## [Electronic Supplementary Information (ESI) to accompany *Nanoscale* manuscript NR-ART-03-2017-001931.R1] The competing effects of core rigidity and linker flexibility in the nanoassembly of trivalent small molecule-DNA hybrids (SMDH<sub>3</sub>s)–a synergistic experimental-modeling study

Vincent Y. Cho,<sup>‡</sup> Bong Jin Hong,<sup>‡</sup> Kevin L. Kohlstedt, George C. Schatz,\* and SonBinh T. Nguyen\*

Department of Chemistry and International Institute for Nanotechnology, Northwestern University, 2145 Sheridan Road, Evanston, Illinois 60208-3113.

## Table of Content

1 avi		
S1.	Materials and instrumentation	SI
S2.	Synthesis of small molecule-DNA hybrids (SMDH <sub>3</sub> s) possessing three DNA arms.	S1
S3.	MALDI-ToF mass spectrometric characterization of SMDH <sub>3</sub>	S2
S4.	Nanoassembly of SMDH <sub>3</sub> s	<i>S6</i>
S5.	Force-field parameters for the coarse-grained (CG) model	S9
S6.	Computational details	S11
S7.	Nanoassembly population and k <sub>0</sub> value	<i>S11</i>
S8.	"Spacer" CG beads in the current pyrSMDH <sub>3</sub> and tpSMDH <sub>3</sub> model versus the non-interacting CG beads in	
	the previous equal-enthalpy (ee) fSMDH <sub>3</sub> model	<i>S12</i>
S9.	Effect of T <sub>n</sub> spacers on assembly populations	S13
S10.	Contribution of different energy terms to the hybridization	S14
S11.	References	S14

## S1. Materials and instrumentation.

Unless otherwise stated, all reagents and reagent-grade solvents were purchased from Sigma-Aldrich (Milwaukee, WI) and Glen Research (Sterling, VA), and used as received. Ultrapure deionized (DI) H<sub>2</sub>O (18.2 MΩ·cm resistivity) was obtained from a Milipore system (Milli-Q Biocel).

DNA syntheses were carried out on an Expedite 8909 Nucleic Acid System. DNA products were purified and analyzed on an Agilent 1100 HPLC equipped with reverse-phase (RP) semi-preparative (Dynamax, 250 × 10 mm, Microsorb 300 Å/10  $\mu$ m/C18, Agilent # R083213C10) and analytical (Dynamax, 100 × 4.6 mm, Microsorb 100 Å/3  $\mu$ m/C18, Agilent # R0080200E3) columns, respectively.

Absorption spectra of DNA materials were recorded on a Varian Cary 300 Bio UV-vis spectrophotometer (Varian, Inc., Palo Alto, CA) using a masked quartz cell (path length = 10 mm, catalog # 29B-O-10-MS, Starna cells Inc., Atascadero, CA).

Matrix-assisted laser desorption/ionization time-of-flight (MALDI-ToF) mass spectrometric data of DNA samples were collected as negative ions using the linear mode on a Bruker AutoFlex III MALDI-ToF mass spectrometer (Bruker Daltonics, Billerica, MA). The instrument was equipped with Smartbeam<sup>™</sup> laser technology operated at 30-40% power with a sampling speed of 10 Hz. One thousand scans were averaged for each mass spectrum. The instrument was operated using the following parameters: ion source voltage 1 = 20 kV, ion source voltage 2 = 18.5 kV, lens voltage = 8.5 kV, linear detector voltage = 0.6 kV, deflection mass = 3000 Da. Data from Agilent HPLC and Bruker MALDI-ToF instruments were processed using MestreNova software version 8.1.1-11591.

### S2. Synthesis of small molecule-DNA hybrids (SMDH<sub>3</sub>s) possessing three DNA arms.

a. Synthesis of tris(4-azidophenyl)methane core 2. This synthesis is modeled after the synthesis of tetrakis(4azidophenyl)methane.<sup>S1</sup> Tris(4-aminophenyl)methane (0.38 g, 1.3 mmol) was dissolved in 2 M aq HCl (25 mL) in a 250 mL round-bottom flask and cooled down to 0 °C. A solution of NaNO<sub>2</sub> (0.32 g, 4.6 mmol) in H<sub>2</sub>O (2.5 mL) was then added drop-wise into the cooled reaction flask with vigorous stirring. The reaction mixture was kept at 0 °C for 30 minutes before being slowly and carefully neutralized with solid CaCO<sub>3</sub> (CAUTION: exothermic reaction!). To this mixture was then added a solution of NaN<sub>3</sub> (0.34 g, 5.2 mmol) in H<sub>2</sub>O (2.5 mL) that has been precooled at 0 °C. The resulting mixture was allowed to stir at 0 °C for an additional 20 min and then filtered. The collected solid was washed with excess H<sub>2</sub>O, airdried, and redissolved in a minimal amount of dichloromethane before further purification by flash chromatography (column dimension =  $3 \text{ cm} \times 20 \text{ cm}$ ), eluting with 33 vol % dichloromethane in hexanes. The desired product was obtained as a purple solid (0.31 g, 64% vield). <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>):  $\delta$  7.05 (d, J = 8.5 Hz, 6H): 6.96 (d, J = 8.5 Hz, 6H): 5.75 (s, 1H). <sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>):  $\delta$  55.1, 119.3, 130.7, 138.6, 140.3. ESI-MS: m/z = 367.14 obsd for M<sup>+</sup>; 367.37 calcd.

b. Synthesis of 1,3,5-tris(4-azidophenyl)benzene core 3. This synthesis is modeled after the synthesis of tetrakis(4azidophenyl)methane.<sup>S1</sup> 1,3,5-Tris(4-aminophenyl) benzene (0.46 g, 1.3 mmol) was dissolved in 2 N aq HCl (25 mL) in a 250 mL round-bottom flask and cooled down to 0 °C. A solution of NaNO<sub>2</sub> (0.32 g, 4.6 mmol) in H<sub>2</sub>O (2.5 mL) was then added drop-wise into the cooled reaction flask with vigorous stirring. The reaction mixture was kept at 0 °C for 30 minutes before being slowly and carefully neutralized with solid CaCO<sub>3</sub> (*CAUTION: exothermic reaction*!). To this mixture was then added a solution of NaN<sub>3</sub> (0.34 g, 5.2 mmol) in H<sub>2</sub>O (2.5 mL) at 0 °C. The resulting mixture was allowed to stir at 0 °C for an additional 20 min and then filtered. The collected solid was washed with excess H<sub>2</sub>O, air-dried, and redissolved in a minimal amount of dichloromethane before further purification by flash chromatography (column dimension = 3 cm × 20 cm), eluting with 25 vol % dichloromethane in hexanes. The desired product was obtained as a purple solid (0.29 g, 52% yield). <sup>1</sup>H NMR (500 MHz, CDCl<sub>3</sub>):  $\delta$  7.69 (s, 3H); 7.68 (d, *J* = 8.5 Hz, 6H); 7.15 (d, *J* = 8.5 Hz, 6H). <sup>13</sup>C NMR (126 MHz, CDCl<sub>3</sub>): 119.7, 124.8, 128.8, 137.8, 139.8, 141.7. ESI-MS: m/z = 429.15 obsd for M<sup>+</sup>; 429.44 calcd.

c. Solid-phase synthesis of pyr-T<sub>n</sub>-SMDH<sub>3</sub> and tp-T<sub>n</sub>-SMDH<sub>3</sub>. In a typical experiment, the as-prepared dry CPG beads containing alkyne-modified DNA (1 µmol) were placed in a 2 mL Eppendorf tube. To the tube were sequentially added solutions of either core 2 or 3 (5-7 µmol) in DMF (0.75 mL), tris(3-hydroxypropyltriazolylmethyl)amine (7 µmol) in DMF (0.2 mL), CuSO<sub>4</sub>·5H<sub>2</sub>O (5 µmol) in DMF (0.05 mL), and L-ascorbic acid (10 µmol) in DMF (0.1 mL). The reaction tube was then filled with dry nitrogen gas before being capped and shaken for 18 h at 25 °C in a Thermomixer R (Eppendorf, Hauppauge, NY) instrument at 1200 rpm (NOTE: the CPG beads should be properly agitated in DMF solution and not allowed to settle at the bottom of the tube during the reaction). The resulting CPG beads were filtered using a oneside fritted 1 µmol Expedite DNA synthesis column (Glen Research, # 20-0021-01), washed with DMF (10 × 1 mL) and acetone ( $10 \times 1$  mL), and dried using a stream of dry nitrogen. The CPG beads containing the products were placed in a vial containing 1 mL of AMA (1:1 v/v 30% ammonium hydroxide solution: methylamine solution; CAUTION: Only freshly made AMA solutions should be used); and the vial was capped and heated at 65 °C for 15 minutes to cleave the SMDHs from the solid supports. The ammonia and methylamine byproducts were then removed by passing a stream of dry nitrogen gas over the content of the vial until the characteristic ammonia smell disappears. The remaining liquid, which contains the crude SMDHs, was collected by pipette and the remaining beads were further extracted with ultrapure deionized water ( $3 \times$  $200 \ \mu$ L). These extracts were combined with the initial solution of crude SMDHs (affording a total volume of 0.8 mL at the end) and filtered through 0.45 μm nylon syringe filter (Acrodisc® 13 mm syringe filter # PN 4426T).

**d. Purification of SMDH**<sub>3</sub>. To identify the different products that were formed in the SMDH preparation, an aliquot of the collected sample of crude SMDHs was first analyzed using an analytical RP-HPLC column (see Section S1) and a gradient method beginning with 95:5 v/v 0.1 M TEAA (aq):MeCN (TEAA (aq) = triethylammonium acetate, aqueous solution), and increasing to 60:40 v/v 0.1 M TEAA(aq):MeCN over 35 minutes (at a ramp of +1 vol% MeCN/minute), with a flow rate of 1 mL/min. Then, the whole sample was subjected to purification using a semi-preparative RP-HPLC column (see Section S1) and a gradient method beginning with 95:5 v/v 0.1 M TEAA (aq):MeCN and increasing to 60:40 v/v 0.1 M TEAA(aq):MeCN over 35 minutes (at a ramp of +0.5 v/v 0.1 M TEAA (aq):MeCN and increasing to 60:40 v/v 0.1 M TEAA(aq):MeCN over 70 minutes (at a ramp of +0.5 vol % MeCN/minute, a slower gradient was employed here to ensure adequate separation of the peaks), with a flow rate of 3 mL/min. The identity of the collected SMDH<sub>3</sub> product was confirmed by MALDI-TOF MS analysis (insets in Fig. S1-S16) and its purity was reassessed using analytical RP-HPLC (Fig. S1-S16) with the aforementioned analytical RP-HPLC solvent program.

## S3. MALDI-ToF mass spectrometric characterization of SMDH<sub>3</sub>

The MALDI-ToF MS matrix was prepared by adding aqueous ammonium hydrogen citrate (0.6  $\mu$ L of a 33.3 wt % solution in water) to a solution of 2-hydoxypicolinic acid (30  $\mu$ L of a 2:15:15 wt/vol/vol mixture of HPA:H<sub>2</sub>O:MeCN). An aliquot (1  $\mu$ L, 10-100 pmol) of isolated SMDH<sub>3</sub> was then mixed with a portion of this matrix (5  $\mu$ L). A small amount (2-3  $\mu$ L) of the resulting mixture was dropped on a steel MALDI-ToF plate and dried at rt before being analyzed. The MALDI-ToF mass spectra of the SMDH<sub>3</sub>s are shown as insets in Fig. S1-S16.

Table S1. List of DNA sequences of SMDH<sub>3</sub> DNA arms used for an experimental study.

Entry	DNA sequences <sup>a</sup>
1	3'-TCC GCC GA-T <sub>n</sub> -core
2	3'- TCG GCG GA-T <sub>n</sub> -core

 $^{a}T_{n}$  represents an oligo  $T_{n}$  spacer between cores and DNA arms, and the spacers are composed of 0, 3, 6, and 15 thymine nucleotides, respectively. Since each single CG bead in the CG model is designed to present three nucleotides, the DNA arms in our CGMD model ended up with one more DNA base in comparison to those in our experimental systems; however, this does not alter the CGMD nanoassembly behaviors of our models. As previously reported,<sup>S2</sup> our CGMD model fully captures the nanoassembly behaviors of the 9-fSMDH<sub>3</sub> experimental system, which is virtually indistinguishable from those for the 8-fSMDH<sub>3</sub>. As such, for the remainder of this manuscript, we will consider the 9-hybridizing-DNA-base arms in the CGMD constructs to be adequate representation of the experimental 8-hybridizing-DNA-base arms.



Fig. S1 Analytical RP-HPLC trace of the pure  $tpSMDH_3$ . The trace is the signal from the diode detector set at 260 nm. Inset shows the MALDI-ToF mass spectrum of the pure product: m/z = 8020 (8018.5 theoretical).



Fig. S2 Analytical RP-HPLC trace of the pure complementary  $tpSMDH_3$ . The trace is the signal from the diode detector set at 260 nm. Inset shows the MALDI-ToF mass spectrum of the pure product: m/z = 8260 (8258.5 theoretical).



Fig. S3 Analytical RP-HPLC trace of the pure pyrSMDH<sub>3</sub>. The trace is the signal from the diode detector set at 260 nm. Inset shows the MALDI-ToF mass spectrum of the pure product: m/z = 7958 (7956.5 theoretical).



Fig. S4 Analytical RP-HPLC trace of the pure complementary pyrSMDH<sub>3</sub>. The trace is the signal from the diode detector set at 260 nm. Inset shows the MALDI-ToF mass spectrum of the pure product: m/z = 8198 (8196.5 theoretical).







Fig. S6 Analytical RP-HPLC trace of the pure complementary tp- $T_3$ -SMDH<sub>3</sub>. The trace is the signal from the diode detector set at 260 nm. Inset shows the MALDI-ToF mass spectrum of the pure product: m/z = 10995 (10996.3 theoretical).



Fig. S7 Analytical RP-HPLC trace of the pure  $pyr-T_3$ -SMDH<sub>3</sub>. The trace is the signal from the diode detector set at 260 nm. Inset shows the MALDI-ToF mass spectrum of the pure product: m/z = 10695 (10694.3 theoretical).



Fig. S8 Analytical RP-HPLC trace of the pure complementary  $pyr-T_3$ -SMDH<sub>3</sub>. The trace is the signal from the diode detector set at 260 nm. Inset shows the MALDI-ToF mass spectrum of the pure product: m/z = 10940 (10934.3 theoretical).



Fig. S9 Analytical RP-HPLC trace of the pure tp- $T_6$ -SMDH<sub>3</sub>. The trace is the signal from the diode detector set at 260 nm. Inset shows the MALDI-ToF mass spectrum of the pure product: m/z = 13470 (13494.1 theoretical).



Fig. S10 Analytical RP-HPLC trace of the pure complementary tp- $T_6$ -SMDH<sub>3</sub>. The trace is the signal from the diode detector set at 260 nm. Inset shows the MALDI-ToF mass spectrum of the pure product: m/z = 13740 (13734.1 theoretical).



Fig. S11 Analytical RP-HPLC trace of the pure  $pyr-T_6$ -SMDH<sub>3</sub>. The trace is the signal from the diode detector set at 260 nm. Inset shows the MALDI-ToF mass spectrum of the pure product: m/z = 13430 (13432.1 theoretical).



Fig. S12 Analytical RP-HPLC trace of the pure complementary  $pyr-T_6$ -SMDH<sub>3</sub>. The trace is the signal from the diode detector set at 260 nm. Inset shows the MALDI-ToF mass spectrum of the pure product: m/z = 13670 (13672.1 theoretical).



Fig. S13 Analytical RP-HPLC trace of the pure  $tp-T_{15}$ -SMDH<sub>3</sub>. The trace is the signal from the diode detector set at 260 nm. Inset shows the MALDI-ToF mass spectrum of the pure product: m/z = 21710 (21707.2 theoretical).



Fig. S14 Analytical RP-HPLC trace of the pure complementary tp- $T_{15}$ -SMDH<sub>3</sub>. The trace is the signal from the diode detector set at 260 nm. Inset shows the MALDI-ToF mass spectrum of the pure product: m/z = 21950 (21947.5 theoretical).



Fig. S15 Analytical RP-HPLC trace of the pure  $pyr-T_{15}$ -SMDH<sub>3</sub> product with a pyramidal core and an oligo  $dT_{15}$  spacer. The trace is the signal from the diode detector set at 260 nm. Inset shows the MALDI-ToF mass spectrum of the pure product: m/z = 21650 (21645.2 theoretical).



Fig. S16 Analytical RP-HPLC trace of the pure complementary pyr- $T_{15}$ -SMDH<sub>3</sub>. The trace is the signal from the diode detector set at 260 nm. Inset shows the MALDI-ToF mass spectrum of the pure product: m/z = 21890 (21885.5 theoretical).

#### S4. Nanoassembly of SMDH<sub>3</sub>s

**a.** Nanoassembly of SMDH<sub>3</sub>s. Following established procedures, <sup>S2-4</sup> equimolar mixtures of the as-prepared SMDH<sub>3</sub>s and its complementary partner in TAMg buffer solution (40 mM Tris, 20 mM acetic acid, and 7.5 mM MgCl<sub>2</sub>; pH = 7.4) were added into 0.5 mL Eppendorf tubes. The resulting solutions were then heated to 90 °C in a heating block (Thermomixer R; Eppendorf, Hauppauge, NY) and kept there for 5 min to remove all initial DNA interactions. The power

to the heating block was then turned off to allow the solution to slowly cool to room temperature over 3 h (for a typical cooling profile of this equipment, please see Fig. S16 in the SI for Yildirim, I.; Eryazici, I.; Nguyen, S. T.; Schatz, G. C. J. *Phys. Chem. B* **2014**, *118*, 2366-2376).

**b.** Native polyacrylamide gel electrophoresis (PAGE). Equal amounts of assembled SMDH<sub>3</sub>s were loaded onto a sheet of native PAGE gel (6 wt %). The gel experiments were carried out in TAMg buffer at 80 V for 4 h during which the gel was kept under 26 °C to prevent de-hybridization of assembled SMDH<sub>3</sub>s. The developed gels were then stained with ethidium bromide (2  $\mu$ g/mL) and their pictures were taken with a Typhoon 9400 scanner (GE Healthcare, Pittsburgh, PA). The acquired gel images were analyzed by ImageJ software (version 1.50e, National Institute of Health, USA).

c. Cryo-scanning transmission electron microscopy (cryo-STEM) imaging. The SMDH<sub>3</sub> assembled mixture (2  $\mu$ L) was mixed with phosphotungstic acid (2  $\mu$ L of a 4 wt % aqueous solution) and an aliquot (3  $\mu$ L) of this solution was then placed on a lacey carbon TEM grid (Ted Pella, Inc., Redding, CA) that was plasma-treated (for 20 sec and generally used within 1 h of treatment) in a PC 2000 Plasma Cleaner (South Bay Technology, San Clemente, CA). After 30 sec, the excess solution was gently wicked away from the grid with a piece of filter paper and the grid was immediately immersed into liquid ethane. Cryo-scanning transmission electron microscopy (STEM) images were acquired using a Hitachi HD-2300A microscope (Hitachi High-Technologies Corp., Tokyo, Japan) operated at 80 and 200 kV.

 Table S2.
 Population distribution of products experimentally obtained from the nanoassemblies of 8-pyrSMDH<sub>3</sub>, 8-tpSMDH<sub>3</sub>, and 8-fSMDH<sub>3</sub> comonomers.

Products	Population $(\%)^a$											
Tioducis -	8-pyrSMDH <sub>3</sub>				8-tpSMDH <sub>3</sub>				8-fSMDH <sub>3</sub>			
	4 μΜ	8 μΜ	16 µM	32 µM	4 μΜ	8 μΜ	16 µM	32 µM	4 μΜ	8 μΜ	16 µM	32 µM
Dimer	29	29	28	30	18	18	17	18	≥95	≥95	≥95	≥95
Tetramer	26	25	23	22	3	3	3	5	0	0	0	0
Hexamer	13	11	8	7	6	8	7	7	0	0	0	0
Octamer	3	4	5	4	25	16	14	10	0	0	0	0
$\geq$ Nonamer	30	32	36	37	48	55	59	60	<5	<5	<5	<5
								9-fSN	MDH <sub>3</sub>			
									4 μΜ	8 μΜ	16 µM	32 µM
Dimer									≥95	≥95	≥95	≥95
Tetramer									0	0	0	0
Hexamer									0	0	0	0
Octamer									0	0	0	0
≥Nonamer									<5	<5	<5	<5

<sup>a</sup>The populations of the dimers and oligomers were calculated from total signal intensity of gel bands in the PAGE gel images (Fig. 2a and 2b for pyrSMDH<sub>3</sub> and tpSMDH<sub>3</sub> in the manuscript, respectively; and Fig. 3a and 4a for 8- and 9-fSMDH<sub>3</sub> in the previous publication<sup>S2</sup>) using ImageJ analysis tool. As discussed in the caption of Fig. 2 in the manuscript, we assume that intermediate-size odd oligomers (trimers, pentamers, etc.) do not form.

**d. On the potential formation of self-pairing cage dimers**. A potential concern in the nanoassembly of SMDH<sub>3</sub> comonomers is the self-pairing of the individual hybrids. For example, 3'- TCG GCG GA-5' is predicted to form a self-paring dimer with  $T_m = 37.1 \,^{\circ}C$  (Fig. S17a) and 3'-TCC GCC GA-5' is predicted to form a self-paring dimer with  $T_m = 4.7 \,^{\circ}C$  (Fig. S17b). However, because both of these are still much less stable than the full-complementary base-pairing configuration between these two DNA sequences ( $T_m = 56 \,^{\circ}C$ , Fig. S17c), they should not be present in any significant amount during our nanoassembly process, which started at 90 °C where all DNA duplexes were dehybridized. In other words, because the mixture of two complementary SMDH<sub>3</sub> comonomers are slowly cooled down to rt from 90 °C, where all the DNA arms started out in single-stranded form, the nanoassembled caged dimers will be favored to form first via full-complementary base-pairing at ~56 °C as any structures formed via the self-pairing interaction would not be stable at this high temperature, a substantial 20 °C above its  $T_m (37 \,^{\circ}C)$ .

The aforementioned assertion is partially supported by the PAGE-gel images shown in Fig. 2, 6 and 9 in the manuscript. The control lanes in these images, where the individual SMDH<sub>3</sub> monomers were separately subjected to the nanoassembly conditions, each shows only one spot corresponding to the respective SMDH<sub>3</sub> monomers. More importantly, no other nanoassemblies except caged dimers could also be observed in the PAGE gel image (Fig. 6) for the assembly of either tp-T<sub>6</sub>-SMDH<sub>3</sub> or pyr-T<sub>6</sub>-SMDH<sub>3</sub>. While the slow time scales of the gel experiment and the associated electrophoretic

#### Cho et al, ESI to accompany Nanoscale manuscript NR-ART-03-2017-001931.R1

forces has usually rendered PAGE-gel impractical for the study of fast nucleation events such as the formation of selfpairing dimers, it can be argued that our nanoassembly protocol, where the components are slowly cooled to rt from a high temperature (90 °C), largely removed this constraint. The full-complementary base-pairing of the DNA arms to form nanoassembled caged dimers (and large-oligomer networks) will be favored before any of the self-pairing dimers and would have consumed the majority of the SMDH<sub>3</sub> monomers. While we cannot exclude the nucleation of large-oligomer networks via self-pairing, the formation of self-pairing cage dimers is probably not a concern in our experiments.



Fig. S17. (a-b) Schematic representations of the most-stable self-pairing configurations for the two DNA sequences used as the SMDH<sub>3</sub> DNA arms in this work. (c) A schematic representation of the full-complementary base-pairing configuration between the two DNA sequences used as the SMDH<sub>3</sub> DNA arms in this work. The relevant thermodynamic data were provided underneath each figure. Simulations were carried out using the MFold nearest-neighbor simulation program.<sup>S5</sup> Simulation conditions: [DNA] = 96  $\mu$ M (to simulate [SMDH<sub>3</sub>] = 32  $\mu$ M), [Mg<sup>2+</sup>] = 7.5 mM, [Na<sup>+</sup>] = 10 mM.<sup>S6</sup>

e. Preference for forming even-numbered nanoassemblies. Since each SMDH<sub>3</sub> comonomer has three DNA arms of the same type, any odd-numbered nanoassemblies (trimer, pentamer, etc.; Fig. S18) will have at least three dangling DNA arms. As fully hybridizing nanoassemblies are energetically more stable than partially hybridizing ones (see Section S10), the partially hybridizing, odd-numbered nanoassemblies are more likely to combine with other assemblies or monomers with unhybridized DNA arms to be transformed into fully hybridizing, even-numbered nanoassemblies (tetramer, hexamer, etc.).



Fig. S18 A schematic illustration of potential even-numbered dimers, tetramers, and hexamers versus odd numbered trimers and pentamers that were observed during the CGMD simulations of the nanoassemblies. As shown in Fig. 2 and 4 in the manuscript, the populations of even-numbered dimers, tetramers, and hexamers are more dominant than those of odd-numbered trimers and pentamers in both experiments and simulations. That the odd-numbered oligomers such as trimer and pentamer showed up in the simulation but were not observed by PAGE-gel analyses (Table S2), can be attributed to the unstable presence of dangling arms. Under the long experimental incubation time (3 h instead of the short (~35 µs) simulation time), most of these species presumably convert to even-numbered nanoassemblies or large-oligomer networks. This difference may also account for the slight excess of large-oligomer networks experimentally observed (vs simulated; see Fig. 5 in the manuscript).

#### S5. Force-field parameters for the coarse-grained (CG) model

In the CG model used herein, DNA oligomers and the core are coarse-grained into spherical beads of various diameters as follows:

- 1) Core 2 and 3 were modeled by four As beads (denoted as "small molecule core" in Fig. 3 in the manuscript) with diameter d = 1 nm. This setup allows us to explore both the pyramidal and trigonal planar geometries of these cores. These are used in addition to the five-As-bead model for the previously described tetrahedral core 1.<sup>S2</sup>
- 2) DNA oligomer backbones and DNA spacers were modeled by As beads with diameter d = 1 nm. Every three DNA oligomer backbones or spacers were grouped into one As bead. Additionally, three Ad beads with diameter d = 2 nm were used as a scaffold to prevent unphysical hybridization and promote directional binding in the hybridizable DNA arms throughout this study. They do not represent the DNA bases and have no DNA sequence information.
- 3) DNA base pairs that form hydrogen bonds were modeled by B and C beads with diameter d = 1 nm. These were chosen such that B and C beads can form hydrogen bonds with each other, but not with themselves. They are essentially As beads that are capable of hydrogen bonding, each represents three DNA bases and when hybridized, their combination (B-C) represents three base pairs. As an improvement to the model employed in the previous study,<sup>S2</sup> the B and C beads now have imposed directionality that prevents face-to-face SMDH<sub>3</sub> dimers to form. Different bead types (B<sub>1</sub>, B<sub>2</sub>, and B<sub>3</sub>, for B; C<sub>1</sub>, C<sub>2</sub>, and C<sub>3</sub> for C) were assigned to each CG bead within a single DNA arm, favoring only hydrogen bonding between specific pairs (B<sub>1</sub>-C<sub>3</sub>, B<sub>2</sub>-C<sub>2</sub>, and B<sub>3</sub>-C<sub>1</sub> pairs). This directionality therefore prohibits unphysical, parallel hybridization arrangement (all DNA hybridization will be between anti-parallel strands).
- 4) To maintain the directionality of the hydrogen bond, smaller F beads (denoted as "flank CG beads" in Fig. 3 in the manuscript) with diameter d = 0.6 nm were used.

**Bond potential**. Two neighboring CG beads form a bond. Two other types of bond were used in this model: (i) the tether bond between the core beads (As-As) and (ii) the link bond between the core and DNA backbone beads (As-Ad). Because the DNA phosphate backbone is regarded as a rigid body, the bond potential among its beads is neglected. The bond potential function is defined as:

$$V_{bond} = \frac{k}{2} (l - l_0)^2$$

with k = 330 ( $\epsilon/\sigma^2$ ), l<sub>0</sub> = 0.84  $\sigma$  for tether bonds, and l<sub>0</sub> = 1.26  $\sigma$  for link bonds with  $\sigma$  = 2 nm setting the length scale.<sup>S7</sup>

**Angle potential**. Three neighboring CG beads form an angle. Two types of angle were used in this model: (i) the tether angle between three core beads (As-As-As) and (ii) the link angle between two core beads and one DNA backbone bead (As-As-Ad). The angle potential function is defined as:

$$V_{angle} = \frac{k_0}{2} (\theta - \theta_0)^2$$

with  $\theta_0 = 1.87$  rad for tether angle in pyramidal cores,  $\theta_0 = 2.09$  rad for tether angle in trigonal planar cores, and  $\theta_0 = \pi$  for link angle. The angle force constant  $k_0$  was systematically varied from 10 to 90 ( $\epsilon$ /rad<sup>2</sup>) for the tether angle, and was set equal to 30  $\epsilon$ /rad<sup>2</sup> for the link angle.

Steric repulsions. Non-bonded interactions of CG beads are modeled by the Weeks-Chandler-Andersen (WCA) potential:

$$V_{WCA} = 4\varepsilon_{ij} \left[ \left( \frac{\sigma_{ij}}{r} \right)^{12} - \left( \frac{\sigma_{ij}}{r} \right)^6 \right] - \varepsilon_{ij}$$

with the cutoff criteria ( $r < 2^{1/6} \sigma$ ), where indices iterate over the aforementioned types of CG beads. Table S3 lists the WCA force parameters.

For the B-C bead types, the cutoff is lengthened to allow for binding and the potential is not shifted.

A cutoff of  $r = 3 \sigma$  was used to model the H-bond attraction of the beads.

As described above, our CGMD model is parametrized to capture the core flexibility and geometry of the SMDH<sub>3</sub> via the angle stiffness  $k_0$  and equilibrium angles, the chemical potential driving force for assembly via the density of monomers, and the length and flexibility of the spacer units on the three DNA arms via the angle potential  $V_{angle}$ . All three of the parameters were varied via a sensitivity-type analysis and importantly can be validated against experimental data. Especially,  $k_0$  was varied by using the population of selected assemblies (dimers, tetramers, and everything beyond hexamers) as the parameter that was to match the experimental values. Other parameters in the CG model are based mainly on empirical findings such as the binding strength of the linkers (based on the  $\Delta G$  of hybridization of free strands), bond strengths based on stiffness of nucleotides, etc.<sup>S7</sup> and were not altered in this study. We note that blindly toggling force-field parameters may lead to unphysical predictions by the CG model. For instance, the hybridization energy of each CG bead cannot be too strong (>>10 k<sub>B</sub>T) or there will be a driving force to form three-body hybridization configurations, a consequence of the fact that hydrogen bond cannot be explicitly modeled as a two-body proton donor and acceptor.<sup>S8</sup>

As the spacers are designed to be non-hybridizable, they were flexible enough (as can be observed from the simulation snapshots) in our simulations that we choose to not optimize their parameters from that described above. However, the bond and angle potential parameters between the spacer beads could potentially be lowered to make the spacers more flexible.

We note in passing that because our simulations were carried out in implicit solvent and all our core and DNA bead potentials are repulsive, our model does not have the capability to capture a hydrophobic collapse driven by the entropic gain by excluding water. Such interactions have also been implied in DNA simulation *in vacuo* in the absence of explicit water where the van der Waals interatomic (or interbead in the CG model) terms get overestimated and mimic hydrophobic interactions.<sup>59</sup> As discussed in the manuscript and detailed in reference 9 in the manuscript, this hydrophobic effect can effectively shorten the  $T_n$  spacer and stabilize dimer formation in experiments more than what is predicted due to the increased confinement. Such a scenario may explain the difference in the experimentally observed ( $T_6$ ) vs predicted ( $T_3$ ) optimal spacer lengths for generating the caged dimers (Fig. 10 in the manuscript).

Bead types	$\varepsilon_{ij} (k_B T)$	σ <sub>ij</sub> (nm)
As - As	1.0	1.0
As - Ad	1.0	1.5
As - B	1.0	1.0
As - C	1.0	1.0
As - F	2.0	0.8
Ad - Ad	2.0	2.0
Ad - B	1.0	1.5
Ad - C	1.0	1.5
Ad - F	2.0	1.3
B - C	6.0	1.0
B - F	1.0	0.8

Table S3. CG force-field parameters, as modified from a previous report.<sup>S7</sup>

Cho et al, ESI to accompany Nanoscale manuscript NR-ART-03-2017-001931.R1



**Fig. S19** A snapshot of the initial CGMD simulation cell for a pyr-T<sub>9</sub>-SMDH<sub>3</sub> assembly. Inset: A magnified view of a CG SMDH<sub>3</sub> molecule.

#### S6. Computational details

The Highly Optimized Object-Oriented Molecular Dynamics (HOOMD)-blue<sup>S8</sup> code was employed for all simulations, which were carried out on the GPUs of either the TARDIS cluster maintained by the Olvera de la Cruz research group or the Polaron cluster maintained by the Schatz research group at Northwestern University. Force-field parameters (see Section S5) were derived from the model proposed by Knorowski et al.,<sup>58</sup> which had been successfully applied to both DNA-Au nanoprism,<sup>S7</sup> and fSMDH systems.<sup>S2</sup> The simulation cell was built using randomized molecular packing program packmol,<sup>S10</sup> where CG constructs representing two complementary SMDH<sub>3</sub> comonomers were randomly put into a cubic box at the target densities of [SMDH<sub>3</sub>] = 16, 24, 32, and 40  $\mu$ M, corresponding to 40, 60, 80, and 100 CG molecules/box, respectively. A box size of L = 20 nm with periodic boundary conditions was used to model the bulk conditions of the experiment. (For an initial simulation cell used for the pyr-T<sub>2</sub>-SMDH<sub>3</sub> assembly, see Fig. S19). The system was equilibrated in a canonical NVT ensemble (i.e., where the number of molecules (N), volume (V), and temperature (T) are conserved) for  $5 \times 10^5$  steps to relax the CG SMDH<sub>3</sub> monomers before the CG beads hybridized. Hybridization and structural data were taken from NVT ensemble runs at a reduced temperature of  $T = 0.6 k_B T$ . At least 3 independent trajectories were run and averaged over the ensemble. Unless otherwise specified, all simulations were run at least for  $2 \times$  $10^8$  time steps (total time ~35 µs). The unitless time step used in all simulation is dt = 0.002 $\tau$ , where  $\tau$  is the characteristic time scale that has units of picoseconds. The salt and water molecules were implicitly parameterized in the force field and were not included in the actual simulations.

#### S7. Nanoassembly population and k<sub>0</sub> value

For the assemblies of comonomers possessing either the pyramidal core **2** or the trigonal planar core **3**, the  $k_0$  force constant of the angle harmonic potential was varied from 10 to 90, as described in Section S5. Fig. S20 below shows the populations of a few dominant species (without any spacers) as a function of  $k_0$  at the SMDH<sub>3</sub> concentration of 32  $\mu$ M. As shown, the population distributions of the dimer and tetramer appear to depend less on  $k_0$  than it does on the core geometry (i.e., angle) and concentration (Fig. 4 in the manuscript), while the larger networks are much more sensitive to the core flexibility. This can be rationalized if we consider that the stability of the small assemblies (dimers, tetramers) is more

dependent by the hybridized DNA arms confinement than by the core flexibility. In other words, the flexibility of the core cannot entirely compensate for the steric repulsion of the linkers during cofacial hybridization to form the smaller assemblies; as a result, their populations are suppressed no matter how stiff is  $k_0$ . In the unphysical limit of  $k_0 = 0$  (i.e., negligible chain stiffness), we do expect the assembled population distributions to be based on the confinement entropy for each of the hybridizing DNA arms in the assemblies.



Fig. S20 Population of selected major species in the assemblies of SMDH<sub>3</sub> comonomers possessing cores 2 (a) and 3 (b) as a function of  $k_0$ . Although the populations of larger networks (" $\geq$  Nonamer" in the plot) for both cores increase along with  $k_0$  over the lower ranges, both trends show significant fluctuations at the higher ranges of  $k_0$  and are far from being monotonic. Notably, the populations of dimer and tetramer are not highly sensitive to  $k_0$ .

# S8. "Spacer" CG beads in the current pyrSMDH<sub>3</sub> and tpSMDH<sub>3</sub> model versus the non-interacting CG beads in the previous equal-enthalpy (ee) fSMDH<sub>3</sub> model

Although we employed "non-interacting" CG beads to represent non-hybridizing segments in the assembly of equalenthalpy (ee) fSMDH<sub>3</sub> comonomer systems in our previous study, <sup>S2</sup> and those beads appear at first hand to be similar to the "spacer" CG beads in the current study, they are not really equivalent. Because these previously used "non-interacting" CG beads were implemented as a part of rod-like DNA arms that can hybridize with their complementary partners, they are rigid and must move together with the hybridizing DNA segments as a whole. In contrast, the "spacer" CG beads used in the current study are independently implemented (i.e., not a part of the hybridizing DNA segments), they can move much more freely and confer more flexibility on the overall SMDH<sub>3</sub> comonomer structure.

We note in passing that while two of the  $T_n$  spacer lengths ( $T_6$  and  $T_{15}$ ) that we employed in this work may appear to be similar to the equal-enthalpy (ee) fSMDH<sub>3</sub> comonomers with 15- and 24-mer DNA arms that we have simulated in our previous study,<sup>S2</sup> the assembly populations are quite different. In addition to the explanation above regarding the difference between the "non-interacting" and "spacer" CG beads in the two respective systems, there is a second, indirect reason for this observation. Although the  $T_n$  spacer CG beads introduce more flexibility to both pyrSMDH<sub>3</sub> and tpSMDH<sub>3</sub> comonomers, they do not affect the intrinsic core flexibility. As a result, these comonomers still have a less flexible core than that of the fSMDH<sub>3</sub> comonomers based on core **1**, and the resulting difference in the assembly formation is not negligible (Fig. 4 in the manuscript).

## **S9.** Effect of T<sub>n</sub> spacers on assembly populations

Fig. S21 below shows the population of various assemblies formed when  $T_n$  spacers of varying length were incorporated between cores 2 (or 3) and the hybridizable DNA segments. Dimers and large-oligomer networks ( $\geq$  nonamers) comprise the two major populations as compared in the manuscript (Fig. 10 in the manuscript).



**Fig. S21** Populations of all assemblies found in the CGMD simulations of pyramidal core **2** (a), and trigonal planar core **3** (b) with  $T_n$  spacers of varying length ( $T_3$ ,  $T_6$ ,  $T_9$ ,  $T_{12}$ , and  $T_{15}$ ), as simulated at [SMDH<sub>3</sub>] = 32  $\mu$ M. Each population was averaged over three repeated simulations. Dimers and large-oligomer networks ( $\geq$  nonamer) are formed as two dominant assemblies, followed by tetramers.



**Fig. S22** Population distribution profiles for dimers, intermediate oligomers (from trimers to octamers), and largeoligomer networks ( $\geq$  nonamers) found in CGMD simulation study for the assembly of pyr-T<sub>n</sub>-SMDH<sub>3</sub> (a) and tp-T<sub>n</sub>-SMDH<sub>3</sub> (b) comonomers pairs (n = 0, 3, 6, 9, 12, and 15). The orange circles, blue rectangles, and gray triangles represent the populations of dimers, intermediate oligomers (from trimers to octamers), and largeoligomer networks ( $\geq$  nonamers), respectively. The correspondingly colored lines were only included as visual guides. The CGMD simulation study was carried out at [SMDH<sub>3</sub>] = 32 µM.

#### Cho et al, ESI to accompany Nanoscale manuscript NR-ART-03-2017-001931.R1

## S10. Contribution of different energy terms to hybridization

The statement that a fully hybridized SMDH<sub>3</sub> has a lower Gibbs free energy ( $\Delta$ G) than a partially hybridized state is quantifiable. While there are multiple contributions that determine  $\Delta$ G, the potential energy of hybridization ( $\Delta$ H) is a balance between energetic gain in hydrogen-bond formation between the DNA arms and the repulsion due to excluded volume (electrostatically based), while the entropic costs of (full) hybridization are losses of the conformational degrees-of-freedom ( $\Delta$ V). In the context of our CG model, the potential energy of hybridization can be logged as an output of the trajectory via the contributions to the internal energy of the simulation cell (Fig. S23) because we do not have a volume change during the reaction. Additionally, we show in Fig. S23 the Lennard-Jones (L-J) interaction that models hydrogenbonding interaction is the dominant contribution to the potential energy and is invariant under changes to the length of the oligonucleotide spacers (T<sub>3</sub>-T<sub>9</sub>). The repulsion due to excluded volume interactions is shown with the bond and angle contributions to potential energy in Fig. S23 and shows a linear increase in energy as the number of spacers increases.

The other quantifiable measure is the free-energy of the population distribution. Weighted histogram analysis methods (WHAM) have been utilized successfully on CG models to calculate changes in entropy, enthalpy, heat capacities of reactions and phase changes.<sup>S11-13</sup> However, that kind of analysis is beyond the scope of the current manuscript since we are not interested in quantifying the thermodynamics of the hybridization of the DNA arms. Nevertheless, the population distributions that we calculated do carry the information of the *relative* thermodynamic stabilities of the assembled structures. In other words, the distributions of the different SMDH<sub>3</sub> populations are directly related to the free-energy relationships between dimers, trimers, etc. via  $\Delta G = -k_B T \ln P_1/P_2$ , where P<sub>1</sub> and P<sub>2</sub> are the properly *normalized* populations. For instance, the ratio between the dimers and trimers of tp-SMDH<sub>3</sub> for d = 40 (taken from Fig. 4b in the manuscript) is ~7.5, leading to  $\Delta G = 1.3$  kcal/mol (or  $\Delta S = 4.0$  cal/K mol). These relative energies give a frame of reference regarding the sizes of the energies that contribute to a particular assembly.



Fig. S23 Plots of the contribution of three different energy terms–namely Lennard-Jones potential, bond harmonic potential, and angle harmonic potential–to hybridization as a function of time steps. To show the trend of these terms across the spacer length, three datasets were imported from the pyr- $T_n$ -SMDH<sub>3</sub> assemblies (n = 3, 6, 9) with all else being equal. The Lennard-Jones potential contributes the most to the hybridization, with absolute values amounting to -4000 in reduced energy units, whereas both bond and angle harmonic potentials contribute less but increase with the number of spacer beads.

#### S11. References

- (S1) Pandey, P.; Farha, O. K.; Spokoyny, A. M.; Mirkin, C. A.; Kanatzidis, M. G.; Hupp, J. T.; Nguyen, S. T. J. Mater. Chem. 2011, 21 (6), 1700-1703.
- (S2) Hong, B. J.; Cho, V. Y.; Bleher, R.; Schatz, G. C.; Nguyen, S. T. J. Am. Chem. Soc. 2015, 137 (41), 13381-13388.
- (S3) Eryazici, I.; Prytkova, T. R.; Schatz, G. C.; Nguyen, S. T. J. Am. Chem. Soc. 2010, 132 (48), 17068-17070.
- (S4) Eryazici, I.; Yildirim, I.; Schatz, G. C.; Nguyen, S. T. J. Am. Chem. Soc. 2012, 134 (17), 7450-7458.
- (S5) http://unafold.rna.albany.edu/?q=mfold, the website of The RNA Institute at University at Albany, State University of New York. Accessed May 6, 2017.

- (S6) While we did not use any NaCl in our experiments, a minimum amount of NaCl (10 mM) is necessary to carry out the simulation.
- (S7) Kohlstedt, K. L.; Olvera de la Cruz, M.; Schatz, G. C. J. Phys. Chem. Lett. 2013, 4 (1), 203-208.
- (S8) Knorowski, C.; Burleigh, S.; Travesset, A. Phys. Rev. Lett. 2011, 106 (21), 215501.
- (S9) Chen, J.; Brooks, C. L. J. Am. Chem. Soc. 2007, 129 (9), 2444-2445.
- (S10) Martínez, L.; Andrade, R.; Birgin, E. G.; Martínez, J. M. J. Comput. Chem. 2009, 30 (13), 2157-2164.
- (S11) Kumar, S.; Rosenberg, J. M.; Bouzida, D.; Swendsen, R. H.; Kollman, P. A. J. Comput. Chem. 1992, 13 (8), 1011-1021.
- (S12) Wong-Ekkabut, J.; Baoukina, S.; Triampo, W.; Tang, I. M.; Tieleman, D. P.; Monticelli, L. *Nat. Nanotechnol.* 2008, *3* (6), 363-368.
- (S13) Knotts, T. A.; Rathore, N.; Schwartz, D. C.; de Pablo, J. J. J. Chem. Phys. 2007, 126 (8), 084901.