Developmental Self-Assembly of a DNA Tetrahedron

John P. Sadowski,[1,2] Colby R. Calvert,[4] David Yu Zhang,[1,3] Niles A. Pierce,[4,5] and Peng Yin[1,3*]

[1] Wyss Institute for Biologically Inspired Engineering, Harvard University, Boston, MA 02115

[2] Department of Chemistry and Chemical Biology, Harvard University, Cambridge, MA 02138

[3] Department of Systems Biology, Harvard Medical School, Boston, MA 02115

[4] Division of Biology and Biological Engineering, California Institute of Technology, Pasadena, CA 91125

[5] Division of Engineering and Applied Science, California Institute of Technology, Pasadena, CA 91125

[*] E-mail address: py@hms.harvard.edu

Contents

S1 Expanded versions of figures

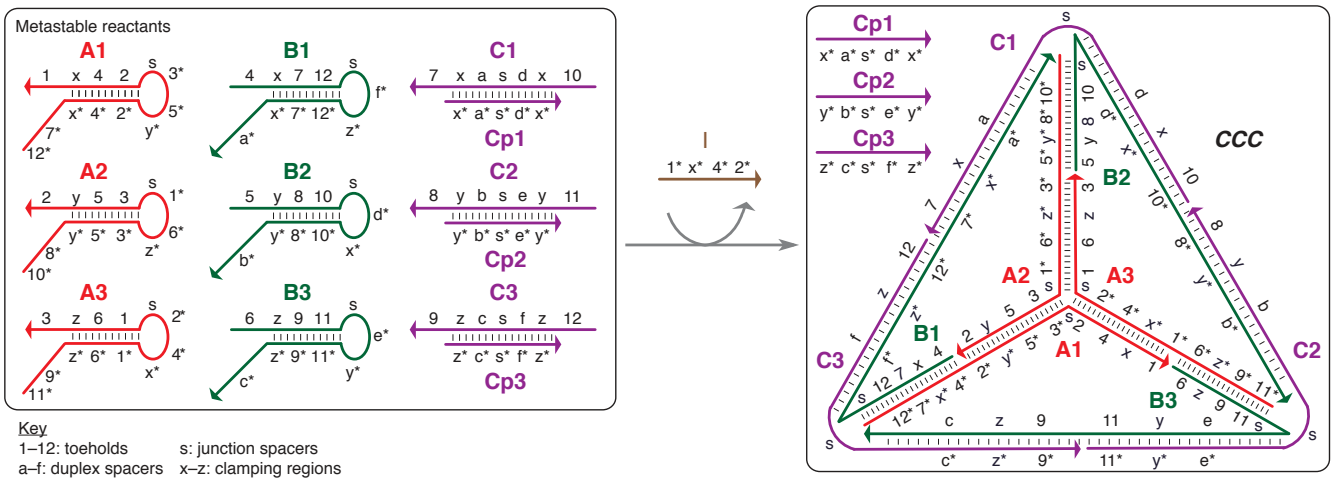

Figure S1. Overview of the reaction with all sequence segments labeled. Toeholds are labeled with numbers and various types of spacers are labeled with letters. This is an expanded version of Fig. 1a.

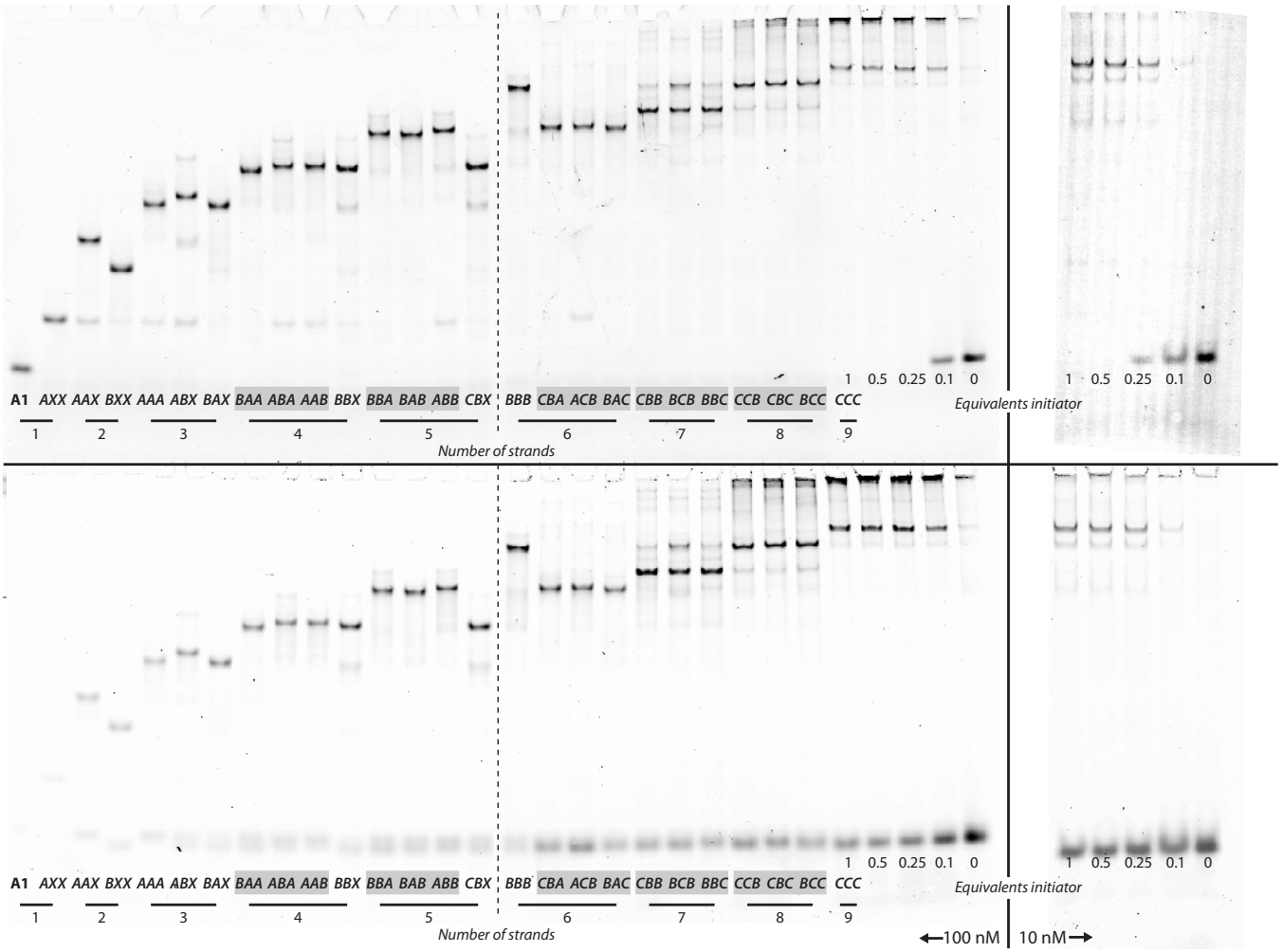

Figure S2. Formation of the tetrahedron. These are the separated channels from the gel in Fig. 2. Top, the scan of FAM fluorescence in the unstained gel; bottom, the scan of SYBR Gold fluorescence after staining.

Name	A1	B1	C1	A2	B2	C2	A3	B3	C3	I	Structure
AXX	+									+	
AAX	+			+						+	
BXX	+	+								+	
AAA	+			+			+				
ABX	+			+	+					+	
BAX	+	+		+						+	
BAA	+	+		+			+				
ABA	+			+	+		+				
AAB	+			+			+	+			
BBX	+	+		+	+					+	
BBA	+	+		+	+		+				
BAB	+	+		+			+	+			
ABB	+			+	+		+	+			
CBX	+	+	+	+	+					+	
BBB	+	+		+	+		+	+			
CBA	+	+	+	+	+		+				
ACB	+			+	+	+	+	+			
BAC	+	+		+			+	+	+		
CBB	+	+	+	+	+		+	+			
BCB	+	+		+	+	+	+	+			
BBC	+	+		+	+		+	+	+		
CCB	+	+	+	+	+	+	+	+			
CBC	+	+	+	+	+		+	+	+		
BCC	+	+		+	+	+	+	+	+		
CCC	+	+	+	+	+	+	+	+	+		

Figure S3. Composition of the intermediates. The strands contained in each species are marked.

S2 Additional characterization of the tetrahedron assembly

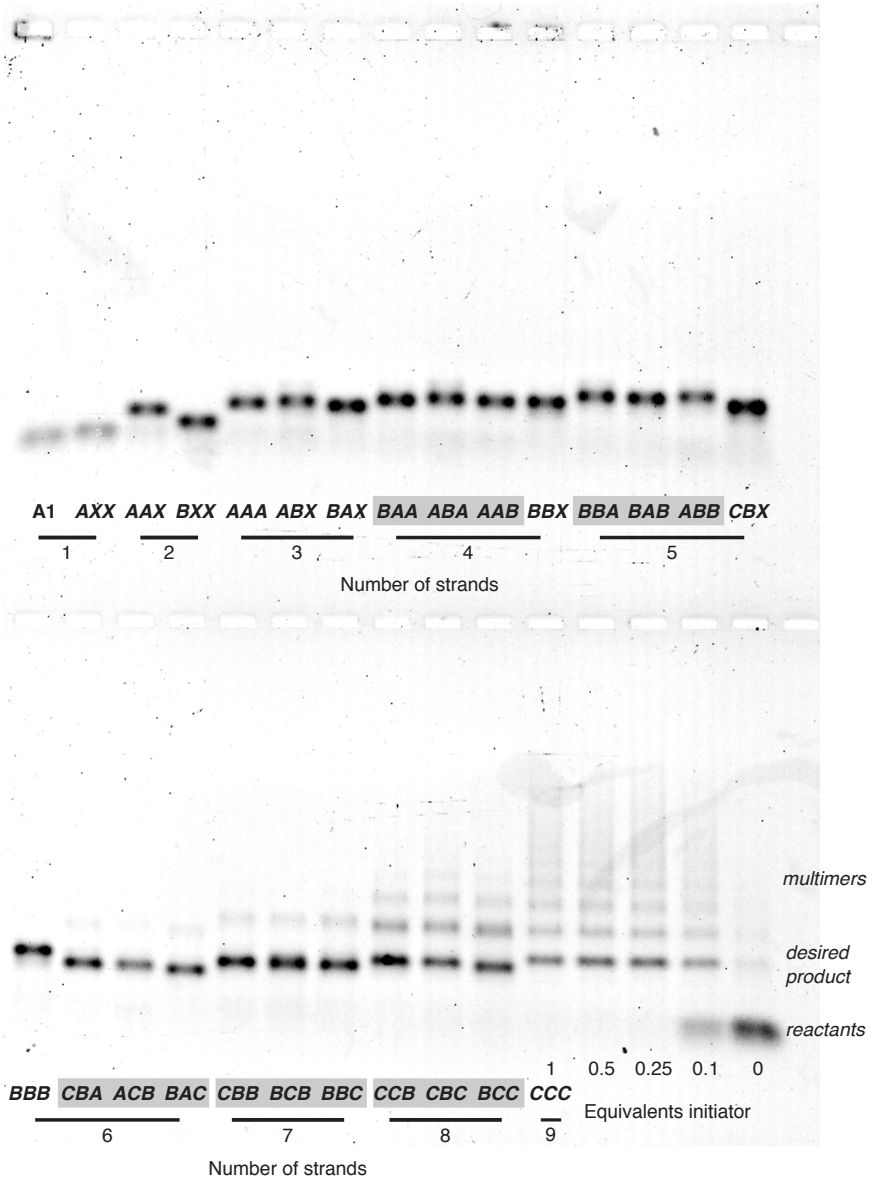

Figure S4. Characterization of side products of higher molecular weight. While the side products of higher molecular weight do not exit the well in the polyacrylamide gel in Fig. 2, in this agarose gel they are well-resolved as a series of discrete bands that might possibly correspond to well-defined multimers of the desired tetrahedron product. This gel uses samples from the same aliquots used in the formation assay in gel in Fig. 2. This is a 1% native agarose gel of a 100 nM assembly reaction.

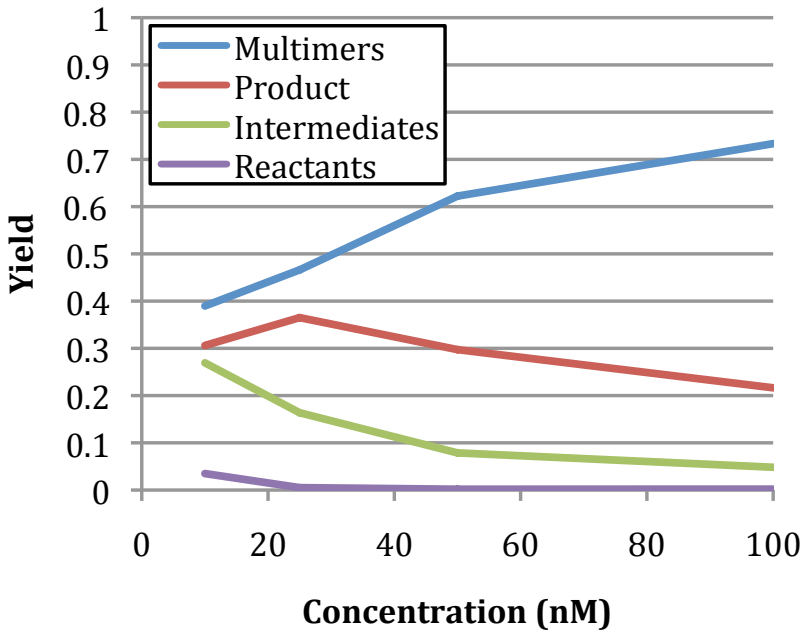

Figure S5. Concentration robustness of the assembly reaction. This graph shows the yield of the reaction at different concentrations of reactants, as calculated through quantitation of gel electrophoresis staining intensity. As concentration increases, the yield of intermediates decreases but the yield of multimeric side products increases. Yield was calculated as the percentage of signal in the appropriate region of the lane divided by the total signal in the entire lane, using the same parameters described in the Experimental Methods section. Each value is the average of three gel lanes each containing the same sample.

a

Toehold lengths	Total base pairs formed	*CBB* forms?
8:8	15	Yes
10:6	15	Yes
12:4	15	Yes
16:0	15	Unstable product
10:4	13	Yes
6:6	11	No reaction
12:0	11	No reaction
4:4	7	No reaction

b

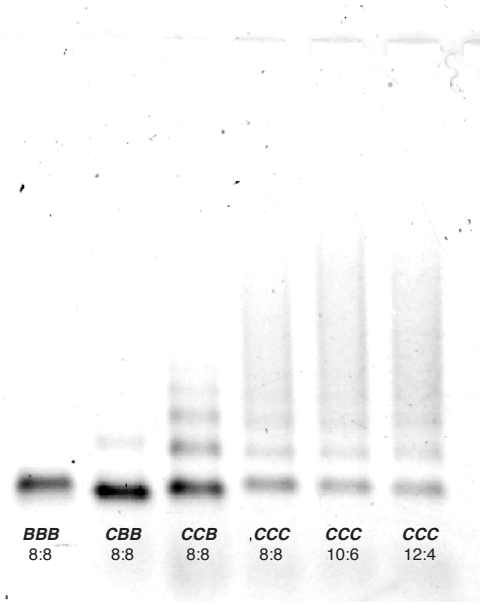

Figure S6. The effects of toehold lengths on the assembly of cooperative hybridization complexes. **a**, The results of gel electrophoresis studies on assembly of the intermediate ***CBB*** using cooperative hybridization complexes with different toehold lengths. The lengths are listed with the 5′ toehold length followed by the 3′ toehold length, separated by a colon (:). The 8:8 complex was the one used in the main body of this paper. In general, the total number of base pairs formed seems to be the main determinant of whether the assembly proceeds, rather than the lengths of the individual toeholds. Note that the junction spacer base becomes unpaired during assembly, lowering the total number of base pairs formed by one. For the failed complexes, the ***BBB*** band was observed, indicating no assembly event involving the cooperative complex had occurred. For the 16:0 complex only, the singly bound intermediate was stable enough to form a streak on the gel above the the ***BBB*** band. We made these complexes by extending or truncating the protector strands as appropriate. **b**, A gel showing the full tetrahedron ***CCC*** where all three cooperative complexes have the specified toehold lengths. No qualitative difference in yield is observed. This is a 1% native agarose gel of a 100 nM assembly reaction.

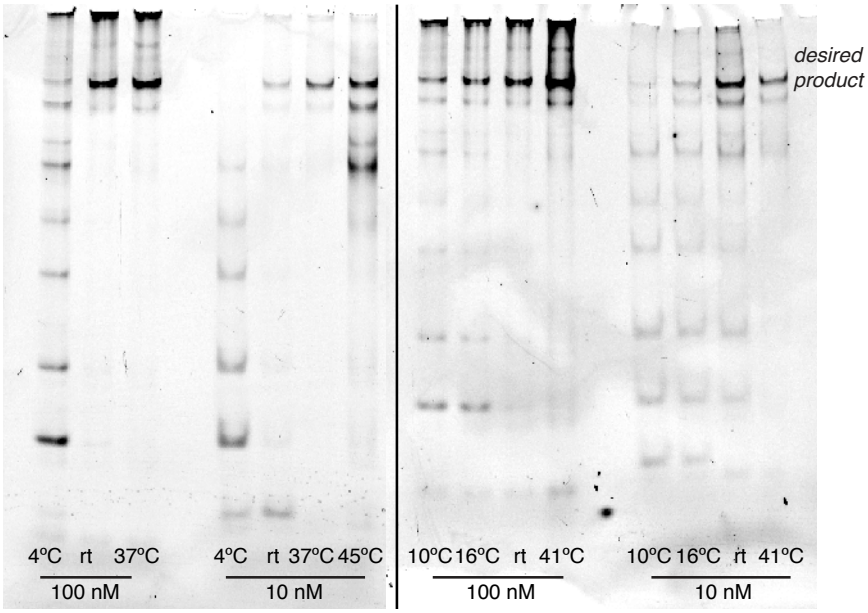

Figure S7. Temperature robustness of the assembly reaction. These gels show the results of assembly at different temperatures, at two concentrations after 18 hr. The reaction appears to go to completion to the same extent as the room temperature samples in roughly the range of 16–41°C. At lower or higher temperatures, the reaction does not go to completion. These are 6% native polyacrylamide gels. The solid line separates two gels that were run at different times.

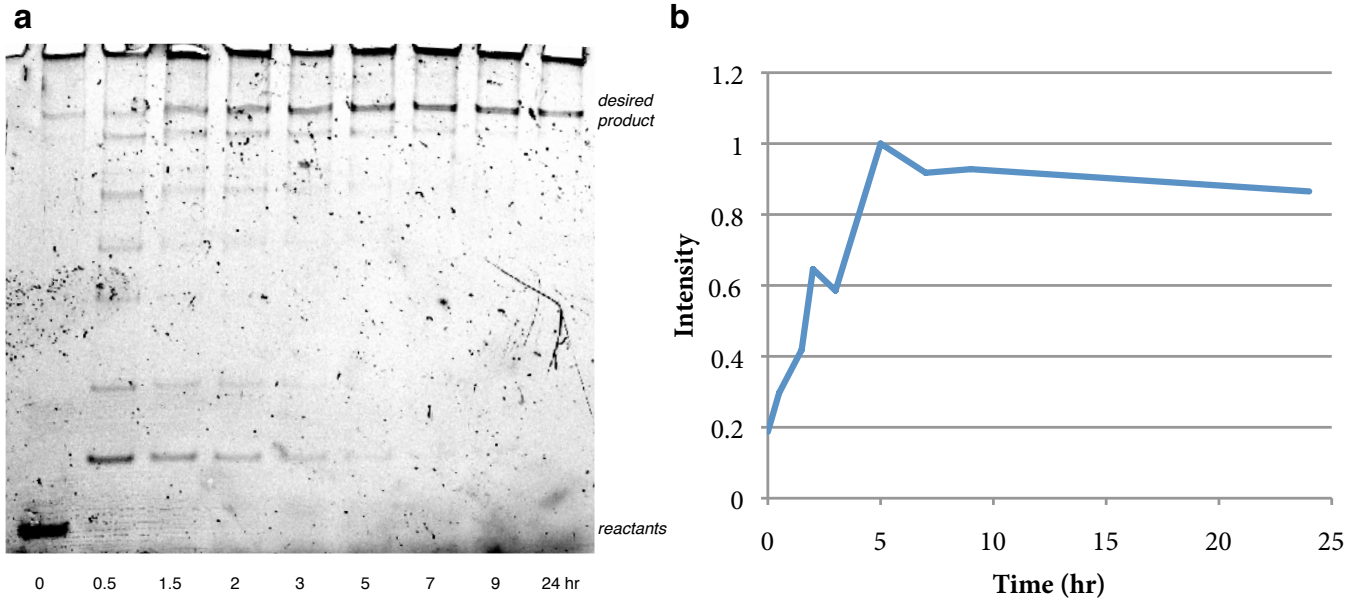

Figure S8. Progress of the reaction over time. a, A gel showing a timecourse of the assembly. These reactions all had the nine reactants mixed 24 hr before gel loading, and the initiator was added at the specified time before gel loading. This is a 6% native polyacrylamide gel of a 100 nM assembly reaction. **b**, A graph showing the relative intensities of the desired product bands in the gel.

S3 Design considerations

S3.1 Motif design

The secondary structure design of the tetrahedron (Fig. S1) required us to consider the placement and dimensioning of the sequence domains within the tetrahedron. Because we desired each edge of the tetrahedron to have the same length, we inserted the spacer domains a–f into the tetrahedron's outer edges to lengthen them. We also added one-nucleotide junction spacers s to increase the flexibility of the junctions. We designed the tetrahedron to have an integral number of turns for each edge with the minor groove facing inwards at each junction. Two-nucleotide clamping domains x–z were placed at the outer edges of the hairpin stems to discourage leakage out of the metastable hairpin state. We avoided placing large complementary segments in the loops, which has been observed to cause leakage due to kissing-loop interactions.

S3.2 Sequence design

The sequence design of the tetrahedron was accomplished using an as-yet unpublished software platform, called Multisubjective, which incorporates the previously described program Domain Design (DD).[1]

Multisubjective minimizes undesired secondary structure in nucleic acid systems. It identifies a minimal set of specific bases which are responsible for undesired secondary structure in a given nucleic acid system, selecting the best design from a set of candidate sequences. Multisubjective then instructs a client designer to redesign only those bases, holding the others constant. Multiple iterations of this analysis–redesign process yields sequences where the undesired secondary structure has been greatly diminished or entirely removed. Multisubjective streamlines this design process by interacting directly with supported client designers, automatically passing the analyzed sequences to them and getting the redesigned candidate sequences back without requiring user intervention. The base pairing probabilities used in this analysis are generated by a local copy of the NUPACK analysis software,[2,3] which are then processed by Multisubjective's internal algorithm.

S3.3 Alternative structures using bridge interactions for ring formation

In the course of developing the tetrahedron presented in the main body of this paper, which we have designated TET9, a number of alternative designs were evaluated. Two of these, TET6 and TET8, used "bridge regions" for the ring formation reactions instead of cooperative hybridization complexes. The bridge binding interaction occurs between two complementary regions that are independently revealed on different hairpins; these are intended to hybridize to each other once both are present within the same complex. In the TET6 and TET8 systems, the bridge interactions are each between regions on a pair of one **B** and one **C** hairpin. (Fig. S9). TET6 and TET8 differ in the length of the region involved in the bridge interaction: for TET6, 12 base pairs are formed, while for TET8, 20 base pairs are formed.

A partial formation assay on TET6 suggested that all nine strands did assemble into a single complex (Fig. S10), but a fluorescence-quenching assay failed to indicate that the rings were forming to any significant degree (Fig. S11a). However, the quenching experiment was successful for a single ring closing event on a variant fragment of TET6 where one bridge region was extended to form 20 instead of 12 base pairs (Fig. S11b). This suggested that the shorter length was not enough to drive the ring formation to completion, and TET8 was designed to incorporate 20 nt bridge regions throughout.

However, TET8 did not form properly (Fig. S12). We believe this was because of an effect we call *segment proliferation*: because segments must be complementary both across the closed hairpin stems and within the final structure, chains of complementarity dependence can cause segments to appear multiple times in the system. In this case, the bridge segments had proliferated so that they appeared in the tails and stems of the three **B** hairpins, causing them to form an off-pathway trimer when open, leading to the formation of aggregates. Further investigation indicated that there was no "Goldilocks" length that was neither too short to form the rings, nor too long to avoid off-pathway interactions leading to aggregation. This result led to the design of TET9 using cooperative hybridization complexes instead of bridge interactions for the ring formations.

In summary, bridge interactions are capable of forming rings if the interaction forms enough base pairs, but their use was found to be incompatible with the specific tetrahedron architecture used here due to issues with segment proliferation. Nevertheless, bridge interactions are still an option for use in future assemblies with different architectures.

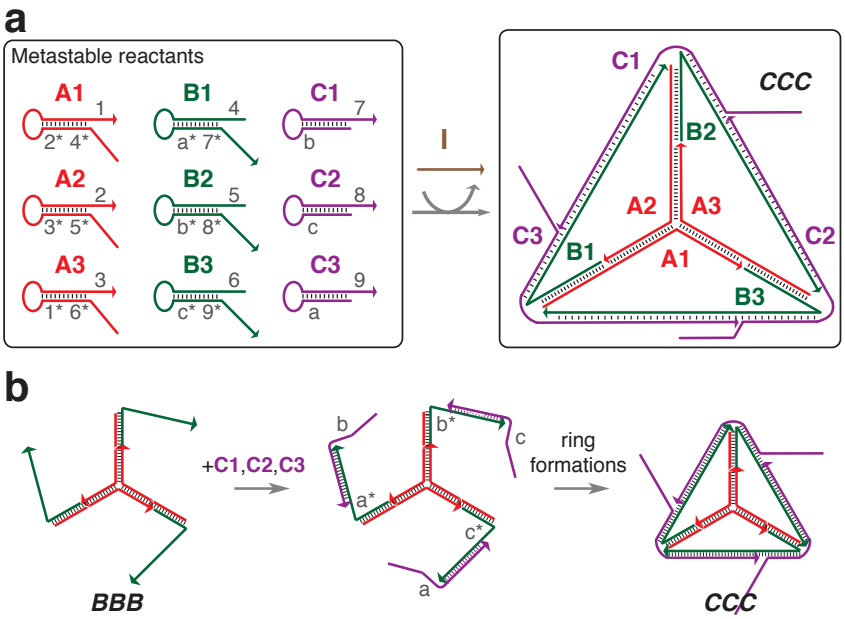

Figure S9. Overview of the TET6 and TET8 systems. a, Nine hairpins self-assemble in the presence of a catalytic initiator to form the tetrahedron. Portions of the **C** strands become single-stranded attachments to the final structure. Only toeholds (numbers) and bridge regions (letters) are marked in this figure. **b**, Schematic showing the bridge interactions used for the ring forming steps. The assembly exposes complementary single-stranded regions called "bridge regions" on three corresponding sets of **B** and **C** hairpins. The hybridization of each set of bridge regions is intended to drive the ring formations to occur. TET6 had 12 nt bridge regions, while TET8 had 20 nt bridge regions. Note that the ring formations do not occur in a synchronized manner, as shown here, but each may occur at any time after the corresponding set of a **B** and a **C** hairpin has assembled.

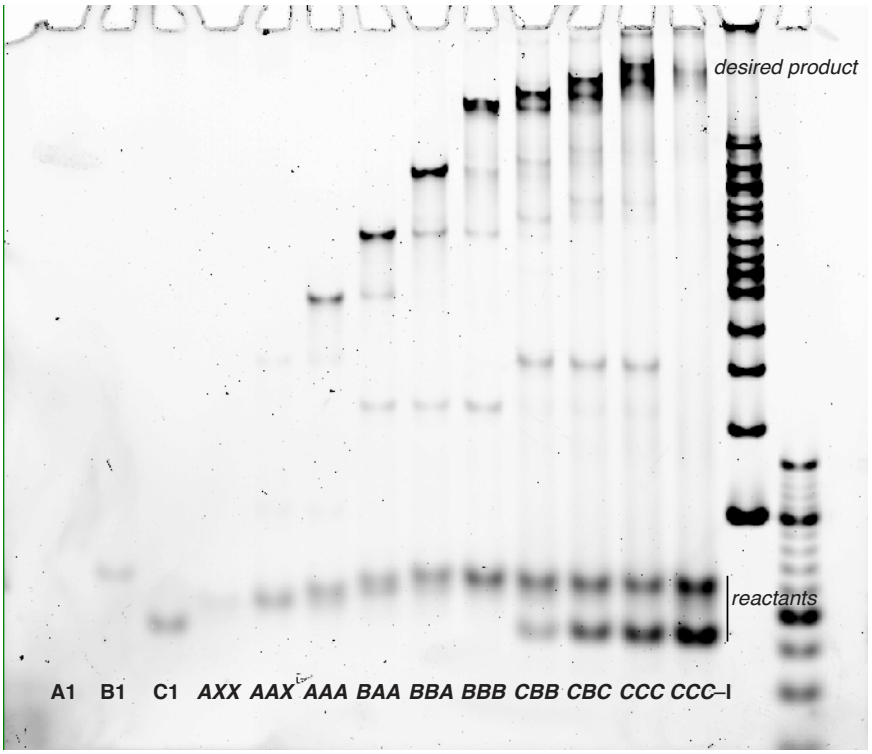

Figure S10. Partial formation assay of TET6. The desired product is observed to form. The ***CCC*–I** lane contains all nine reactants without the initiator. This is a 6% native polyacrylamide gel of a 100 nM assembly reaction with 0.5× initiator.

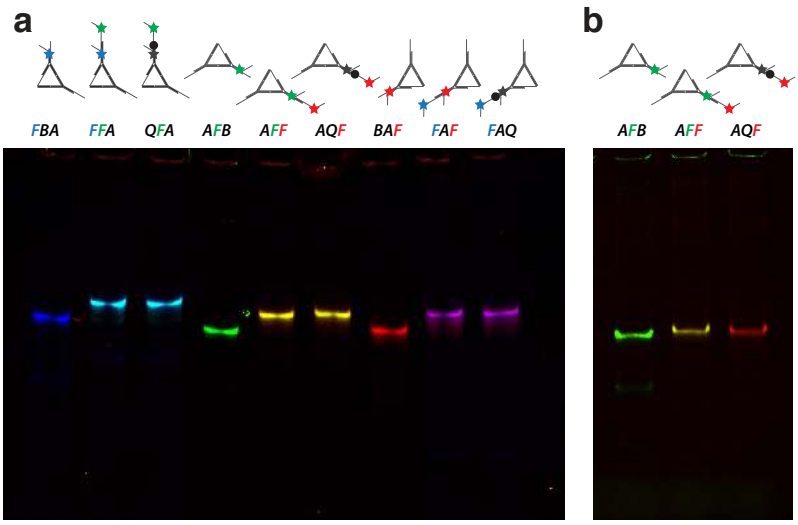

Figure S11. Fluorescence quenching assay of TET6. **a**, This quenching experiment uses the same principles as that in Fig. 3, but this gel shows no significant quenching in the lanes for ***QFA***, ***AQF***, and ***FAQ***. This suggests that the rings have not formed in any of these structures, meaning that the full complex is likely to mostly exist as a large three-arm junction corresponding to the center structure in Fig. S9b. The diagrams above the gel represent the desired rather than the experimentally determined structures. **b**, A quenching experiment using an altered set of TET6 strands that had a bridge region with a length of 20 nt rather than 12 nt for one of the bridge interactions. The desired quenching in the ***AQF*** lane is observed, indicating that a longer bridge region length is necessary to drive the desired interaction to completion. These are 6% native polyacrylamide gels of 100 nM assembly reactions with 0.5× initiator.

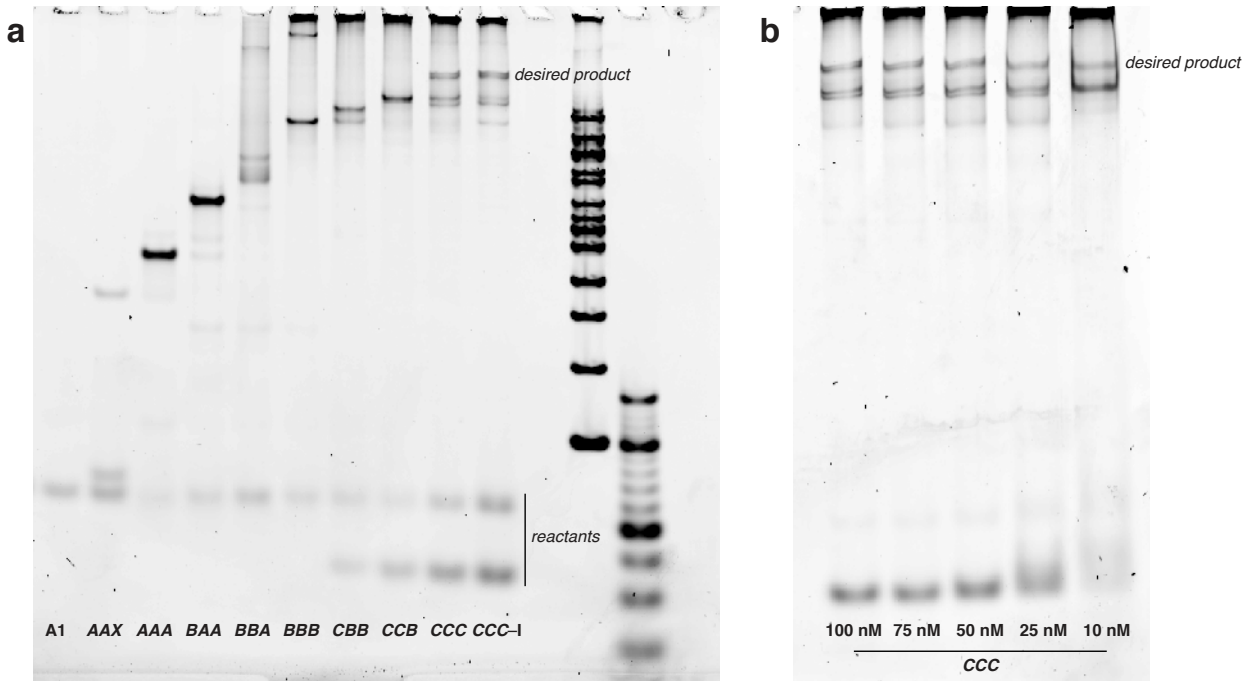

Figure S12. Partial formation assay of TET8. **a**, Undesired higher molecular weight products are seen in the ***BBA*** lane and dominate in the ***BBB*** lane and subsequent lanes. These side products are thought to be aggregates resulting from undesired interactions involving open **B** hairpins; these undesired interactions are believed to be the result of segment proliferation. This is a 6% native polyacrylamide gel of a 100 nM assembly reaction with 1× initiator. **b**, A gel showing the full tetrahedron at different assembly concentrations. The aggregate remains the dominant product at concentrations as low as 10 nM. This is a 6% native polyacrylamide gel where different volumes of sample were loaded in each lane so that the amount of sample was constant.

S3.4 An alternative structure using three-output hairpins for ring formation

TET10 is a structure designed so that the ring-closing interactions occur within one edge rather than overlapping two edges of the tetrahedron, leading to ring formation reactions that are modular with respect to the edges (Fig. S13a,b). This necessitates a novel three-output hairpin motif for the three **A** hairpins, each having an extra initially-accessible output region on its tail. Because these open output regions assemble to cooperative hybridization complexes, no assembly can occur until that cooperative complex's other initiator becomes available (Fig. S13c). A partial formation gel suggests that this design works qualitatively as well as TET9, although some of the intermediates seem to be of lower quality (Fig. S14). Nevertheless, this indicates that the three-output motif is a viable option for future assemblies.

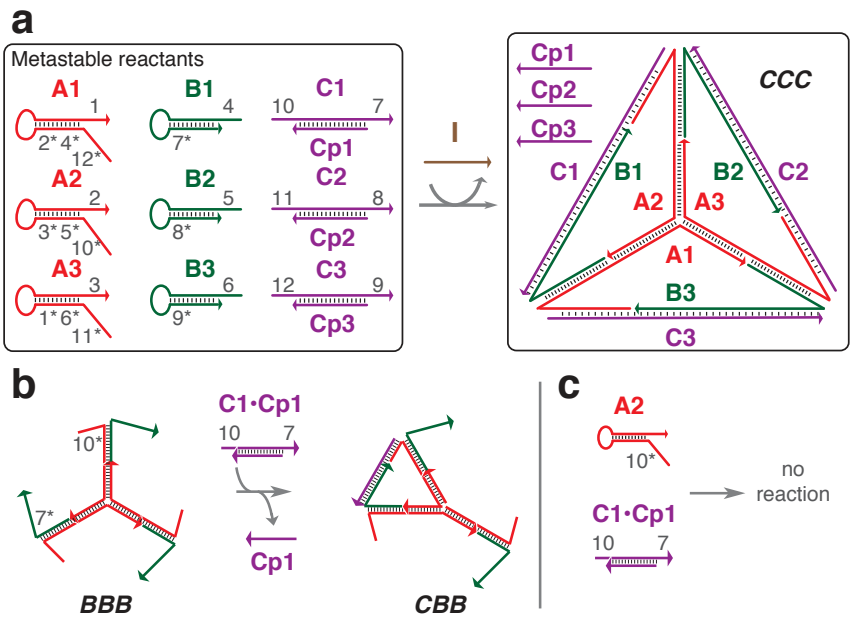

Figure S13. Overview of the TET10 system. **a**, Six hairpins and three cooperative hybridization complexes combine in the presence of a catalytic initiator to form the tetrahedron. Relative to TET9, the **A** hairpins are elongated and the **B** hairpins are truncated. This results in the **C** strands being in a different position, with each strand corresponding to one edge of the tetrahedron rather than overlapping two edges each. Only toeholds are marked in this figure. **b**, Schematic of a ring formation reaction. **c**, Even though the 10 toehold of the cooperative complex and 10* toehold on the tail of the hairpin are both initially accessible, no assembly occurs because the 7* toehold (in the **B1** hairpin) is not available until it is exposed by assembly events elsewhere.

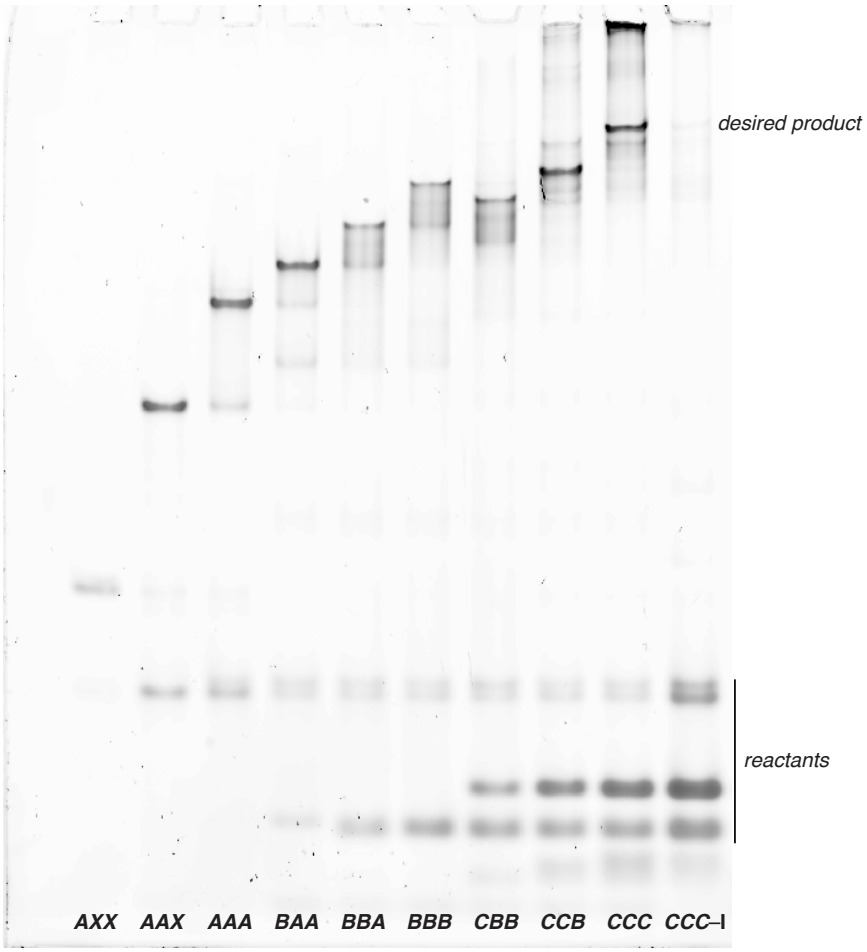

Figure S14. Partial formation assay of TET10. The desired product is observed to form with a qualitatively similar yield to TET9, but the quality of the ***BBA***, ***BBB***, and ***CBB*** intermediates appears to be lower than that of TET9. This is a 6% native polyacrylamide gel of a 100 nM assembly reaction with 1× initiator.

S4 Nodal abstraction and reaction graphs

The nodal abstraction emphasizes the function of each reactant complex rather than its structure, providing a concise notation for a prescriptive molecular program that can be compiled into a domain-level molecular design. The nodal abstraction specifies a molecular program that is executed as the assembly proceeds, with the output ports switching states from closed to open as upstream input ports are activated, and then switching back to closed as they activate downstream input ports. Readers are referred to previous work describing the nodal abstraction in detail.[4]

Each molecular species is represented by a node appearing as a black ring with shapes representing input and output ports. Each port corresponds to a physical region on the strand or complex, containing one or several sequence domains. Triangles represent input ports while circles represent output ports. Solid ports are inaccessible at the beginning of the assembly and become accessible only when the input ports on that reactant becomes assembled, while outlined ports are accessible at the beginning of the assembly. The color of the ports represents their position on the strand.

TET6 and TET8 contain a new type of ports that we call bridge ports. Bridge ports contain no toehold segments and their interactions are represented by converging double arrows; these assembly events occur only when both output ports are simultaneously accessible. Cooperative hybridization complexes are represented here as having two input ports and no output ports; assembly occurs only when both ports are activated simultaneously.

Assembly reactions are represented by solid arrows and disassembly reactions by dashed arrows; these connect the nodes into a reaction graph that specifies the overall assembly order. The nodal reaction graphs for each of the tetrahedra described here appear in Fig. S15.

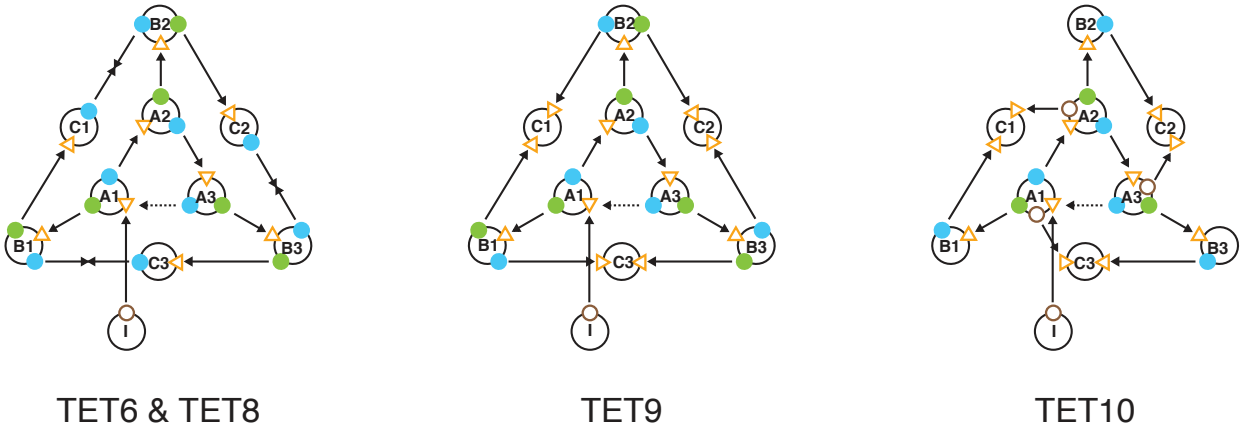

Figure S15. Nodal reaction graphs of the tetrahedra.

S5 Strand sequences

I: TGGTTCGTCGTGTCCGGGATATGGGT
A1: AGATGAATTCTTAACTCGTGTCCGGGATATGGGTCTATAAGTGTGATCGGTGTACCCATA
TCCCGGACACGACGAACCA
A2: GATATGTGATTTAGAACTATAAGTGTGATCGGTGGCTATGAGGATGGTTCGTTCACCGAT
CACACTTATAGACCCATAT
A3: TAGAATTTGATGACTAGCTATGAGGATGGTTCGTCGTGTCCGGGATATGGGTTACGAACC
ATCCTCATAGCCACCGATC
B1: CCCGGACACGAGTTAAGAATTCATCTTAAGGTGACATTGTTCGGCAGATGAATTCTTAAC
TCGGTAGTATACGTTAGAG
B2: ACACTTATAGTTCTAAATCACATATCTCTACTGTAGACTTTGACGGATATGTGATTTAGA
ACTTAGTCGCGATAACTTA
B3: TCCTCATAGCTAGTCATCAAATTCTATTGTACCCTTGGATTAGCTTAGAATTTGATGACT
AGCGGAAGTACTGGATCGT
C1: CACATATCCGTCAAAGTCTACAGTAGTCTCTAACGTATACTACCGAGTTAAGA
C1p: CGGTAGTATACGTTAGAGACTACTGTAGACTTTGACG
C2: AAATTCTAAGCTAATCCAAGGGTACATTAAGTTATCGCGACTAAGTTCTAAAT
C2p: CTTAGTCGCGATAACTTAATGTACCCTTGGATTAGCT
C3: ATTCATCTGCCGAACAATGTCACCTTTACGATCCAGTACTTCCGCTAGTCATC
C3p: GCGGAAGTACTGGATCGTAAAGGTGACATTGTTCGGC

The following are the halves of the **C** strands that were ligated together to make the reactants for the fluorescence quenching experiment.
C1aF: /5TYE665/CACATATCCGTCAAAGTCTACAGTAG
C1bQ: /5Phos/TCTCTAACGTATACTACCGAGTTAAGA/3IABkFQ/
C1b: /5Phos/TCTCTAACGTATACTACCGAGTTAAGA
C2aF: /56-TAMN/AAATTCTAAGCTAATCCAAGGGTACA
C2bQ: /5Phos/TTAAGTTATCGCGACTAAGTTCTAAAT/3IAbRQSp/
C2b: /5Phos/TTAAGTTATCGCGACTAAGTTCTAAAT
C3aF: /56-FAMK/ATTCATCTGCCGAACAATGTCACCTT
C3bQ: /5Phos/TACGATCCAGTACTTCCGCTAGTCATC/3IAbRQSp/
C3b: /5Phos/TACGATCCAGTACTTCCGCTAGTCATC

References

1. Zhang., D. Y. Towards Domain-Based Sequence Design for DNA Strand Displacement Reactions. *Lect. Notes Comp. Sci.* **2011**, *6518*, 162–175.
2. Dirks, R. M.; Bois, J. S.; Schaeffer, J. M.; Winfree, E.; Pierce, N. A. Thermodynamic Analysis of Interacting Nucleic Acid Strands. *SIAM Rev.* **2007**, *49*, 65–88.
3. Zadeh, J. N.; Steenberg, C. D.; Bois, J. S.; Wolfe, B. R.; Pierce, M. B.; Khan, A. R.; Dirks, R. M.; Pierce, N. A. NUPACK: Analysis and Design of Nucleic Acid Systems. *J. Comput. Chem.* **2011**, *32*, 170–173.
4. Yin, P.; Choi, H. M. T.; Calvert, C. R.; Pierce, N. A. Programming Biomolecular Self-Assembly Pathways. *Nature* **2008**, *451*, 318–322.