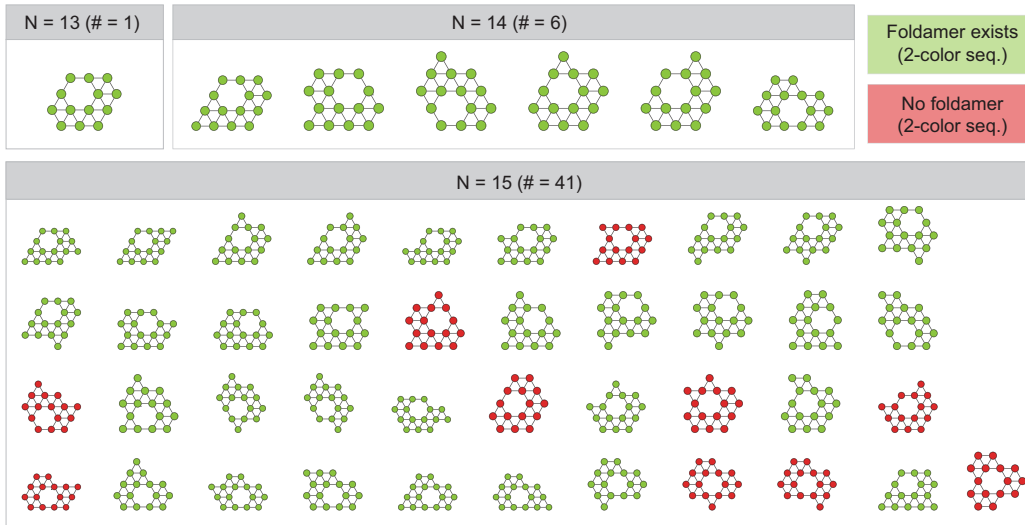
Extended Data Fig. 5 | Non-compact clusters in a hexagonal lattice in 2D. The first non-compact cluster arises at $N = 13$; for $N = 14$ there are 6 different clusters, and for $N = 15$ we identify 41. All non-compact clusters contain a single hexagonal hole. For those structures colored in green, there exists at least one protocol in random sequence space that can fold them uniquely.