Supporting Information

Nano-Fabrication Yields. Hybridization and Click-Fixation of Polycyclic DNA Nano-Assemblies

Erik P. Lundberg[†], Calin Plesa[†], L. Marcus Wilhelmsson[†], Per Lincoln[†], Tom Brown[‡] and Bengt Nordén[†]*

[†]Department of Chemical and Biological Engineering/Physical Chemistry, Chalmers University of Technology, SE-41296 Gothenburg, Sweden.

[‡]School of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, U.K

DNA sequences of oligonucleotides

1	5'	GAC	GCT	AAT	C	N	GAT	GCT	GTG	G	3'
	5'	CCT	CTC	TGG	T						
2	5'	CCA	TAC	ATA	C	N	CCA	CAG	CAT	C	3'
	5'	CAG	CTT	GAG	G						
3	5'	GTA	TGT	ATG	G	N	CAT	CCT	CCT	C	3'
	5'	CTT	GCA	GTA	C						
4	5'	GGC	TCT	ACA	G	N	GAG	GAG	GAT	G	3'
	5'	CGA	AAC	TGG	C						
5	5'	CTG	TAG	AGC	C	N	GAT	ACC	ATC	G	3'
	5'	GTG	TGA	GAA	G						
6	5'	GAT	TAG	CGT	C	N	CGA	TGG	TAT	C	3'
	5'	GTC	CTA	ACG	G						
7	5'	CCG	TTA	GGA	C	N	CTT	GTT	GGC	T	3'
	5'	GGT	GAC	AGT	G						
8	5'	AGA	AGA	AGA	G	N	GTT	ACA	CGC	C	3'
	5'	CAC	TGT	CAC	C						
9	5'	CGT	CGA	ATG	G	N	CAC	GTA	GCA	G	3'
	5'	CTC	TTC	TTC	T						
10	5'	ACC	AGA	GAG	G	N	CAA	TGA	AGG	C	3'
	5'	CCA	TTC	GAC	G						
11	5'	TCT	GAA	CCA	C	N	GGC	GTG	TAA	C	3'
	5'	CTA	CCT	CGG	A						
12	5'	TCC	GAG	GTA	G	N	TCC	GCA	TGT	C	3'
	5'	AGA	CAC	TTC	G						
13	5'	TGG	AGC	AAC	G	N	GAC	ATG	CGG	A	3'
	5'	CGA	ACA	GAT	G						
14	5'	CGT	TGC	TCC	A	N	CTG	CTA	CGT	G	3'
	5'	TCA	TAG	TCG	G						
15	5'	CTA	CCT	CGG	A	N	GCC	TTC	ATT	G	3'
	5'	ACG	AGT	CCA	G						
16	5'	AGT	TGT	CGC	G	N	GAC	ATG	CGG	A	3'
	5'	CCT	CAA	GCT	G						

Figure S1. Sequences of three-way oligonucleotides used in assembly of four-ring structure. The 1,3,5-trisubstituated-benzene node[1] is positioned in the middle (N).

Additional evidence of formation of DNA-phenalene

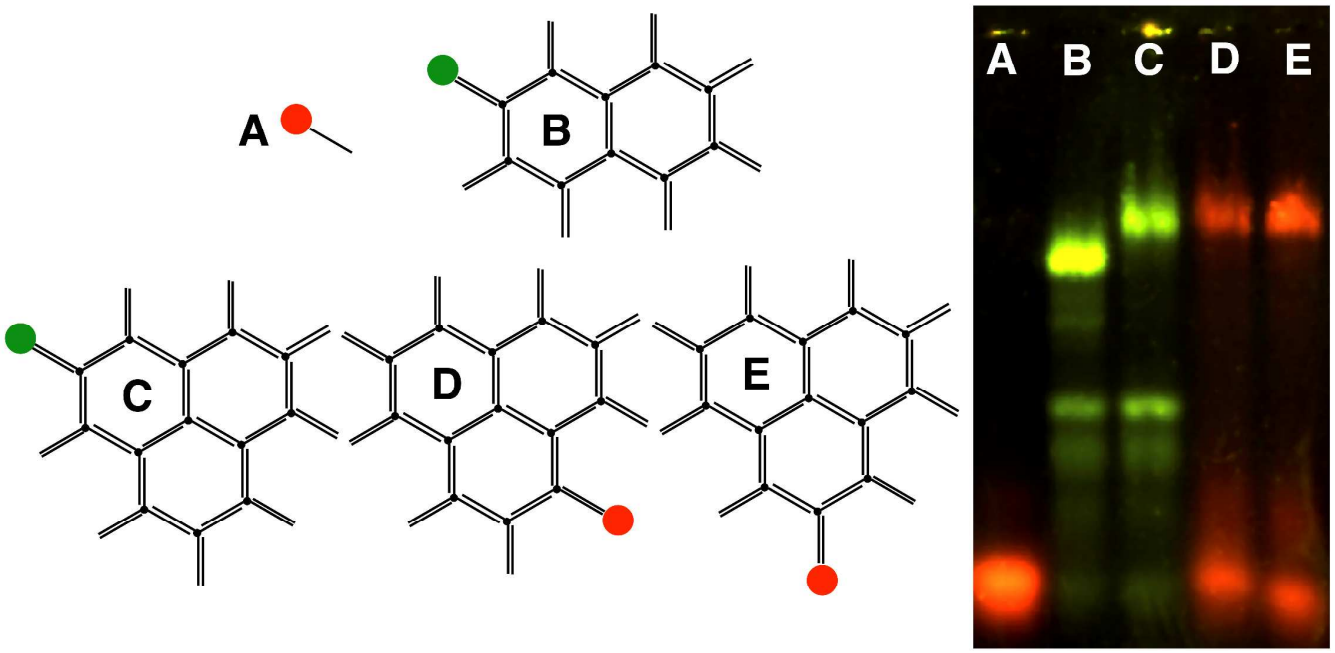
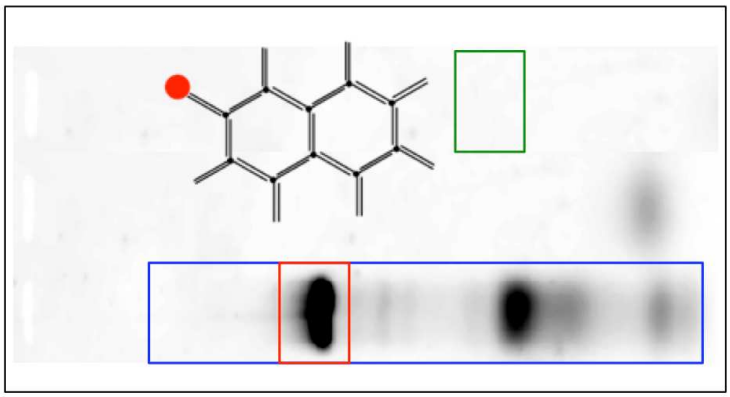


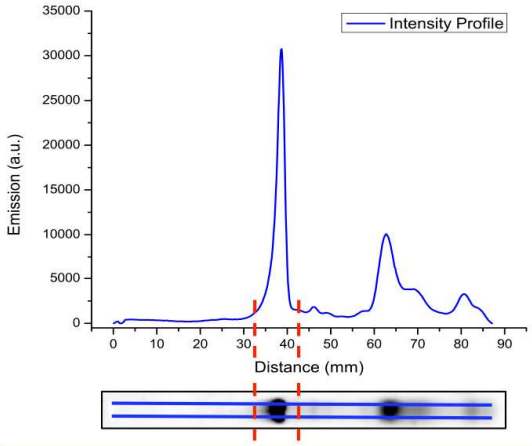
Figure S2. Three-cell construct DNA-phenalene with fluorophore label following the schematic to the left. Analysis was made using 4.5% MetaPhor agarose gel electrophoresis. Red and green dots indicate the position of the ROX and Cy3 label, respectively, used for visualization in gel scanner (Typhoon 9410, GE Healthcare). The two-color scan was done using excitation at 532 nm (SYAG laser) and with a 580 nm and 610 nm bandpass filters (580 BP 30 & 610 BP 30) to collect the emission for Cy3 and ROX, respectively.

Because the stepwise formation of DNA-phenalene presented in Figure 3 of the main text lacked unambiguous evidence that the second node in sample E was bound to the structure, the additional experiment presented in Figure S2 was conducted. A secondary fluorophore label (ROX) was positioned on the DNA-phenalene structure as indicated by the red dots (sample D and E in Figure S2). The main band in for sample C, D and E have the same migration in the gel, thus proving that the missing node in Figure 3 of main text, indeed is bound to the structure when all nodes are present. This means that DNA-phenalene is formed.

Yield Analysis of Gel Electrophoresis Results



Box	Ave. Em.	Area (a.u.)	Int. Em.	Corr. Em.	Yield
DNA-Naphthalene (red)	8551	5676	48535476	38948712	41.7%
Total (blue)	4025	39996	160983900	93430656	
Background (green)	1689	5676	9586764	0	



Profile Peak	Integral (Background corrected)	Yield
DNA-Naphthalene (red)	24844075	42.4%
Total	58550861	

Figure S3. Illustration of yield analysis using emission data from gel electrophoresis results. Emission was collected using a Typhoon 9410 gelscanner (GE Healthcare). Cy3 (red dot) is the fluorophore used; excited at 532 nm (SYAG laser) and with a 580 nm bandpass filter (580 BP 30) to collect the emission. The analysis is made on a DNA-naphthalene sample. (Left) Analysis based on integration of intensity volume. (Right) Analysis based on integration of intensity profile.

The yield data presented in Figure 5 of the main text is based on emission analysis from gel electrophoresis results. As illustrated in Figure S3 (left), the integrated emission from the desired band (red box) is compared with the total intensity of the lane (blue box), with correction being made for the background emission (green box). To strengthen the estimated yield results a secondary analysis method was used on the same sample. By plotting the intensity profile of a lane, as shown in Figure S3, it is possible to integrate the desired peak and compare it with the total integral to get a second estimation of the yield. The example in Figure S3 is a DNA-naphthalene structure and the analysis results in an estimated yield of 42% for the desired construct, using both analysis techniques, strengthening the reliability of the analysis. To obtain statistics, the same analysis was performed on multiple experiments carried out under the same conditions, resulting in the data presented in Figure 5 of the main text.

Cy3 is used as fluorescent probe in all cases and it is attached covalently on a 10mer oligonucleotide, positioned as illustrated in the figure (red dot). The main advantage of using a covalently attached fluorophore, compared to post-staining using dyes such as SybrGold or EtBr, is the increase in detection sensitivity and band resolution. Furthermore, using dyes such as EtBr would render an uneven selection between ssDNA and dsDNA. There might be concerns about the use of only one fluorophore to visualize the DNA structures. Since the Cy3-label only sits at one position, all possible substructures will not be visualized. However, regardless of what substructure is formed, the decrease of the desired structure will be the same (less intense band on the gel). Meaning that the visualized distribution of substructures may not be the correct one but the fraction of the desired structure is correct. Since the yield of the desired structure is that of interest, the analysis is valid. Another concern may be regarding the stoichiometric ratio of the Cy3-labelled oligonucleotide. If too much is added it would appear as a strong band of excess labelled oligonucleotide, which would result in underestimation of the desired structure. This does not seem to be a problem in any of the gels presented in the manuscript and can thus be disregarded. If too little of the Cy3-labelled oligonucleotide is added, on the other hand, less DNA would be seen but there are no reasons to believe that this would result in another distribution then the "true" one. A significant bias between structures in this aspect is unlikely. Even though the total intensity would be lower, the distribution would be the same, resulting in no influence on the analysis.

Detailed Analysis of Stepwise Formation of DNA-Anthracene

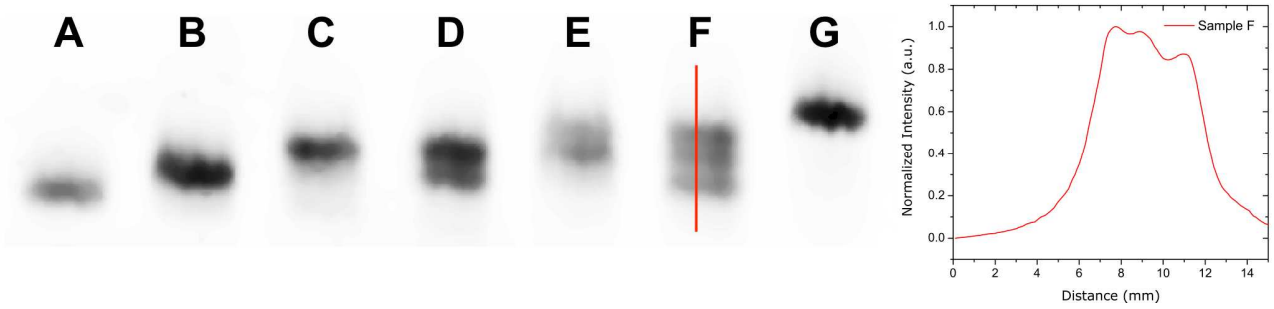


Figure S4. (Left) Zoom of bands of interest from Figure 2 in the main text. Red line gives the intensity profile of sample F (Right). Three strongly overlapping bands can be seen, corresponding to binding of one, two and three nodes to the two-cell construct.

Thermodynamic Effect of Ring-Closure

Following the treatment in Ref. 48 (main text), but with a slight change of notation, let x denote the cyclic oligomer formed by joining of n subunits A, B, ... , W. The equilibrium of ring formation is:

$$A + B + \mathsf{K} + W \leftrightarrow x \qquad \text{(S0)}$$

Each subunit can form one bond to a unique subunit to the left and a further bond to the right, *e.g.* W-A and A-B. The bonds between the subunits are in thermodynamic equilibrium governed by the thermodynamic binding constants K_i:

$$K_a = \frac{[A-B]}{[A-][-B]}, K_b = \frac{[B-C]}{[B-][-C]}, \mathsf{L}\ , K_w = \frac{[W-A]}{[W-][-A]} \qquad \text{(S1)}$$

where [A-] denotes the concentration of A not in x, in which the right, B-binding part is unbound, the status of the left part being otherwise unspecified, and [-A] denotes the concentration of A, not in x, in which the left, W-binding, part is unbound; [A-B] is the concentration of A-B bonds not in the cyclic oligomer x, and [A] is the total concentration of subunit A. Thus

$$[A-] = [A] - [A-B] - [x]$$

$$[-A] = [A] - [W-A] - [x] \quad \text{(S2)}$$

The concentration of free subunits is the concentration of subunit with the right part free, times the fraction of the same subunit (not in x) which has the left part free:

$$[-A-] = \frac{[A-][-A]}{[A] - [x]} \quad \text{(S3)}$$

In terms of the individual equilibrium constants, the concentration of the cyclic oligomer x is:

$$[x] = [-A-]K_a[-B-]K_b \text{L} \quad [-W-]K_w K_x \quad \text{(S4)}$$

where the ring closure equilibrium constant K_x is to be interpreted as the "effective" concentration of one end of the linear n-mer relative to the other end.

Given [x], the concentration of [A-B] is obtained by insertion of S2 into S1 and solving the quadratic, where *a* and *b* denote [A]-[x] and [B]-[x], respectively, and k_a denote K_a^{-1}:

$$[A-B] = \frac{1}{2}\left(a + b + k_a - \sqrt{\left(a + b + k_a\right)^2 - 4ab}\right) \quad \text{(S5)}$$

Expanding the square, and neglecting k_a^2 in the limit of k_a approaching zero, the expression can be rewritten as

$$[A-B] = \frac{1}{2}\left(a + b + k_a - \sqrt{2k_a(a+b) + (a-b)^2}\right) = \frac{1}{2}\left(a + b - |a-b|\right)\left(1 - \frac{k_a}{|a-b|}\right) \quad \text{(S6)}$$

if $a>b$, then

$$[A-B] = b\left(1 - \frac{k_a}{a-b}\right) \quad \text{(S7a)}$$

and if $a<b$, then

$$[A-B]=a\left(1-\frac{k_a}{b-a}\right) \tag{S7b}$$

insertion of S2 into S3, and rearranging to pair together factors with common bond, gives

$$[x]=\frac{(a-[A-B])(b-[A-B])}{a}K_a\frac{(b-[B-C])(c-[B-C])}{b}K_b\cdots\frac{(w-[W-A])(a-[W-A])}{w}K_wK_x \tag{S8}$$

In the limit of k_a approaching zero, insertion of S7 into S8 gives

$$[x]=\left(\frac{\min(a,b)}{a}\right)\left(\frac{\min(b,c)}{b}\right)\cdots\left(\frac{\min(w,a)}{w}\right)K_x \tag{S9}$$

Let

$$[A]=m(1+r_1),\ [B]=m(1+r_2),\cdots,[W]=m(1+r_n)$$

where m is the average concentration in the sample, thus $\Sigma r_i=0$. The relative concentration deviations r_i are assumed to belong to a normal distribution with standard deviation σ. By dividing by m, equation S9 gives an expression for the yield of x (Y=x/m):

$$Y=1-\alpha=\kappa\left(\frac{\left(1+\frac{\min(r_1,r_2)}{\alpha}\right)}{\left(1+\frac{r_1}{\alpha}\right)}\right)\left(\frac{\left(1+\frac{\min(r_2,r_3)}{\alpha}\right)}{\left(1+\frac{r_2}{\alpha}\right)}\right)\cdots\left(\frac{\left(1+\frac{\min(r_n,r_1)}{\alpha}\right)}{\left(1+\frac{r_n}{\alpha}\right)}\right) \tag{S10}$$

where the dimensionless ring closure equilibrium constant $\kappa = K_x/m$. By taking logarithms on both sides, expressing the logarithms as power series and collecting like powers of the r_i/α one get the following series of sums:

$$\ln(1-\alpha)-\ln(\kappa)=\frac{1}{\alpha}\sum\left(\min(r_i,r_{i+1})-r_i\right)-\frac{1}{2\alpha^2}\sum\left(\min(r_i,r_{i+1})^2-r_i^2\right)+\frac{1}{3\alpha^3}\sum\left(\min(r_i,r_{i+1})^3-r_i^3\right)+\cdots \tag{S11}$$

Rearranging the terms in the sums allow them to be evaluated as means, *e.g.*:

$$\sum\left(\min\left(r_i,r_{i+1}\right)-r_i\right)=\sum\left(\min\left(r_i,r_{i+1}\right)-\frac{1}{2}\left(r_i+r_{i+1}\right)\right)=-\frac{1}{2}\sum\left|r_i-r_{i+1}\right|=-\frac{1}{2}n\left\langle\left|r_i-r_{i+1}\right|\right\rangle$$
$$\sum\left(\min\left(r_i,r_{i+1}\right)^2-r_i^2\right)=\sum\left(\min\left(r_i,r_{i+1}\right)^2-\frac{1}{2}\left(r_i^2+r_{i+1}^2\right)\right)=0 \qquad \text{(S12)}$$
$$\sum\left(\min\left(r_i,r_{i+1}\right)^3-r_i^3\right)=\sum\left(\min\left(r_i,r_{i+1}\right)^3-\frac{1}{2}\left(r_i^3+r_{i+1}^3\right)\right)=-\frac{1}{2}\sum\left|r_i^3-r_{i+1}^3\right|=-\frac{1}{2}n\left\langle\left|r_i^3-r_{i+1}^3\right|\right\rangle$$
K

where index n+1 = 1 and the averaging is done over indices i, =1,2,..n and j=i+1.

The average means of the absolute values (i.e. averages over all indices i, j=1,2,3...n) can readily be calculated for a normal distribution of deviations r_i by transformation to polar coordinates $r_i = t\cos(u)$ and $r_j = t\sin(u)$ and integrating between $\pi/4$ and $5\pi/4$, the interval where $\sin(u) > \cos(u)$:

$$\left\langle\left|r_i^{2k+1}-r_j^{2k+1}\right|\right\rangle=\frac{1}{2\sigma^2\pi}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\left|r_i^{2k+1}-r_j^{2k+1}\right|e^{-\frac{r_i^2+r_j^2}{2\sigma^2}}dr_i dr_j$$
$$=2\int_{\pi/4}^{5\pi/4}\left(\sin^{2k+1}(u)-\cos^{2k+1}(u)\right)du\,\frac{1}{2\sigma^2\pi}\int_0^{\infty}t^{2k+2}e^{-\frac{t^2}{2\sigma^2}}dt \qquad \text{(S13)}$$

an expression which, however, will be exact only in the limit of n, the number of subunits, going to infinity. For finite systems, the actual means $\langle| r_i^{2k+1} - r_j^{2k+1}|\rangle$ for a sample with a certain set of deviations r_i will be progressively less well approximated by the average means of S13 as k increases, and the series in S11 is better truncated to the first 3 non-zero terms.

Evaluating the standard integrals of S13 for k=0, 1 and 2, insertion into S11 and taking the exponential gives S14 (Eq.4 in main text).

$$Y=1-\alpha=\kappa e^{-\frac{n}{\sqrt{\pi}}\left(\frac{\sigma}{\alpha}+\frac{5}{6}\left(\frac{\sigma}{\alpha}\right)^3+\frac{43}{20}\left(\frac{\sigma}{\alpha}\right)^5\right)} \qquad \text{(S14)}$$

Taking κ to be unity, Eq.4 simplifies to Eq.3 for α large enough that the higher order terms can be neglected.

Finally, k=0 in S13, gives

$$\left\langle \left| r_i - r_j \right| \right\rangle = \langle \Delta \rangle = \frac{2\sigma}{\sqrt{\pi}} \tag{S15}$$

as in Eq. 2 (in main text).

Analysis of Binomial Distributions

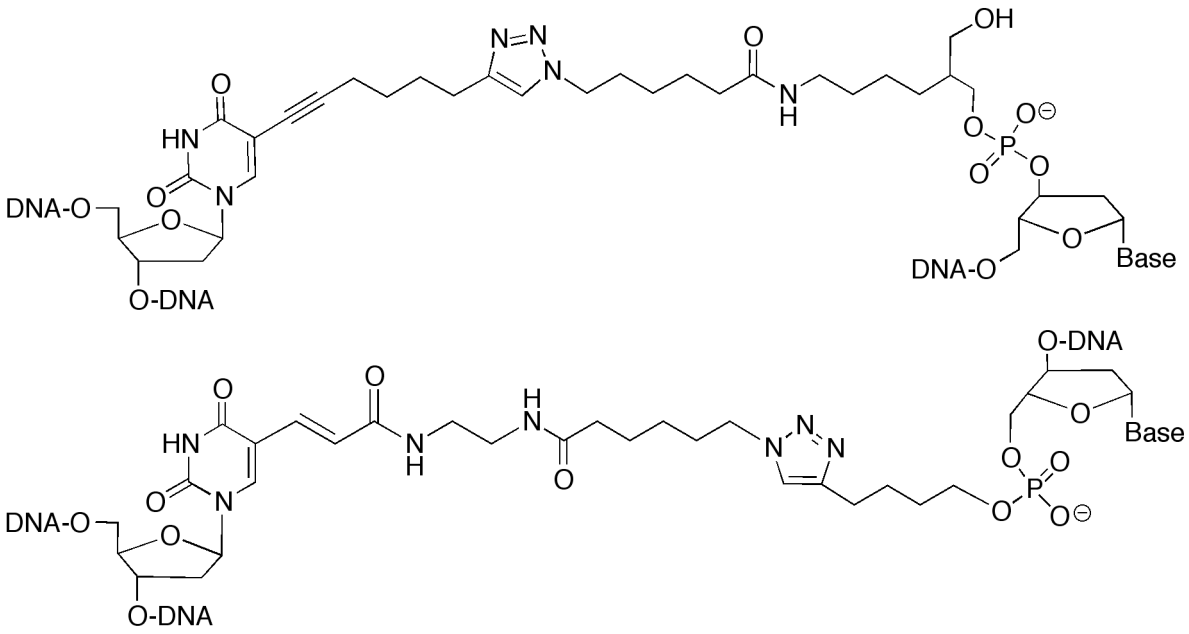
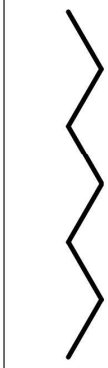
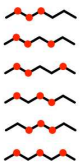
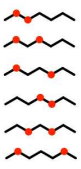
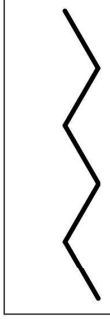
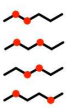
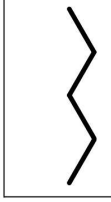


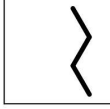
Figure S5. The two resulting triazole cross-links of the site-specific click reactions, from an unpaired thymine to either 3' or 5' phosphate of a complementary oligonucleotide. The relatively short linkers ensure reaction specificity.

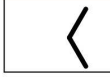
Table S1. Overview of possible combinations of click reactions (red dots) that can take place on different starting material and what substructures result from a specific combination.

Start	No. of Click	Position	Degeneracy	Total No.	Substructures
	5		1	1	L(6)
	4		2 2 1	5	5, 1 4, 2 3, 3
	3		2 2 2 2 1 1	10	4, 1, 1 3, 2, 1 3, 2, 1 3, 2, 1 2, 2, 2 4, 1, 1
	2		2 2 2 2 1 1	10	3, 1, 1, 1 2, 2, 1, 1 2, 2, 1, 1 3, 1, 1, 1 2, 2, 1, 1 2, 2, 1, 1
	1		2 2 1	5	2, 1, 1, 1, 1 2, 1, 1, 1, 1 2, 1, 1, 1, 1
	0		1	1	1, 1, 1, 1, 1. 1

Start	No. of Click	Position	Degeneracy	Total No.	Substructures
	4		1	1	5
	3		2 2	4	4, 1 3, 2
	2		2 2 1 1	6	3, 1, 1 2, 2, 1 3, 1, 1 2, 2, 1
	1		2 2	4	2, 1, 1, 1 2, 1, 1, 1
	0		1	1	1, 1, 1, 1, 1

Start	No. of Click	Position	Degeneracy	Total No.	Substructures
	3		1	1	4
	2		2 1	3	3, 1 2, 2
	1		2 1	3	2, 1, 1 2, 1, 1
	0		1	1	1, 1, 1, 1

Start	No. of Click	Position	Degeneracy	Total No.	Substructures
	2		1	1	3
	1		2	2	2, 1
	0		1	1	1, 1, 1

Start	No. of Click	Position	Degeneracy	Total No.	Substructures
	1		1	1	2
	0		1	1	1, 1

Start	No. of Click	Position	Degeneracy	Total No.	Substructures
	0		1	1	1

Table S2. Summary of the contribution of binomial terms to the different substructures.

Structure	Binomial Terms	Values
	$\frac{6}{6}\left(\frac{1}{1}Q(5)\right)$	0.3791
	$\frac{5}{6}\left(\frac{2}{5}Q(4)\right)$	0.1353
	$\frac{4}{6}\left(\frac{2}{5}Q(4)+\frac{3}{10}Q(3)\right)$	0.1430
	$\frac{3}{6}\left(\frac{2}{5}Q(4)+\frac{6}{10}Q(3)+\frac{4}{10}Q(2)\right)$	0.1407
	$\frac{2}{6}\left(\frac{2}{5}Q(4)+\frac{9}{10}Q(3)+\frac{12}{10}Q(2)+\frac{5}{5}Q(1)\right)$	0.1224
	$\frac{1}{6}\left(\frac{2}{5}Q(4)+\frac{12}{10}Q(3)+\frac{24}{10}Q(2)+\frac{20}{5}Q(1)+\frac{6}{1}Q(0)\right)$	0.0795

Structure	Binomial Terms	Values
	$\frac{5}{5}\left(\frac{1}{1}R(4)\right)$	0.4602
	$\frac{4}{5}\left(\frac{2}{4}R(4)\right)$	0.1577
	$\frac{3}{5}\left(\frac{2}{4}R(3)+\frac{3}{6}R(2)\right)$	0.1562
	$\frac{2}{5}\left(\frac{2}{4}R(3)+\frac{6}{6}R(2)+\frac{4}{4}R(1)\right)$	0.1367
	$\frac{1}{5}\left(\frac{2}{4}R(3)+\frac{9}{6}R(2)+\frac{12}{4}R(1)+\frac{5}{1}R(0)\right)$	0.0892

Structure	Binomial Terms	Values
	$\frac{4}{4}\left(\frac{1}{1}S(3)\right)$	0.5588
	$\frac{3}{4}\left(\frac{2}{3}S(3)\right)$	0.1795
	$\frac{2}{4}\left(\frac{2}{3}S(2)+\frac{3}{3}S(1)\right)$	0.1581
	$\frac{1}{4}\left(\frac{2}{3}S(2)+\frac{6}{3}S(1)+\frac{4}{1}S(0)\right)$	0.1037

Structure	Binomial Terms	Values
	$\frac{3}{3}\left(\frac{1}{1}T(2)\right)$	0.6784
	$\frac{2}{3}\left(\frac{2}{2}T(1)\right)$	0.1937
	$\frac{1}{3}\left(\frac{2}{2}T(1)+\frac{3}{1}T(0)\right)$	0.1279

Structure	Binomial Terms	Values
	$\frac{2}{2}\left(\frac{1}{1}U(1)\right)$	0.8237
	$\frac{1}{2}\left(\frac{2}{2}U(0)\right)$	0.1763

Structure	Binomial Terms	Values
	$\frac{1}{1}V(0)$	1.0000

Table S3. Summary of binomial distributions with different starting point.

n	Function	k = n-1						
6	P(k)	0.3122	0.4011	0.0716	0.0695	0.0633	0.0512	0.0311
5	Q(k)	-	0.3791	0.1353	0.1430	0.1407	0.1224	0.0795
4	R(k)	-	-	0.4602	0.1577	0.1562	0.1367	0.0892
3	S(k)	-	-	-	0.5588	0.1795	0.1581	0.1037
2	T(k)	-	-	-	-	0.6784	0.1937	0.1279
1	U(k)	-	-	-	-	-	0.8237	0.1763
0	V(k)	-	-	-	-	-	-	1.0000

Table S4. Hybridization yield data for different sub-structures in click-fixation system.

Structure	Hybridization Yield
Hexagon	0.2693
6mer	0.0973
5mer	0.1286
4mer	0.1050
3mer	0.1051
2mer	0.0806
Monomer	0.2141

Footnote: Data originating from Lundberg *et al* 2010[2].

References

1. Tumpane, J.; Sandin, P.; Kumar, R.; Powers, V. E. C.; Lundberg, E. P.; Gale, N.; Baglioni, P.; Lehn, J. M.; Albinsson, B.; Lincoln, P.; Wilhelmsson, L. M.; Brown, T.; Nordén, B., Addressable High-information-density DNA Nanostructures. *Chem. Phys. Lett.* **2007,** *440*, 125-129.

2. Lundberg, E. P.; El-Sagheer, A. H.; Kocalka, P.; Wilhelmsson, L. M.; Brown, T.; Norden, B., A New Fixation Strategy for Addressable Nano-network Building Blocks. *Chem. Commun.* **2010,** *46*, 3714-3716.