Synthesis of crystals with a programmable kinetic barrier to nucleation

Rebecca Schulman and Erik Winfree
Computer Science and Computation & Neural Systems, California Institute of Technology, Pasadena, CA, USA

SI Appendix

S1 Sequences and Synthesis

S1.1 Design methodology

While several of the double-stranded (core) sequences for tiles were identical to those used in earlier experiments [1], we designed all the sticky end sequences and the sequences for tiles Z9-Z12 (see Section S1.2) needed for ZZ5 and ZZ6 for the experiments described here. Design of sequences occurred in the following order:

1. The sticky ends for ZZ3 and ZZ4 ribbons.
2. The cores for the 4 additional tiles used in ZZ5 and ZZ6.
3. The sticky ends for the additional tiles.

At each step, we designed sequences that were predicted to behave well in combination with existing sequences.

The goal of sticky end design was to create sticky ends with similar energies and low crosstalk (energy of binding between non-complementary sticky ends). Sticky end energies were designed to be approximately the same to facilitate uniform treatment in modeling and so that the design for the ribbons was as simple as possible and could be generalized and used elsewhere.

The design process relied on the prevailing model of the energetics of DNA strand hybridization, the nearest neighbor model [2], for predicting the energy of potential sticky ends. In the nearest neighbor model, the energy of hybridization is a linear sum of the energies of pairs of adjacent bases within a double-stranded helix. For example, in the sequence

```
>attgc>
 |||||
<taacg<
```

the pairs of adjacent bases are `at`, `tt`, `tg`, and `gc`. The ΔG° of this conformation is the sum of the ΔG° of each of the 4 pairs. At 37 °C in 1 M sodium buffer, these energies have been measured as -1.10, -0.93, -2.11 and -3.42 kcal/mol, respectively [3]. Following this model, we assumed that the sticky

end hybridization energy was equal to the energy of the new nearest neighbor base pairs formed when two sticky ends hybridized, including pairs with a nick between the bases (a total of six new nearest neighbor pairs.) Although our experiments were performed in buffer with low sodium concentration and 12.5 mM Mg^{++}, we used the default nearest neighbor model parameters for DNA in 1 M sodium buffer because known parameter adjustments for salt conditions are not sequence dependent [4, 5]. We found pairs of sticky end sequences with hybridization energies within a pre-specified range (ΔG° between -8.23 and -8.89 kcal/mol at 37 $^\circ$C) such that the predicted binding energy between all non-matching sticky ends was less than a certain value ($\Delta G^\circ < 3.2$ kcal/mol at 37 $^\circ$C). Sticky ends were then arranged on the tiles by hand using sequence symmetry minimization [6, 7] between the sticky ends and the cores of the tiles they were placed on. The sequence symmetry minimization process selects sequences with the shortest possible spurious complementary regions between strands.

To design the new cores, we started with randomly chosen sequences that both obeyed the required Watson-Crick complementarity rules imposed by the geometry of the tile and had sequences `cacc` or `gtgg` around the crossover point. The terminal base pair of helix ends was constrained to be either cytosine (`c`) or guanine (`g`) to reduce fraying (`g-c` pairs are stronger than `a-t` base pairs [3]). To find the desired sequences, we randomly permuted a few bases of the initial sequences at a time, keeping the changes that increased the "goodness" of the sequences. "Good" sequences, according to our objective function, did not contain any subsequences where stacking is known to be somewhat irregular [8] or more than 4 `a-t` or `g-c` base pairs in a row. The objective function penalized spurious strand matches in a single strand, within a single tile, and within the collection of strands progressively more weakly. Penalties increased exponentially in proportion to the number of base pairs in the base pairing. Pre-existing sequences and the sequences around crossover points were not permuted in the optimization process. Permuting a few bases at a time to improve the quality of the sequences produced a local minimum of the objective function. To ensure that there wasn't a much better sequence in another region of sequence space, the process was repeated several times and the best design was chosen.

S1.2 Zig-zag ribbon tile sequences

For each tile, the colored diagram representing the logic of the tile (from the main paper) is shown in the same orientation as the tile diagram sequences. In the tile sequence diagrams, a "$>$c" denotes a c at the 5' end of a sequence and "g$>$" denotes a g at the 3' end of a sequence.

Tile Z1:

```
     2  >cagagtgg-acgaaagc\/agtgccgt-ccgatgtc<  3
1 <aagga-gtctcacc tgctttcg/\tcacggca ggctacag-aagag>  4
                | |       Z1       | |
1 >cttgt-caaacgca ggaacctg\/tatgaacc tgctcaac-tcgta<  4
     2  <gtttgcgt-ccttggac/\atacttgg-acgagttg>  3
```

Strands:

```
1  cttgtcaaacgcaccactctgaggaa
2  cagagtggacgaaagctcacggcaccaagtatcaggttcctgcgtttg
3  ctgtagcctgccgtgagctttcgtggaacctgatacttggacgagttg
4  atgctcaactcgtggctacagaagag
```

Tile Z1W:

```
     2  >cagagtgg-acgaaagc\/agtgccgt-ccgatgtc<  3
1 <tggtt-gtctcacc tgctttcg/\tcacggca ggctacag-aagag>  4
                | |       Z1W      | |
1 >cttgt-caaacgca ggaacctg\/tatgaacc tgctcaac-tcgta<  4
     2  <gtttgcgt-ccttggac/\atacttgg-acgagttg>  3
```

Strands:

```
1  cttgtcaaacgcaccactctgttggt
```

Strands 2, 3 and 4 have the same sequences as strands 2, 3, and 4 of tile Z1.

Tile Z2:

```
      2  >gatgatgt-ccttgtaa\/tgaagcgg-acaacgag<  3
1  <ttagc-ctactaca ggaacatt/\acttcgcc tgttgctc-gctaa>  4
                | |       Z2       | |
1  >aagtc-gaacgacc tgattgcg\/taatctca ggcattcg-cacta<  4
      2  <cttgctgg-actaacgc/\attagagt-ccgtaagc>  3
```

Strands:

```
1  aagtcgaacgaccacatcatccgatt
2  gatgatgtccttgtaaacttcgccactctaatcgcaatcaggtcgttc
3  gagcaacaggcgaagtttacaaggtgattgcgattagagtccgtaagc
4  atcacgcttacggtgttgctcgctaa
```

Tile Z3:

```
      2  <gtcggtca-ggctcgtc\/acgacacc-tgagacgg>  3
1  >tggaa-cagccagt ccgagcag/\tgctgtgg actctgcc-gaaca<  4
                | |       Z3       | |
1  <ttctc-gaggatgg acgcttag\/tctgtagt ccgcattg-aatcg>  4
      2  >ctcctacc-tgcgaatc/\agacatca-ggcgtaac<  3
```

Strands:

```
1  tggaacagccagtggtaggagctctt
2  ctcctacctgcgaatctctgtagtggtgtcgtctgctcggactggctg
3  caatgcggactacagagattcgcaccgagcagacgacacctgagacgg
4  acaagccgtctcaccgcattgaatcg
```

Tile Z3W:

```
      2  <gtcggtca-ggctcgtc\/acgacacc-tgagacgg>  3
1  >accaa-cagccagt ccgagcag/\tgctgtgg actctgcc-gaaca<  4
                | |       Z3W      | |
1  <ttctc-gaggatgg acgcttag\/tctgtagt ccgcattg-aatcg>  4
      2  >ctcctacc-tgcgaatc/\agacatca-ggcgtaac<  3
```

Strands:

```
1  accaacagccagtggtaggagctctt
```

Strands 2, 3 and 4 have the same sequences as strands 2, 3, and 4 of tile Z3.

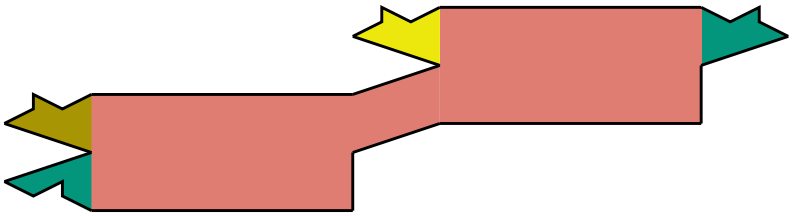
Tile Z4:

```
      2  <ccgttagg-acattgca\/cggcttgt-ccgttcgc>  3
1  >agcat-ggcaatcc tgtaacgt/\gccgaaca ggcaagcg-ttcag<  4
                | |       Z4       | |
1  <cgatt-cgccaaca ggttgaat\/ccagatcc tgtagagc-gacat>  4
      2  >gcggttgt-ccaactta/\ggtctagg-acatctcg<  3
```

Strands:

```
1  agcatggcaatccacaaccgcttagc
2  gcggttgtccaacttaccagatccacaagccgacgttacaggattgcc
3  gctctacaggatctggtaagttggtgtaacgtcggcttgtccgttcgc
4  gacttgcgaacggtgtagagcgacat
```

Tile Z42:



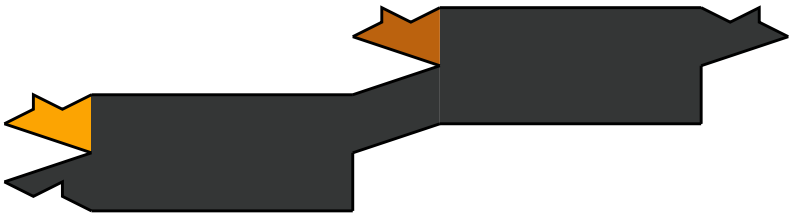
```
                                          5  >gatgatgt-ccttgtaa\/tgaagcgg-acaacgag<  3
                                       6  <ttagc-ctactaca ggaacatt/\acttcgcc tgttgctc-gctaa>  4
                                                  | |       Z42      | |
     2  <ccgttagg-acattgca\/cggcttgt-ccgttcgc-aagtc-gaacgacc tgattgcg\/taatctca ggcattcg----cacta-cc-tt\  4
1  >agcat-ggcaatcc tgtaacgt/\gccgaaca ggcaagcg-ttcag-cttgctgg-actaacgc/\attagagt-ccgtaagc>3 >gtgat-gg-tt/
               | |       Z42      | |
1  <cgatt-cgccaaca ggttgaat\/ccagatcc tgtagagc-gacat>  5
     2  >gcggttgt-ccaactta/\ggtctagg-acatctcg-ctgta<  6
```

Strands:

```
1  agcatggcaatccacaaccgcttagc
2  gcggttgtccaacttaccagatccacaagccgacgttacaggattgcc
3  gagcaacaggcgaagtttacaaggtgattgcgattagagtccgtaagc
4  gtgatggttttccatcacgcttacggtgttgctcgctaa
5  gatgatgtccttgtaaacttcgccactctaatcgcaatcaggtcgttcgacttgcgaacggtgtagagcgacat
6  atgtcgctctacaggatctggtaagttggtgtaacgtcggcttgtccgttcgcaagtcgaacgaccacatcatccgatt
```

Note : Strands 1 and 2 have the same sequences as strands 1 and 2 of tile Z4. Strand 3 has the same sequence as strand 3 of Z2.

Tile Z56:

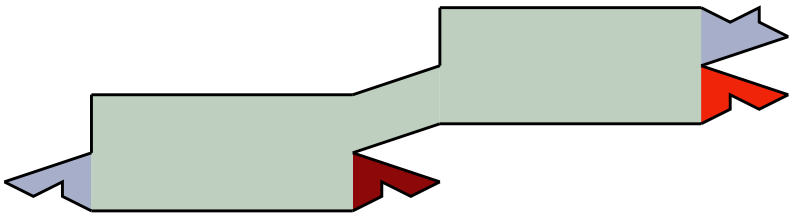


```
                                          5  <gcaagcgt-ccacttgg\/gcagtagg-acgcctcg>  3
                                       6  >gtgat-cgttcgca ggtgaacc/\cgtcatcc tgcggagc-gcaat<  4
                                                  | |       Z56      | |
      2  >gtttgagg-acgctatg\/ttgtaggt-ccatgagc-acgaa-cgaaagcc tgagctag\/tccagaca ggtcatcg----aaggc-cc-tt\  4
1  <ctgta-caaactcc tgcgatac/\aacatcca ggtactcg-tgctt-gctttcgg-actcgatc/\aggtctgt-ccagtagc<3 <ttccg-gg-tt/
               | |       Z56      | |
1  >cgtta-gctcggca ggtgtctc\/acgaatcc tggttacg-aaggc<  5
      2  <cgagccgt-ccacagag/\tgcttagg-accaatgc-ttccg>  6
```

Strands:

```
1  cgttagctcggcacctcaaacatgtc
2  gtttgaggacgctatgaacatccacctaagcagagacacctgccgagc
3  cgatgacctgtctggagatcgagtggtgaaccgcagtaggacgcctcg
4  taacgcgaggcgtggtcatcgaaggccctttgggccttt
5  cggaagcattggtggtactcgtgcttgctttcggactcgatctccagacacctactgcggttcacctgcgaacg
6  gtgatcgttcgcaccgaaagcaagcacgagtacctggatgttcatagcgtggtgtctctgcttaggaccaatgcttccg
```

Tile Z78:



```
                                                    5  <gaagcagg-acaagcga\/ctcagtgt-ccgattgg>  3
                                                6  >catac-cttcgtcc tgttcgct/\gagtcaca ggctaacc-tccaa<  4
                                                      | |       Z78      | |
1   /tt-gg-tcttg  2>cattctgg-tgaccata\/tctatcct-ccgatgac-gacag-ccgtgcca ggtagcgg\/tacactcc tgcttctg-ttcct>  4
    \tt-cc-agaac----gtaagacc actggtat/\agatagga ggctactg-ctgtc-ggcacggt-ccatcgcc/\atgtgagg-acgaagac<  3
                         | |       Z78      | |
          1  >aggtt-ctaccgca gcctattc\/tgacgtgg tgcctagc-accttt<  5
                2  <gatggcgt-cggataag/\actgcacc-acggatcg>  6
```

Strands:

```
1  aggttctaccgcaccagaatgcaagacctttttggtcttg
2  cattctggtgaccataagataggaggtgcagtgaataggctgcggtag
3  cagaagcaggagtgtaggcgatggtgttcgctctcagtgtccgattgg
4  aacctccaatcggtgcttctgttcct
5  ttccacgatccgtggctactgctgtcggcacggtccatcgcctacactccacactgagagcgaacaggacgaag
6  cataccttcgtccaccgtgccgacagcagtagcctcctatcttatggtcagcctattcactgcaccacggatcg
```

Tile Z9:

```
          3  >gagcgagt-ccatatca\/tacttagg-acgactgg<  2
    4  <aagga-ctcgctca ggtatagt/\atgaatcc tgctgacc-agaac>  1
                     | |       Z9       | |
    4  >tgagt-gttcatcc tgcgacgt\/tcaacgca ggctattg-aatgg<  1
          3  <caagtagg-acgctgca/\agttgcgt-ccgataac>  2
```

Strands:

```
1  ggtaagttatcggtgctgaccagaac
2  ggtcagcaggattcattgatatggtgcgacgtagttgcgtccgataac
3  gagcgagtccatatcaatgaatccacgcaactacgtcgcaggatgaac
4  tgagtgttcatccactcgctcaggaa
```

Tile Z10:

```
      3  >gaatgagg-actgagta\/tccgctgt-cccaaatc<  2
4  <attgg-cttactcc tgactcat/\aggcgaca gggtttag-tgagt>  1
                | |      Z10       | |
4  >aatgg-ctagaaca ggaaacgc\/tagatgcc tgaagacg-tggtt<  1
      3  <gatcttgt-cctttgcg/\atctacgg-acttctgc>  2
```

Strands:

```
1  ttggtgcagaagtgggtttagtgagt
2  ctaaaccctgtcgccttactcagtggaaacgcatctacggacttctgc
3  gaatgaggactgagtaaggcgacaccgtagatgcgtttcctgttctag
4  aatggctagaacacctcattcggtta
```

Tile Z11:

```
      3  <gagaatca-ggctctca\/atctagcc-tggttcgg>  2
4  >tggaa-ctcttagt ccgagagt/\tagatcgg accaagcc-actca<  1
                | |      Z11       | |
4  <tcttg-gctgaagg actattgt\/tttgtagt ccctttag-taacc>  1
      3  >cgacttcc-tgataaca/\aaacatca-gggaaatc<  2
```

Strands:

```
1  actcaccgaaccacccttagtaacc
2  ctaaagggactacaaatgttatcaccgagagtatctagcctggttcgg
3  cgacttcctgataacatttgtagtggctagatactctcggactaagag
4  tggaactcttagtggaagtcggttct
```

Tile Z12:

```
      3  <caaagcgg-acaacgta\/ttccatgt-ccttagac>  2
4  >ttacc-gtttcgcc tgttgcat/\aaggtaca ggaatctg-ttacc<  1
                | |      Z12       | |
4  <actca-gctctaca ggcattag\/gttatgcc tgtatcgc-accaa>  1
      3  >cgagatgt-ccgtaatc/\caatacgg-acatagcg<  2
```

Strands:

```
1  ccattgtctaaggtgtatcgcaccaa
2  gcgatacaggcataacgattacggtgttgcatttccatgtccttagac
3  cgagatgtccgtaatcgttatgccacatggaaatgcaacaggcgaaac
4  ttaccgtttcgccacatctcgactca
```

S1.3 Tiles that comprise each ribbon

3 tile wide ribbon : Z78, Z1, Z3, Z42.

4 tile wide ribbon : Z78, Z1, Z3, Z2, Z4, Z56.

5 tile wide ribbon : Z78, Z9, Z10, Z11, Z12, Z1W, Z3W, Z42.

6 tile wide ribbon : Z78, Z9, Z10, Z11, Z12, Z1W, Z3W, Z2, Z4, Z56.

S1.4 Sequences of tiles without sticky ends

Strands not listed are the same those in the original tiles. (noSE stands for “no sticky ends”.)

Tile Z1-noSE:

```
1  caaacgcaccactctg
4  caactcgtggctacag
```

Tile Z2-noSE:

```
1  gaacgaccacatcatc
4  gcttacggtgttgctc
```

Tile Z3-noSE:

```
1  cagccagtggtaggag
4  ccgtctcaccgcattg
```

Tile Z4-noSE:

```
1  ggcaatccacaaccgc
4  gcgaacggtgtagagc
```

Tile Z42-noSE:

```
1  ggcaatccacaaccgc
4  gcttacggtgttgctc
```

Tile Z56-noSE:

```
1  gctcggcacctcaaac
3  gccttcgatgacctgtctggagatcgagtggtgaaccgcagtaggacgcctcg
4  cgaggcgtggtcatcgaaggc
```

Tile Z78-noSE:

```
1  ctaccgcaccagaatg
4  ccaatcggtgcttctg
```

Tile Z9-noSE:

```
1  gttatcggtgctgacc
4  gttcatccactcgctc
```

Tile Z10-noSE:

```
1  gcagaagtgggtttag
4  ctagaacacctcattc
```

Tile Z11-noSE:

```
1  ccgaaccacccttag
4  ctcttagtggaagtcg
```

Tile Z12-noSE:

```
1  gtctaaggtgtatcgc
4  gtttcgccacatctcg
```

Z1W-noSE is the same tile as Z1-noSE, and Z3W-noSE is the same tile as Z3-noSE.

S1.5 Crystal seed sequences

The sequences of the crystal seed are shown in tile-like pieces, but it is not expected that assembly of the seed proceeds by forming these pieces first.

```
CS-1
1   cataccttcgtccaccgtgccgacag
2   ttccacgatccgtggctactgctgtcggcacggtccatcgcctacactccacactgagagcgaacaggacgaag
3   cagaagcaggagtgtaggcgatggtgttcgctctcagtgtccgattgg
4   aacctccaatcggtgcttctgttcct

CS-2
1   cattctggtgaccataagataggaggtgcagtgaataggctgcggtagttttctaccgcaccagaatgcaagacctttggtcttg
2   cagtagcctcctatcttatggtcagcctattcactgcaccacggatcgtggaacagccagtggtaggagctctt

CS-3
1   atgctcaactcgtggctacagaagagctcctacctgcgaatctctgtagtggtgtcgtctgctcggactggctg
2   caatgcggactacagagattcgcaccgagcagacgacacctgagacgg
3   acaagccgtctcaccgcattgaatcg

CS-4
1   cagagtggacgaaagctcacggcaccaagtatcaggttcctgcgtttgttttcaaacgcaccactctgaggaacctttggttcct
2   ctgtagcctgccgtgagctttcgtggaacctgatacttggacgagttgagcatggcaatccacaaccgcttagc

CS-5
1   atcacgcttacggtgttgctcgctaagcggttgtccaacttaccagatccacaagccgacgttacaggattgcc
2   gctctacaggatctggtaagttggtgtaacgtcggcttgtccgttcgc
3   gacttgcgaacggtgtagagcgacat

CS-6
1   gatgatgtccttgtaaacttcgccactctaatcgcaatcaggtcgttcttttgaacgaccacatcatccgattccttttggaatcg
2   gagcaacaggcgaagtttacaaggtgattgcgattagagtccgtaagcgtgatcgttcgcaccgaaagcaagca

CS-7
1   cggaagcattggtggtactcgtgcttgctttcggactcgatctccagacacctactgcggttcacctgcgaacg
2   cgatgacctgtctggagatcgagtggtgaaccgcagtaggacgcctcg
3   taacgcgaggcgtggtcatcgaaggccctttgggccttt

CS-8
1   gtttgaggacgctatgaacatccacctaagcagagacacctgccgagcttttgctcggcacctcaaacatgtcggttttccgacat
2   cgagtacctggatgttcatagcgtggtgtctctgcttaggaccaatgcttccg
```

S1.6 DNA synthesis and quantitation

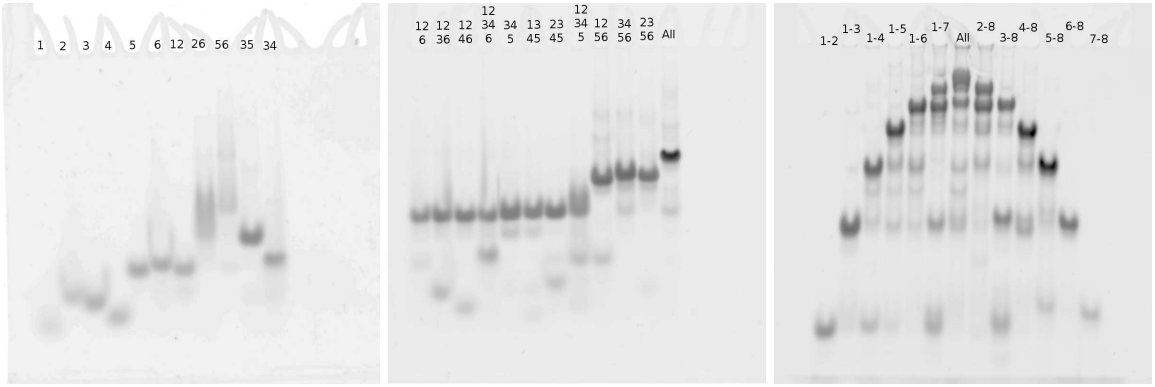
All DNA was ordered in lyophilized form from Integrated DNA Technology (IDT, Coralville, IA). IDT purified the DNA using polyacrylamide gel electrophoresis (PAGE) to >90% purity. We added water to the DNA and quantitated the resulting solution by measuring the absorbance of 260 nm UV light (A_{260}) on an Eppendorf Biophotometer, typically using a 50-fold dilution. The concentration of each solution was determined using the formula $A_{260} = \log\left(\frac{I_{transmitted}}{I_{received}}\right) = C\ e_{260}\ l$ where C is the concentration, e_{260} is the molar absorbance coefficient of the strand at 260 nm (we used a nearest neighbor model for predicting extinction coefficients [3]) and l is the path length of the cuvette the absorbance was measured in.

S1.7 Non-denaturing gels of tiles and crystal seed

Non-denaturing gel electrophoresis verified that all the tiles used in the paper formed as designed. Gels were 6% polyacrylamide, polymerized by the addition of 1% by volume of 10% ammonium persulfate and cross-linked by the addition of 0.075% by volume tetramethylethylenediamine (TEMED). We ran the gels in Tris Acetate EDTA buffer with 12.5 mM added MgCl for 3-4 hours at 100V at either 25 °C or 30 °C and stained them with Sybr Gold (Invitrogen Corporation) for 35 minutes.

In this work, we used two DNA structures that had not been previously synthesized: the double tile and the crystal seed structure. Figures S1(a) and S1(b) show that the strands for one of the double tiles we used, Z56, and the crystal seed form assemble into a single structure with high yields. Similar gels for the other double tiles (data not shown) confirmed that their component strands also assembled into single structures.

Because the strands of the seed structure are highly interleaved, we tested the theory that folding the structure in steps rather than combining all of the strands at once would produce the seeds at a higher yield. We first annealed the strands that comprise each tile-like piece as shown in Section S1.5 from 90 °C to 20 °C at 1 °C / minute to form 8 pieces. Starting with piece 1 (or piece 8), we then added the adjacent piece (either 2 or 7) and let the samples sit at 40 °C for 20 minutes. For each sample we added each subsequent piece in the same manner until the samples each contained all 8 pieces. The yield of structures assembled in this manner was no higher than what is shown below.



(a) Z56 double tile

(b) Crystal seed structure

Figure S1: **Non-denaturing gels of new DNA structures. (a)** Two gels confirm that Z56, a double tile, forms as designed. Numbered labels indicate strand numbers. The rightmost lane in the second gel shows all strands together, and forms a clean single band. While strand stoichiometry imbalances sometimes produced extra bands, all strands interact according to the rules imposed by their design. Each strand was at 1 μM in the standard annealing buffer. **(b)** Gel showing formation of the crystal seed structure. Each strand was at 1 μM in the standard annealing buffer. Number labels indicate the numbers of tile-like pieces. The corresponding strands are shown in Section S1.5. Successively adding the strands for each piece produced successively heavier structures. The complete structure forms with $>$50% efficiency with most of the remaining structures being seeds missing one or two pieces. Incomplete structures can also be expected to lower the barrier to ribbon nucleation, so we concluded it was unnecessary to purify the full-formed seed structure.

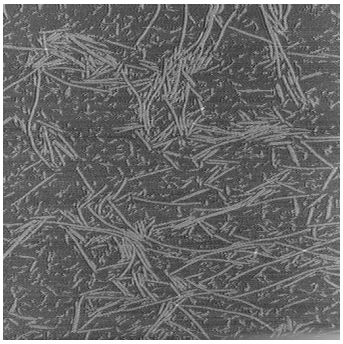
S2 Annealing is required to form long ribbons

In order to determine whether long ribbons can form without annealing, we compared ribbons formed from tiles at room temperature with those formed by two annealing methods. We first combined the strands for each of the 6 tiles in ZZ4 in separate tubes. We then annealed the six tubes from 90 °C to 20 °C at 1 °C / minute.

After annealing, the tiles were mixed together in three samples such that each sample contained 100 nM of each tile. The first sample was annealed from 90 °C to 40 °C at 1 °C / minute and from 40 °C to 20 °C at 1 °C / hour, so that the tiles were re-formed from strands during re-annealing but furthermore ribbons could also form. The second sample was annealed from 40 °C to 20 °C at 1 °C / hour, which should not have melted tiles but should have melted any ribbons that may have formed during mixing, thus resulting in fresh ribbon formation from well-formed tiles during re-annealing. The third sample was left at room temperature during the same period. The results (Figure S2) show that while both methods of annealing produce ribbons, tiles do not assemble into long ribbons at room temperature.

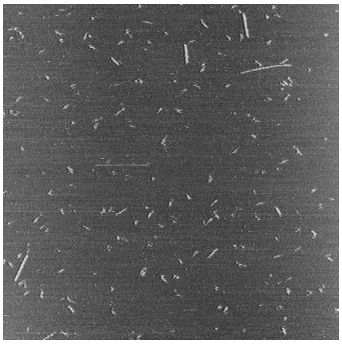
Both annealed ribbon samples contained ribbons. In the experiments described in this paper, ribbons were formed by mixing all the strands for all tiles and annealing them together. Here, we compared the quality of ribbons that results when tiles are annealed separately first with those annealed in the manor we study in this paper. The results are shown in Figures S2(a) and (b).

While both annealed ribbon samples contains long ribbons, qualitative observation suggests that ribbons in the tube where tiles were annealed separately (and not then re-melted) are longer than those in the tube where all tiles are annealed together. This judgement must be qualified by technical issues in sample preparation. The samples shown here are pipetted into a droplet of buffer solution that covers the mica substrate. Transport to the surface where they are adsorbed may occur through diffusive and/or convective processes. Thus it is possible that more ribbons land nearer to the point where samples are pipetted than land further away, and that longer ribbons diffuse a shorter distance than shorter ones. Small ribbons may not stick to the surface at all. Additionally, ribbons often tear during the deposition process, which changes their lengths and leaves fragments on the surface. Therefore, we could not rigorously study which kind of annealing method produces longer ribbons using this technique.



(a) Ribbons annealed from strands

(b) Ribbons annealed from pre-formed tiles



(c) Ribbons prepared at room temperature

Figure S2: **Ribbon formation in annealed samples and in samples prepared at room temperature.** Atomic force microscopy images (5 μm scan size) of a 100 nM solution of ZZ4 ribbon tiles prepared in three different ways.

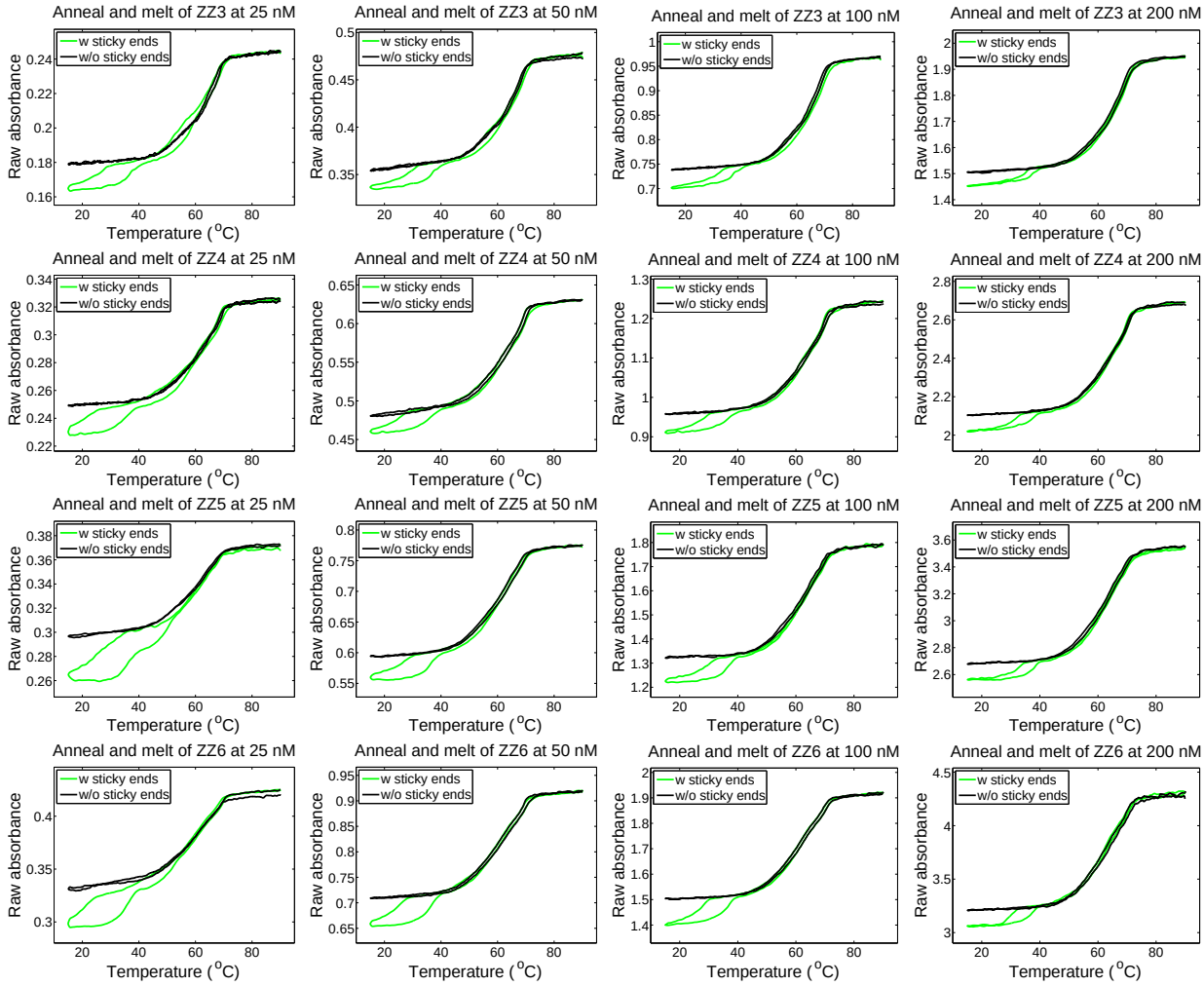


Figure S3: **Temperature ramp experiments for ZZ3-ZZ6 at 25, 50, 100 and 200 nM.** All absorbance data was taken every 0.1 °C smoothed with 10 neighbors to reduce instrument noise. The path length used for the 200 nM and 6 tile wide 100 nM melts was 1/2.5 the other path lengths, so the measured absorbances are multiplied by 2.5. Data for tiles without sticky ends was rescaled as in Figure 3(a). Tiles are designed so that about 5 times as many base pairs form during tile formation as during ribbon formation (see Figure 1(a)). The size of the two transitions is roughly consistent with this ratio, although the ratio increases with decreasing concentration. We're not sure what might cause this increase.

S3 Measurement of ribbon joining rates

In this section we describe how we measured ribbon joining rates at room temperature. Our results suggest a joining rate of about $k_j = 35,000$ /M/s, which is high enough that we expect that joining affects the kinetics of ribbon growth as observed in the temperature-ramp and temperature-hold experiments.

To measure joining, we decorated a subset of the tiles in the 4 tile wide ribbon (ZZ4) with hairpins perpendicular to the plane of the tile (Figure S4). Tiles containing hairpins provided more contrast in AFM imaging than tiles without hairpins [9]. Two types of decorated ribbons were prepared: "dotted" ZZ4 ribbons utilizing T1-hp and T3-hp (Figure S5(a)), and "striped" ZZ4 ribbons utilizing T3-hp and T4-hp (Figure S5(b)). These two ribbon types can be readily distinguished in AFM images.

Our joining experiment proceeded as follows. First, dotted and striped ribbons were separately annealed at 100 nM (each strand) from 90 °C to 20 °C over 8 hours. AFM imaging of these samples showed that the ribbon lengths were similar to those shown in Figure 2(b) (data not shown). The two samples were then mixed and left at room temperature. After 4, 24, and 48 hours, the mixed ribbons were examined by taking 30 atomic force microscopy images at random non-intersecting locations on the the mica surface. Joining events between ribbons with different patterns were visible as a change in the hairpin pattern along the ribbon. For each sample, we tabulated the number of rows of ribbons seen in the images and the number of visible joining events. The results are shown in Table S1.

Time after mixing	Fraction observed joined rows	95% confidence interval
4 hours	0.0039	[.0030-.0048, N=3341 rows total]
24 hours	0.0067	[.0056-.0078, N=3852 rows total]
48 hours	0.0078	[.0063-.0094, N=3210 rows total]

Table S1: **Ribbon joining rates at room temperature.** The percentage of rows that were observed to be joined to another ribbon at 4, 24 and 48 hours after mixing at room temperature. The error in this measurement was determined by bootstrapping: 1000 randomly-generated sample data sets (each of the same size as the full data set, but drawn with replacement) of observed ribbons were chosen, for each the fraction of observed joined rows was computed, and the variation in these estimates was recorded.

From this data, we estimated the rate constant for ribbon joining. Because sites where ribbons with identical patterns joined are not visible, we assumed the actual ribbon joining rates were twice the observed rate. The rate constant for joining, k_j can be estimated based on the following equation, where $[r]$ is the concentration of ribbons: $\frac{d[r]}{dt} = -k_j[r][r]$. (While a more accurate model would account for the widely varying diffusion constants for short and long ribbons, thus requiring distinct rate constants depending on the size of the ribbon, this complexity was not warranted for our purposes. Similarly, the temperature dependence of k_j was neglected.) If we assume that at $t = 0$, the average ribbon length is L repeat units, then the initial ribbon concentration is $[r]_0 = 100/L$ nM. Integrating this equation yields $[r](t) = \frac{[r]_0}{1+[r]_0 k_j t}$ and the fraction of joined rows at time t is $f(t) = \frac{1}{L}\left(1 - \frac{[r](t)}{[r]_0}\right)$. The data in Table 1 was fit by minimizing mean squared error (MSE) between the formula and the (doubled) observed fraction of joined rows, yielding $L = 60$ and k_j =35,000 /M/s. The error in this measurement was determined by bootstrapping, using the same method as for estimating the confidence intervals for Table S1. For each random sample, we fit L and k_j to the fraction of observed joined rows by minimizing MSE, yielding a 95% confidence interval between 11,000 and 80,000 /M/s for k_j. While this is a large range, the qualitative result that nucleation rates decrease with increasing ribbon width holds for all values in this range. Our estimate for k_j is only about 30 times smaller than the typical association rate of short oligos [4] which are much smaller and diffuse more quickly.

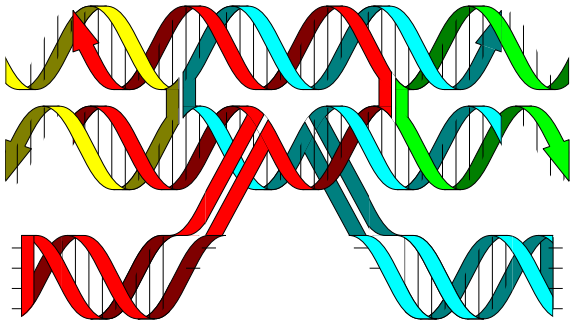
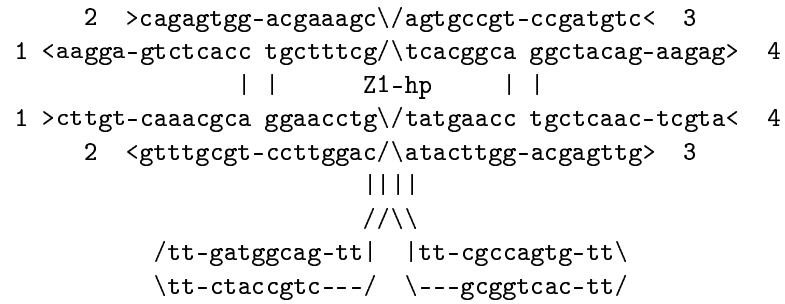


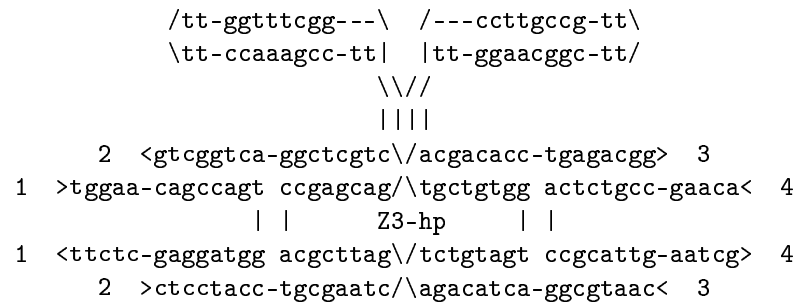
Figure S4: **An example tile with hairpins.** Modifications were made to tiles Z1, Z3, and Z4 for experiments to measure joining rates. In each modification, two hairpins were added to the tile in order to increase their contrast in atomic force microscopy. The illustration here shows Z4-hp. Z1-hp and Z3-hp have similar looking hairpins but they are attached to the other helix of the tile.

Tile Z1-hp :



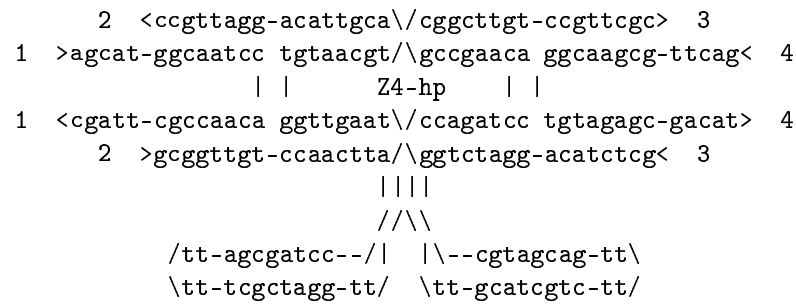
Sequences:
2 cagagtggacgaaagctcacggcaccaagtatctgccatctttgatggcagttcaggttcctgcgtttg
3 ctgtagcctgccgtgagctttcgtggaacctggcggtcacttttgtgaccgcttatacttggacgagttg

Tile Z3-hp :



Sequences:
2 ctcctacctgcgaatctctgtagtggtgtcgtggctttggttttccaaagccttctgctcggactggctg
3 caatgcggactacagagattcgcaccgagcagccttgccgttttcggcaaggttacgacacctgagacgg

Tile Z4-hp :



Sequences:
2 gcggttgtccaacttacctagcgatttttcgctaggttccagatccacaagccgacgttacaggattgcc
3 gctctacaggatctggcgtagcagttttctgctacgtttaagttggtgtaacgtcggcttgtccgttcgc

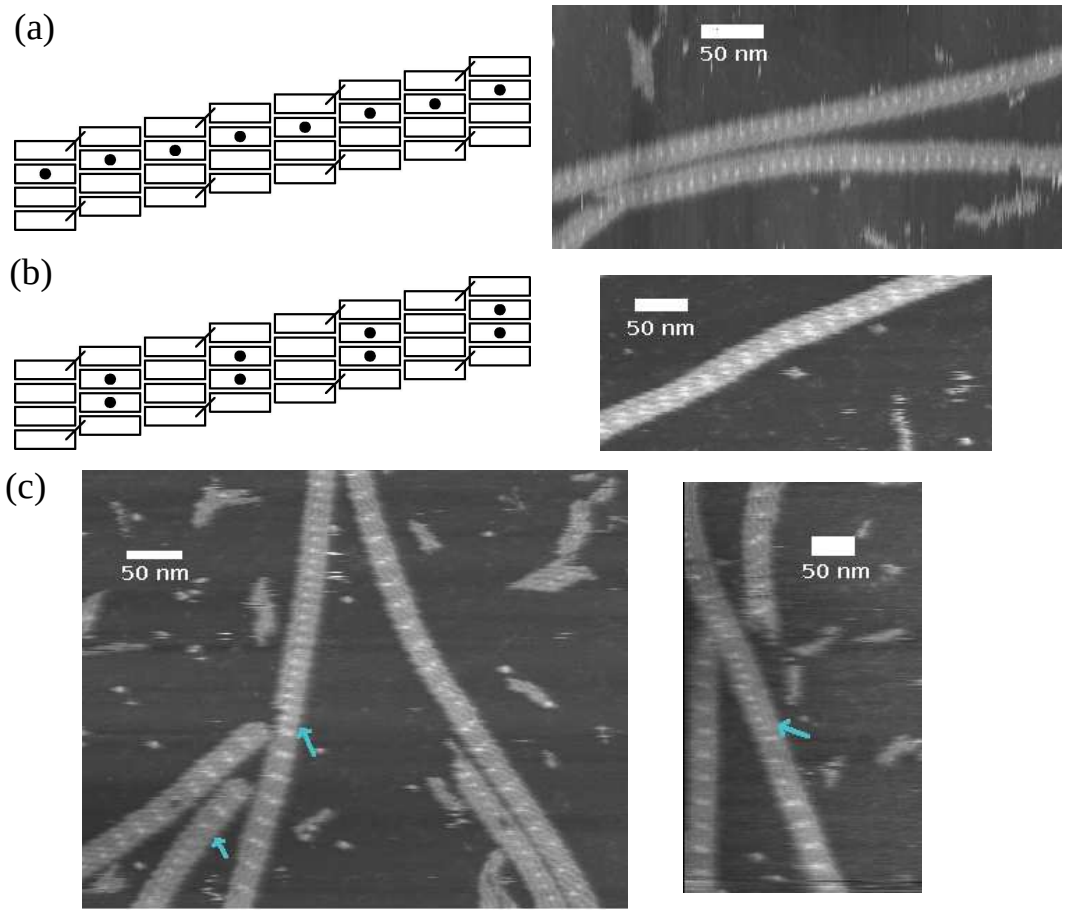


Figure S5: **Joining rates were measured by counting the fraction of attachments between zig-zag ribbons of with different patterns at three time points after the two ribbon types were mixed.** **(a)** Ribbons with Z1-hp and Z3-hp. The hairpins on these tiles form a dotted pattern under the atomic force microscope. **(b)** Ribbons with Z3-hp and Z4-hp. These hairpins form stripes in alternating columns and can be distinguished from those in (a). **(c)** When the dotted and striped ribbons are mixed, joining reactions between ribbons with different patterns leave a visible transition point in the ribbon. Two sample atomic force microscopy images of the mixtures of the two ribbon types are shown. Blue arrows point to sites where two ribbons joined.

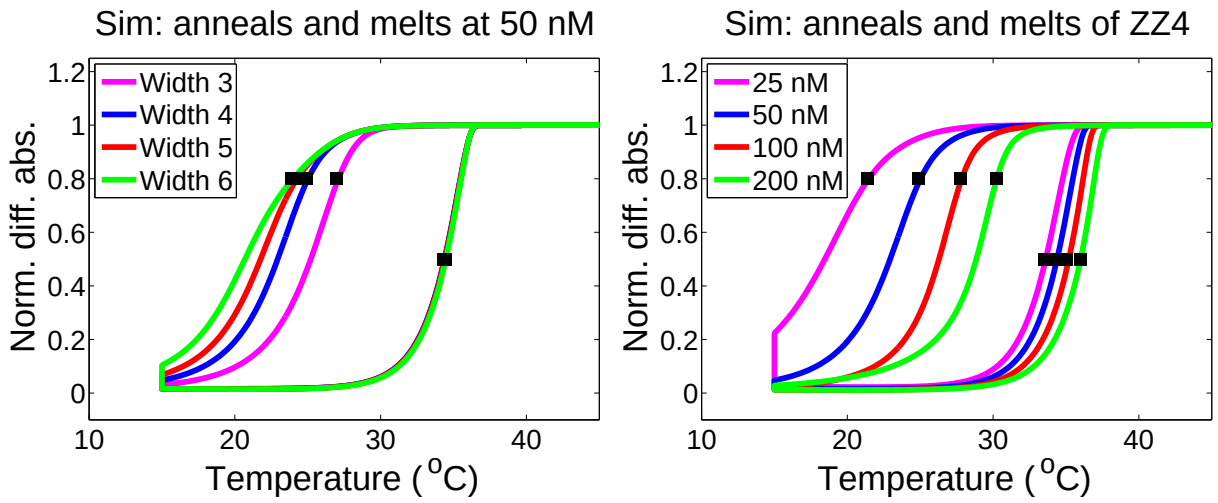


Figure S6: **Simulation of the standard sequence model for temperature-ramp experiments. (a)** Compare to Fig. 3(b) **(b)** Compare to Fig. 3(c).

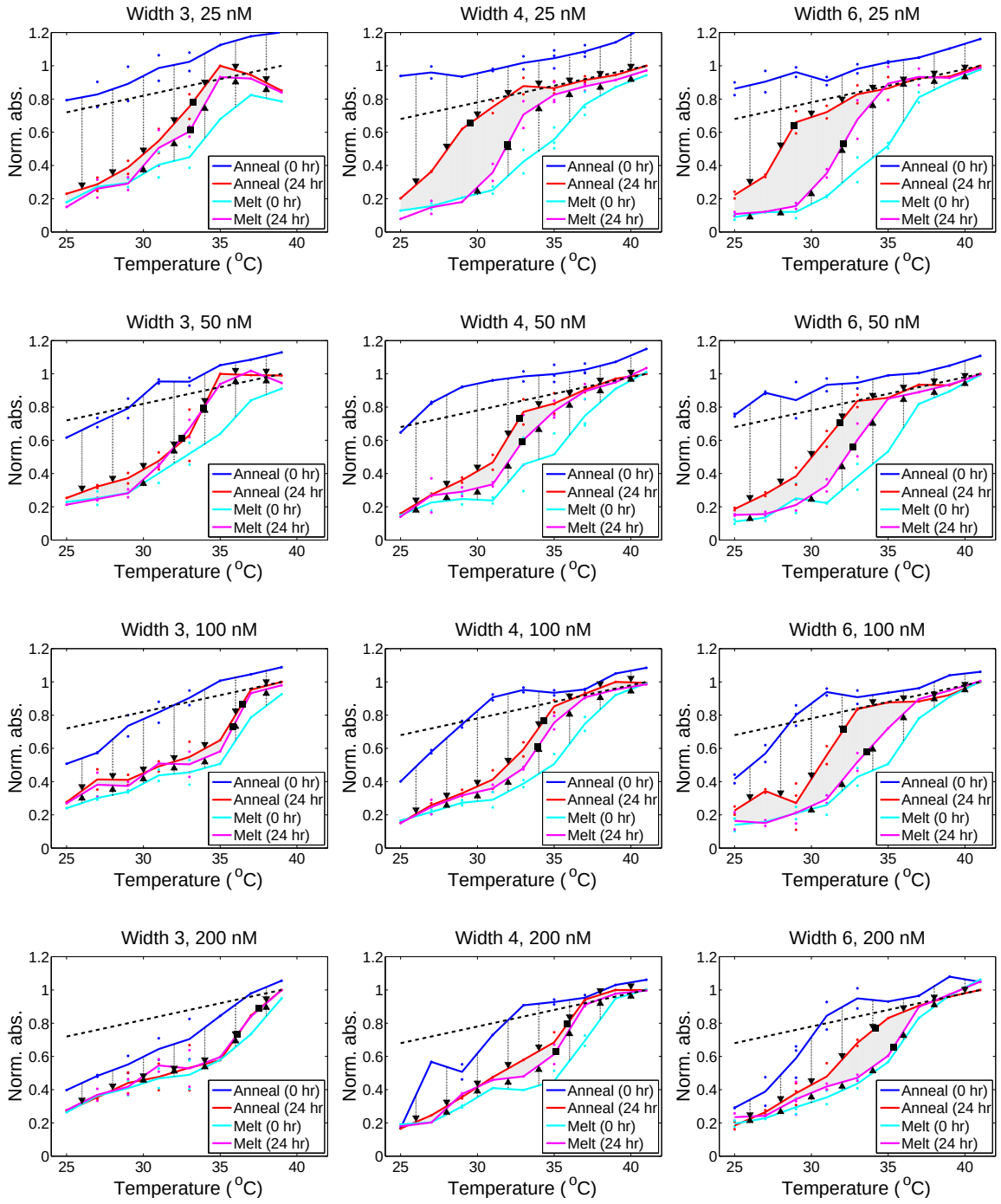


Figure S7: **Additional plots of temperature-hold experiments.** Data are plotted as in Figure 4(d-i). The critical temperature for crystal growth is ≈ 35 °C for 25 nM and 50 nM solutions for all ribbon widths, while for 100 nM and 200 nM it is ≈ 37 °C. The solution becomes supersaturated below this temperature.

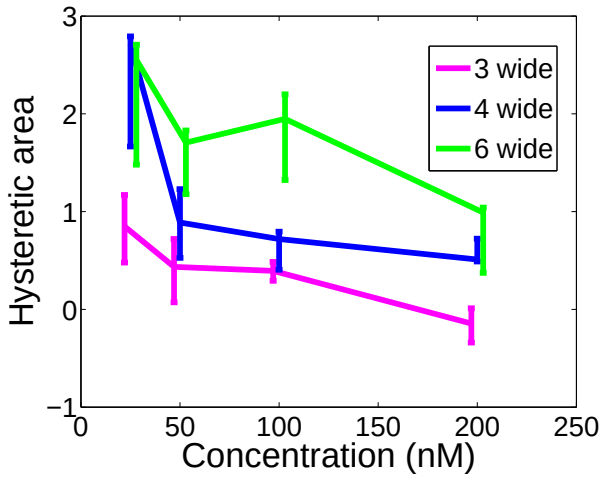


Figure S8: **Hysteretic area.** The area of the gray regions in Figure 4 above (the temperature hold experiments) for each ribbon width and concentration. Temperature-hold hysteresis grows with increasing ribbon width and decreasing concentration.

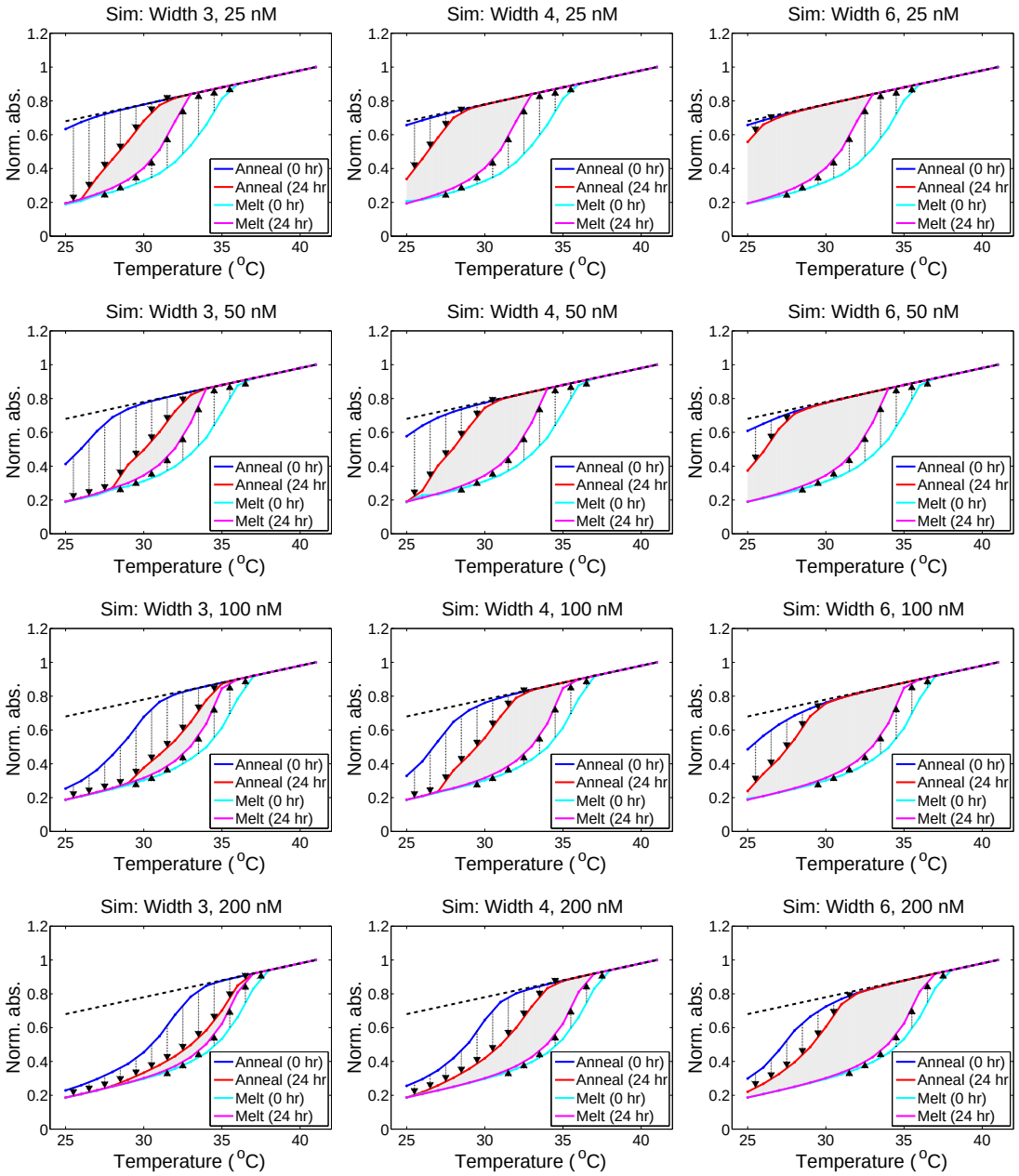


Figure S9: **Simulations of the standard sequence model for temperature-hold experiments.** See Fig. 4 for explanation of line colors and other notation. Simulated absorbance values assume that absorbance is due to (1) the total concentration of free sticky ends in solution, modeled as independent of assembly size, and (2) temperature-dependent absorbance of double-helical DNA, using the empirical slope calculated in Fig. 4. Linear normalization used reference point 0 for the absorbance at 15 °C and 1 for the absorbance at 41 °C. Note that absorbance changes due to continuing tile formation were not modeled, so the blue line ("absorbance at the beginning of the anneal hold") does not go above the dashed black line ("100% free monomer tile").

S4 Estimation of nucleation rates

Because classical nucleation could not be used to predict nucleation and growth uniformly across the ribbon widths, temperatures, and concentrations studied in this work, we chose to estimate the nucleation rate directly for suitable individual temperature-hold runs. As a basis for these estimations, we chose a simplified ordinary differential equation model of ribbon nucleation and growth that imposes no assumptions about the width-, temperature-, or concentration-dependence of the nucleation rate n_r. Thus we can use this model to directly estimate n_r under conditions where these factors are approximately constant.

For the unbound tile concentration to remain roughly constant, we consider only runs during which the total amount of tile depletion (ribbon growth) is small, ie where nucleation rates do not change significantly over time. As discussed below, we also cannot estimate nucleation rates when too little tile depletion has taken place. To accommodate these constraints, for each ribbon width and initial tile concentration, we only use the temperature-hold data corresponding to the melting temperature of the ribbons.

In our simplified model, the concentration of ribbons is given by $[r]$, the tile monomer concentration by $[m]$, the initial tile concentration by $[m]_0$, the number of tile types by N ($2w-2$ where w is the ribbon width), a forward rate constant for tile attachment k_f, a rate constant for ribbon joining k_j, the energy of attachment of a tile to a ribbon by two sticky ends, $\Delta G^\circ = \Delta H^\circ - T\Delta S^\circ$, and a reverse rate constant for tile detachment $k_r = k_f e^{\Delta G^\circ / RT}$.

$$\frac{d[r]}{dt} = n_r - k_j[r][r]$$
$$\frac{d[m]}{dt} = \frac{2}{N}\left(k_r - k_f[m][r]\right)$$

Assuming $[r] = 0$ and $[m] = [m]_0$ at $t = 0$, these equations can be solved analytically as a function of t:

$$[r]_t = \sqrt{\frac{n_r}{k_j}}\tanh\left(\sqrt{k_j n_r}t\right)$$
$$[m]_t = [m]_{eq} - \left([m]_{eq} - [m]_0\right)\left[\cosh\left(\sqrt{k_j n_r}t\right)\right]^{-\frac{2k_f}{Nk_j}}$$

where $[m]_{eq} = e^{\Delta G^\circ / RT}$ is the equilibrium concentration of free tiles. From this, we can express the nucleation rate n_r for an experiment with constant nucleation rate from $t = 0$ to $t = t_{end}$ as:

$$n_r = \frac{1}{k_j t_{end}^2}\left(\cosh^{-1}\left[\left(\frac{[m]_{eq} - [m]_0}{[m]_{eq} - [m]_{end}}\right)^{\frac{Nk_j}{2k_f}}\right]\right)^2, \tag{1}$$

where $[m]_{end}$ is the concentration of unbound tiles at $t = t_{end}$.

To use this formula to estimate nucleation rates from experimental data, we needed values for N, k_f, k_j, $[m]_{eq}$, and $[m]_{end}$. N is known by design, we used the same values for k_f and k_j as in the standard sequence model, and $[m]_{eq}$ can be calculated from the ΔH° and ΔS° determined in Figure 4(k). To estimate $[m]_{end}$ from absorbance measurements at $t_{end} = 24$ hours, we assumed that at the end of a 24 hour hold experiment, tile formation (as opposed to ribbon formation) was at equilibrium, so that absorbance was linearly proportional to absorbance. Thus, $[m]_{end}$ was determined by linear interpolation between the absorbance of $[m]_0$ and $[m]_{eq}$, where the former was taken to be the the estimated temperature-dependent absorbance for unbound tiles in solution (black dashed lines in Figures 4(d–i) and S7) and the latter was taken to be the final absorbance of the melt hold at 24 hours (magenta lines in Figures 4(d–i) and S7).

We used measured values of absorbance given in Section S4.1 to calculate nucleation rates. We calculated the nucleation rate at the melting temperature for each concentration, which were computed as described in Section S4.2.

As an example, for ZZ4 at 200 nM at the melting temperature 35.62 °C, the absorbance value along the dashed black line in Figure S7 is 0.8924. The value along the magenta line, corresponding to $[m]_{eq}$ = 100 nM, is 0.7049. The absorbance value along the red line is 0.7646, which using linear interpolation corresponds to a final free tile concentration of 124.7 nM. The resulting nucleation rate (obtained by plugging these values into Equation 1) is $\approx 1 \times 10^{-6}$ nM/s.

The absorbances for $[m]_{eq}$, $[m]_0$ and $[m]_{end}$ must be significantly different from each other in order to measure the nucleation rate accurately (see Equation 1— otherwise either the numerator or denominator of the term $\frac{[m]_{eq}-[m]_0}{[m]_{eq}-[m]_{end}}$ would be 0, in which cases the $cosh^{-1}$ of this term is undefined.) For the case where $[m]_0$=200 nM, instrument noise prevents us from reliably discriminating concentrations that are different by less than about 2% (the average deviation of our absorbance measurements from the average). Therefore, assuming the k_j and k_f used above, we cannot measure nucleation rates below about 1×10^{-8} nM/s (because $[m]_0$ and $[m]_{end}$ would be too close) or above about 7×10^{-6} nM/s (because $[m]_{eq}$ and $[m]_{end}$ would be too close.)

S4.1 Numerical normalized absorbance values

These numbers are also plotted in Figures 4(d–i) and S7.

Normalized absorbance values used to determine nucleation rates of ZZ3 at 200 nM

Temperature (Centigrade)	No ribbons	After anneal hold	After melt hold
37	.9600	.8455	.8402
35	.9200	.5775	.5964

Normalized absorbance values used to determine nucleation rates of ZZ4 at 200 nM

Temperature (Centigrade)	No ribbons	After anneal hold	After melt hold
37	.9200	.9421	.9107
35	.8800	.6848	.6124

Normalized absorbance values used to determine nucleation rates of ZZ6 at 200 nM

Temperature (Centigrade)	No ribbons	After anneal hold	After melt hold
37	.9200	.9052	.8978
35	.8800	.8308	.6041

S4.2 Calculation of Ribbon Melting Temperatures

At equilibrium, tile monomer ($[m]$) addition and release from ribbons ($[r]$) should occur at the same rate. That is, the following chemical reaction should be at equilibrium:

$$m + r \underset{k_r}{\overset{k_f}{\rightleftharpoons}} r \tag{2}$$

where $k_r = k_f e^{\Delta G^\circ / RT}$. Thus, at equilibrium

$$\frac{[r]}{[m][r]} = \frac{k_f}{k_r} = e^{-\Delta G^\circ / RT} \tag{3}$$

$$[m] = e^{\Delta G^\circ / RT}. \tag{4}$$

At the melting temperature, the equilibrium monomer concentration is half the total initial concentration. Algebraic manipulation of Equation 4 yields

$$T_m = \frac{\Delta H^\circ}{\Delta S^\circ + R \ln([m])}. \tag{5}$$

We used Equation 5 and the measured values $\Delta H^\circ = -102.4$ kcal/mol and $\Delta S^\circ = -0.300$ kcal/mol/K to determine the melting temperatures for each tile concentration. The temperatures are given in Table S2.

Concentration	Melting temperature
25 nM	31.84 °C ± .50 °C
50 nM	33.09 °C ± .31 °C
100 nM	34.35 °C ± .21 °C
200 nM	35.62 °C ± .32 °C

Table S2: **Melting temperatures used for each concentration in the calculation of the nucleation rate.** 95% confidence intervals are determined using bootstrapping: We calculated the melting temperature for each of the twelve samples assuming each possible combination of redundant data points were removed. We then used a random melting temperature from each of these sets of possible melting temperatures to calculate 1000 possible ΔH° and ΔS° values, and therefore melting temperatures, for each concentration. Each confidence interval is twice the standard deviation of this set of melting temperatures.

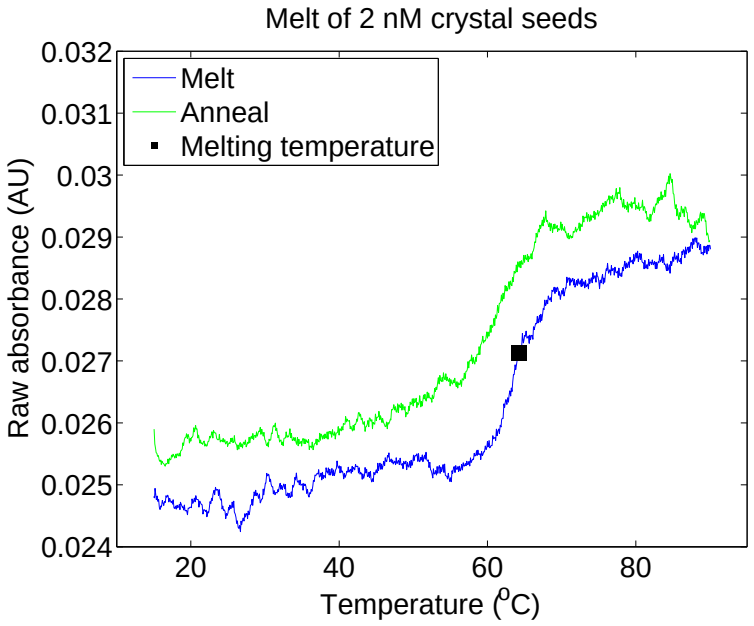


Figure S10: **Melt and re-anneal of ribbon seeds.** Seeds were prepared by annealing all strands together at 50 nM (each strand) in the standard annealing buffer. The annealed sample was then diluted to 2 nM (each strand), melted and re-annealed in the spectrophotometer at about 1 °C per 7.5 minutes. The melting curve indicates a melting temperature around 62 °C (marked with a square). This shows that the crystal seed structure is stable at both the ribbons' melting temperature (≤ 36 °C) and the temperature at which seeds were added to the ribbon strand bouillabaisse (40 °C). Even though we found that the seed structure formed with acceptably high yield (Figure S1(b)), the folding of the seed structure is not reversible at the speed of the melt shown here. The seeds were present in the crystal seeding experiments at concentration shown here (2 nM). The noise in the data is due to the low concentration of seeds and hence low absolute signal.

References

[1] Rothemund, P. W. K, Papadakis, N, & Winfree, E. (2004) *PLOS Biology* **2**, 424–436.

[2] SantaLucia, Jr., J, Allawi, H. T, & Seneviratne, P. A. (1996) *Biochemistry* **35**, 3555–3562.

[3] Bloomfield, V. A, Crothers, D. M, & Tinoco, Jr., I. (2000) *Nucleic Acids: Structures, Properties, and Functions*. (University Science Books).

[4] Wetmur, J. G. (1991) *Critical Review in Biochemistry and Molecular Biology* **36**, 227–259.

[5] SantaLucia, Jr., J. (1998) *Proc. Nat. Acad. Sci. USA* **95**, 1460–1465.

[6] Seeman, N. C. (1990) *Journal of Biomolecular Structure & Dynamics* **8**, 573–581.

[7] Dirks, R, Lin, M, Winfree, E, & Pierce, N. A. (2004) *Nucleic Acids Research* **32**, 1392.

[8] McCall, M, Brown, T, & Kennard, O. (1985) *Journal of Molecular Biology* **183**, 385–396.

[9] Winfree, E, Liu, F, Wenzler, L. A, & Seeman, N. C. (1998) *Nature* **394**, 539–544.