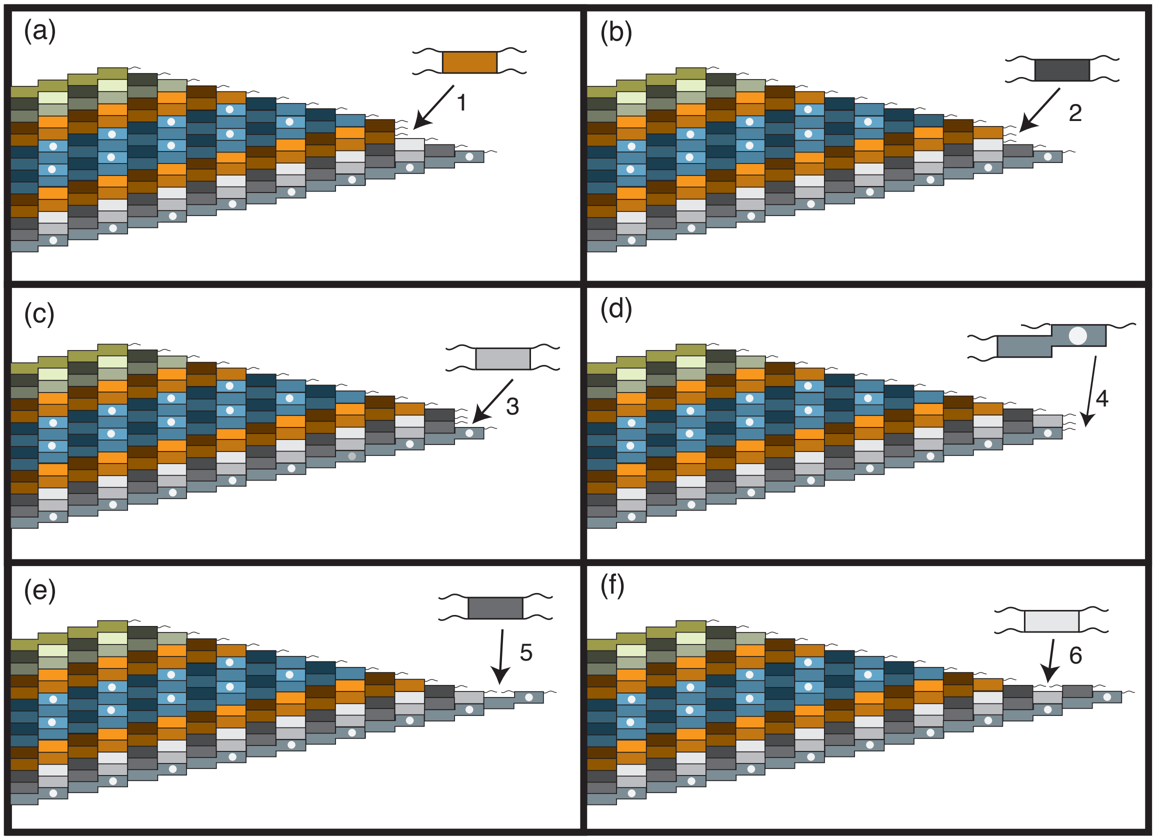
Robust self-replication of combinatorial information via crystal growth and scission

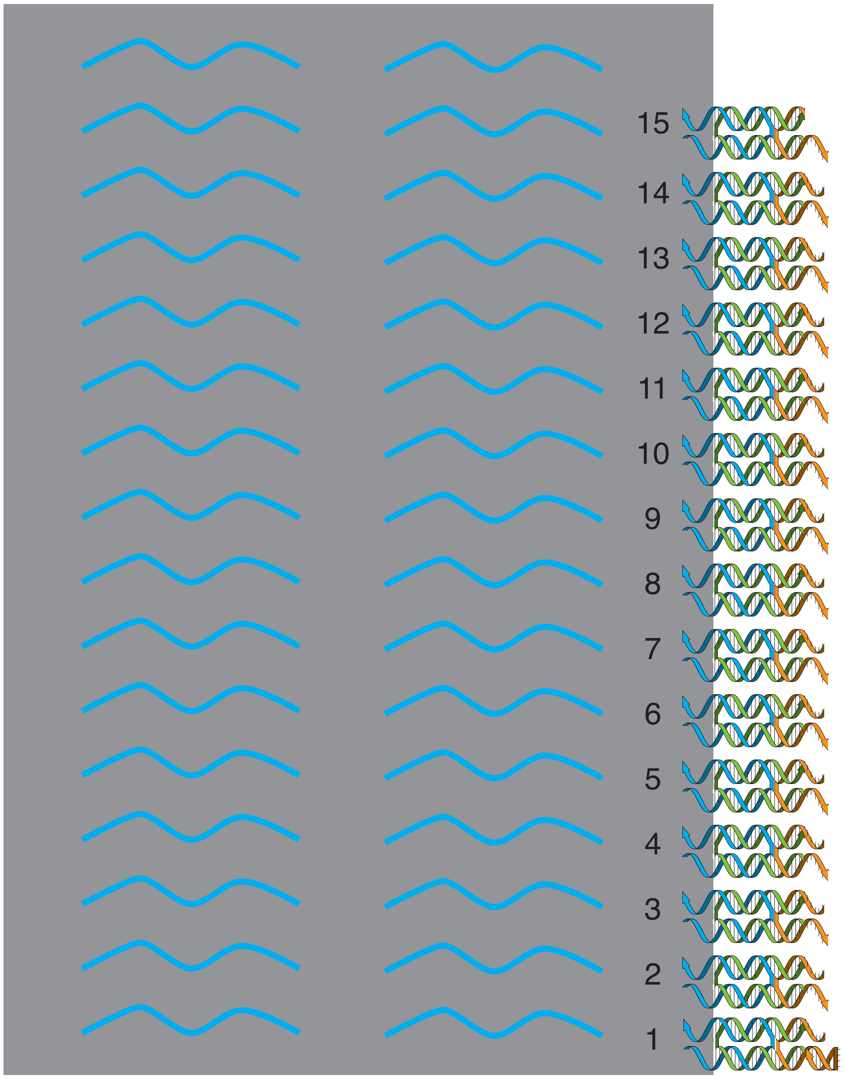
Supplemental Information

Rebecca Schulman[1,*,†], Bernard Yurke[2,3], Erik Winfree[4,5,6]
[1]Computer Science, California Institute of Technology
[2]Materials Science and Engineering, Boise State University
[3]Electrical and Computer Engineering, Boise State University
[4]Computer Science, California Institute of Technology
[5]Computation and Neural Systems, California Institute of Technology
[6]Bioengineering, California Institute of Technology
[*]To whom correspondence should be addressed; Email: rschulm3@jhu.edu.
[†]Current address: Department of Chemical and Biomolecular Engineering, Johns Hopkins University

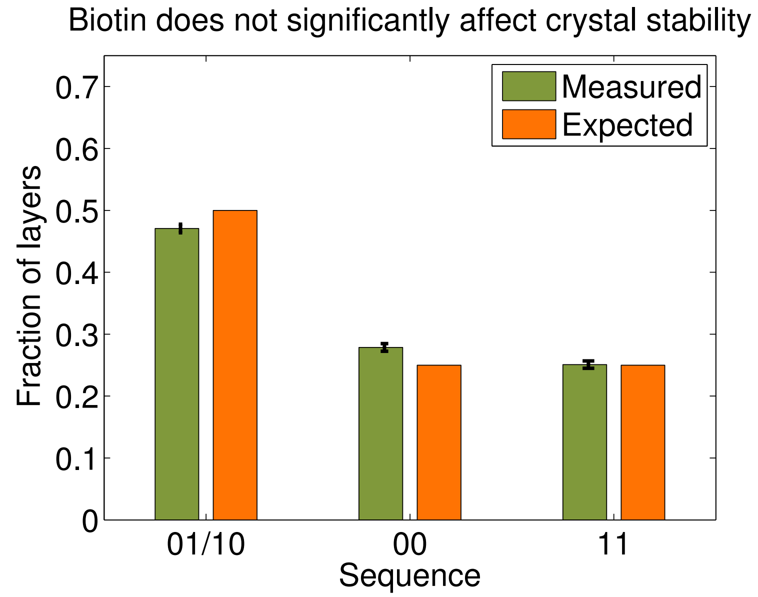
Supplemental Figures



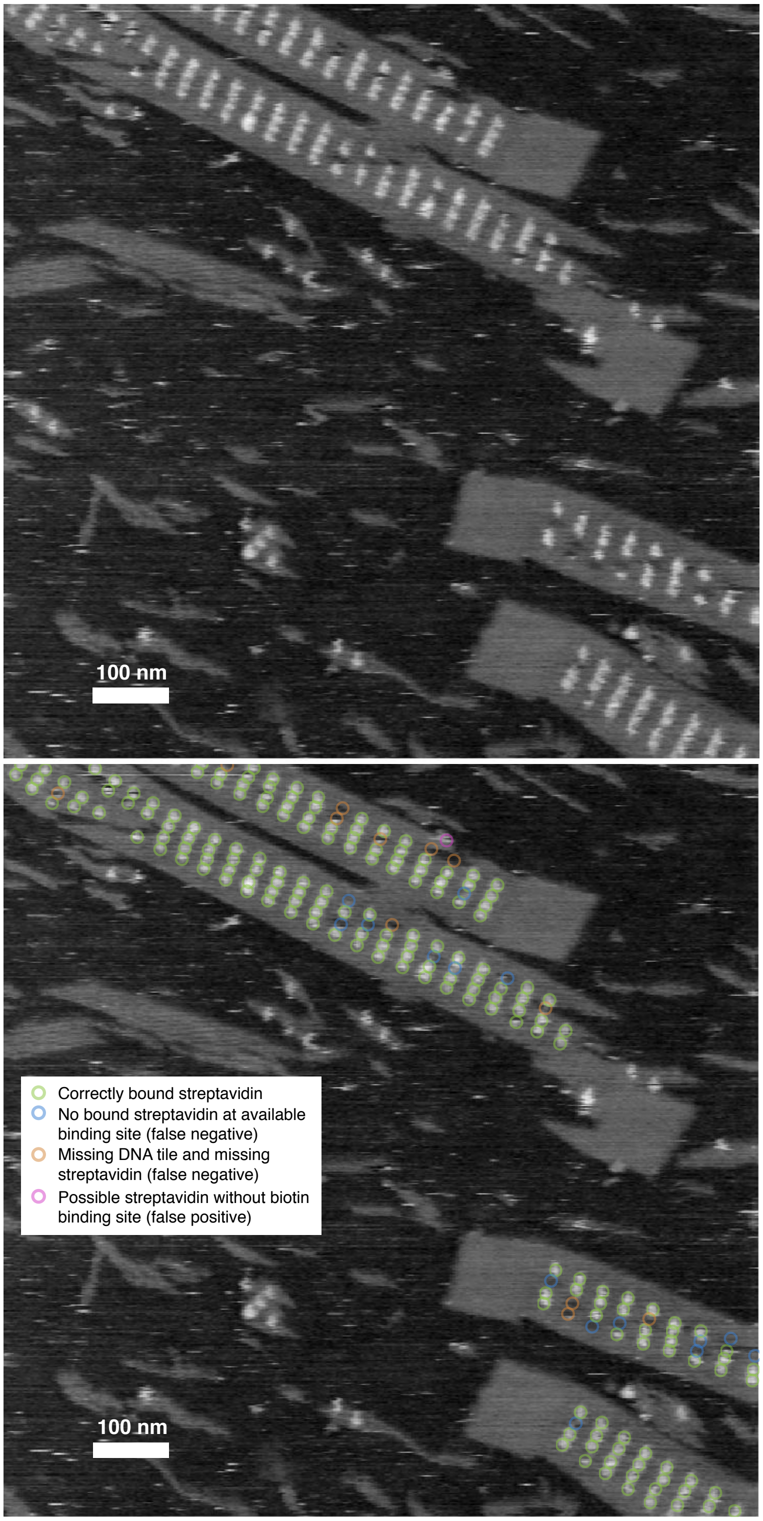
Supplemental Figure 1: **Steps in the zig-zag pathway for ribbon growth.** Tile attachment is only favorable at sites where a tile can attach by more than one binding site. The tile set is designed such that there is a pathway for growth consisting exclusively of favorable attachments. When growth follows this pathway there is exactly one site on a growth front where a tile can favorably attach at any given time. A sequence of growth events at these sites follows a zig-zag path. (a)-(f) show six successive steps along this zig-zag growth path. Growth proceeds down the facet (a-d), and when it reaches the bottom edge in (d), the addition of an edge tile creates a new site from which growth can then proceed upwards toward the top edge (e-f). The favorable growth pathway continues up to the top edge then back down.



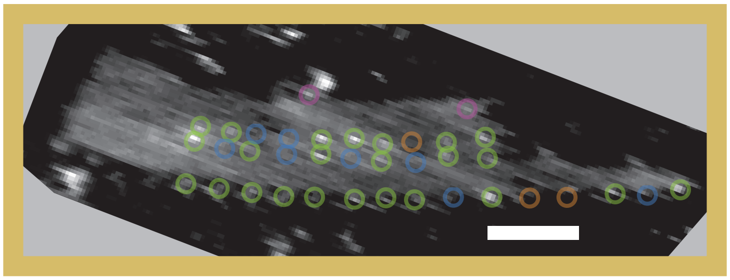
Supplemental Figure 2: Adapter tiles, which together emulate the structure of a DNA tile ribbon growth front, are arranged vertically on a 32-helix "DNA origami" seed. Each adapter tile contains 2 strands (green and orange) that assemble on a scaffold strand that winds through the whole seed structure. The sequences of each adapter tile are unique, as they must bind a) to the unique sequence region on the scaffold strand and b) possess fully programmable sticky end sequences. The sequences of the adapter tiles are given in Supplemental Methods 12.9.



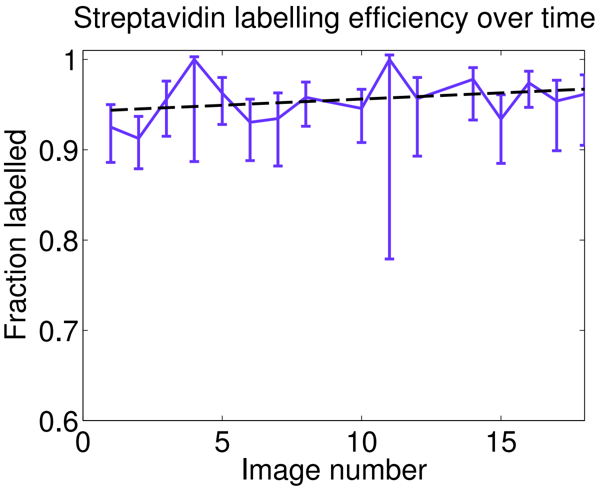
Supplemental Figure 3: **Biotin-labelled and unlabelled tiles assort randomly.** Previously, it has been observed that DNA tiles labelled with hairpins [1, 2, 3, 4] were less thermodynamically stable in crystals than those without hairpin labels. Differences in stability led to assembly errors, specifically growth or nucleation errors in which a tile without a hairpin label bound in place of a tile with a hairpin label. To test whether there was a difference in stability between biotin-labelled tiles and those without biotin labels, we designed 6-tile-wide crystals in which 2 embedded 2-tile-wide blocks could either contain a block with a biotin label (1) or a block without a biotin label (0). Equal concentrations of blocks with and without biotin labels were mixed with edge tiles and were annealed without seeds so that all crystal forms could spontaneously nucleate. Nucleation rates of particular crystal forms are heavily dependent on thermodynamic stability, so that if biotin labels affected stability, it would be expected that crystals without biotin would nucleate first, then blocks without biotin, leading to an over-representation of 00 blocks and some over-representation of 11 blocks, which nucleated later. 01 blocks, which mixed biotin-labelled and unlabelled blocks, would be expected to be underrepresented. If biotin does not affect stability, it would be expected that tile blocks randomly assort. We observe that the fraction of crystals with mixed labelled and unlabelled blocks is approximately the fraction expected by random chance, suggesting that biotin-labelled tiles have approximately the same stability as unlabelled tiles. The experiments shown here were repeated with the biotin labels reversed with similar results, suggesting that this assortment is not due to idiosyncrasies of particular tile sequences.



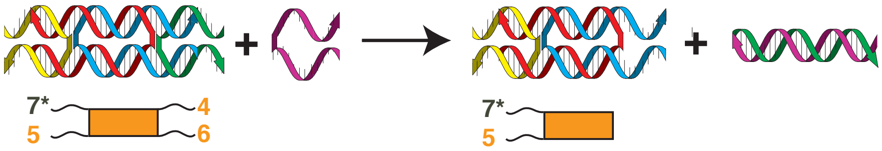
Supplemental Figure 4: **Determination of streptavidin binding efficiency in control experiments.** Unlabelled atomic force microscopy image of nucleated OOOO, biotin-labelled crystals imaged with streptavidin (top) and the same image with labels indicating indicating how these sites were tabulated in experiments determining streptavidin binding efficiency. Sample preparation and imaging protocol is described in Supplemental Methods 9 and the results of binding efficiency experiments are described in detail in Supplemental Note 3.



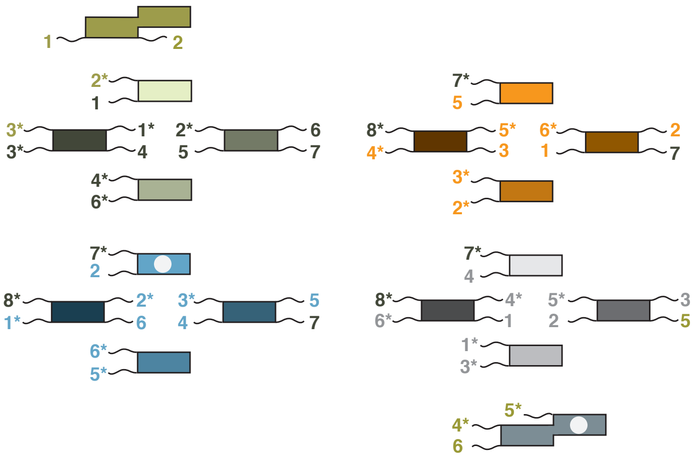
Supplemental Figure 5: **Sequence determination using streptavidin labels.** In some images, unbound streptavidin, missing tiles or imaging artifacts prevented every element of the sequence in every layer from being read. In addition, during G2, the mixture of the two labelling types produced a mixture of apparent labelling patterns. However, because of the redundancy of the displayed sequences (and the low error rate in layer to layer copying), the sequences for most layers could still be deduced reliably. Here we show an example crystal where the sequence was inferred when some labels were incorrect and there was lattice tearing from imaging. The unlabelled version of this image is in the main text in Figure 2. Circle colors indicate identification according to the legend in Supplemental Figure 4. Detailed notes on sequence determination procedures are provided in Supplemental Methods 3. Scale bar is 100 nm.



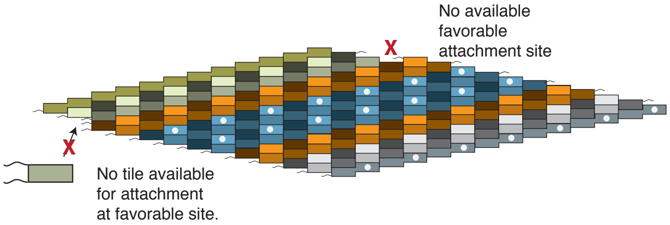
Supplemental Figure 6: **Binding efficiency of streptavidin onto biotinylated tiles in DNA crystals.** OOOO crystals with biotin labels on the O block tiles were used to test the efficiency of streptavidin binding visualized via AFM imaging. The blue shows the measured fraction of O blocks that are properly labelled along with 95% confidence intervals. There were on average 175 sites where streptavidin should have bound per image, although some images contained many fewer. The black line is a linear best fit to the data, suggesting that streptavidin binding continues during imaging, although efficiency is already quite high in the first image.



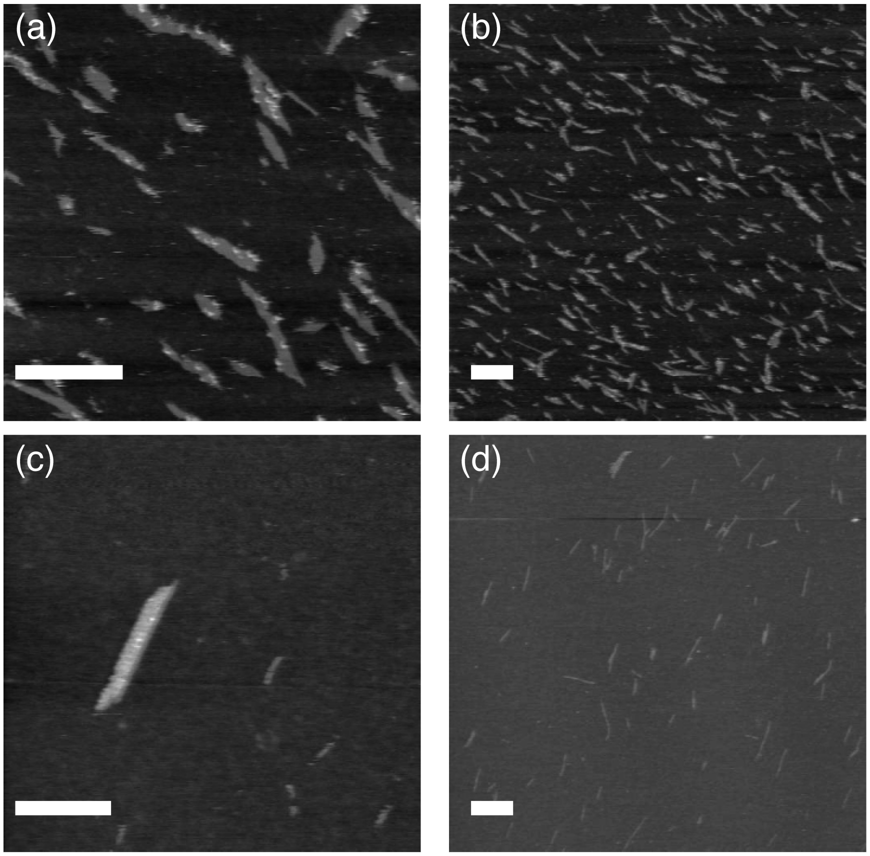
Supplemental Figure 7: **Watson-Crick complementary strands can inactivate DNA tile sticky ends via branch migration.** Before imaging crystals via AFM, we added "guard" strands with sequences that are the Watson-Crick complements of some of the DNA tile short strands. Guard strands remove these complementary strands from the corresponding DNA tile as shown.



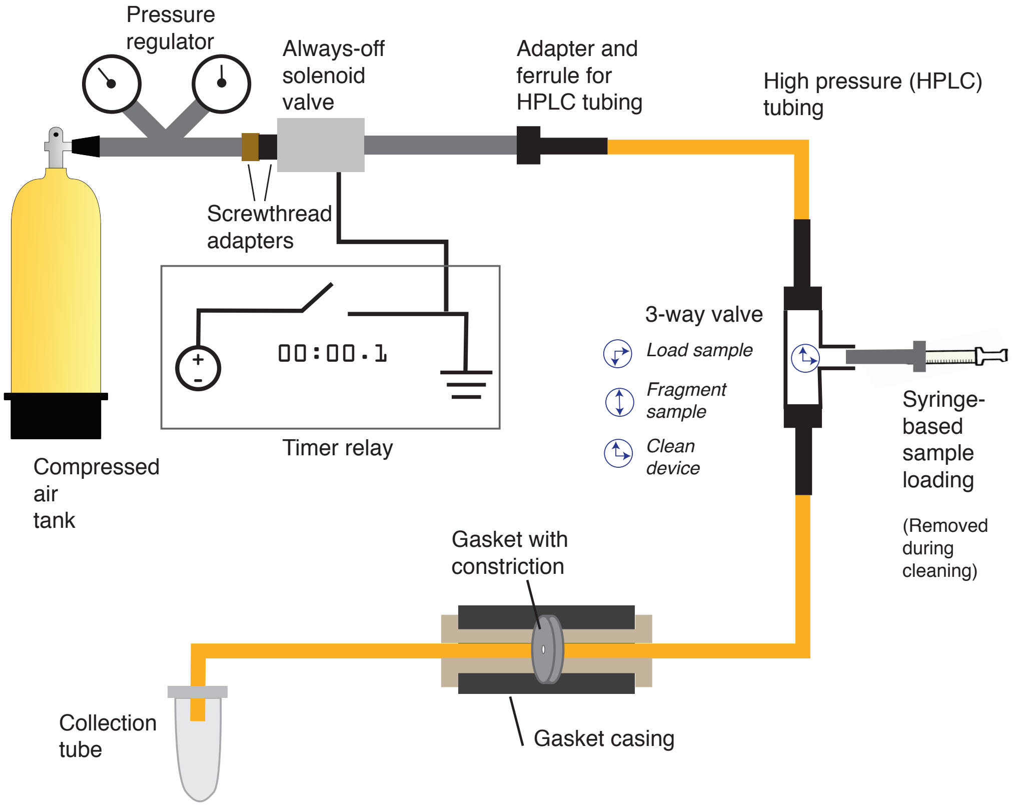
Supplemental Figure 8: **The sticky ends of tiles after guard strands react.** We used 10 guard strands (Sequences listed in Supplemental Methods 12.10) which removed 10 pairs of sticky ends from tiles.



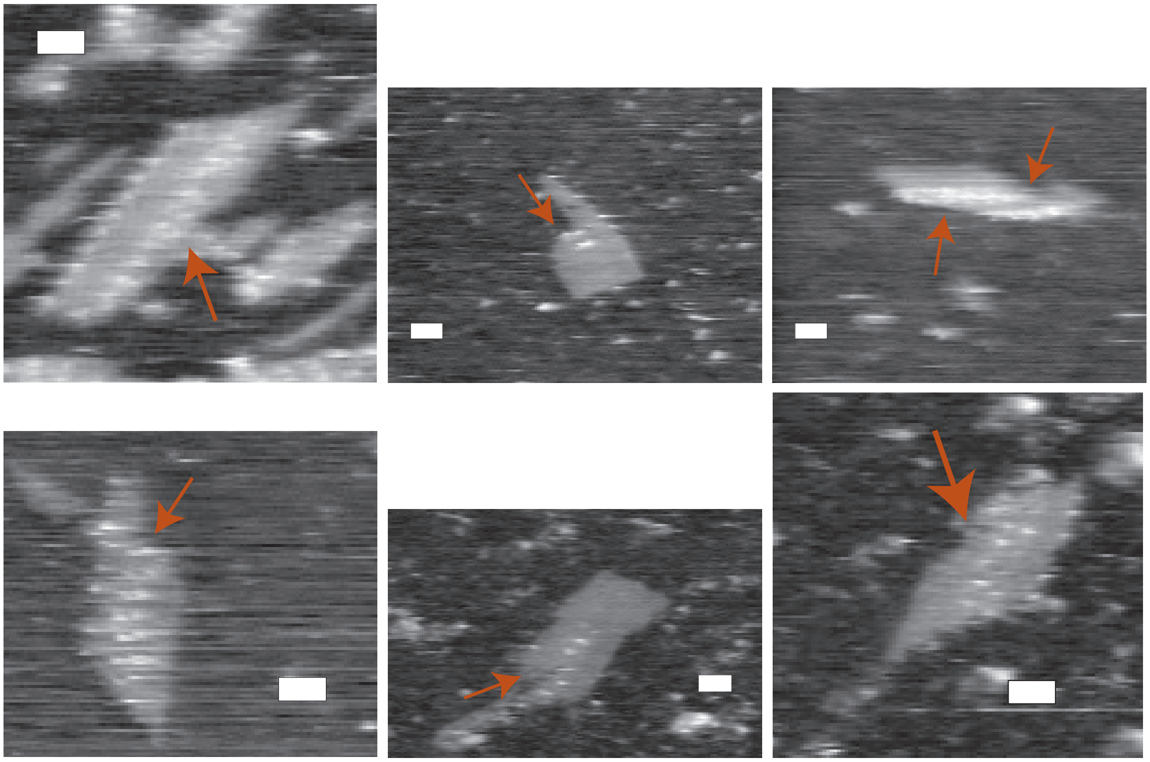
Supplemental Figure 9: **Guard strand modification of tiles prevents new growth.** 12 guard strands remove sticky ends from the tiles shown in Supplemental Figure 8. The diagram shows how the right edge of the crystal will not be able to grow because guard strands remove some of the available sticky ends on the growth front. On the left edge, the crystal will not be able to grow because guard strands remove the sticky ends needed for attachment from a set of tiles essential for growth. The strands are removed from free tiles and from tiles attached to crystals for which at least one of the sticky ends is unbound.



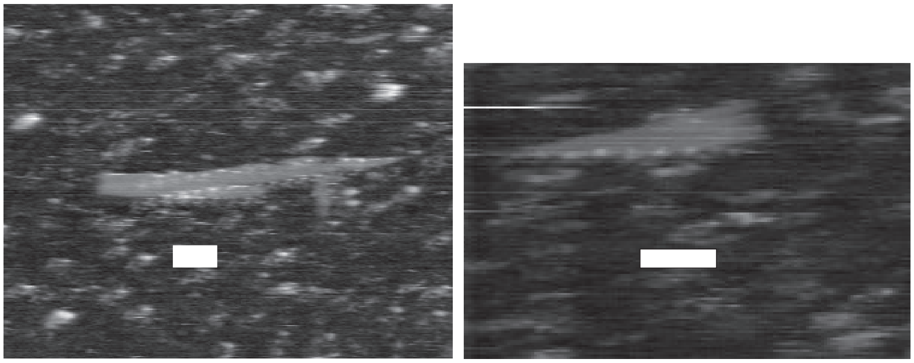
Supplemental Figure 10: **Guard strands reduce crystallite presence on the mica.** AFM images; scale bars are 250 nm. **(a)–(b)** Samples from G2 prepared as described in Supplemental Methods 6, except without the use of guard strands. Small crystallites form at room temperature on the mica that were not present in solution, occluding imaging of other crystals. **(c)–(d)** The same sample prepared identically as in (a),(b) but with guard strands, as described in Supplemental Methods 6. Many fewer crystallites are visible. These experiments should be considered qualitative evidence for the effectiveness of guard strands, as the heterogeneity of mica surface also influences the number of visible structures.



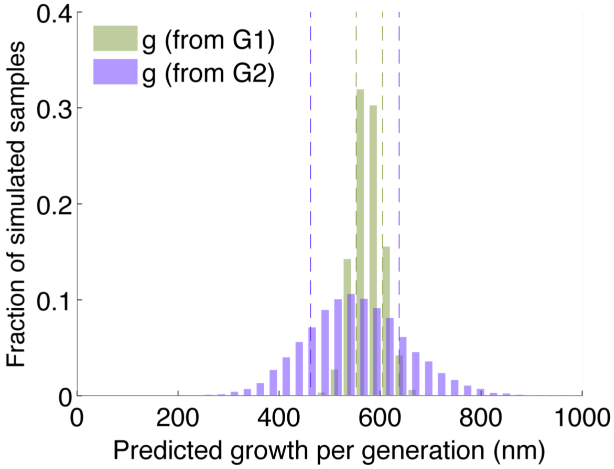
Supplemental Figure 11: **Schematic diagram of the device used for crystal scission.** A description of the components is given in Supplemental Methods 11.



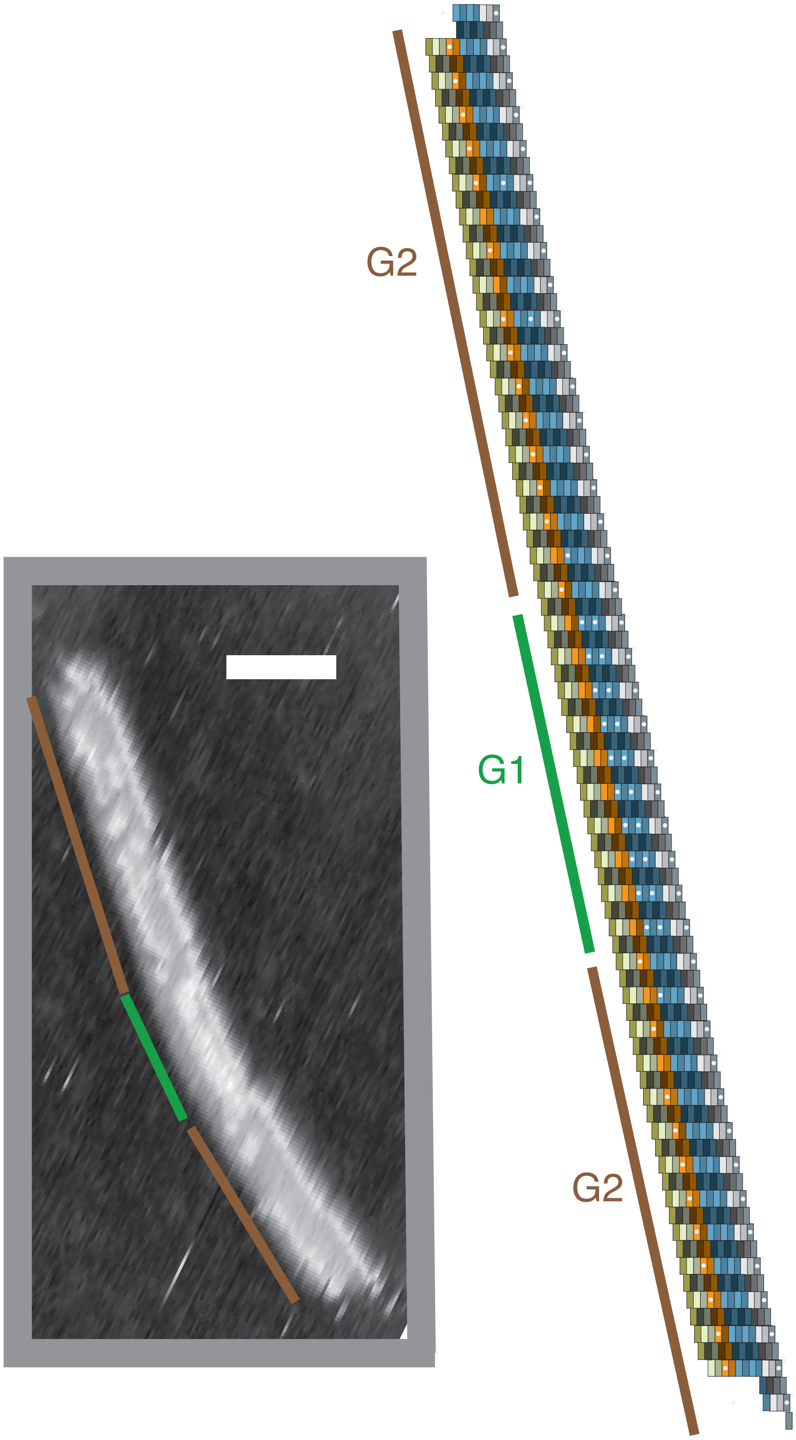
Supplemental Figure 12: **Frayed crystal edges observed after scission.** After scission, many growth front shapes deviated significantly from the diagonal facet produced by zig-zag growth. Many of these growth fronts could return to zig-zag growth, which correctly copies the sequence carried by the crystal, via a series of favorable tile additions. However, by qualitative inspection about 20% were visibly frayed, *i.e.* they had either facet shapes where favorable growth could produce mutations that would then be propagated, or shapes where unfavorable growth steps were needed to return to zig-zag growth. Examples of frayed growth fronts are shown at arrows. Scale bars 50 nm.



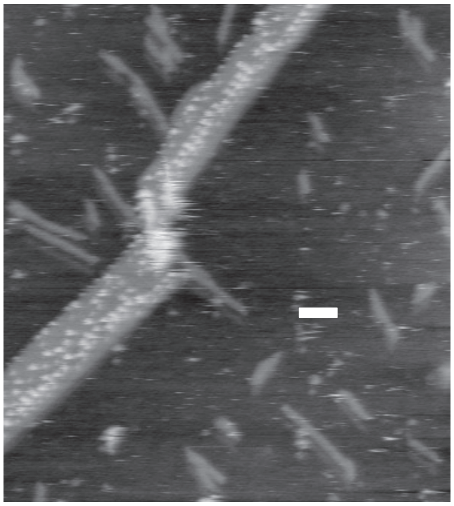
Supplemental Figure 13: **Crystals fragmented across every copy of their sequence rather than in between sequence copies,** *i.e.* **parallel to the long axis of the ribbon.** Scale bars 100 nm.



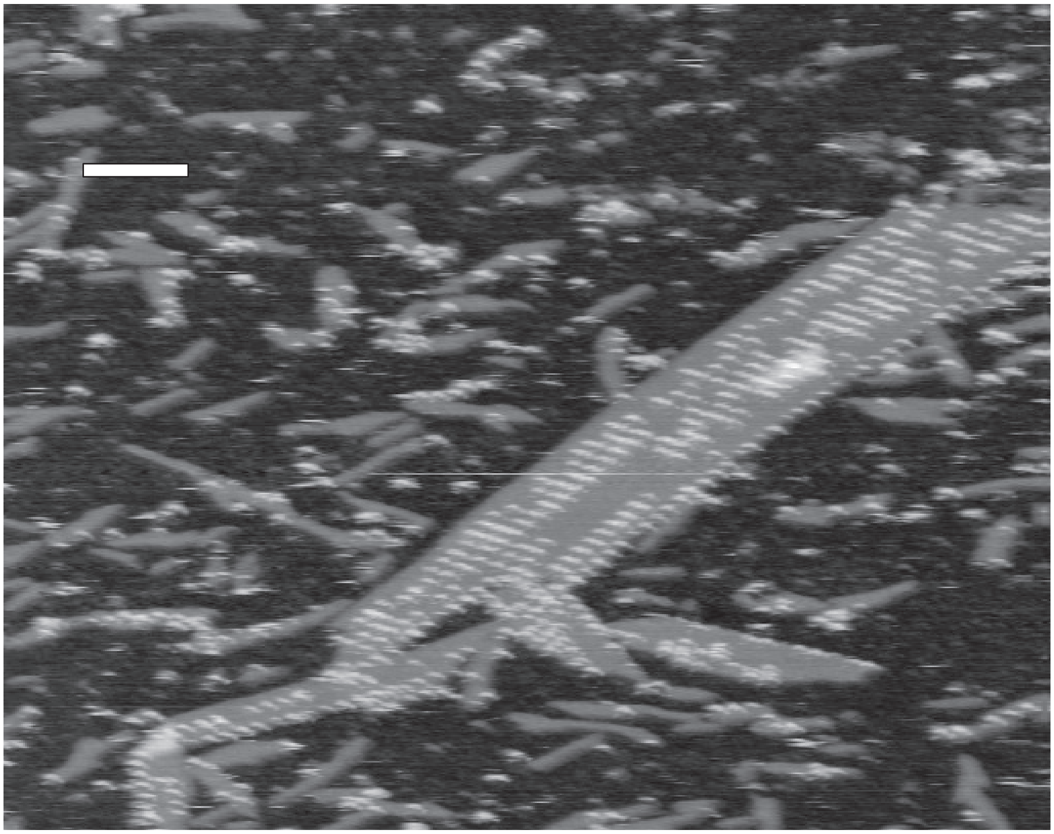
Supplemental Figure 14: A histogram showing the probability distribution of two predicted values of g based on Supplemental Equations 37 and 38, which correspond to our measurements of the amount of growth occurring during G1 and G2, respectively. Dashed lines show the region of the center 63% of the histogram, or one standard deviation assuming that the distributions are Gaussian. While the predicted value of the amount G1 growth is much more accurate that the predicted value of the amount of G2 growth, they are statistically indistinguishable. Thus, the assumption we made in our model of replication that the amount of growth (in layers) per growth front is the same for each generation is consistent with our experimental findings.



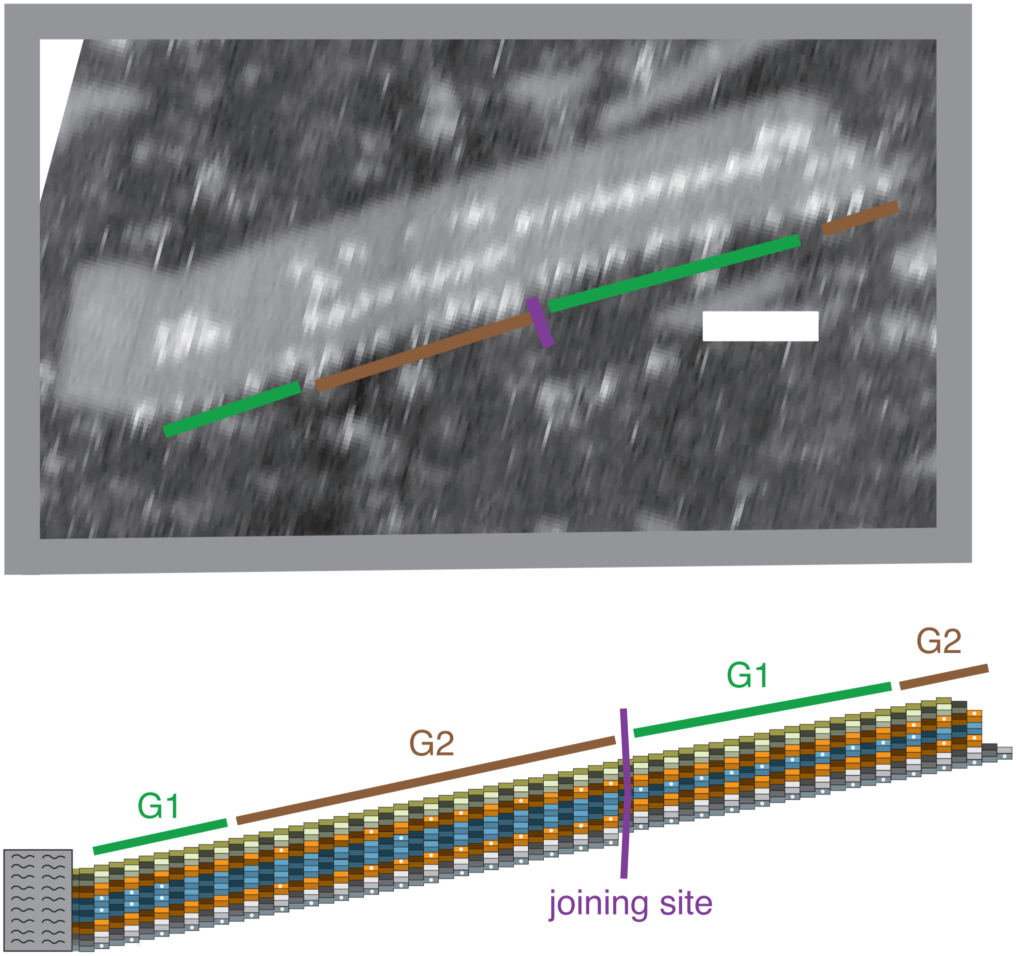
Supplemental Figure 15: **Replication of sequence OII.** While only the sequence OIIO was templated by the origami seeds, sequence errors during nucleation and growth created new, “mutant” sequences. These sequences, although much rarer than the OIIO sequence, were also replicated. An example is shown here. Left, AFM image of the crystal. Right, schematic interpretation of the crystal showing the inferred tile types given the location of the biotin labels. Scale bar 200 nm.



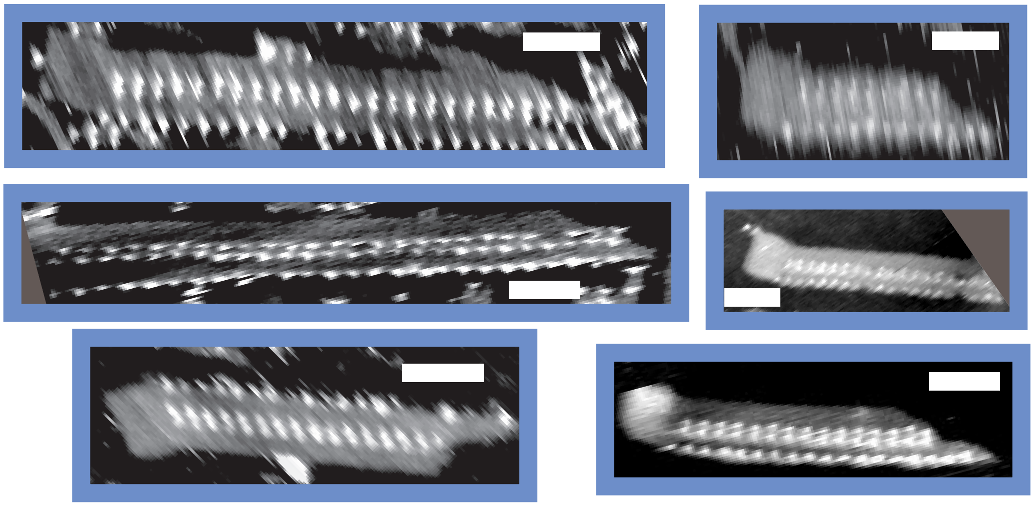
Supplemental Figure 16: **Damage to growth fronts sustained during scission can induce errors.** An AFM image of a crystal from G2. While the G2 growth on the growth front at the top of the image proceeds correctly, an error occurs before G2 growth starts on the other side of the crystal that increases the crystal's width. The twisting of the crystal at the point where the error occurs may suggest lattice defects. The slight doubling of the image visible beneath the long thin crystallites (doubling is not visually detectable on the large crystal) is mostly likely the result of an AFM tip that has splintered, causing it to contact the same point on the surface more than once. Scale bar 100 nm.



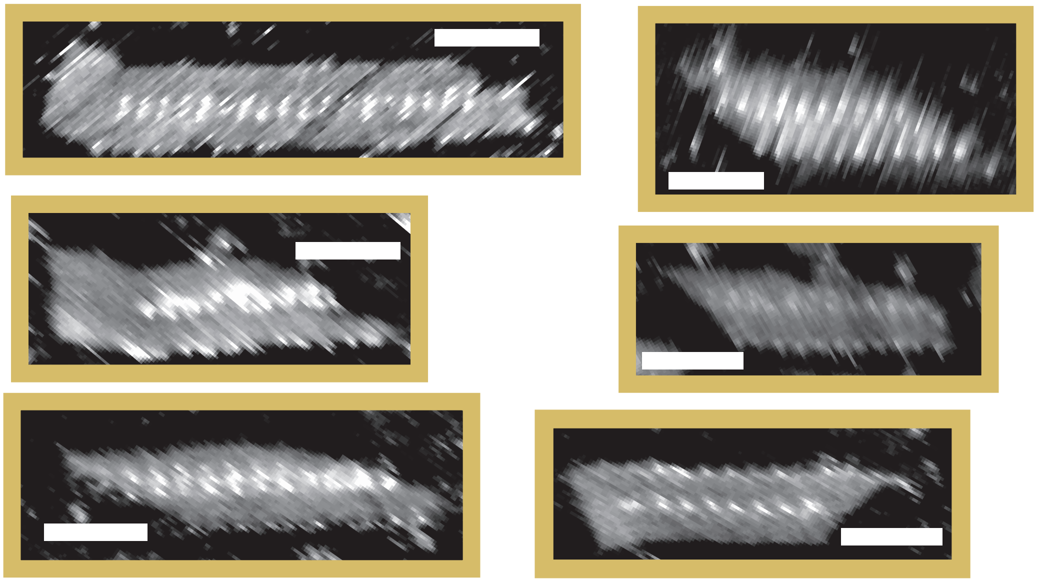
Supplemental Figure 17: **Sequence extension or side-to-side joining along the long axis of crystals during G2.** After scission experiments where the samples were potentially cooled due to escaping air, crystals with sequences $\gg 4$ bits were sometimes observed during G2. Many crystals had damaged edges after these scission experiments, and it is possible that when the sample was cooled, facet growth proceeded on these edges, producing wide crystals. Alternatively, two crystals with damaged edges may have joined along their long edges. Scale bar 200 nm.



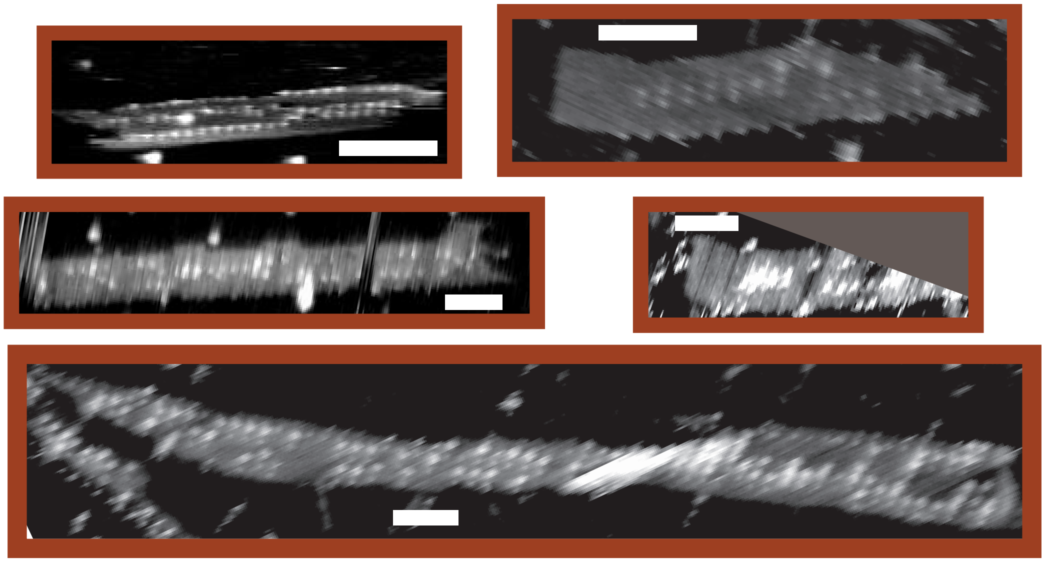
Supplemental Figure 18: **End-to-end joining of crystals during G2.** During G1, templated crystals had only one available growth front, but during G2, many crystals had two growth fronts available. Crystals with 2 growth fronts have complementary DNA sequences and can join at their growth fronts. Note that the two crystals that joined here carried different sequences, OIIO and OIO respectively. Joining can take place for crystals with the same or different sequences, so long as there are enough aligned complementary sticky ends to produce a favorable reaction. Scale bar 100 nm.



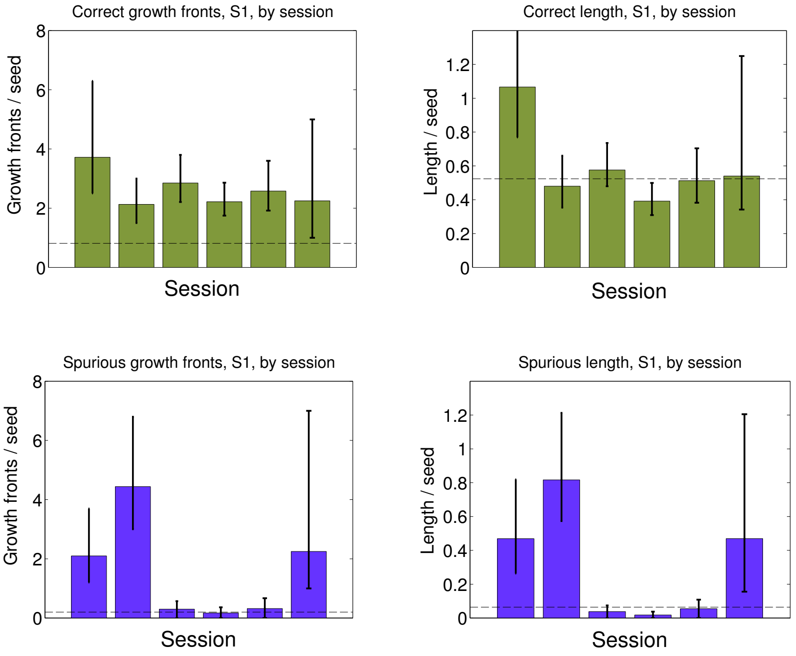
Supplemental Figure 19: **Additional images of crystals after G1.** Scale bars are 100 nm.



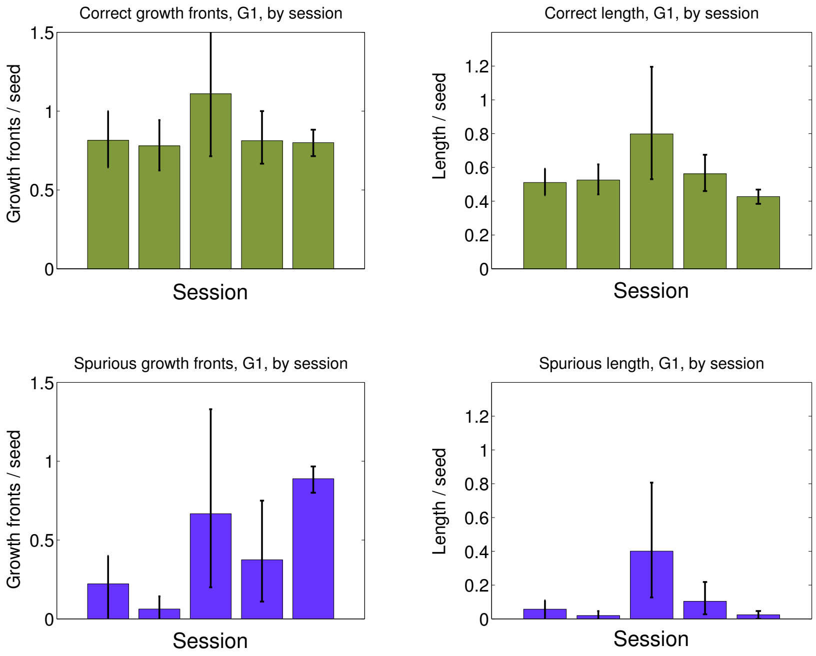
Supplemental Figure 20: **Additional images of crystals after S1.** Scale bars are 100 nm. The crystal on the bottom right has a nucleation error.



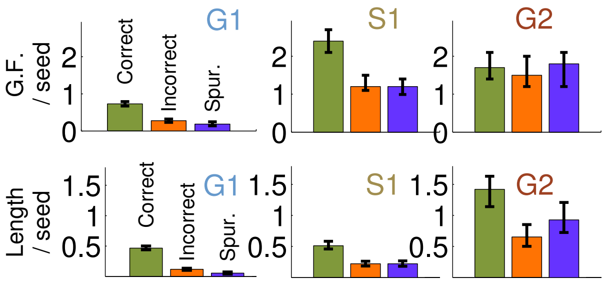
Supplemental Figure 21: **Additional images of crystals after G2.** Scale bars are 100 nm.



Supplemental Figure 22: **Session-to-session variation in the crystals produced by the scission process.** The number of growth fronts and crystal length (in μm) of correct and spuriously nucleated crystals produced in 6 different scission experiments. Averages of the same quantities measured after growth, *i.e.* right before scission, are given by the dotted lines. The length and number of correct crystal growth fronts vary little from experiment to experiment, with the exception of the first experiment. Notably, this experiment was performed in a preliminary version of the scission apparatus, which lacked a timed shut off of the air powering the system. Other variations appear consistent with the small sample sizes (approximately 20-50 crystals) in each experiment. In contrast, the number of spurious nuclei clusters into 2 groups, with one group having mean numbers of growth fronts and layers significantly less than the average number of spuriously nucleated growth fronts (0.20) and crystal length (65 nm) per seed observed during the growth phase, and the other group having a mean number of growth fronts and layers significantly above the means. Scission experiments where the number of spuriously nucleated growth fronts and layers are significantly higher than the mean correspond with scission experiments where cold air audibly escaped from the scission device onto the sample, resulting in brief cooling of the sample.



Supplemental Figure 23: **Session-to-session variation in crystal growth rates.** The number of growth fronts and crystal length (in μm) of correct and spuriously nucleated crystals produced in 5 different G1 experiments. The variation in crystal length and number of correct crystal growth fronts is consistent with the small sample sizes (approximately 20-50 crystals) in each experiment. The rate of spurious nucleation and the length of spuriously nucleated crystals varies markedly, but because of the very small number of spuriously nucleated crystals observed during each growth experiment (1-10), it is difficult to characterize this variation quantitatively. Because samples were characterized immediately after preparation and AFM characterization was performed over several hours, it was not possible to characterize the same sample before scission (*i.e.* after growth) and after scission. Thus, the samples characterized after growth and scission are different. We characterized 6 samples after scission (Supplemental Figure 22) and 5 samples after growth, shown here.



Supplemental Figure 24: Worst-case numbers of correct, incorrect and spuriously nucleated growth fronts and layers after G1, S1 and G2. The number of growth fronts and length in μm (proportional to the number of layers) with a correct sequence (green), incorrect sequence (orange) or no sequence/spurious (blue) over the three stages of replication. In contrast to the corresponding Figure 4 in the main text, here crystal growth fronts and layers that are not legible are considered to be incorrect. These values should therefore be considered lower bounds on the number of correct layers and growth fronts observed and upper bounds on the number of incorrect layers and growth fronts observed at each stage of growth and replication.

Supplemental Tables

Replication Stage	Growth Front Type	Measured values	Worst-case correctness values
G1	Correct	0.81 ± 0.07	0.73 ± 0.06
	Incorrect	0.26 ± 0.09	0.28 ± 0.05
	Spuriously nucleated	0.20 ± 0.07	0.19 ± 0.06
S1	Correct	2.6 ± 0.4	2.4 ± 0.3
	Incorrect	0.65 ± 0.16	1.2 ± 0.3
	Spuriously nucleated	1.3 ± 0.3	1.2 ± 0.2
G2	Correct	2.2 ± 0.6	1.7 ± 0.4
	Incorrect	0.62 ± 0.23	1.5 ± 0.5
	Spuriously nucleated	2.9 ± 0.7	1.8 ± 0.6

Supplemental Table 1: Growth fronts per seed. Measured values are the values reported in the paper, where illegible growth fronts are excluded from analysis. Worst-case correctness values are those where illegible growth fronts are are included in analysis and assumed to be incorrect.

Replication Stage	Layer Type	Measured values	Worst-case correctness values
G1	Correct	520 ± 40 nm	470 ± 30 nm
	Incorrect	79 ± 15 nm	117 ± 19 nm
	Spuriously nucleated	65 ± 25 nm	53 ± 20 nm
S1	Correct	560 ± 60 nm	510 ± 70 nm
	Incorrect	130 ± 30 nm	220 ± 40 nm
	Spuriously nucleated	240 ± 50 nm	220 ± 40 nm
G2	Correct	1800 ± 400 nm	1400 ± 300 nm
	Incorrect	470 ± 180 nm	650 ± 200 nm
	Spuriously nucleated	1200 ± 400 nm	930 ± 300 nm

Supplemental Table 2: Crystal length in nanometers (proportional to the number of layers) per seed. Measured values are the values reported in the paper, where illegible layers are excluded from analysis. Worst-case correctness values are those where illegible layers are are included in analysis and are assumed to be incorrect.

Supplemental Methods

1 Tile and Assembly Pathway Design. The DNA tile set in this paper employs logic (in addition to chemistry and molecular structure per se) to reduce assembly errors. The design principles have been described previously in detail [4] but are summarized here.

Briefly, the design of the tiles is informed by our understanding of crystal growth processes. Specifically, the tiles are designed so that a) if every attachment is favorable, crystal growth copies the desired sequence perfectly, and b) multiple unfavorable attachments must occur successively for either a growth error or spurious nucleation to occur. Details about these mechanisms have been described previously [5, 6].

Ribbon crystal growth occurred under conditions where tile attachment is energetically favorable only at sites where a tile can attach by more than one sticky end simultaneously (*i.e.* crystal kinks). The tiles are designed so that there is a pathway for growth consisting of exclusively favorable attachments. When growth follows this pathway, there is exactly one site on a growth front where a tile can attach favorably. The correct attachment of a tile at this site creates one new site where a tile can attach favorably. This pathway zig-zags up and down the growth front (Supplemental Figure 1). When growth follows this pathway, the sequence information encoded in the crystal structure is correctly propagated.

Each bit in the crystal is encoded by a block of tiles rather than by a single tile. This design choice ensures that if a mismatched tile attaches, there is no tile that can favorably attach at the new kink site created by the incorrect attachment [5]. Only two consecutive unfavorable attachments can allow exclusively favorable growth to resume.

A similar design feature ensures that there is a large energy barrier to spurious nucleation. Spurious nucleation happens if enough tiles attach to each other that they form a super-critical structure that can continue to grow by purely favorable attachments. Any full-width fragment of a tile ribbon (containing both a top edge tile and a bottom edge tile) is a super-critical structure, but any fragment that is less than full width cannot grow to full width without unfavorable attachments. For the tile set we use in this work, the thinnest full-width assembly is 6 tiles wide and requires 5 unfavorable addition steps to grow to full width. Under low supersaturation, where our experiments were performed, the spurious nucleation rate is expected to decrease exponentially with the number of unfavorable attachments required to reach full width [6, 7]. As a result, spuriously nucleated crystals are generally thin, *i.e.* 6 or 8 tiles wide, as described in the main text.

2 Growth Simulations. Inadvertently, the same sticky end sequence pair was used twice in our tile set, once between two nucleation barrier block tiles and once between two I block tiles. This reuse could have enabled a single nucleation barrier tile to replicate an I block tile or vice versa (resulting in either an addition of an I to the end of the sequence being propagated or termination of the sequence at the point of error respectively.) But because the tiles that had the same sticky end sequences were at different positions within tile blocks, we conjectured that these events could not have occurred without at least 3 consecutive mismatches, making it unlikely to have contributed to the error rate during crystal growth: even 2 consecutive mismatches were rarely observed in our experiments. However, to ensure that there was not some more likely pathway by which the reuse of sticky ends would have compromised the accuracy of sequence propagation, we compared the performance in simulation of the tile set with the erroneous sticky end repetition to the performance of a tile set where the two sticky end pairs were different, and thus non-interacting.

For this comparison we employed Monte Carlo simulations of DNA tile assembly via the program `xgrow` (available at `http://www.dna.caltech.edu/software/xgrow`). `xgrow` simulates the growth of a DNA tile crystal via stochastic single tile attachments and detachments. The forward rate for attachments was $k_f[t]$, where k_f is a tile-independent rate constant and $[t]$ is tile concentration. We used k_f=100,000 /M/s, midway between estimates of k_f made previously [8, 9, 10]. In the simulation, a tile could attach at any site where it matched at least one sticky end. (While in principle tiles can attach at sites without any matching sticky ends, these tiles almost never remain attached long enough to be stabilized by further attachments and thus become part of a final crystal. For performance reasons these attachments were therefore not included in our simulation.) The detachment rate was $k_f e^{-\Delta G^\circ/RT}$, where ΔG° is the average sticky-end energy previously predicted [11] for a sticky end at the simulation temperature, times the number of sticky ends by which the tile is attached to a crystal. R is the universal gas constant and T is absolute temperature. The simulation used 25 nM of nucleation barrier and edge tiles and 100 nM of indigo and orange block tiles, the same concentrations used in our experiments, and followed nucleation and growth from a seed structure with the same shape and adapter strand logic as in our experiments. In order to efficiently observe any differences in error rates between the two tile sets, we simulated assembly at a higher level of supersaturation than where our experiments took place (*i.e.* where the error rate should be high); the parameters of the simulations corresponded to 28 °C. The mismatch rate in simulation per bit copied was $0.036\% \pm 0.006$ for the tile set used in our experiments, and $0.033\% \pm 0.006$ for the tile set without a repeated sticky end pair. These values suggest no significant difference in error rates for the two tile sets.

3 Sequence Design. This section describes our sequence design methods. A list of the sequences used in this work is given in Supplemental Methods 12.

The DNA origami technique [12], in a which a long "scaffold" strand is folded into a pre-designed shape by a large number of short, designed "staple" strands, was used to produce a DNA origami seed for DNA tile ribbon crystals. This method has been detailed elsewhere [4] and was applied with only minor changes; no significant sequence design was required. The scaffold strand was the 7249-base single-stranded M13mp18 viral genome. The 192 short (32 base) staple strands were the same as those used previously [12, 4]. As was performed previously [4], the origami structure was augmented by adapter strands that attach to the origami seed and have local structure that is identical to the DNA tile motif of the ribbon crystals. The adapter strands used are identical to those in previous work [4] except that the single-stranded "sticky end" regions were changed to match the sticky ends of the ribbon crystals used here.

Other sequences were designed as described previously [4, 7]. Briefly, tile strand sequences and the regions of adapter strand sequences not complementary to the scaffold strand were optimized to minimize the amount of complementarity between subsequences that were not intended to hybridize. The optimization was performed using MATLAB scripts available at `http://www.dna.caltech.edu/DNAdesign/`. Sticky end sequences were designed so that complementary sticky end pairs had similar hybridization energies, subject to the constraint that the hybridization energies of non-complementary sticky ends were small. We used the DNA hybridization energy model in [11] and energy model parameters from [13]. Sticky end pairs were arranged on tiles by hand according to the logic of the tiles, with the goal of minimizing complementarity between the sticky end sequence and the sequence of the tile it was placed on.

4 Crystal Growth. Unless otherwise stated, all reaction steps were performed in "standard buffer": TAE (40 mM Tris-Acetate, 1 mM EDTA) with 12.5 mM of added magnesium acetate. All water for solutions was purified by a Milli-Q unit (Millipore) and all solutions were filtered using a 0.22 μm filter before use.

For G1, 25 nM of the edge tiles (Supplemental Methods 12.2), 25 nM of the nucleation barrier tiles (Supplemental Methods 12.3–12.4), 100 nM of the indigo tiles (Supplemental Methods 12.6 with the relevant biotin-labelled strands, Supplemental Methods 12.7) 100 nM of the orange tiles (Supplemental Methods 12.5, no biotin-labelled strands), 25 pM of the M13MP18 scaffold strand for the seed (Supplemental Methods 12.8), 50 nM of each of the 192 staple strands (Supplemental Methods 12.8) and 50 nM of each of the adapter tile strands (Supplemental Methods 12.9) were combined in a total volume of 100 μL standard buffer. (Supplemental Methods 5 describes the protocol for mixing and storing strands.) In order to maximize the yield of fully formed tiles (as opposed to incomplete tile assemblies missing one or more of the component strands), we included the listed concentration of each of the "core" strands that lacked sticky ends and a 5x excess of sticky ends strands. Thus, 25 nM of a tile means 25 nM of the 48 (or more) base pair strands for that tile and 125 nM of the two 26 base pair strands for that tile.

Three identical such G1 samples were prepared. The monomers used during G2 were also prepared at this time. The G2 mixture included 25 nM of the edge tiles, 25 nM of the nucleation barrier tiles, 100 nM of the indigo tiles (no biotin-labelled strands) and 100 nM of the orange tiles (with the relevant biotin-labelled strands, sequences in Supplemental Methods 12.7) in 50 μL of standard buffer.

4–5 G2 samples were prepared. The G1 and G2 samples were placed in a PCR machine (Eppendorf MasterCycler) inside of a temperature-controlled glove box (Coy Labs, (`http://www.coylab.com`, 4 foot basic polymer glove box with a heated fan with PID controller) set to 31 °C. The glove box's temperature controller was turned on 12 to 24 hours before the start of an experiment and all equipment used during the replication process was stored in the box at this time to ensure that the temperature of the sample remained constant throughout the growth and scission processes. To anneal the samples, we used a PCR machine program that first held the samples at 90 °C for 5 minutes to melt the mixture into single strands Next, a temperature ramp decreased the temperature from 90 °C to 37 °C at

1 °C per minute to allow the tiles and seeds to form (tiles are expected to form between 70 °C and 45 °C, from [7]). The PCR machine then held the sample at 37 °C for 6 hours in order to ensure the seeds and tiles folded completely. Afterwards, the temperature was decreased from 37 °C to 31 °C at 1 °C per 10 minutes; this annealing schedule produced fewer nucleation errors than an annealing schedule that cooled the sample from 37 °C to 31 °C over 6 minutes. The temperature of the sample was then held at 31 °C for the duration of growth, 8 hours.

After growth, the G1 samples were removed from the PCR machine into the thermally equilibrated 31 °C glove box. The G2 monomers were left in the PCR machine and heated to 40 °C for 30 minutes to melt any crystals in the G2 samples that may have spuriously nucleated during G1. (The melting temperature of the crystals is approximately 35 °C at the tile concentrations used here [7].) The G2 monomers were then cooled from 40 °C to 31 °C at 1 °C per minute.

The G1 sample continued to grow in the glove box for the duration of the melting of the G2 monomers. Afterwards, we fragmented the G1 crystals with the scission device as described in Supplemental Methods 11. 12.5 μl of the resulting fragmented crystal mixture was then added to each of 4-5 tubes of 50 μl of G2 monomers in the PCR machine that had been cooled back down to 31 °C. pipetted to mix, then held at 31 °C for another 6–8 hours to complete G2 growth.

5 Strand Preparation. With the exception of the M13MP18 single-stranded DNA, which was obtained from New England Biolabs (`http://www.neb.com`, Product N4040S), all DNA was synthesized by Integrated DNA Technologies (IDT, `http://www.idtdna.com`). DNA for the tiles and adapter strands was purified by polyacrylamide gel electrophoresis (PAGE), except for strands with biotin labels, which were purified via high-pressure liquid chromatography (HPLC). Staple and guard strands were not purified. All strands were delivered in a lyophilized state.

Purification was performed by IDT. IDT's website states that $>90\%$ strand purity is typically achieved after PAGE purification and $> 85\%$ purity is typically achieved after HPLC purification. Before use, water was added to each lyophilized strand to make a 50 μM solution according to IDT's stated yield. To correct for errors in yield determination or for the loss of DNA before the lyophilized DNA was hydrated, we then measured the 260 nm optical density of the solution to within 1% error, and used this value and the predicted extinction coefficient of the molecule [13] to compute the concentration of the solution.

However, the effective concentrations of each strand still varied in our experiments because of factors that we expect included pipetting error, differences in strand purity, hydrolysis of strands stored for weeks or months in a refrigerator, repeated freezing and defrosting of strand solutions, and the limited accuracy of the quantitation process. At low supersaturation (where our experiments were performed) even small variations in concentration could cause significant differences in the rates of crystal growth, defect generation and spurious nucleation. Under our experimental conditions we observed significant experiment-to-experiment variation in the number of spurious nuclei observed and the lengths of templated crystals.

We adapted our methods to attempt to control for as much of this variation as possible. To reduce the effect of pipetting error, we premixed the strands for each block of tiles (16 strands) and the strands for the top and bottom edge tiles (the tiles in Supplemental Methods 12.2) in large batches and made our samples using these premixed batches rather than individual strand solutions. To control for damage to DNA resulting from repeated freezing and defrosting or refrigerator storage over an extended period of time, we divided each of these large batches into aliquots and froze all of the aliquots except for one, which was stored in the refrigerator and used over 4–10 experiments, typically a few weeks.

These practices resulted in consistent crystal sizes, defect rates and spurious nucleation rates from experiment to experiment. However, occasionally a batch of strands needed to be replenished. Crystals assembled using this new batch sometimes differed in size and quality from those of previous experiments. We attribute these differences to differences in the effective concentration of strands between the strand batches and hypothesized that by slightly changing the amount of this mixture that was used, we could replicate the performance of earlier experiments.

Thus, when introducing a new batch of strands into our experiment, we empirically determined the "concentration" of these strands for which crystal growth showed the fewest defects and lowest rates of spurious nucleation. To determine the best such concentration, we mixed crystal components and grew seeded crystals as described in Supplemental Methods 4 for G1. After 6-8 hours of growth, we imaged samples of crystals as described in Supplemental Methods 6–7 and qualitatively measured the rates of error and of spurious nucleation. If either of these rates appeared high (greater than approximately twice what was reported) or if crystals did not grow well (e.g. achieved lengths of just 10 layers or so, or if one type of block did not appear to attach efficiently, so that the crystals did not have the typical facet shape), the concentrations of the new batch used in our experiments was adjusted up or down by up to 25%. Typically, trying just 1 and never more than 3–4 adjustments in concentration following this procedure was necessary to achieve good growth. The stated concentrations of tiles therefore differed from those used by up to 25% from experiment to experiment. However, we expect that these differences reflect limitations in our ability to measure tile concentrations accurately, rather than any inherent unrepeatability of the crystal growth process. The adapter strands and staple strands were similarly grouped and aliquoted, but their concentrations were never adjusted and the concentration of the scaffold strand was likewise never adjusted.

6 Sample Preparation for Imaging. Sample preparation for AFM imaging was performed in a temperature-controlled glove box set to 31 °C (Supplemental Methods 4). Equipment and solutions for sample preparation were thermally equilibrated to 31 °C before use.

40 μl of TAE buffer with 12.5 mM magnesium acetate and 100 mM NaCl was added to an approximately 1 cm diameter area of mica (Ted Pella, `http://www.tedpella.com`, Product 56) glued to an AFM metal specimen puck (Ted Pella Product 16218). Imaging took place at room temperature, where crystals could continue to grow and the free tiles in solution could easily form small assemblies. The NaCl in the sample preparation solution reduced the adsorption of these small (approximately < 20 tile) assemblies to the mica [14]. 5–20 μl of the solution to be imaged was added on top of the buffer. To reduce the formation rate of small assemblies at room temperature and prevent further crystal growth during imaging, 10 μl of 1 μM of "guard" strands with sequences complementary to the short strands of tiles were added to the sample on the mica. Guard strands were designed to prevent further tile binding by removing many of the short strands, and thus the sticky ends, of unbound tiles and tiles along the perimeter of crystals (Supplemental Methods 8). We let this mixture stand for 3 minutes in the glove box so that sticky ends were removed by guard strands before the sample was exposed to room temperature. After this short incubation, the mica puck was removed from the glove box into a room temperature environment and remained there for the duration of imaging. 1 μl of 1 μM streptavidin (Rockland Immunochemicals for Research, `http://www.rockland-inc.com`, Product S000-01), was then added to the sample, after which the sample was placed on the AFM and images were collected as described in Supplemental Methods 7.

7 AFM Imaging. Images were taken in fluid using a Digital Multimode Nanoscope scanner with a Nanoscope IIIa controller from Veeco (Santa Barbara, CA) using silicon nitride cantilevers (DNP variety, Veeco) in tapping mode at tip resonant frequencies, generally 8-10 kHz.

As described in Supplemental Methods 6, the addition of NaCl to the imaging buffer, because of the presence of Na^+ ions, decreased the surface adsorption of DNA crystals to the mica. Since mica surfaces differ from sample to sample and domain to domain, the amount of sodium required to both maintain adsorption of large crystals and prevent nucleation or adsorption of small ones varied from experiment to experiment. Thus, during imaging TAE buffer with added sodium and magnesium were added as needed to adjust the amount of crystal adsorption to the mica.

8 Guard Strands. Imaging of crystals was performed at room temperature, where hybridization energies are larger than at 31 °C. As a result, the critical crystal size (the size at which crystal growth rather than melting is energetically favored) is much smaller at room temperature than at 31 °C, and could be as low as 3–4 tiles. Because the critical size is so small, spurious nucleation could occur at high rates. Existing crystals can also grow at room temperature and would incorporate many more errors than they would during growth at 31 °C. Since tile assembly can occur on the mica surface and in the buffer above the surface during the imaging process, we worried that assembly during imaging would produce large numbers of new errors that were not part of the assembly process at 31 °C we were characterizing. It was therefore desirable to prevent growth at room temperature by removing or inactivating free tiles.

Previously devised methods for removing or inactivating free tiles before imaging have involved removing free tiles via filtration after assemblies were stabilized via enzymatic ligation [15, 4]. Because the ligation process requires several hours, it was impractical for characterization of crystals that are still growing on a a timescale much faster than hours. We therefore devised a mechanism for inactivating tiles that took only a few minutes to complete. This mechanism employs strands we termed "guard strands." Each guard strand has a sequence exactly complementary to one of the short (26 base pair) strands that contain a tile's sticky ends. Guard strands can remove this short strand via toehold-mediated branch migration [16] as shown in Supplemental Figure 7. Once a short strand is removed from a tile, the sticky ends that were on that strand are no longer available.

Toehold-mediated branch migration will only proceed if one or both of the sticky ends on the strand complementary to the guard strand are not bound to other tiles. Thus, guard strands will only remove sticky ends from crystal edges and free tiles, so that while new growth on the perimeter of the crystal is prevented, the interiors of existing crystals will not be disassembled. However, the guard strands effectively ensure that new tiles cannot attach to existing crystals at all. Because growth is reversible, the nonzero tile off-rate is expected to cause the slow melting of crystals in the presence of guard strands. (Presumably, for crystals attached to a mica surface, detachment is slower than in solution, but still nonzero.)

At room temperature, however, this melting process is slow: using existing data on sticky end strength [7] and an estimated on rate between 100,000/M/s (from [10]) and 5,000,000/M/s (from [8]), we expect that k_{off} is less than 2 tiles per hour. But it is possible that using a guard strand for every sticky end would hasten damage to crystals around crystal defects introduced by either deposition on the mica or the AFM probe. To minimize this damage, we used a set of guard strands that could remove only a subset of the short strands sequences in our tile set.

The sticky ends removed by guard strands are shown in Supplemental Figure 8. This set of sticky ends is sufficient to prevent sustained zig-zag growth in either direction (Supplemental Figure 9). To ensure that most sticky end strands complementary to a guard strand were removed, an approximately 100-fold stoichiometric excess of guard strands were mixed with tiles.

The fact that both few errors within templated crystals and few spuriously nucleated crystals were observed during imaging supports the effectiveness of guard strands in preventing incorrect growth processes during imaging. Additionally, we qualitatively compared the number of small crystals present on the mica surface when a crystal sample was imaged without added guard strands to the number of small crystals present on a mica surface when the same sample was prepared for imaging using guard strands. The resulting images suggest that fewer crystals nucleated on or adhered to the mica when guard strands are present (Supplemental Figure 10). Not enough data was available to determine whether guard strands reduced the number of errors (which presumably would have formed during imaging rather than during G1 or G2) in templated crystals on the mica surface.

9 Measurement of Streptavidin Binding Efficiency. In order to determine how well streptavidin bound to biotin-labelled DNA tiles and thus functioned as a label that would allow us to read sequences, we performed a set of control experiments with the goal of measuring streptavidin binding efficiency. To separate the issue of sequence from labelling, no I tiles were included and O tiles were all labelled with biotin; other tiles were the same as in G1, S1 and G2. Because there were no I tiles, only O sequences with biotin labels could be grown and thus reading the sequences was not required to determine streptavidin binding fidelity. The seeds encoded the sequence OOOO. In addition, a higher concentration of tiles and seeds was used in these experiments to increase the density of crystals on the mica.

50 nM of edge tiles, 50 nM of nucleation barrier tiles, 200 nM of biotin-labelled orange (O) block tiles, 2 nM of M13MP18 scaffold strand for the seed, 50 nM of each of the 192 staple strands and 100 nM of each of the adapter tile strands were combined in a total of 50 μL standard buffer. In these experiments, the edge tiles were not biotin labelled. The samples were annealed and prepared for atomic force microscopy following [4], except that 2 μl of 1 μM streptavidin in distilled water was added to the solution before atomic force microscopy. In this procedure, to improve image quality, NaCl was not included in the imaging buffer, which allowed crystals to bind more tightly to the surface and prevented smearing due to crystal slipping. We found that the binding efficiency of streptavidin to biotin labelled tiles was 0.91 ± 0.02. More detailed results and an analysis of how this efficiency of binding allowed us to accurately measure mutation and replication rates are presented in Supplemental Note 3.

10 Data Collection and Tabulation. In order to accurately characterize the population of crystals at each stage of growth and scission, it was necessary to record images of random crystals in the population, *i.e.* to avoid selection bias in choosing crystals to image.

For this reason, we recorded images at randomly chosen, non-overlapping locations on a mica puck using a predetermined series of image coordinates. Most images we collected were 2 μm on a side. In some cases during G2 when the density of the crystals on the puck was particularly low, a modified image collection algorithm was employed:

images of larger regions (usually 5 μm on a side) were taken, and when a crystal was observed, an overlapping image was taken that was higher resolution (and thus detailed enough to read a crystal's sequence and identify errors) and that included the entire crystal. This more detailed image was used for tabulation. In cases where more than one crystal was only partially observable in a large region, it was necessary to choose which crystal to take a second image (because of stage drift, it was not always possible to return to the original image to locate a second crystal). To avoid bias, the crystal observed closest to the upper left hand corner of the image (using lexicographic coordinates) was chosen to image in greater detail. Generally, samples were collected as long as imaging conditions remained favorable, on average 2–6 hours after sample preparation. At least 75 images with crystals were collected and quantified for each of G1, S1 and G2 over multiple sessions of growth, scission and continued growth.

We characterized the number of correct, incorrect and spuriously nucleated crystals in each phase by hand. Structures which did not appear to be well-defined ribbons were not counted, and layers and growth fronts which were not clearly legible were not included in statistics, although their numbers were not large enough to qualitatively affect our conclusions (Supplemental Note 14).

Unless otherwise described, error bars and confidence intervals described are determined by a bootstrapping process where we computed the described quantity using half the raw data chosen at random. To calculate a particular confidence interval, this procedure was repeated 1000 times, and range comprising 95% of these samples was considered the error, with the confidence interval adjusted for the bootstrapping sample size.

11 Scission. This section describes the construction of the scission device and the protocol for its operation. The analysis that underlies the design of the device, *i.e.* how we calculated the dimensions and pressures necessary to achieve the desired frequency of scission, are described in Supplemental Note 4.

Essentially, the scission device is a channel with a constriction where the fluid inside the channel is propelled by the release of compressed gas. The components of the device had to be rated for the high pressures (approximately 800 psi) we used to fragment ribbons. As a result, the tubing and junctions were primarily HPLC (high-pressure liquid chromatography) equipment. A diagram of the device is given in Supplemental Figure 11.

Compressed air was supplied by a scuba tank (13.2 ft^2 pony scuba tank, Luxfer AL 13) small enough to fit into the temperature-controlled glove box where our experiments were performed (Supplemental Methods 4). The scuba tank was connected to a 4000 psi gas regulator (Johnstone Electronics). Gas release was controlled electronically via an always-off (*i.e.* the valve is off unless current is applied) solenoid valve (Peter Paul H22G9DGV) and a timer relay that emitted an 0.1 second pulse to activate the valve (Crouzet TMR-48). Screwthread adapters (Orchard Supply Hardware) were used to connect the regulator's output to high-pressure rated HPLC adapters and tubing ferrules (Upchurch parts F-294, F-290x) and then HPLC tubing (0.02" ID PEEK tubing, Upchurch Scientific part 1532).

Before scission, samples were loaded into the circuit downstream of the always-off solenoid valve and upstream of the constriction via a syringe using a high-pressure 3-way manual valve (Rheodyne 3-way 2-position Switching Valve part 7030). We created a constriction by loading a 1-cm radius steel gasket with a 40±4 μm diameter, 125±50 μm deep laser-cut hole (Lenox Laser SS-4-VCR-2-40) into a high-pressure filter assembly (Upchurch A-411, with inline filter removed) to which tubing was fitted with adapters at the entrance and exit. Because the diameter of the gasket was slightly smaller than the diameter of the filter the housing was designed for, we used a custom-machined O-ring to keep the gasket and its constriction aligned with the tubing. The end of the output tubing was inserted into a 2 ml Eppendorf tube coated with Parafilm to prevent splashing.

Performance of the scission device was hampered by the propensity of the constriction to become clogged with salt from the buffer between experiments, despite the fact that immediately after scission, several milliliters of distilled water were passed through the system via syringe to clear any residual salt. Clogging occurred much more frequently when very small constriction sizes were used. We found that a constriction diameter of 40 μm to be the smallest constriction that was clogged relatively infrequently.

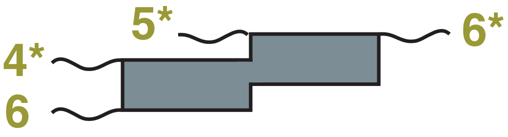
12 Sequences

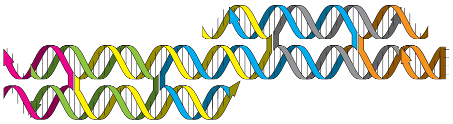
12.1 Tile Sequences

For each tile, we illustrate the sequences by giving the corresponding diagram from Figure 1d in the main text along with a ribbon strand diagram of the assembled tile. In the ribbon diagrams, the end with an arrow is the 3' end of the strand. Each strand in the ribbon diagram is a different color; the sequences for each colored strand are given below the diagrams along with the relevant color code. Sequence names are nonstandard because many of the tiles and strands were used in previous work, and their names are either carried over from that work or modified to indicate their relationship to it.

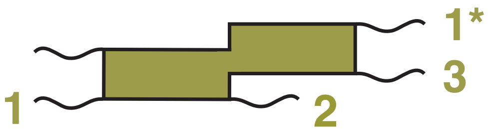
The sequences for tile strands are given in Supplemental Methods 12.2–12.6, Supplemental Methods 12.7 describes how tile strand sequences were labelled with biotin and lists biotin-labelled sequences.

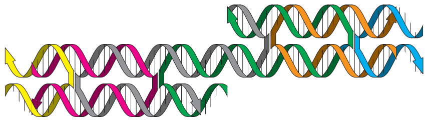
12.2 Edge Tiles





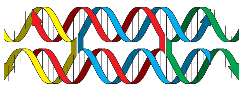
Magenta strand	Z56-1	cgttagctcggcacctcaaacatgtc
Green strand	Z56-2	gtttgaggacgctatgaacatccacctaagcagagacacctgccgagc
Gray strand	Z56-3	cgatgacctgtctggagatcgagtggtgaaccgcagtaggacgcctcg
Orange strand	Z56-4h	taacgcgaggcgtggtcatcgaaggcccttttgggcctt
Blue strand	Z56-5	cggaagcattggtggtactcgtgcttgctttcggactcgatctccagacacctactgcggttcacctgcgaacg
Yellow strand	Z56-6	gtgatcgttcgcaccgaaagcaagcacgagtacctggatgttcatagcgtggtgtctctgcttaggaccaatgcttccg



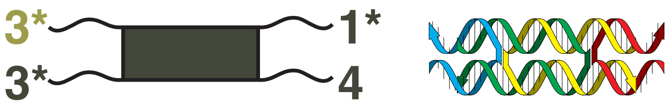


Yellow strand	Z78-1	aggttctaccgcaccagaatgcaaga
Magenta strand	Z78-2n	cattctggtgaccataagataggaggtgcagtgaataggctgcggtag
Orange strand	Z78-3n	cagaagcaggagtgtaggcgatggtgttcgctctcagtgtccgattgg
Blue strand	Z78-4vs	aacctccaatcggtgcttctgaccaa
Green strand	Z78-5vs	ttggtcgatccgtggctactgctgtcggcacggtccatcgcctacactccacactgagagcgaacaggacgaaggtatg
Gray strand	Z78-6n	catacttcgtccaccgtgccgacagcagtagcctcctatcttatggtcagcctattcactgcaccacggatcg

12.3 Top Nucleation Barrier Block Tiles

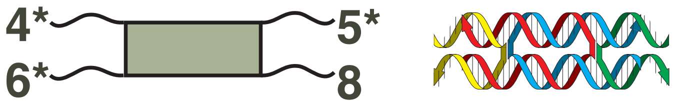


Yellow strand	ZEnp-1	accaactgcatctgtcctcagtacga
Red strand	ZE-2nh	ctgaggactctcgacgggtcaggtgatccgaactccagcgagatgcag
Blue strand	ZE-3nh	cctgtgagacctgacccgtcgagacgctggagttcggatctggcaacg
Green strand	ZE-4	tccttcgttgccactcacaggcttac



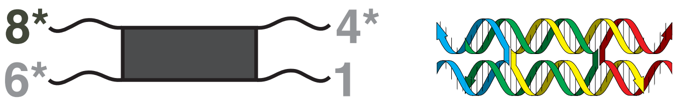
Blue strand	ZFnp-1	aaggactgacagacggcacagttggt
Green strand	ZF-2nh	ctgtgccgagcatcgccagcactaccacgcctctcgtagctctgtcag
Yellow strand	ZF-3nh	caaggctctagtgctggcgatgctgctacgagaggcgtggaccatgag
Red strand	ZF-4	caaacctcatggtgagccttgtcgta

Blue strand	ZG-1	gcttagccaaggcaggacttcgtaag
Green strand	ZG-2nh	gaagtcctcgagactgtagctggcaatccgtcggtgctcagccttggc
Yellow strand	ZG-3nh	gcctggaagccagctacagtctcgtgagcaccgacggattcgactagc
Red strand	ZG-4vs	ttccagctagtcgttccaggcatgac



Yellow strand	ZH-1	gtttggacactccacctcaacgtcat
Red strand	ZH-2nhp	gttgaggtccagttggcacagtccaccgatctgcctcgtaggagtgtc
Blue strand	ZH-3nhp	gtgagacaggactgtgccaactggtacgaggcagatcggtcaaggacc
Green strand	ZH-4vs	taagcggtccttgtgtctcacttcct

12.4 Bottom Nucleation Barrier Block Tiles



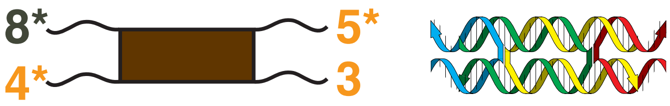
Blue strand	Z1-1v	cttgtcaaacgcaccactctgaggaa
Green strand	Z1-2	cagagtggacgaaagctcacggcaccaagtatcaggttcctgcgtttg
Yellow strand	Z1-3	ctgtagcctgccgtgagctttcgtggaacctgatacttggacgagttg
Red strand	Z1-4	atgctcaactcgtggctacagaagag

Blue strand	Z2-1	aagtcgaacgaccacatcatccgatt
Green strand	Z2-2	gatgatgtccttgtaaacttcgccactctaatcgcaatcaggtcgttc
Yellow strand	Z2-3	gagcaacaggcgaagtttacaaggtgattgcgattagagtccgtaagc
Red strand	Z2-4	atcacgcttacggtgttgctcgctaa

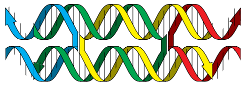
Yellow strand	Z3-1v	tggaacagccagtggtaggagctctt
Red strand	Z3-2n	ctcctacctgcgaatctctgtagtggtgtcgtctgctcggactggctg
Blue strand	Z3-3n	caatgcggactacagagattcgcaccgagcagacgacacctgagacgg
Green strand	Z3-4	acaagccgtctcaccgcattgaatcg

Yellow strand	Z4-1	agcatggcaatccacaaccgcttagc
Red strand	Z4-2	gcggttgtccaacttaccagatccacaagccgacgttacaggattgcc
Blue strand	Z4-3	gctctacaggatctggtaagttggtgtaacgtcggcttgtccgttcgc
Green strand	Z4-4	gacttgcgaacggtgtagagcgacat

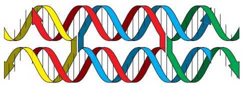
12.5 O Tiles



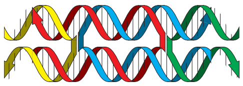
Blue strand	Z9-4n	tgagtgttcatccactcgctcaggaa
Green strand	Z9-3	gagcgagtccatatcaatgaatccacgcaactacgtcgcaggatgaac
Yellow strand	Z9-2	ggtcagcaggattcattgatatggtgcgacgtagttgcgtccgataac
Red strand	Z9-1n	ggtaagttatcggtgctgaccagaac



Blue strand	Z10-4n	aatggctagaacacctcattcggtta
Green strand	Z10-3	gaatgaggactgagtaaggcgacaccgtagatgcgtttcctgttctag
Yellow strand	Z10-2	ctaaaccctgtcgccttactcagtggaaacgcatctacggacttctgc
Red strand	Z10-1v	ttccagcagaagtgggtttagtgagt

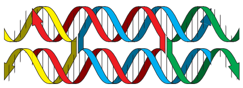


Yellow strand	Z11-4	tggaactcttagtggaagtcggttct
Red strand	Z11-3	cgacttcctgataacatttgtagtggctagatactctcggactaagag
Blue strand	Z11-2	ctaaagggactacaaatgttatcaccgagagtatctagcctggttcgg
Green strand	Z11-1	actcaccgaaccaccctttagtaacc

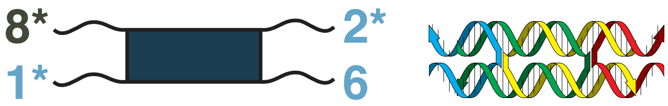


Yellow strand	Z12-4	ttaccgtttcgccacatctcgactca
Red strand	Z12-3	cgagatgtccgtaatcgttatgccacatggaaatgcaacaggcgaaac
Blue strand	Z12-2	gcgatacaggcataacgattacggtgttgcatttccatgtccttagac
Green strand	Z12-1v	ccattgtctaaggtgtatcgcttcct

12.6 | Tiles

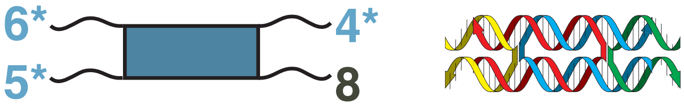


Yellow strand	ZA-1	tggaactgaccttgcagcacgatgac
Red strand	ZA-2nh	cgtgctgctcaggctcagtaggttgtccgaatggctggtgaaggtcag
Blue strand	ZA-3nh	cgaagctgaacctactgagcctgacaccagccattcggacttggaggc
Green strand	ZA-4	gagatgcctccaacagcttcgagcat



Blue strand	ZB-1	`atctcgtgactccatcgcatcaggaa`
Green strand	ZB-2nh	`gatgcgatccgaccatcctagcacagcagcctgcagagtaggagtcac`
Yellow strand	ZB-3nh	`gtggttgagtgctaggatggtcggtactctgcaggctgctccgtatcc`
Red strand	ZB-4	`atgacggatacggtcaaccacgtcat`

Blue strand	ZC-1	`tcagacgttccgaccacactcatgct`
Green strand	ZC-2nh	`gagtgtggaacctgtccaagacgaccgcacctcagccacctcggaacg`
Yellow strand	ZC-3nh	`cagtcgcctcgtcttggacaggttggtggctgaggtgcggataggagc`
Red strand	ZC-4v	`ttccagctcctatggcgactgacaca`

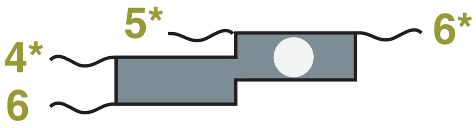


Yellow strand	ZD-1	`gtcatgctacgccacaagcagtgtgt`
Red strand	ZD-2NoHP	`ctgcttgtcctcgaacggaatgccaccagacctctgcgaaggcgtagc`
Blue strand	ZD-3NoHP	`gtccatcaggcattccgttcgaggttcgcagaggtctggtccgttgcg`
Green strand	ZD-4v	`tctgacgcaacggtgatggacttcct`

12.7 Biotin-Labelled Sequences

To label sequences with biotin we added an extra thymine (T) base halfway between the two cross-over points to which a biotin group was covalently linked (Integrated DNA Technologies internal biotin modification). We used two biotin labels per tile, with the two labels being at the same position along a helix on opposite strands. The biotin label did not appear to significantly affect tile stability compared with the same strand without an additional biotin-labelled thymine base (Supplemental Figure 3). In the sequences below, `/iBiodT/` denotes a biotin-labelled thymine base.

To produce the labelled bottom edge tile



the following two strands were used

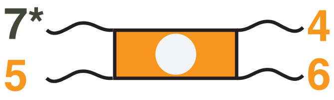
Green strand	Z56-2bio	`gtttgaggacgctatgaacatccacctaagca/iBiodT/gagacacctgccgagc`
Yellow strand	Z56-6bio	`gtgatcgttcgcaccgaaagcaagcacgagtacctggatgttcatagcgtggtgtctc/iBiodT/tgcttaggaccaatgcttccg`

in place of

Green strand	Z56-2	`gtttgaggacgctatgaacatccacctaagcagagacacctgccgagc`
Yellow strand	Z56-6	`gtgatcgttcgcaccgaaagcaagcacgagtacctggatgttcatagcgtggtgtctctgcttaggaccaatgcttccg`

The biotin-labelled versions of Z56 strands (as opposed to the unlabelled strands Z56-2 and Z56-6) were used in all experiments.

To produce the labelled O tile



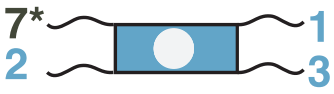
the following two strands were used

Red strand	Z11-2b	`ctaaagggactacaaa/iBiodT/tgttatcaccgagagtatctagcctggttcgg`
Blue strand	Z11-3b	`cgacttcctgataaca/iBiodT/tttgtagtggctagatactctcggactaagag`

in place of

Red strand	Z11-2	`ctaaagggactacaaatgttatcaccgagagtatctagcctggttcgg`
Blue strand	Z11-3	`cgacttcctgataacatttgtagtggctagatactctcggactaagag`

To produce the labelled I tile



the following two strands were used

Red strand	ZA-2nhb	`cgtgctgctcaggctc/iBiodT/agtaggttgtccgaatggctggtgaaggtcag`
Blue strand	ZA-3nhb	`cgaagctgaacctact/iBiodT/gagcctgacaccagccattcggacttggaggc`

in place of

Red strand	ZA-2nh	`cgtgctgctcaggctcagtaggttgtccgaatggctggtgaaggtcag`
Blue strand	ZA-3nh	`cgaagctgaacctactgagcctgacaccagccattcggacttggaggc`

12.8 Seed Sequences

The crystal seed consists of a folded 7249-base single-stranded DNA bacteriophage, M13mp18, and "staple" strands which fold the vector into a rectangle shape [12], and "adapter" strands which serve as binding sites for the nucleation of a sequence. Only strands nucleating OIIO were used in this work, but simple modifications to the adapter strands can nucleate arbitrary I/O sequences [4].

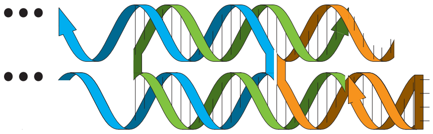
The **seed scaffold** was obtained from New England Biolabs (Massachusetts, US), N4040S. The sequence is available from Genbank at `http://www.ncbi.nlm.nih.gov/nuccore/58253`

The **seed staples** are the same as those used to fold the seed in a previous work [4], which lists the sequences of each strand used.

12.9 Seed Adapter Sequences

Fifteen sets of **adapter strands** attach to the right side of the seed structure to nucleate the OIIO sequence. Each set of strands is termed an "adapter tile" as it acts as a tile to which other (non-adapter) tiles attach. The following illustrations show the adapter strands assembled on a segment of the scaffold strand, shown in blue. The adapter tiles are arranged on the seed as shown in Supplemental Figure 2.

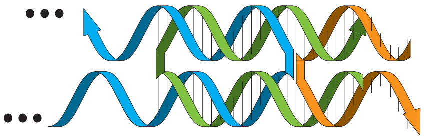
Adapter Tile 1



The blue strand is the region of the scaffold strand to which the adapter tile attaches on the seed.

Green strand	AT1-2	`ccaagcggtgaaagtattaagaggctattattctgaaacagagtcggg`
Orange strand	AT1-3VWCOPY	`taacgcccgactcccgcttggcaagaccttttggtcttg`

Adapter tiles 2–14 have the following structure:



The blue strand is the region of the scaffold strand to which the adapter tile attaches on the seed.

Adapter Tile 2

Green strand	AT2-2	`ggcaccctgtcagacgattggcctcaggaggttgaggcagtgcgggct`
Orange strand	AT2-3V	`acaagcgtaaccgatctggacaatcg`

Adapter Tile 3

Green strand	AT3-2	`gtccagatgcgtcagactgtagcgatcaagtttgcctttacggttacg`
Orange strand	AT3-3V	`ccattatgcgatgccgtgtgtttcct`

Adapter Tile 4

Green strand	AT4-2	acacacggagacaaaagggcgacaggtttaccagcgccaacatcgcat
Orange strand	AT4-3V	actcagtaaggtggcttccgctaacc

Adapter Tile 5

Green strand	AT5-2	gcggaagcgcagatagccgaacaatttttaagaaaagtaacaccttac
Orange strand	AT5-3V	ccattaactccacggtcctccttcct

Adapter Tile 6

Green strand	AT6-2	ggaggaccaacgtcaaaaatgaaaaaacgatttttttgtttgtggagtt
Orange strand	AT6-3VWCOPY	tctgaaactccacggtcctccttcct

Adapter Tile 7

Green strand	AT7-2	gtcctagtgcttatccggtattctaaatcagatatagaagtgggcagc
Orange strand	AT7-3VWCOPY	gagatgctgcccaactaggacagcat

Adapter Tile 8

Green strand	AT8-2	ctgtgactacgcgcctgtttatcagttcagctaatgcagaggggtggg
Orange strand	AT8-3VWCOPY(1-bit)	tctgacccaccccagtcacagttcct

Adapter Tile 9

Green strand	AT9-2	`caatgtgcgaaaaagcctgtttagggaatcataattactacggagccc`
Orange strand	AT9-3VWCOPY(1-bit)	`gagatgggctccggcacattgagcat`

Adapter Tile 10

Green strand	AT10-2	`gcatgtcccataggtctgagagacgtgaatttatcaaaatacgcgagc`
Orange strand	AT10-3V	`ccattgctcgcgtggacatgcttcct`

Adapter Tile 11

Green strand	AT11-2	`gtacaaccgaagatgatgaaacaaaattacctgagcaaaactacgagg`
Orange strand	AT11-3V	`actcacctcgtagggttgtactaacc`

Adapter Tile 12

Green strand	AT12-2	`gggtactgacttctgaataatggatgattgtttggattatggtgagtc`
Orange strand	AT12-3VS	`taagcgactcacccagtacccttcct`

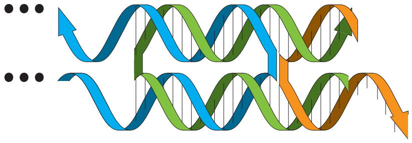
Adapter Tile 13

Green strand	AT13-2	`agtggtaggccgtcaatagataatcaactaatagattagacatcgaga`
Orange strand	AT13-3VS	`tcctttctcgatgctaccactcttac`

Adapter Tile 14

Green strand	AT14-2	`gacccggcccagcagaagataaaaaataccgaacgaaccagtgtcgaa`
Orange strand	AT14-3VS	`aacctttcgacacgccgggtcaccaa`

Adapter Tile 15



The blue strand is the region of the scaffold strand to which the adapter tile attaches on the seed.

Green strand	AT15-2	`ctcggggactacattttgacgctcacgctcatggaaatactcgtgcaa`
Orange strand	AT15-3VS	`ttgcacgatccccgagtcttg`

12.10 Guard Strand Sequences

Z56-4h-C `aaggcccaaaagggccttcgatgaccacgcctcgcgtta`

Z78-4vs-C `ttggtcagaagcaccgattggaggtt`

ZE-4-C `gtaagcctgtgagtggcaacgaagga`

ZH-4vs-C `aggaagtgagacacaaggaccgctta`

Z3-4-C `cgattcaatgcggtgagacggcttgt`

Z4-4-C `atgtcgctctacaccgttcgcaagtc`

Z11-1-C `ggttactaaagggtggttcggtgagt`

Z12-1v-C `aggaagcgatacaccttagacaatgg`

ZA-4-C `atgctcgaagctgttggaggcatctc`

ZD-4v-C `aggaagtccatcaccgttgcgtcaga`

Supplemental Notes

1 Combinatorial Replication Allows for the Replication of an Arbitrary Number of Species. In this work we are concerned with the capacity for combinatorial information replication. We define combinatorial information replication as the capacity to replicate an arbitrary number of different species. An example of a system with this capacity is a system that can replicate a chemical sequence of any length. Examples of chemical systems without this capacity are an autocatalytic system that can replicate only a single species or a system that can replicate chemical sequences of a fixed length n.

Combinatorial information replication is of interest because systems with this capacity have the potential for "open-ended evolution," a process of Darwinian evolution that can continue, and potentially produce sequences of increasingly complexity, without a clear stopping point. In biology, for example, evolution has proceeded from an initial state consisting of a system simple enough to arise spontaneously to the creation of organisms of today that have billion-bit genomes, without evidence that the complexity of organisms that exist today represents any fundamental limit.

2 Conversion of Crystal Length to Number of Layers. Most images of crystals from G1 were of sufficiently high quality that we were able to see both the sequence and count the number of individual layers in a crystal. As such, we were able to report the mean number of layers in templated crystals in the main text. The mean length, 620±40 nm, and the mean number of layers, 39±3, suggest that each layer was approximately 16 nm in length. As we measured, the length of each layer is expected to be approximately the diagonal length across a tile, previously determined to be approximately 12.6 × 4–6 nm in size [17], which would be around 13–14 nm, in close agreement.

3 Analysis of Streptavidin Binding Efficiency and Determination of Sequences in Crystals Using Streptavidin Labels. Binding efficiency experiments were performed as described in Supplemental Methods 9. We defined

$$\text{Binding efficiency} = \frac{\text{Number of bound streptavidin}}{\text{Number of bound streptavidin + number of sites with tiles but no streptavidin}}. \tag{1}$$

In the binding efficiency experiments, the measured binding efficiency was 0.91 ± 0.02. We also found that at some sites where streptavidin was expected, a hole in the assembly was present, suggesting that the tile (and possibly a bound streptavidin) had been separated (or ripped) during AFM scanning of the sample. We defined

$$\text{Fraction of ripped binding sites} = \frac{\text{Number of ripped sites}}{\text{Number of ripped sites} + \text{Number of sites with either tile or streptavidin present}}. \tag{2}$$

We found that 0.043 ± 0.008 of sites were ripped in the same experiments, suggesting that ripping had a small but significant effect on the binding efficiency we would expected to observe.

We also measured the rate at which "false positive" signals in streptavidin binding, *i.e.* where streptavidin appeared on the assembly but could not have been attached to a biotin label. In the binding efficiency experiments, the regions where streptavidin labels were not expected were in the edge and nucleation barrier tiles. We defined

$$\text{Fraction of false positive binding} = \frac{\text{Number of false binding events}}{\text{Number of total blocks where streptavidin is not expected to bind}}. \tag{3}$$

We found that 0.04 ± 0.02 of sites without biotin labels showed false positives in binding.

To determine whether the binding efficiency changed over the course of imaging, we also measured these same quantities for each image and plotted the resulting values as a function of the image number (Supplemental Figure 4). Images were taken at a steady pace, with each image taking 3-5 minutes of scan time. We fit the fractions using a linear fit and found that the fraction of sites that potentially had a biotin labelled but were unoccupied decreased as imaging proceeded (slope = 0.003), suggesting that some small amount of streptavidin binding occurred during imaging. The fraction of ripped sites or false positives did not change significantly with image number (slopes = -0.0006, 0.0009 respectively).

To ensure that the differences between the protocol used to measure streptavidin binding rates and the protocol used to image crystals after growth or replication did not affect the measured rates of streptavidin binding, we measured streptavidin binding efficiency in 25 randomly chosen images from the G1 growth stage that were clear enough to allow the identification of each individual streptavidin binding site. In these images, we found that the binding efficiency was 0.90 ± 0.01, and the rate at which tiles containing streptavidin were ripped was 0.04 ± 0.01, and the false positive binding rate was 0.05 ± 0.02. These rates are statistically indistinguishable from the rates we measured in the binding efficiency experiment.

To determine the fidelity of sequence copying, we included in our measurements of the fraction of correct growth fronts only crystals whose sequences were legible at the growth front. When tabulating statistics about the fraction of correct layers, only layers whose sequence could be reliably determined using the layer itself and nearby layers within the same crystal were included.

When an error or mutation occurred in a crystal, in general that error is propagated over many rows. There is some chance that an error would occur and then revert back to the original sequence, but such an event should occur much more rarely than single errors. As a result, it was not necessary that the streptavidin sequence in every layer be completely legible. Instead *what was necessary was an ability to accurately infer the sequence of every layer based on the labels on the layer itself and those nearby.*

In inferring the sequence of layers and growth fronts, we assumed that layers near growth errors are no more likely to contain errors than other layers and that errors were no more likely to occur at the end of crystals rather than in the middle, which is consistent with the growth model and with previous experimental investigations of algorithmic self-assembly [9, 7, 6, 4]. In cases where not every sequence was completely readable, we then inferred the the *most likely* interpretation of sequence consistent. One important consequence of our assumptions and the measured rate of errors clearly propagated over many rows, is that *no errors followed by reversions to the original sequence after 1-3 layers would be expected* in our sample size of crystals. We would also not expect to observe more than 1 error near the end of a crystal (where the absence of redundant rows can make it hard to determine whether a deviation from the displayed sequence is a result of mislabelling or a true error). To be conservative in our estimates, if there was a case where the labelling near the end of a crystal was ambiguous, this growth front was not counted in our statistics, and the ambiguous layers were ignored. An example of how a crystal with imperfect labelling can be clearly interpreted is shown in Supplemental Figure 4.

If despite the redundant display of sequence information we could not deduce the sequence of a crystal clearly and unequivocally, we discarded the crystal from the sample. The discarded set represented a relatively small portion of crystals (<20%) for each growth stage. Many of the crystals in this illegible subset were crystals for which we could see only a few layers on the captured image, making it difficult to use the principle of redundancy to be sure of the sequence. Because the position of the crystal in a scan, whose coordinates are randomly chosen, is essentially independent of the crystal's sequence, there should be little to no difference in sequence between the crystals that we considered in our sample and those we deemed illegible.

Nevertheless, because some crystals were excluded from our sample, there is a possibility that this exclusion introduces a bias in sequence counts. The analysis of our data given the assumption that every illegible crystal contained sequences other than the OIIO sequence is given in Supplemental Note 14: even in this most conservative case of analysis, the qualitative conclusion that replication occurs is unaffected. Further, other statistics such as the fraction of correct growth fronts at each stage are not altered by more than a factor of approximately 2.

4 Design of the Ribbon Scission Device. The design for the ribbon scission device relies on a previously developed model of elongational-flow-induced DNA nanotube scission [18]. This model was successfully used to quantitatively predict the distribution of lengths of DNA nanotubes after scission as a function of elongational flow rate and the lengths of nanotubes before scission.

We assumed that the mechanism of ribbon scission in elongational flow would be the same as the mechanism of DNA nanotube scission – force-induced melting of sticky ends. Here we adapt the previously developed model [18] to account for the geometric differences between the DNA tile ribbons used here and the DNA nanotubes used previously [18]. We then use this modified model to calculate the amount of elongational flow required to fragment ribbons and the corresponding requirements for the geometry of the scission device and the rate at which to flow ribbon solution through the device.

Following the previous model [18], DNA ribbons will fragment if they reach a critical tension of

$$T_{crit} = n f_c, \tag{4}$$

where n is the number of helices in the ribbon, 28 for an OIIO ribbon, and f_c is the tension required to break a single DNA helix, which we assumed to be equal to a previously measured value (at room temperature) of approximately 65 pN [19].

Stress induced by differences in fluid velocity (and thus drag force) along the length of the crystal applies tension to the crystal. The line tension at the center of the crystal is dependent on the crystal's length, L, and reaches a high enough value to fragment the crystal at a critical length L_{crit}. The relationship between the critical length, critical tension T_{crit} and elongational flow (the flow velocity gradient) is [18]:

$$T_{crit} = \frac{\pi \mu \dot{\epsilon}_{max} L_{crit}^2}{4 \log(L_{crit}/2R)}, \tag{5}$$

where μ is the viscosity of the crystal solution, $\dot{\epsilon}_{max}$ is the maximum elongational flow achieved in the channel and R

is the effective “radius” of the ribbon crystal, which is modelled, roughly, as a cylindrical rod[1].

Rearranging Supplemental Equation 5 gives the maximum elongational flow $\dot{\epsilon}_{max}$ corresponding to a given critical length:

$$\dot{\epsilon}_{max} = \frac{4T_{crit}\log(L_{crit}/2R)}{\pi\mu L_{crit}^2}. \tag{6}$$

Our goal was to achieve high elongational flow in a device consisting of a flow channel with a constriction. Intuitively, the volumetric flux through the device must be constant along the flow path because the crystal solution is incompressible. If the constriction has a much smaller cross-sectional area than the rest of the flow channel, the average velocity through the constriction must be much faster than the velocity in the rest of the flow channel. This difference in velocities produces elongational flow.

By definition, the elongational flow $\dot{\epsilon}$ is

$$\dot{\epsilon} = \frac{\partial u}{\partial \rho}, \tag{7}$$

where u is fluid velocity and ρ is the distance from the center of the constriction’s entrance.

We model the constriction as a cylindrical channel with radius r. The flow inside the constriction must satisfy

$$\dot{V} = \pi r^2 \bar{u} \tag{8}$$

where $\bar{u}$ is the mean fluid velocity across the constriction and $\dot{V}$ is the volumetric flux through the constriction. We assume that the flow velocity in the device at points far from the constriction is negligible compared with $\bar{u}$. Near the entrance to the constriction, the flow is radial. The fluid velocity is therefore a function only of the distance to the channel entrance ρ and the volumetric flux $\dot{V}$ across the surface consisting of all points distance ρ from the entrance to the channel does not change with ρ (for $\rho > r$). This surface is a half-sphere with radius ρ, so that

$$\dot{V} = 2\pi\rho^2 u_\rho \tag{9}$$

where u_ρ is the fluid velocity at a distance ρ from the channel entrance. Because this volumetric flux must be the same as the volumetric flux inside the constriction,

$$u_\rho = \frac{r^2}{2\rho^2}\bar{u} \tag{10}$$

when $\rho > r$.

The scaling of u_ρ with respect to ρ given in Supplemental Equation 10 breaks down near the constriction entrance, since in this region the fluid transitions from the radial flow field outside the constriction to the parallel flow field inside the constriction. Hence, for a constriction whose length l is large compared to r, the elongational rate will reach a maximum near the constriction entrance and rapidly fall to zero within the interior of the constriction. An estimate of the maximum elongational rate is obtained by evaluating Supplemental Equation 7 at $\rho = r$ using Supplemental Equation 10:

$$|\dot{\epsilon}_{max}| \approx \left[\frac{\partial u_\rho}{\partial \rho}\bigg|_{\rho=r} \right] = \left[\frac{r^2}{\rho^3}\bar{u}\bigg|_{\rho=r} \right] = \frac{\bar{u}}{r}. \tag{11}$$

We used compressed gas to propel the crystals through the device. For gas with pressure P, the pressure drop ΔP across the constriction is $P - P_0$, where P_0 is atmospheric pressure. The flow velocity in a long channel (*i.e.* where $l \gg r$ for channel length l) with pressure drop ΔP is [20]:

$$u_z = \Delta P \frac{r^2}{4\mu l}\left(1 - \frac{z^2}{r^2}\right) \tag{12}$$

where z is the radial distance from the center of the channel.

Averaging the fluid velocity over z gives the mean channel velocity, *i.e.*

$$\bar{u} = \Delta P \frac{r^2}{8\mu l}. \tag{13}$$

---

[1] Supplemental Equation 5 holds asymptotically for $L_{crit} \gg 2R$. In the case where L_{crit} is comparable to $2R$, this expression should be considered only a semiquantitative guide. Here we consider the case where $L_{crit} \approx 10(2R)$.

Thus, for a constriction with radius r and pressure drop ΔP, the maximum elongational flow achieved within the device is

$$\dot{\epsilon}_{max} = \frac{\bar{u}}{r} = \frac{\Delta P r}{8\mu l}. \tag{14}$$

To maximize $\dot{\epsilon}_{max}$, a large pressure drop and a short constriction are optimal[2]. The constriction must also be narrow enough such that the flow rate outside the constriction is negligible compared to the flow rate inside the constriction.

Combining Supplemental Equations 6 and 14 gives:

$$\frac{4T_{crit}\log(L_{crit}/2R)}{\pi\mu L_{crit}^2} = \frac{\Delta P r}{8\mu l}. \tag{15}$$

For crystals to fragment, the critical fragment length L_{crit} needs to be significantly shorter than the mean length of the crystals after growth. The average length of the nucleated crystals we observed was $L = 620$ nanometers and the additional length of the seeds was approximately 90 nm, for an initial average crystal length of $L_0 = 710$ nm. We designed the system to produce fragments about L_{crit}=200 nm in size. the fragment size would correspond to a critical length of 400 nm.

It is was easier to change the pressure drop across the device than to change the geometry of the channel and constriction of an existing device. We therefore built a constriction setup that was capable of withstanding a wide range of pressures with the idea that we could change the applied pressure if our calculations suggested a pressure that in experiments turned out to be too low or too high. However, the pressure we estimated initially produced fragment sizes in line with what we designed the device to do.

The small amount of sample available for scission (300 μl) required that both the device's flow channel and the constriction itself be narrow, minimizing total volume. The flow channel was tubing with an 0.02" (500 micron) inner diameter. The constriction consisted of a $r = 20 \pm 2$ μm radius, $l = 125 \pm 50$ μm long laser-cut hole in a steel disk. (The majority of the disk was approximately 1 mm thick; the constriction was cut through a depression in the disk's center.). A ribbon's height is the height of a DNA helix (2 nm from [13]) and a ribbon's width is the average measured width of OIIO ribbon crystals, which we found to be approximately 120 nm. Approximating the perimeter of the ribbon as a cylinder, we have $2(120 + 2) = 2\pi R$, which gives an effective radius of approximately R=39 nm. We assumed that μ, the viscosity of the solution, was the viscosity of water at 30 °C, 0.7978 mPa·s.

Given these values, the pressure required to achieve L_{crit}=200 nm is approximately 300 psi. Taking into account the large fabrication variance in the length of the gasket constriction, the range of possible pressures required for an individual gasket is at least ΔP = 200–450 psi. In the experiments we report, we used $P = 800$ psi (pounds per square inch) of gas pressure, although experiments with 600 psi and 1200 psi of applied pressure also fragmented crystals successfully.

5 Quantifying the Number of Observed Growth Fronts and Layers. The replication process was repeated many times, and it was not possible to take a large sample of images of crystals at each stage: imaging was subject to, among other things, the exigencies of atomic force microscopy tips, which were variable in quality, and some imaging sessions produced more usable images than others. Thus, raw data was collected from several experiments and pooled.

Crystal seeds were neither created nor destroyed in the experiment, and are imaged on the same surface as crystals. The number of crystal seeds is also the number of initial growth fronts in the system. Therefore, to track changes in the concentration of growth fronts and layers through the stages of replication, we compared the number of growth fronts and layers we observed via AFM to the number of seeds we observed via AFM.

The energetic barrier to spurious nucleation grows with ribbon width [6, 7]. As a result, when no template for ribbon growth is provided, ribbons overwhelmingly contain 0, or occasionally 1, bit of sequence information [4]. Spuriously nucleated ribbons are therefore easily distinguishable from ribbons that were templated with an OIIO sequence because spuriously nucleated crystals are much thinner. In this work, any crystal with 0 bits or 1 bit of

---

[2] Supplemental Equation 14 also suggests that for a constant pressure drop, the elongational flow decreases as constriction size decreases. This may seem counterintuitive, because for the same flux through the circuit a smaller constriction could increase the flow rate through the constriction, thus increasing the elongational flow and applying more tension to crystals. However, as constriction size decreases, the pressure drop required for a particular flux increases quadratically.

sequence information was assumed to have been spuriously nucleated. No crystals with 2 bits of sequence information were observed, so errors which reduced sequence lengths were unlikely to have been frequent enough that any of the crystals carrying 0 or 1 sequence bit could have descended from a templated crystal.

6 The Dynamics of Growth and Scission Exponentially Replicate Crystal Sequences. Equations 1–2 in the main text present a simple model of sequence replication by crystal growth and scission:

$$L_s[n+1] = L_s[n] + gF_s[n] \tag{16}$$
$$F_s[n+1] = F_s[n] + 2fL_s[n+1] \tag{17}$$

where $L_s[n]$ is the number of layers of sequence s in generation n and $F_s[n]$ the number of growth fronts of sequence s at generation n. Solving Supplemental Equation 16 for $F_s[n]$ and for $F_s[n+1]$ in terms of L_s and substituting those expressions into Supplemental Equation 17 yields the expression

$$L_s[n+2] - 2(1+fg)L_s[n+1] + L_s[n] = 0. \tag{18}$$

To determine the asymptotic replication rate, we compute the Z-transform [21] of Supplemental Equation 18 which gives us an equation in Z-transform space for the replication rate r:

$$r^2 - 2(1+fg)r + 1 = 0. \tag{19}$$

(Note that r is used for the replication rate for clarity, but the variable z is used by convention in the discrete frequency domain.)

Solving Supplemental Equation 19 yields

$$r = (1+fg) \pm \sqrt{(1+fg)^2 - 1} \tag{20}$$

or

$$r_+ = (1+fg) + \sqrt{(1+fg)^2 - 1} \tag{21}$$
$$r_- = (1+fg) - \sqrt{(1+fg)^2 - 1}. \tag{22}$$

Taking the inverse Z-transform of Supplemental Equation 19, we have

$$L_s[n] = c_+ r_+^n + c_- r_-^n \tag{23}$$

where c_+ and c_- are constants of integration.

At time 0 we had no layers and quantity 1 of nuclei, of which the fraction $(1 - e_{nuc})$, where e_{nuc} is the nucleation error rate from the programmable seeds, grew the correct sequence. That is,

$$L_s[0] = 0 \tag{24}$$
$$F_s[0] = (1 - e_{nuc}). \tag{25}$$

To put Supplemental Equation 25 in terms of L_s, we substitute Supplemental Equations 24 and 25 into Supplemental Equation 16 for $n = 0$ to get

$$L_s[1] = gF_s[0] + L_s[0] = g(1 - e_{nuc}). \tag{26}$$

To solve for c_+ and c_-, we substitute $n = 0$, $L_s[0]$ to get

$$0 = c_+ + c_- \tag{27}$$
$$c_- = -c_+, \tag{28}$$

and $n = 1$, $L_s[1]$ to get

$$g(1 - e_{nuc}) = c_+ r_+ + c_- r_- \tag{29}$$

$$g(1 - e_{nuc}) = c_+ (r_+ - r_-), \tag{30}$$

so that

$$c_+ = \frac{g(1 - e_{nuc})}{r_+ - r_-} \tag{31}$$

$$= \frac{g(1 - e_{nuc})}{2\sqrt{(1 + fg)^2 - 1}}. \tag{32}$$

Thus,

$$L_s[n] = c_+ \left(r_+^n - r_-^n \right). \tag{33}$$

Supplemental Equation 33 isn't exactly exponential, but since $r_+ > 1$ and $r_- < 1$, it predicts asymptotically exponential growth at replication rate r^+, the value we report in the text as the replication rate,

$$\lim_{n \to \infty} L_s[n] = c_+ r_+^n. \tag{34}$$

7 Determination of the Replication Rate. In the model of replication in Supplemental Equations 16–17, two parameters f and g determine the rate of sequence replication. Initially (see Supplemental Note 6),

$$F_s[0] = 1 - e_{nuc} \tag{35}$$

$$L_s[0] = 0. \tag{36}$$

The measured values for $F_s[1]$, $F_s[2]$ and $L_s[1]$ constrain the values of f and g via the following expressions

$$L_s[1] = gF_s[0] = g(1 - e_{nuc}) \tag{37}$$

$$L_s[2] = gF_s[1] + L_s[1] \tag{38}$$

$$F_s[1] = 2fL_s[1] + F_s[0]. \tag{39}$$

We used a maximum likelihood approach to fit f and g given these constraints. We assumed that the measured parameters were uncertain and that their true value could be sampled by probability distributions where the standard deviation is half the reported 95% uncertainty.

To estimate the best fit for f and g, we sampled the values of e_{nuc}, $F_s[1]$, $L_s[1]$ and $L_s[2]$ according to this distribution and then solved for f and g via least squares. We repeated this process 10,000 times. The reported values, *i.e.* $f = 0.0016 \pm 0.0003$ per nm per generation and $g = 540 \pm 70$ nm per generation, and the asymptotic replication rate $r = (1 + fg) + \sqrt{(1 + fg)^2 - 1} = 3.5 \pm 0.4$ are the mean values determined by the simulations. The 95% confidence intervals are the ranges within which at least 95% of the fits fell.

To check that our simple model is consistent with our experimental observations, we note that we inferred only two parameters f and g but assumed the three constraints on the measured parameters given in Supplemental Equations 37–39. If all three constraints can be satisfied simultaneously for the values of f and g we fit, this provides at least some confidence in our model. Two of the constraints involve just the parameter g. If they are both satisfied simultaneously,

$$\frac{L_s[1]}{1 - e_{nuc}} \approx \frac{L_s[2] - L_s[1]}{F_s[1]}. \tag{40}$$

Supplemental Figure 14 shows overlaid histograms of the probability distributions of $\frac{L_s[1]}{1 - e_{nuc}}$ and $\frac{L_s[2] - L_s[1]}{F_s[1]}$ as determined by sampling. The distributions are not significantly different from one another.

8 A Simplified Model of Scission Does Not Qualitatively Alter The Predicted Replication Rate. The model of replication that we used to infer the replication rate makes several simplifications of the replication process. Most notably, the model assumes that per layer fragmentation rates are independent of each other, which is not consistent with the physics of the scission process (Supplemental Note 4). We expect that for values where $g \gg L_c$, such that g is the average length of new layers added per growth front per generation and L_c is the critical length at which crystals break, the predictions of this model would be similar to one which modeled the physics of breakage more accurately: in this case, most crystals would be broken into fragments of approximately equal lengths L_c, for an effective breakage rate $f = \frac{1}{L_c}$. However, in our system, while g is larger than L_c it is of roughly the same order, so this assumption does not hold. In the region of parameter space relevant to our experiments, both models predict exponential growth of the number of layers and growth fronts. But the two models predict different crystal length distributions after scission, and as a result, asymptotically different scission rates. Thus, here we examine whether our assumed distribution of crystal scission sites significantly affected the replication rate we report.

Estimation of L_c using a more physically accurate model of scission would require information about the distribution of crystal lengths (as opposed to the average crystal length). Unfortunately, information about crystal length distributions could not be obtained from our experiments: In order to be able to clearly distinguish individual bits of crystal sequences, we used fairly small AFM image sizes during sample characterization. These image sizes were small enough that entire crystals were often not visible in a single image. When a crystal was only partially visible, instrument limitations meant that we could often not reconstruct an image of the entire crystal, in which case we could not measure its length. As a result, we knew the lengths of only some of the crystals in our sample. The crystals whose length we could determine were more likely to be short enough to fit within a single small image, so they were not an accurate sample of the true distribution.

Nevertheless, we can use our limited information about crystal lengths to estimate a plausible length distribution and use this distribution to estimate (rather than predict) the critical length at which crystals broke in our experiments. Here we do this calculation and then compare this estimated critical length with the scission rate predicted in Supplemental Note 7. If the models are compatible, we would expect $2\langle L_c \rangle > \frac{1}{f} > \langle L_c \rangle$, where $\langle L_c \rangle$ is the mean critical length and the allowed range accounts for end affects in the lengths produced by fragmentation. We find that this is true. Thus, our use of a model which simplified the physics of scission did not significantly affect the reported replication rate.

The critical length at which a crystal undergoes scission in a channel with a constriction is dependent on where the crystal passes through the constriction: the smallest critical length $L_{c,min}$ is at the center of the channel, with the critical length diverging at the channel edge. Here we infer the most likely $L_{c,min}$ and from this value compute the average critical length at which scission was likely to occur, $\langle L_c \rangle$.

We assume a Gaussian (*i.e.* a binomial if we measured layers discretely) distribution for the lengths of crystals, which would be expected for crystals that grow from a seed, then find the most likely $\langle L_c \rangle$ for this distribution. Our best estimate for the mean of the distribution is the average length of correct crystals after G1, 520 nm plus 90 nm of length for the crystal seed, and our best estimate for the standard deviation is 236 nm, which is the standard deviation of the lengths of nucleated G1 crystals that appeared entirely on a single AFM image.

To find the most likely $L_{c,min}$, we simulated the scission of 1000 crystals with lengths sampled from this Gaussian distribution for a range of possible values of $L_{c,min}$ and determined the number of new ends created by scission. For each $L_{c,min}$ we ran 100 of these simulations and took the expected number of new growth fronts for that $L_{c,min}$ as the average of the number of new growth fronts found in the 100 simulations. The simulated $L_{c,min}$ that produced the same number of new growth fronts as we observed in our experiments was assumed to be the most likely value.

The critical length $L_{c,z}$ for a crystal passing through the constriction a distance z from the constriction's center, *i.e.* with velocity u_z, satisfies [18]

$$\sqrt{u_z \ln\left(\frac{L_{c,min}}{2R}\right)} L_{c,z} = \sqrt{u_0 \ln\left(\frac{L_{c,z}}{2R}\right)} L_{c,min} \tag{41}$$

where u_0 is the velocity at the center of the channel and R is the effective radius of the ribbon as it goes through the crystal, assumed to be $R = 39$ nm as described in Supplemental Note 4.

For each simulated crystal scission event we randomly selected a value of u_z using the velocity distribution within the channel given by Supplemental Equation 12 and the distribution of area within the cylindrical channel. Given the sampled value of u_z, we computed the resulting $L_{c,z}$. Given $L_{c,z}$, the number of fragments the crystal would split into is $\lfloor \frac{2L}{L_{c,z}} \rfloor$, where L is the length of the crystal undergoing scission.

In our experiments we observed 3.2 ± 0.5 times more correct growth fronts after S1 than after G1. This increase corresponded to a simulated $L_{c,min} = 275$ nm, with a 95% confidence interval of 230–330 nm.

Taking into account the dependency of L_c on both z and u_z,

$$\langle L_c \rangle = \frac{\int_0^r L_c(z) u_z z dz}{\int_0^r u_z z dz} = \frac{4}{u_0 r^2} \int_0^r L_c(z) u_z z dz \tag{42}$$

such that the most likely $\langle L_c \rangle$ is 430 nm, with a 95% confidence interval of 370–490 nm. $\frac{1}{f} = 590 \pm 80$ nm is thus between the predicted values of $\langle L_c \rangle$ and $2 \langle L_c \rangle$.

9 Estimation of the Amount of End-to-End Joining during G2. End-to-end joining of ribbons reduces the number of available growth fronts. Thus, if end-to-end joining occurred frequently it would change the dynamics of replication. End-to-end joining was observed in our experiments (Supplemental Figure 18) and has been observed in previous experiments with DNA tile ribbons [7] and DNA tile nanotubes [22]. Here we compare the rate of joining we observed in our experiments with previously measured rates of DNA nanostructure joining and determine whether a) these rates are similar to what we observed and b) whether joining might have a significant effect on the replication process. We find that even though the joining rates we observed were as high or higher than previously measured rates of DNA tile ribbon joining, joining is not expected to significantly affect the rate of replication in a crystal growth and scission process.

Previously, the rate of end-to-end joining between ribbons was measured to be $k_j = 35,000$ /M/s at room temperature [7]. Although this measurement was imprecise (confidence interval 11,000 to 80,000 /M/s), we can use it to estimate how much joining between templated ends would be predicted during G2. (During G1, joining between templated ends could not occur because joining can only occur between a "left" growth front and a "right" growth front, and the "left" ends were attached to seeds.)

We assume that since joining is an effectively irreversible reaction and the forward rate of DNA hybridization is not strongly temperature-dependent, the end-to-end joining rate at 31 °C is about the same as the end-to-end joining rate at room temperature. After scission, 2.6±0.4 correct and 0.65±0.16 incorrect growth fronts per seed were observed. Since seeds were present at a concentration of 25 pM, the effective concentration of templated growth fronts was thus 81±14 pM. These growth fronts were quickly diluted by a factor of 5 for the start of G2 to a final concentration of 16±3 pM. We assume that no joining took place in the brief time before dilution.

Growth fronts were either of type "left" or of type "right," and there were 5 pM of the growth fronts of type "left" that could not join because they were attached to origami. There were thus effectively 5.5±1.5 pM of "left" growth fronts L and 10.5±1.5 pM of "right" growth fronts R. Joining is a reaction of the form

$$L + R \xrightarrow{k_j} \varnothing. \tag{43}$$

The kinetics of this reaction are

$$\frac{dL}{dt} = -k_j [L][R] \tag{44}$$

$$\frac{dR}{dt} = -k_j [L][R], \tag{45}$$

$$\tag{46}$$

The constraint $[L] = [R] - 5$ pM gives a single differential equation

$$\frac{dR}{dt} = -k_j [R]([R] - 5). \tag{47}$$

This equation has the solution

$$t = \frac{1}{5k_j} \left(\ln \left(\frac{[R]_0}{[R]} \right) - \ln \left(\frac{[R]_0 - 5}{[R] - 5} \right) \right) \tag{48}$$

where t is time in seconds, $[R]_0$ is the concentration of "right" ends at time 0, *i.e.* 10.5 pM, $[R]$ is the concentration of "right" ends after time t seconds and k_j is the /pM/s end-to-end joining rate.

Supplemental Equation 48 and the measured value of k_j predict that after 8 hours, on order 0.018-0.13 pM of templated growth fronts joined, or roughly 0.1%–0.8% of templated growth fronts. Only 2 joining events were observed out of a total of about 300 growth fronts, or approximately 1%±1 of the growth fronts.

10 Prediction of Replication Dynamics over Multiple Generations. To determine whether the replication and mutation (error) rates we measured are sufficient to permit accurate replication to be sustained over many generations, we created a model of a multi-generation replication process that (unlike the simple model of replication in Equations 1 and 2 in the main text) takes into account mutation. Here we use this model to simulate a replication process with the measured rates of crystal growth, mutation and scission and characterize the expected yield of this simulated process.

For our simulation, we chose a many-generation replication process that is essentially an extension of the protocol we followed in the experiments we describe in this work. We model a process that begins with a population of crystal seeds bearing sequence s in a solution of free monomers. The crystals grow (G1), undergo scission (S1) and then are added to a larger quantity of monomers for further growth in G2. After G2, we simulate another scission process S2 and then dilution into a still larger bath of fresh monomers for G3. The simulation continues with cycles of scission, dilution of the fragmented mixture into a larger volume of fresh monomers and further growth.

In principle, the replication process could also be performed without the dilution step: we would simply start out with a very large bath of monomers containing a tiny fraction of seeds and continually fragment the whole solution every so often. This method would also produce replication. We chose to include dilution because it would minimize the amount of spurious nucleation during the replication process: By slowly adding fresh monomers as they are needed via dilution, fewer monomers are free in solution at any given time where they are possible substrates for spurious nucleation. Because with or without dilution monomers are not significantly depleted, the growth and scission rates of seeded crystals are the same with or without dilution. Since dilution is expected to decrease the replication rate of spurious nuclei but not affect the replication rate of seeded crystals, it is expected to optimize the yield of templated crystals[3].

The simulated replication process began with quantity 1 of nuclei. At the start of the simulated replication process, *i.e.* "generation 0", we simulated the nucleation of crystals on the templates with nucleation error rate e_n, producing $I_0 = e_n$ incorrect growth fronts and $C_0 = 1 - e_n$ correct growth fronts. The initial number of spurious nuclei was set at $S_0 = 0$.

After this initial nucleation stage, the simulated model of replication alternated between growth and scission stages: G1, S1, G2, S2, etc. During a growth stage, spurious nuclei were produced at rate s_n per stage, and correct growth fronts mutated into incorrect growth fronts at a growth error rate e_g per stage. After the simulation of G1, the number of correct growth fronts C_{G1} was $C_{G1} = C_0(1 - e_g)$, the number of incorrect growth fronts I_{G1} was $I_{G1} = I_0 + C_0 e_g$ and the number of spurious nuclei are S_{G1} was $S_{G1} = s_n$.

Each simulated scission stage amplified the number of correct growth fronts by the factor a_c, the number of incorrect growth fronts by the factor a_i and the number of spurious growth fronts by the factor a_s, such that

$$C_{Sx} = a_c C_{Gx} \tag{49}$$

$$I_{Sx} = a_i I_{Gx} \tag{50}$$

$$S_{Sx} = a_s S_{Gx}, \tag{51}$$

where Gx refers to the growth stage of generation x and Sx refers to the scission stage of generation x. We ignored mutation during scission (as opposed to mutation during growth) because we assumed the mutation rate during scission

---

[3] "Serial dilution" is a standard laboratory technique for studying biological replication and evolution and for techniques such as directed evolution. However, "spontaneous generation" is not generally a concern in these processes, so serial dilution is therefore not generally used to optimize replication yield. It is also generally performed a bit differently than the serial dilution we describe here: usually every so often a small portion of the population is diluted into a mixture of fresh raw material such that the total volume of the new mixture is the same as the volume of the original mixture. That is, the population decreases and the volume of the reaction stays the same, in contrast to our protocol, where the population size stays the same and the volume of the reaction increases. The standard serial dilution process removes individuals from the population so that less fit species become "extinct" over time. We expect that a similar technique could be used to study Darwinian evolution of DNA tile crystal sequences, but such a process would have different evolutionary dynamics than those described in this section.

was small: in our experiments we found that there was no statistically significant increase in the number of incorrect growth fronts or layers between G1 and S1.

After the first generation, a growth stage operated on the fragmented sample from the previous generation, such that for $x > 0$

$$C_{G(x+1)} = (1 - e_g)C_{Sx} \tag{52}$$

$$I_{G(x+1)} = I_{Sx} + e_g C_{Sx} \tag{53}$$

$$S_{G(x+1)} = S_{Sx} + a_c^{(x-1)} s_n \tag{54}$$

$$\tag{55}$$

The $a_c^{(x-1)}$ factor reflects a a_c-fold dilution at each generation. In our experiments we used a 5-fold dilution. 5 is larger than a_c, so in our experiments, the concentration of correct crystals decreased from G1 to G2, and the increase in the number of spuriously nucleated crystals was larger than what is predicted in these simulations. Subject to the constraint that the concentration of crystals does not increase over time, the minimum amount of spurious nucleation would occur with an a_c-fold dilution at each stage.

In our experiments, we measured the nucleation error rate $e_n = 0.90 \pm 0.03$, the growth error rate $e_g = 0.81 \pm 0.06$, and the spurious nucleation rate $s_n = 0.19 \pm 0.06$ in G1. We measured the rate at which correct ends are amplified $a_c = 3.2 \pm 0.5$ and the rate at which incorrect ends are amplified $a_i = 3.1 \pm 1.0$ between G1 and G2 and the rate of increase in the number of spuriously nucleated growth fronts between G1 and S1 for the experiments where cool air did not cause spurious nucleation after scission $a_s = 1.3 \pm 0.7$. Each of these values was subject to statistical uncertainty. The goal of this simulation was to use what we knew about each of these experimentally measured parameters to infer the most likely dynamics of our system over many generations. Small changes in any of these parameters would cause large changes in the simulation results, so we wanted to understand the range of likely results of an amplification process by sampling each of these parameters subject to what we know as a result of measurement.

The uncertainty in the parameters can be divided into two components: statistical uncertainty in our measurements of the 6 rates and variation in the rates that would be expected between growth and scission stages over multiple generations. We did not consider variation between stages because we assumed this variation would be small compared to our measurement uncertainty, and that this variation would not be amplified exponentially.

For each replication simulation we therefore sampled each of the 6 rates from a probability distribution and measured the yield of replication in simulation. We assumed that the error in each of these parameters was Gaussian, and the standard deviation was half the reported 95% uncertainty in the reported value. For each of 10,000 simulated replication processes, we simulated replication until we reached generation x for which $C_{Gx} > 1000$, *i.e.* the number of correct growth fronts had reached at least 1000 times the initial value. At the end of each simulated replication process, we computed the fraction of templated crystals that were correct, *i.e.* $\frac{C_{Gx}}{C_{Gx}+I_{Gx}}$ and the fraction of all crystals that were spuriously nucleated, $\frac{S_{Gx}}{C_{Gx}+I_{Gx}+S_{Gx}}$. The reported confidence interval was the range of values that include 95% of the simulated replication processes.

11 Determination of Spurious Nucleation Rates. During the replication process, spurious nucleation of new crystals was an important source of mutation. In this section we estimate the rate of spurious nucleation during G1 and G2 in order to estimate the impact of this process on the yield of the replication process.

A spuriously nucleated crystal's sequence should be considered a mutation because it is not copied from an existing sequence but is instead determined by the nucleation process. After nucleation, spuriously nucleated crystals can grow and fragment in the same way as existing crystals. Here we use the observed number of spuriously nucleated crystals per seed to estimate the rate of spurious nucleation during G1 and G2.

Spurious nucleation occurs at a rate dependent only on the concentration of monomers (as opposed to the concentration of existing crystals.) Because only a small proportion of monomers are used up during growth, we assume as in previous work [7] that the monomer concentration, and thus the spurious nucleation rate, is effectively constant during growth. We therefore determine the average rate of spurious nucleation by dividing the measured increase in spurious nuclei by the time during which those nuclei may have appeared.

During G1, seeds were present at 25 pM and there were 0.20±0.07 spuriously nucleated growth fronts (or 0.10±0.035 crystals) observed per seed during imaging. The effective concentration of spurious nuclei after G2 was therefore

2.5±0.8 pM. G1 proceeded on average for 8 hours, for an effective spurious nucleation rate of $9 \pm 3 \times 10^{-5}$ pM/s.

During G2, seeds were present at 5 pM because the S1 mixture, in which seeds were present at 25 pM, was mixed with 4 parts of G2 tiles, which contained no seeds. 2.9±0.7 spuriously nucleated growth fronts, or 1.45±0.35 spuriously nucleated crystals per origami seed, were observed at the end of G2. The concentration of spurious nuclei was therefore 7.5±1.5 pM.

To calculate the increase in spuriously nucleated crystals, we must subtract the amount of spurious nuclei that were already present at the start of G2 from this final concentration. In the portion of the solution from S1 (20% of the total) there were 25 pM of seeds and 1.3±0.4 spuriously nucleated growth fronts observed per seed, so that the concentration of spurious nuclei in the sample added from S1 was approximately 16±5 pM. (Because the time for the scission process varied, we cannot estimate the spurious nucleation rate during scission.) Since this sample is one fifth of the G2 solution, the concentration of spurious nuclei at the start of G2 was 3.2±1 pM. Thus, 4.3±1.5 pM of nuclei were created during the 6-8 hours of G2. Given the average G2 growth time of 7 hours, the effective spurious nucleation rate during G2 was $17 \pm 6 \times 10^{-5}$ pM/s.

12 Predicted Amplification Rates of Spuriously Nucleated Crystal Without Scission. In a process where we successively add fresh monomers with each generation, the number of spuriously nucleated crystals would increase exponentially even if the crystals were not fragmented, simply because in our protocol the reaction volume, and thus the amount of material available to spuriously nucleated new crystals, exponentially increases. To consider how the cycles of growth and scission as we performed them replicated spuriously nucleated crystals, it is thus necessary to first understand how the number of spuriously nucleated crystals would be expected to increase if scission had no effect on the crystals.

Let $L_{\varnothing}[n]$ be the number of spuriously nucleated layers after generation n and $F_{\varnothing}[n]$ be the number of growth fronts after generation n, where at the end of the n^{th} generation there have been n growth stages each h hours long and n scission stages after which the entire stock of crystals is diluted q-fold into a fresh solution of monomers. After following this protocol for n generations, the volume of solution would be q^{n-1} times an initial volume of V_0 liters. In our experiments, there were initially no spuriously nucleated growth fronts or layers, *i.e.* $F_{\varnothing}[0] = 0$ and $L_{\varnothing}[0] = 0$.

We assume that spuriously nucleated growth fronts are produced by spurious nucleation at a rate of x M/hr and that spuriously nucleated layers grow on a spuriously nucleated growth front at a rate of y layers/hr. Because the volume at generation n is $V_o q^{n-1}$, the spurious nucleation rate at generation n is xV_0q^{n-1} moles / hr.

Each generation lasts h hours, so the number of growth fronts $F_{\varnothing}[n]$ (in moles) for $n > 0$ is

$$F_{\varnothing}[n] = xV_0hq^{n-1} + F_{\varnothing}[n-1] \tag{56}$$

$$= xhV_0 \left(\sum_{i=0}^{n-1} q^i \right). \tag{57}$$

where the first term of Supplemental Equation 56 is the rate of spurious nucleation given the volume for the n^{th} generation, and the second term is the spuriously nucleated crystals that arose in previous generations.

In our experiments, the increase in volume was a factor of $q = 5$, so that

$$\frac{F_{\varnothing}[2]}{F_{\varnothing}[1]} = 1 + q = 6. \tag{58}$$

To derive the an analogous expression for the expected increase in the number of spuriously nucleated layers, we first note that the rate of increase in the number of layers at a given time depends on the number of growth fronts in solution. Consider growth during the first generation. At time t, the number of spuriously nucleated crystals is xV_0t. The rate of increase in the number of layers at time t during the first generation is therefore

$$\frac{dL_{\varnothing}}{dt} = xV_0ty, \tag{59}$$

so that at time t, the number of spuriously nucleated layers is $xV_0yt^2/2$. Thus, after the initial h-hour growth stage, the number of spuriously nucleated layers is $L_{\varnothing}[1] = xV_oh^2y/2$.

The number of spuriously nucleated layers after n generations ($n > 0$) is given by

$$L_{\varnothing}[n] = \frac{xh^2y}{2}V_0q^{n-1} + F_{\varnothing}[n-1]yh + L_{\varnothing}[n-1], \tag{60}$$

where the first term counts the new layers that result from spurious nucleation occurring during this generation, the second term counts the new layers resulting from the growth on spurious nuclei that grew in previous generations and the third term counts the layers that grew in previous generations.

Solving Supplemental Equation 60 gives

$$L_{\varnothing}[n] = \frac{xhV_0^2y}{2}\left(\sum_{i=0}^{n-1}(2i+1)q^{n-1-i}\right), \tag{61}$$

so that

$$\frac{L_{\varnothing}[2]}{L_{\varnothing}[1]} = \frac{q+3}{1} = 8. \tag{62}$$

These results demonstrate that the rate at which the number of spuriously nucleated crystals are amplified with or without scission is strongly dependent on the dilution factor. In contrast, the amplification rate of templated crystals does not depend on the dilution factor, so long as the reaction volume is large enough at each generation that crystal growth does not significantly deplete the supply of free monomers, *i.e.* the concentration of crystals does not grow too large over time. Thus, to maximize yield, the dilution factor should be approximately equal to the total crystal replication rate.

13 The Predicted Critical Length of Spurious Nuclei Is Consistent With Layer Melting During Crystal Scission. In this section we adapt our prediction of the critical length for OIIO crystals after scission given in Supplemental Note 4 to predict the critical length of spuriously nucleated crystals under the same conditions. We show that because spuriously nucleated crystals are much thinner than OIIO (*i.e.* templated) crystals, the predicted critical length of spuriously nucleated crystals is much shorter than the predicted critical length for OIIO crystals. For our scission device and pressure drop, this critical length is just a few tile layers long. Such a short critical length is consistent with the idea that spuriously nucleated fragments often melt and so do not survive to provide growth fronts where their sequences can be propagated.

The critical length is dependent on the critical amount of tension needed to fragment the crystal and the radius of the crystal undergoing fragmentation. While OIIO crystals are 14 tiles, or 28 helices wide, most spurious nuclei are 6 tiles or 12 helices wide. As a result, while the critical tension (written as T_{crit}) for OIIO crystals is predicted to be $28f_c$ (f_c is 65 pN, the previously measured [19] force required to melt a single helix at room temperature), in contrast, $T_{crit,sn}$, the critical tension for spurious nuclei, is predicted to be $12f_c$. Adapting the calculation of the radius of the crystal undergoing scission for OIIO crystals in Supplemental Note 4 to the approximately 40 nm width of spuriously nucleated crystals, the approximate "radius" of a spuriously nucleated crystal, R_{sn}, is 14 nm.

Given a crystal's critical tension and radius, Supplemental Equation 15 relates the geometry of and pressure drop across the scission device we used in our experiments to the expected critical length for a crystal after scission. Since both spurious nuclei and templated crystals undergo scission in the same device under the same conditions, we used the same scission device geometry and physical parameters as in Supplemental Note 4, and set $T_{crit} = T_{crit,sn}$ and $R = R_{sn}$. Supplemental Equation 15 predicts for crystals passing through the center of the constriction, *i.e.* those that undergo maximal elongational flow, the critical length would be approximately 30 nm, or about 2 tile layers. A 2-tile-layer assembly would tend to melt at low supersaturation [6].

Crystals that pass through the constriction away from its center will experience less tension and thus will be fragmented into larger segments than those that pass through the constriction's center. Supplemental Equation 41 is used to predict the critical length of crystals that pass through the constriction away from its center. Applying this Equation to spuriously nucleated crystals, we would predict that 50% of spurious nuclei entering the constriction have an expected critical length of 3 tiles long or less. These calculations thus suggest that a significant proportion of spurious nuclei may be fragmented into such small pieces that they melt and are not able to serve as new growth fronts.

It is important to note however that quantitative predictions of critical length in this regime should be viewed as estimates, since Supplemental Equations 15 and 41 are applicable for the case where $L_{crit} \gg r$. Here $L_{crit} \approx r$.

14 Excluding Illegible Crystals From Analysis Does Not Qualitatively Alter Conclusions about Ribbon Replication. In counting crystal growth fronts and layers, we excluded growth fronts or layers (length of crystal) where the sequence was not legible in our AFM images. One might be concerned that this exclusion introduced bias into our samples because correct crystals might be easier to deem legible, and that this bias might change our conclusions about the replication rate or the yields of replication. However, the number of crystals excluded because they were illegible was small. To illustrate this, here we examine the impact of this exclusion by evaluating the performance of replication when we assume that excluded growth fronts and layers are all incorrect. We term these values "worst-case correctness" rates.

Supplemental Figure 24 shows the number of growth fronts and layers tabulated by including the illegible layers and growth fronts and assuming that they are all incorrect. A comparison of these values with the measurements made in which illegible crystals were excluded are given in Supplemental Tables 1 and 2. (Because the number of seeds in the sample change when we include illegible crystals, the relative numbers of spurious nucleated layers and growth fronts per seed also change.)

When the analysis described in Supplemental Note 6 is performed using the worst-case correctness values instead of the measured values, the best fit for the scission rate $f_{\text{wc}} = 0.0017 \pm 0.0003$ per nanometer per generation and the best fit for the growth rate is $g_{\text{wc}} = 530 \pm 60$ nanometers per generation, for an effective replication factor of $r_{\text{wc}} = 3.5$ (95% confidence interval 3.1-4.0). These values are almost identical to the best fit for the measured values when illegible crystals are excluded, $f = 0.0016 \pm 0.0003$ per nanometer per generation, $g = 540 \pm 70$ nanometers per generation, $r = 3.5 \pm 0.4$.

Simulation of 1000-fold replication of the templated sequence using the worst-case correctness values performed as described in Supplemental Note 10) showed a yield of 8.4% correct crystals with a 95% confidence interval between 0.8% and 35% among all templated crystals. The mean fraction of spuriously nucleated crystals after 1000-fold replication of correct growth fronts was 1% with a 95% confidence interval between 0.1% and 5%. As we report in the main text, when illegible crystals are excluded from our analysis, the predicted yield of correct crystals among templated crystals was 47% with a 95% confidence interval between 12% and 82% and the mean fraction of crystals that were spuriously nucleated was predicted to be 7% with a 95% confidence interval between 2% and 21%. The predicted yield of correct crystals for 1000-fold replication given the worst-case correct values is therefore somewhat worse than the predicted yields when illegible crystals are excluded from our analysis, but even in this "worst case," the simulations still indicate appreciable yields of correct crystals and a low fraction of spuriously nucleated crystals.

References

[1] Rothemund PWK, Papadakis N, Winfree E (2004) Algorithmic self-assembly of DNA Sierpinski triangles. *PLoS Biology* 2:424–436.

[2] Barish RD, Rothemund PWK, Winfree E (2005) Two computational primitives for algorithmic self-assembly: Copying and counting. *Nano Lett.* 5:2586–2592.

[3] Chen HL, Schulman R, Goel A, Winfree E (2007) Reducing facet nucleation during algorithmic self-assembly. *Nano Lett.* 7:2912–2919.

[4] Barish RD, Schulman R, Rothemund PWK, Winfree E (2009) An information-bearing seed for nucleating algorithmic self-assembly. *P. Natl. Acad. Sci.* 106:6054–6059.

[5] Winfree E, Bekbolatov R (2004) Proofreading tile sets: Error-correction for algorithmic self-assembly. *DNA Computing 9* eds Chen J, Reif J (Springer-Verlag, Berlin Heidelberg), Vol. LNCS 2943, pp 126–144.

[6] Schulman R, Winfree E (2009) Programmable control of nucleation for algorithmic self-assembly. *SIAM J. Comput.* 39:1581–1616.

[7] Schulman R, Winfree E (2007) Synthesis of crystals with a programmable kinetic barrier to nucleation. *P. Natl. Acad. Sci.* 104:15236–15241.

[8] Wetmur JG, Fresco J (1991) DNA probes: Applications of the principles of nucleic acid hybridization. *Crit. Rev. Biochem. Mol. Bio.* 26:227–259.

[9] Winfree E (1998) Simulations of computing by self-assembly., (Caltech), Technical Report CS-TR:1998.22.

[10] Hariadi R (2011) Ph.D. thesis (California Institute of Technology, Pasadena, California).

[11] SantaLucia, Jr. J, Allawi HT, Seneviratne PA (1996) Improved nearest-neighbor parameters for predicting DNA duplex stability. *Biochemistry* 35:3555–3562.

[12] Rothemund PWK (2006) Folding DNA to create nanoscale shapes and patterns. *Nature* 440:297–302.

[13] Bloomfield VA, Crothers DM, Tinoco I (2000) *Nucleic acids: structures, properties, and functions* (University Science Books, Mill Valley, Cal.).

[14] Pastré D, et al. (2006) Anionic polyelectrolyte adsorption on mica mediated by multivalent cations: a solution to DNA imaging by atomic force microscopy under high ionic strengths. *Langmuir* 22:6651–6660.

[15] O'Neill P, Rothemund PWK, Kumar A, Fygenson DK (2006) Sturdier DNA nanotubes via ligation. *Nano Lett.* 6:1379–1383.

[16] Yurke B, Turberfield AJ, Mills, Jr. AP, Simmel FC, Nuemann JL (2000) A DNA-fuelled molecular machine made of DNA. *Nature* 406:605–608.

[17] Winfree E, Liu F, Wenzler LA, Seeman NC (1998) Design and self-assembly of two-dimensional DNA crystals. *Nature* 394:539–544.

[18] Hariadi R, Yurke B (2010) Elongational-flow-induced scission of DNA nanotubes in laminar flow. *Phys. Rev. E* 82:046307.

[19] Smith SB, Cui Y, Bustamante C (1996) Overstretching B-DNA: The elastic response of individual double-stranded and single-stranded molecules. *Science* 271:795–799.

[20] Truskey GA, Yuan F, Katz DF (2004) *Transport Phenomena in Biological Systems* (Prentice Hall, New Jersey).

[21] Siebert WM (1986) *Circuits, Signals and Systems* (MIT Press).

[22] Ekani-Nkodo A, Kumar A, Fygenson DK (2004) Joining and scission in the self-assembly of nanotubes from DNA tiles. *Phys. Rev. Lett.* 93:268301.