



This document is the Accepted Manuscript version of a Published Work that appeared in final form in *ACS Nano*, © American Chemical Society after peer review and technical editing by the publisher. To access the final edited and published work see <http://dx.doi.org/10.1021/nn202568g>

(Article begins on next page)

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[†]*

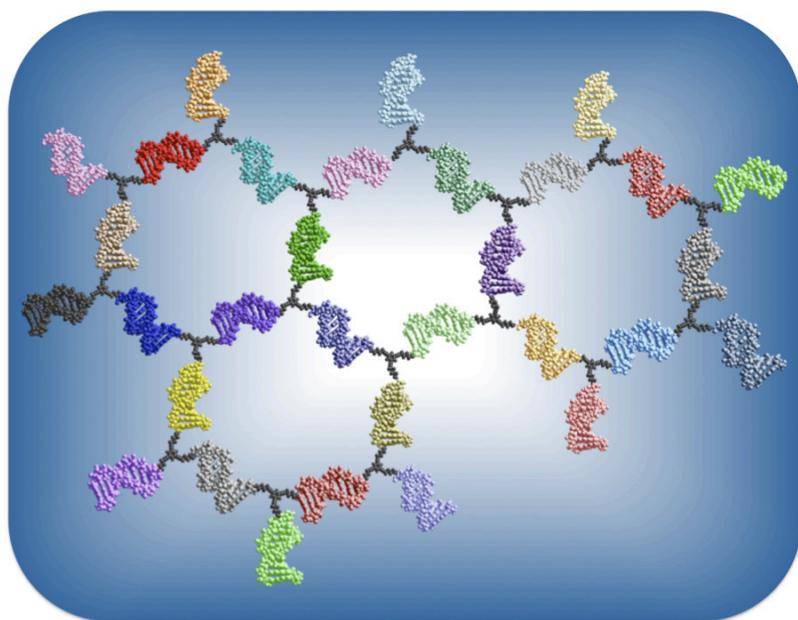
[†]Department of Chemical and Biological Engineering/Physical Chemistry, Chalmers University of Technology, SE-41296 Gothenburg, Sweden.

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

Email: erik.p.lundberg@chalmers.se

RECEIVED DATE (to be automatically inserted after your manuscript is accepted if required according to the journal that you are submitting your paper to)

*Corresponding author: (E.P.L.) Phone: +46 (0)31 772 3053, Fax: +46 (0)31 772 3858



Abstract

We demonstrate the stepwise assembly of a fully addressable polycyclic DNA hexagon nano-network for the preparation of a four-ring system, one of the biggest networks yet constructed from tripodal building blocks. We find that the yield exhibits a distinct upper level <100%, a fundamental problem of thermodynamic DNA assembly that appears to have been overlooked in the DNA nanotechnology literature. A simplistic model based on a single step-yield parameter y can quantitatively describe the total yield of DNA assemblies in one-pot reactions as $Y=y_{\text{duplex}}^n$ with n the number of hybridization steps. Experimental errors introducing deviations from perfect stoichiometry and the thermodynamics of hybridization equilibria contribute to decreasing the value of y_{duplex} (on average $y=0.96$ for our 10-basepair hybridization). For the four-ring system ($n=31$) the total yield is thus less than 30%, which is clearly unsatisfactory if bigger nano-constructs of this class are to be designed. Therefore, we introduced site-specific click chemistry for making and purifying robust building blocks for future modular constructs of larger assemblies. Although the present yield of this robust module was only about 10%, it demonstrates a first step toward a general fabrication approach. Interestingly, we find that the click yields follow quantitatively a binomial distribution, the predictability of which indicates the usefulness of preparing pools of pure and robust building blocks in this way. The binomial behavior indicates that there is no interference between the six simultaneous click reactions but that step-yield limiting factors such as topological constraints and Cu(I) catalyst concentration are local and independent.

Keywords: DNA Nanostructure, Click Chemistry, Supramolecular Assembly, Yield Analysis, Fixation Technology

DNA has become a reliable material for nanoscale construction, a technological application well outside the realms of biology and its natural use as information storage of the genetic code. In his pioneering work Seeman has demonstrated that the unique structural features of DNA may be utilized to fabricate complex constructs by sheer self-assembly on the nanometer scale, proving the potential of applying a molecular perspective to nanoscience with a bottom-up approach. With the development of different robust crossover motifs, oligonucleotides may be assembled into remarkable crystal-like structures, demonstrating that it is possible to create large-scale nanostructures with small DNA building blocks.¹⁻⁴ Since then, various strategies in DNA nanotechnology have emerged and a large variety of geometric shapes have been created using DNA molecules, in 3D as well as in 2D.⁵⁻¹¹ One strategy that has become successful for nanoscale fabrication is “DNA origami”.¹²⁻²² This approach involves the controlled folding of a long viral DNA molecule into a desired structure. The folding process is regulated by numerous short oligonucleotides, complementary to specific parts of the genomic sequence. Adopting this strategy it has been possible to fold the several kilobase long single-stranded DNA molecule of M13 bacteriophage into various shapes. After initial folding into patterns in 2D,¹³⁻¹⁵ researchers have recently been able to create 3D shapes using the origami strategy.^{12, 16-21} The field of DNA nanotechnology has not only seen the advance of architecture strategies for static nanoscale construction but also the development of nano-mechanical devices involving dynamic processes, exploring the possibility of creating DNA machines.²³⁻³¹ There are several examples where structural shifts of a DNA construct have been utilized to repetitively switch a device between geometrically different states, a process driven by *e.g.* an ionic gradient²⁵ or the free energy of hybridization, using oligonucleotides added as “fuel”.^{23, 26} The repetitive motion is based on back-extraction of the fuel oligonucleotides by addition of complementary oligonucleotides forming more stable duplexes.^{27, 28} In this way contractile force walking devices have been constructed, moving unidirectionally along a track.²⁹⁻³¹

Evidently, there is great potential in using DNA as a material for programmable nanoscale fabrication. Using crystal-like arrays of crossover motifs it is possible to create quite large structures from a bottom-up perspective. However, one feature of biological DNA is lost in these artificial repetitive structures, *viz.* the inherently information-rich high complexity of the polymer, a feature that could be important in certain applications. One can envision future applications requiring a spatial precision that is not achievable with repetitive crystal-like assemblies, and in which each position in the assembly is unique and may be addressed individually. The concept of an addressable grid has been demonstrated before in various contexts, *e.g.* with the construction of a DNA-tile based “molecular pegboard”^{32, 33} or more recently using the DNA origami approach.³⁴ With an alternative fabrication path, working with smaller subunits it would, in principle be possible to create structures with similarly distinct features but with higher variability and resolution on a truly molecular scale, compared with the origami strategy which is restricted by the several kilo-bases large genomic template sequence. Our strategy is to assemble a fully addressable DNA network in two dimensions with the smallest practical unit of DNA, *i.e.* one turn of the helix, as a basis for the assembly. By using a small synthetic node with D_{3h} -symmetry (Figure 1, bottom) and orthogonal protection-group chemistry we have previously reported the assembly of a cyclic hexamer, “DNA-benzene”³⁵ and a bicyclic construct, “DNA-naphthalene”,³⁶ where each edge has a unique sequence and, thus, can be specifically addressed. The addressability may allow specific functionalisation patterns with precise positioning of *e.g.* nanoparticles³⁷⁻⁴⁰ or chemical reaction centers.^{34, 41}

Here we present the formation of an unsymmetrical four-ring network assembled in a one-step annealing process (Figure 1, top). As in our previous reports, each side in the hexagonal network constitutes only 10 nucleotides, corresponding to a side-length of 3.4 nm assuming B-DNA helical-structure. The C_4 -linker, connecting the oligonucleotides to the benzene node should create sufficient flexibility to avoid possible distortions due to the 10.5 bp helical pitch. Each side in the network has a unique sequence, giving specific addresses over the entire structure. This is achieved by designing 10 base sequences that are complementary to only one other sequence in the system and orthogonal to all

others. The orthogonality condition is here defined as a mismatch criterion of ≥ 4 (for any strand alignment) between two non-complementary 10 base sequences. There are two sides in the present construct that are not unique. One three-way branched oligonucleotide is deliberately used twice in the assembly, the reason being that we want to investigate assembly robustness and the possibility to incorporate generic nodes in the network and still maintain structural integrity. Finally we considered the yields obtained for the primary hybridization steps, and of the total assembled structure, as well as that of the product from a click-fixation technology,⁴² from a theoretical point of view with the objective of paving the way for alternative approaches to conventional one-step assembly of DNA nanostructures. We find that the yields of these types of reaction may be described in simple statistical terms. Models explaining the fundamental behavior of nano-systems based on DNA self-assembly have so far been absent in the literature. Thus, answering questions concerning the thermodynamic aspects of DNA-nanotechnology is of vital importance for efficient and reliable fabrication of intelligent nanomaterials.

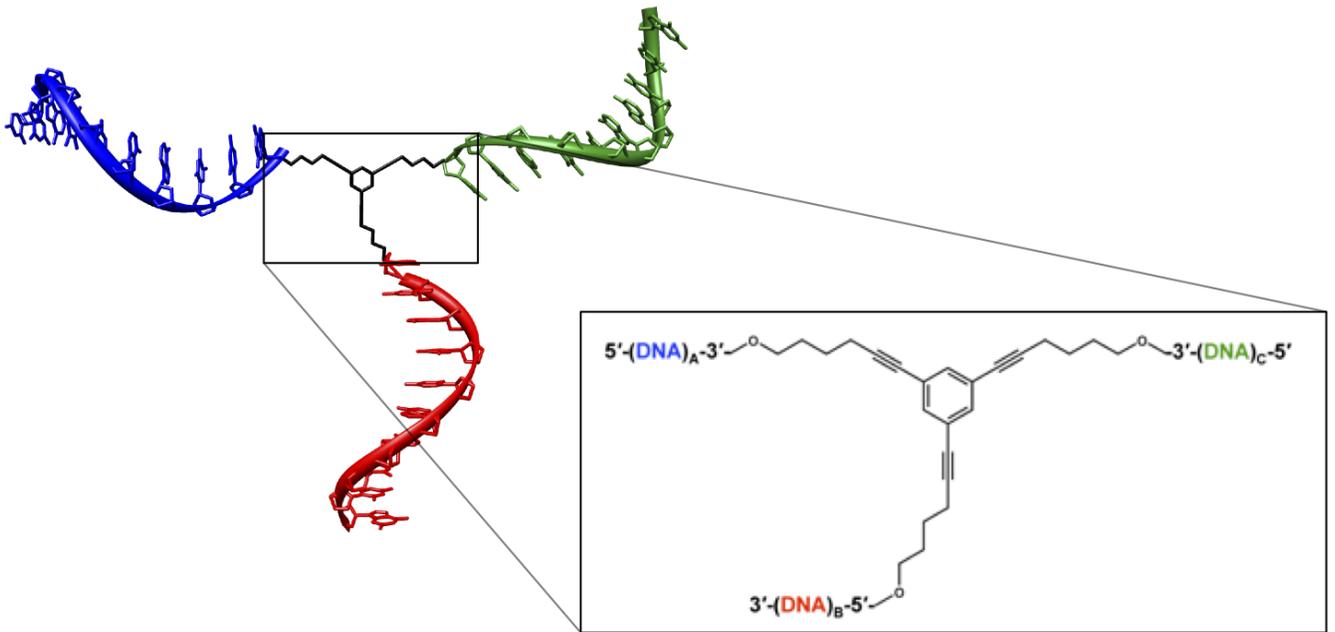


Figure 1. Top. Schematic of the four-ring DNA nanostructure. The structure consists of 16 different synthetic three-way oligonucleotides, where number 12 is used twice in the assembly to investigate the possibility of using generic building blocks. Each side corresponds to a 10 nucleotide long DNA duplex and the color-coding indicates unique, orthogonal sequences. In addition, complementary 10-mer oligonucleotides are added to all blunt arms, making all sequences in the system double-stranded. **Bottom.** Schematic of the synthetic three-way oligonucleotide, with the 1,3,5-trisubstituted-benzene node in the box to the right.

Results & Discussion

We have previously demonstrated the assembly of two hexagonal cells fused together, denoted “DNA-naphthalene”.³⁶ This serves as a reference structure in the current study and the starting point for stepwise assembly of the new structures to be proven. By starting with a known structure and adding stepwise one building block at a time, it is possible to follow the assembly of larger constructs by the sequential retardation of bands in gel electrophoresis. In this study we construct a network of four hexagonal cells fused together, assembled by 17 nodes and 31 duplexes formed in a one-step self-assembly.

Before formation of the four-cell network we demonstrate the stepwise assembly of the two possible three-cell constructs, denoted “DNA-anthracene” and “DNA-phenalene”, respectively. The stepwise assembly of DNA-anthracene is shown in Figure 2. The starting point is the two-cell structure DNA-naphthalene (A), seen in the left lane on the gel. The following lanes contain a sequential addition of new three-way oligonucleotides (B-F), following the schematic to the left of the gel, ending with the DNA-anthracene (G). Each addition of an oligonucleotide causes a small but distinguishable retardation of the band on the gel, due to the change in geometry giving a more rapidly increasing hydrodynamic friction coefficient compared to increased DNA charge-force when building larger constructs. Based on the stepwise decrease in mobility of each sequential construct, we can conclude that the DNA-

anthracene has successfully been assembled. The intermediate structure F exhibits three separate bands corresponding to binding of one, two and three nodes to the two-cell construct (a blow-up of the bands with increased resolution is shown in SI, together with an intensity profile analysis). A similar pattern can be seen for sample D and E on the gel. Both display two adjacent bands corresponding to one node being bound or not. This poor hybridization yield of intermediate structures has been seen in this kind of system before and is believed to originate from the single stranded sequences of the three-way oligonucleotides (see below). By contrast, the totally ring-closed product of DNA-anthracene (G) displays one distinct band with no other large constructs present; and the same can be seen for the two-cell starting point, DNA-naphthalene (A). The band midway between the monomer (H) and the main band, present in all samples corresponds to the one-cell structure or DNA-benzene, presented previously.³⁵

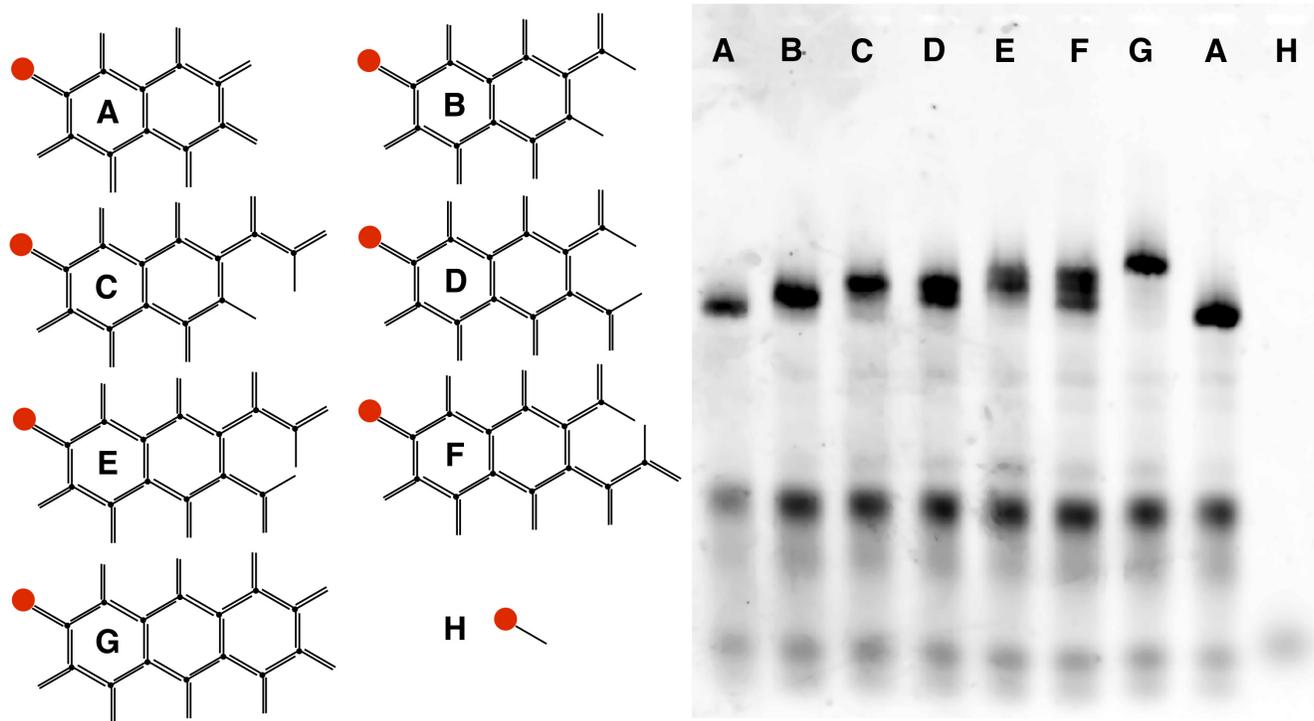


Figure 2. Stepwise formation of the three-cell construct DNA-anthracene following the schematic on the left. Analysis was done using 4.5% MetaPhor agarose gel electrophoresis. Red dots indicate the

position of the Cy3 label used for visualization of product in gel scanner. For an intensity profile of the overlapping bands for sample F, see Figure S4 in SI.

The other three-cell structure, DNA-phenalene, is verified in the same way as DNA-anthracene, by starting with the known two-cell structure and adding one three-way oligonucleotide in a stepwise fashion until the final three-cell structure is reached. The stepwise build-up is presented in Figure 3, starting from DNA-naphthalene (A) on the left and ending on DNA-phenalene (F). The same stepwise retardation pattern can be seen in this series of constructs, each addition of another three-way oligonucleotide yields a slower moving band in the gel. Sample (E) is the exception to this pattern. One would expect this construct to have similar mobility as sample (D), having the same number of three-way oligonucleotides in slightly different orientation. However, the band that would follow this logic is missing; instead the sample displays similar bands as sample (B) and (C), indicating that there is poor binding of the added three-way oligonucleotide. The reason for this poor binding could be some interference effect of secondary order within the specific sequence, *i.e.* semi-stable hairpin-like structures of a few base pairs. This emphasizes the importance of maintaining totally double stranded sequences within the system; even though there might be undesired side-products that are energetically favorable, the free energy of these structures cannot compete with the lower free energy structures when all sequences are perfectly matched. As seen in the final DNA-phenalene sample (F), there is only one major band for large constructs visible indicating that the designed structure is formed without major interference from undesired substructures. To verify that this band corresponds to the desired DNA-phenalene and that the missing node in sample E indeed appears and binds to the structure when all building blocks are present, a gel electrophoresis experiment was performed in which a fluorophore label was positioned on this node. A band with the same migration rate thus appears in all samples irrespective of label position, verifying that the missing node of sample E coordinates when all building blocks are present (Figure S2 in Supporting Information). This observation also confirms that DNA-

phenalene is indeed formed. The unstable construct is always the one with the maximum number of uncyclized flexible branches, *i.e.* E in Figure 3, and E & F in Figure 2. As the fully cyclized versions are more stable this must constitute a driving force for cyclization. It is possible that constructs like E (Figure 3), in which there is considerable conformational flexibility, are destabilized by the close proximity of non-complementary single-stranded (and also double-stranded) regions that are not conformationally locked by base pairing, *i.e.* this construct has dangling single- and double-stranded regions. Such regions have a high density of negative charge and as they are not locked in place they will be mutually repulsive. In structures such as F (Figure 3), which is fully base-paired, the specific organized binding of metal ions to the duplex, *i.e.* to the major and minor grooves, might better be able to neutralize the high negative charge density alongside the free energy gain of hybridization.

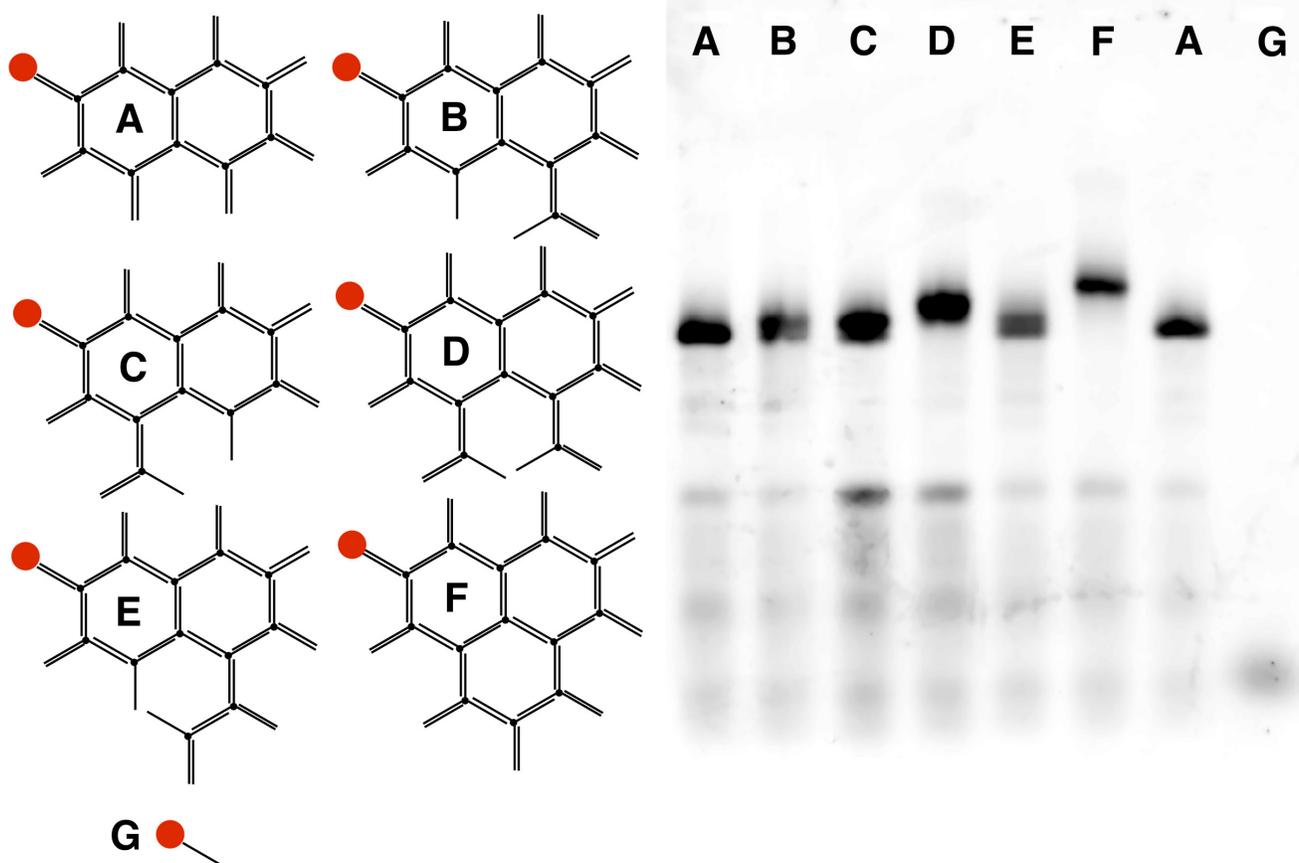


Figure 3. Stepwise formation of the three-cell construct DNA-phenalene following the schematic to the left. Analysis was made using 4.5% MetaPhor agarose gel electrophoresis. Red dots indicate the position of the Cy3 label used for visualization in gel scanner.

Having demonstrated the assembly of the two different three-ring systems separately, we move forward to show the assembly of the entire four-ring structure displayed in Figure 1. Figure 4 shows the different completely ring-closed network structures that have been assembled in this study. Starting from DNA-naphthalene (A), followed by the two variants of three-ring structures: DNA-anthracene (B) and DNA-phenalene (C); and finally the complete four-ring network (D).

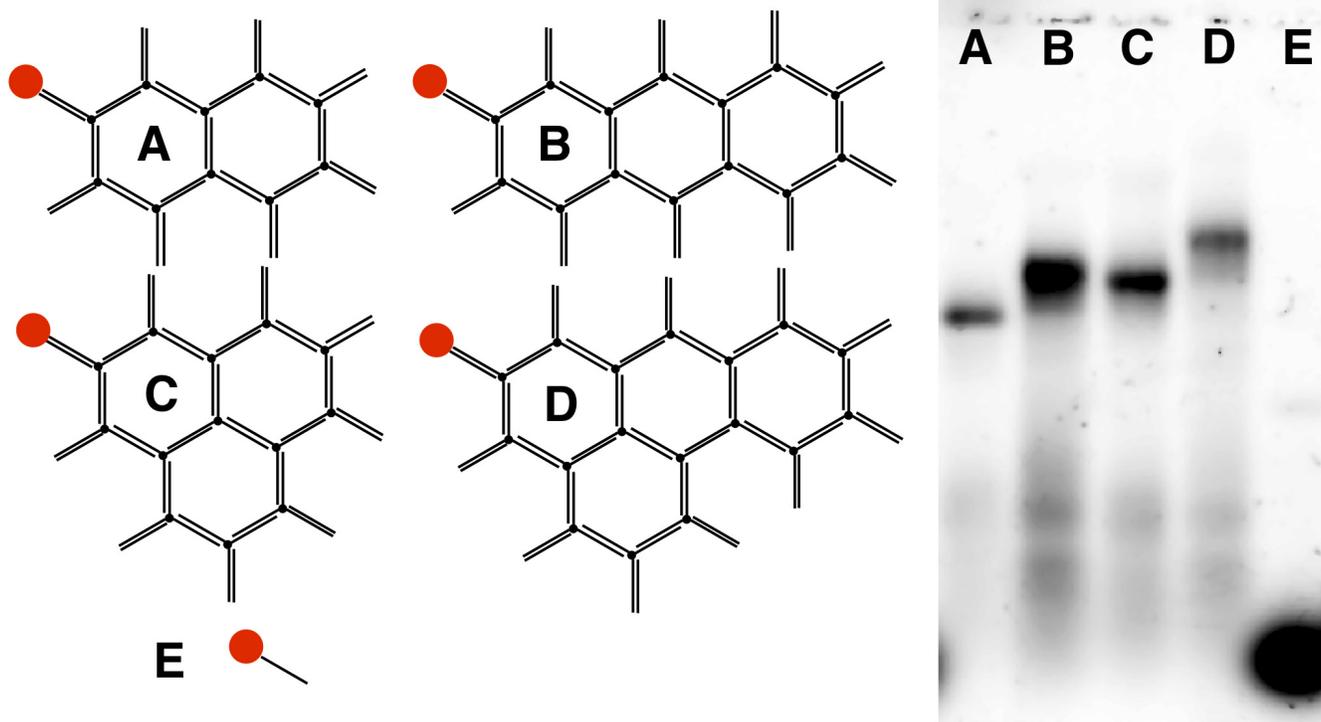


Figure 4. Four different completely ring-closed DNA nanonetworks of varying size, following the schematic outline to the left of the gel. DNA-naphthalene (A), DNA-anthracene (B), DNA-phenalene (C), four-ring structure (D) and Cy3-labelled 10-mer reference (E). The gel is 4.5% MetaPhor agarose and visualization made using Cy3 dye as indicated by the red dot in the scheme to the left.

The gel in Figure 4 clearly shows that all the different structures are formed as major products of each individual assembly process. There is no substantial contribution from any specific by-product in any of the four samples. However, there are traces of other structures in all assemblies, visible in all gels (Figures 2-4). A systematic study of yields of the desired construct for each individual assembly process can give us information about what to expect as the network grows bigger. This can be done by measuring the fluorescence intensity from the band corresponding to one specific construct, and normalizing with respect to the total intensity of the sample in one lane on the gel. The relative fluorescence intensity corresponds to the molecular ratio of that construct relative to all other constructs in the same lane, thus giving the yield (detailed information about the method in Supporting Information). This analysis was made for all ring-closed structures investigated in the current study, since all gels indicate a consistent difference between fully ring-closed structures and the others. Also included in the analysis is the yield from the one-ring structure, DNA-benzene, which we presented in a previous paper.³⁵ These yield data are plotted, in Figure 5, against the number of DNA duplexes involved in forming each specific structure (n), *i.e.* DNA-benzene ($n=12$), DNA-naphthalene ($n=19$), DNA-phenalene ($n=24$), DNA-anthracene ($n=26$) and the final four-ring network ($n=31$). To obtain statistics, the same analysis was performed in multiple experiments carried out under the same conditions, resulting in the error bars (standard deviation) in the figure.

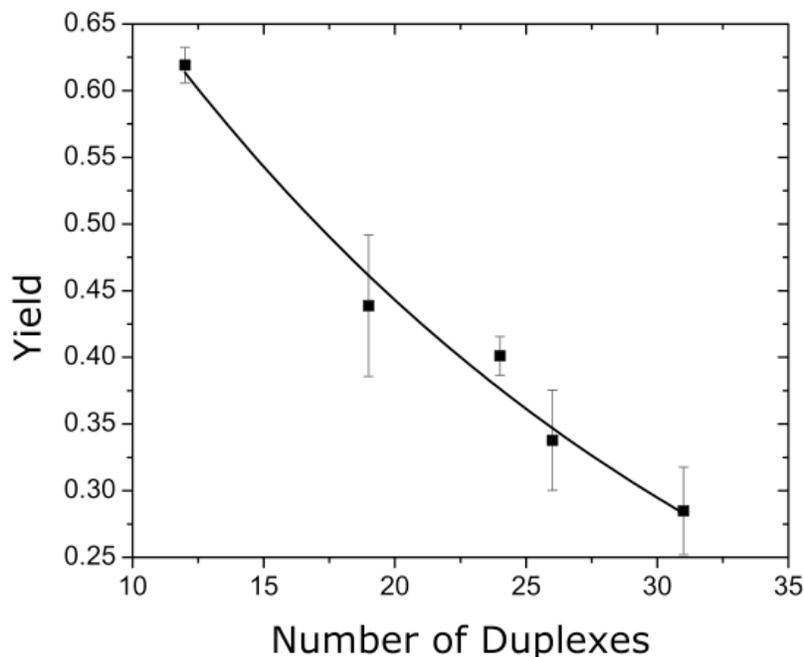


Figure 5. Yields for the ring-closed structures, plotted against number of duplexes involved in forming one specific structure. DNA-benzene (n=12), DNA-naphthalene (n=19), DNA-phenalene (n=24), DNA-anthracene (n=26) and the final four-ring network (n=31). Error bars show standard deviation for the particular structure, from repeated experiments carried out under the same conditions. The number of repeated experiments is 21 for DNA-naphthalene, 6 for DNA-anthracene, and 3 for each of DNA-benzene, DNA-phenalene and the final four-ring network. The solid curve corresponds to the statistical yield model (Eq.1).

The yield result in Figure 5 displays as expected a monotonously decreasing trend as the number of duplexes increases, *i.e.* the size of the network is getting larger. This behavior can be understood using a simple experimental model. Assuming that the probability of a hybridization event to occur, *i.e.* that two complementary sequences form a DNA-duplex, is uniform over the system (y_{duplex}), then the final yield (Y) is this probability raised to the power of number of events, *i.e.* the number of duplexes in the system (n).

$$Y = y_{duplex}^n \quad (\text{Eq.1})$$

By fitting the unknown parameter y_{duplex} of this model to the experimental yield data presented in Figure 5, an average hybridization reaction yield of 96% was obtained ($y_{duplex}=0.96$, $R^2 = 0.98$). To our knowledge this is the first effort to systematically study and explain the outcome of nanofabrication through processes of DNA directed self-assembly. There are numerous reported yields of different DNA nanoassemblies based on oligonucleotides in the literature with a wide spread, ranging from 1%^{7, 43} to 95%.⁴⁴ In the field of DNA origami, several of the remarkable assembled structures have impressively high yields,^{13, 15, 17, 21} but because the assembly protocols vary, such as the amount of excess of “staple” strands compared to the M13 scaffold, it is hard to compare with assemblies based on stoichiometric (equimolar) DNA strand ratios as used here. In addition, it would be interesting to know the effect of and potential tolerance to missing staple strands on the final construct and the possibility of being able to resolve such constructs from fully assembled ones. Whatever assembly approach is chosen, the wide spread of effective yields emphasizes the necessity for a fundamental investigation to gain insight into the reaction processes and facilitate the prediction of the outcome of a specific assembly design. It is highly probable that the effective yields of different assembly strategies of DNA nanofabrication may vary due to different thermodynamic and kinetic fundamentals. For example, the crystal-like DNA-tile assemblies lack the entropic effect of more flexible structures. Topology is, without doubt, a major issue when dealing with these kinds of processes, as well as the dynamics of actual assembly events. One can imagine differences in topological constraints that need to be overcome when comparing fabrication in 2D and 3D, in addition to how and in what order different parts of a construct come together. There have been some studies of these aspects of DNA nanoassembly with FRET-based assays, studying the thermodynamic stability of DNA-tiles⁴⁵ and furthermore the effect of multivalent interactions.⁴⁶ From our laboratory we have reported on the thermodynamic aspects of a hexagonal DNA nanostructure as well as structural dynamics of the same system.^{47, 48}

Looking at our system, based on hierarchic assembly, is there any fundamental thermodynamic reason for the yield of duplex formation to be significantly less than 100%? The equilibrium constant of this reaction can be estimated using thermodynamic data of a hexagonal DNA nanostructure presented in a previous report.⁴⁸ Assuming that an average of the six edges in that hexagon represent a characteristic 10mer sequence in our system, the average equilibrium constant of hybridization can be estimated to approximately $1.2 \times 10^{12} \text{ M}^{-1}$ (at 4°C). With an oligonucleotide concentration in the range of 10^{-7} M , this entails a single-stranded ratio of less than 0.5%. This is obviously not the main reason for the missing, on average, 4% of unformed DNA duplex. There must be other sources of error and those of experimental origin are worth considering.

A potential deviation from perfect 1:1 stoichiometry is an issue close at hand. A deviation from equimolar stoichiometry may originate from volumetric uncertainty and imprecise oligonucleotide concentration determination. The latter is based on ultraviolet absorption, using the nearest-neighbor approximation (NNA) to deduce the extinction coefficient for a given nucleobase sequence. This approximation has an inherent uncertainty of about 10%.⁴⁹ Whatever origin, an uncertainty in oligonucleotide concentration resulting in deviation from equimolar stoichiometry, is an issue that needs to be addressed to determine to what extent it influences the overall yield.

If one considers an assembly of building blocks with a concentration distribution of a given variance, the way the assembly occurs may have a major effect on how the distribution influences the final product. In a sequential addition of building blocks, deviations from the average concentration will deplete the cohort of building blocks available for the desired assembly. The reason for this is that any difference between two consecutive building blocks in the chain will create alternative structures that will compete for the subsequent building blocks. Assuming no difference in adhesion between the desired structure and alternative structures, the yield of the desired structure will decrease based upon the average deviation of concentration. In this case the situation will worsen the larger the assembly, in the same manner as our suggested yield model, *i.e.* one may consider an average error in oligonucleotide concentration decreasing the yield of a hybridization reaction, which accumulates with

increasing assembly size. In fact, the total assembly yield of a chain (Y) decreases with the average of the relative concentration differences (Δ) and system size (n). This average difference is related to the standard deviation from average concentration (σ) in the following way (details in Supporting Information).

$$Y = \left(1 - \frac{\langle \Delta \rangle}{2}\right)^n = \left(1 - \frac{\sigma}{\sqrt{\pi}}\right)^n \quad (\text{Eq. 2})$$

Applied to our system, this error would suggest that a standard deviation of 7% from 1:1 stoichiometry could explain the missing 4% in each hybridization reaction step. However, it is also important to take the effect of ring-closure into account. Although ring-closure is expected to increase electrostatic repulsion and decrease conformational entropy, compared to linear polymerization, it also provides a higher apparent concentration of the two ends closing the circle, in the end leading to a thermodynamically favored process, as found for a closely related single hexagon system.⁴⁸ The equilibrium model of hexagon formation from ref. 48 may be expanded to a ring of any number of building blocks (n). With each oligonucleotide building block being attached to the previous one, from A to W, the last one closing the ring, the overall ring yield (Y) may be expressed in terms of an equilibrium equation involving all building blocks, by relating the amount of formed rings (x) to the mean oligonucleotide concentration (m)

$$Y = \frac{x}{m} = \frac{(\min([A]_0, [B]_0) - x)}{([A]_0 - x)} \frac{(\min([B]_0, [C]_0) - x)}{([B]_0 - x)} \dots \frac{(\min([W]_0, [A]_0) - x)}{([W]_0 - x)} \kappa \quad (\text{Eq. 3})$$

where the concentration distribution of oligonucleotides is characterized by the lowest concentration of each stoichiometric couple (*i.e.* A-B, B-C...W-A) limiting the ring formation (x). The dimensionless parameter κ is the ratio between equilibrium constant of ring-closure and the average oligonucleotide concentration, thus, gauging the strength of ring-closure. Assuming that deviations of oligonucleotide concentration from the mean value follow a normal distribution with standard deviation σ , the overall yield can be expressed as an exponential relation (see Supporting Information).

$$Y = 1 - \alpha = \kappa e^{-\frac{n}{\sqrt{\pi}} \left(\frac{\sigma}{\alpha} + \frac{5}{6} \left(\frac{\sigma}{\alpha} \right)^3 + \frac{43}{20} \left(\frac{\sigma}{\alpha} \right)^5 \right)} \quad (\text{Eq. 4})$$

A large κ , meaning an effective thermodynamic driving-force for ring-closure, could counteract the influence of stoichiometric deviations on the overall yield. However, we note that the model considers one ring only (one parameter κ), while the polycyclic system is considerably more complex with multiple branching points, for which one ring-closing parameter may be inappropriate. Anyhow, looking at the gel electrophoresis results in Figure 2 and 3, the higher yields of all ring-closed structures indeed indicate that ring-closure is an important factor in the studied system that needs to be taken into account. Although the model is crude it may give some indication how to interpret the behavior of our system. Therefore, treating the polycyclic system as rings of different sizes, it is possible to estimate the magnitude of the ring-closure efficiency if one assumes a standard deviation of 10% (given by the uncertainty in NNA) and fit this model to the experimental data in Figure 5, $\kappa=3.6$ for this system.

Furthermore, one cannot disregard the potential interference effects of base sequence matching of secondary order. Even though the sequences are designed to be orthogonal, one should not rule out the possibility of forming undesired metastable secondary structures such as hairpins leading to a smaller effective concentration of the building blocks and even to structural arrest. Analogous to the issue with stoichiometry, this problem is expected to scale with increasing structural size in a similar way as our yield model. There could also be topological problems related to the structural design of the three-way oligonucleotide building block. Even though the system is designed to have swivel flexibility, distortions of the planar conformation might lead to some assembly problems. It would be interesting to make systematic studies with varied nodes and tethers as well as DNA sequence lengths to address this.

From a nano-technological point of view there are two conclusions that can be drawn from our results.

(1) A small deviation from unity for the probability of each hybridization event will have a big impact on the final result as the number of events increases. Hence, we may be rapidly reaching the limit of how large a network we can assemble in one step, with a meaningful yield.

(2) A more general conclusion is that a one-pot reaction of multiple, simultaneous assembly events may be treated in the same way as a sequence of reaction steps in a synthesis protocol.

There are two different possible strategies to increase the yield of the final product. (i) Improve the yield of the single duplex reaction. The closer to total conversion of one hybridization reaction (y_{duplex} approaching 1), the larger the network that can be assembled in a one step process. However, this does not solve the underlying problem that any deviation from total conversion at the single-duplex level will decrease the yield of the end product. The larger the network gets, the more substantial will be the decrease. (ii) A second solution to the one-pot issue for larger DNA nano-constructs could be to adopt a different approach to self-assembly nanofabrication, *viz.* by instead adopting a stepwise process with purification and work-up possibilities, *i.e.* a modular build-up.

We have previously reported a click-fixation strategy for creating a larger robust building block assembled from short oligonucleotides.⁴² By incorporating azide and alkyne functionalities at specific positions, a totally covalently cross-linked DNA nanostructure can be formed using Cu(I)-catalyzed Azide-Alkyne Cycloaddition (CuAAC), *i.e.* click chemistry.^{50,51} The model system is a hexagonal DNA nanostructure assembled from six 22-mer oligonucleotides, with an edge length of 10 bases. Each 22-mer oligonucleotide has two 10 base sequences separated by two thymines. The stretch of 10 bases has one unique partner in the system, forming one side of the hexagon. The two thymines are unpaired, functioning as flexible hinges. Upon formation of the hexagonal nanostructure, azide and alkyne modifications are positioned to enable the six site-specific click reactions of the fixation step (reaction scheme in Figure 6, more information can be found in Ref. 42). We envision that such a covalently-fixed nanostructure could function as basis for assembly of larger constructs in a modular manner, a strategy that would in the future allow us to assemble very large structures by relying on prefabricated modules instead of using a considerable number of smaller building blocks assembled in one step. Solubility could be a problem when the building blocks become very large and only a small number of connectors will require rather high concentrations for good yields. However, if the size is balanced by

an appropriately increased thermodynamic stability of the final assembly through a large number of hybridization sites, the assembly should be possible using dilute solutions of reactants.

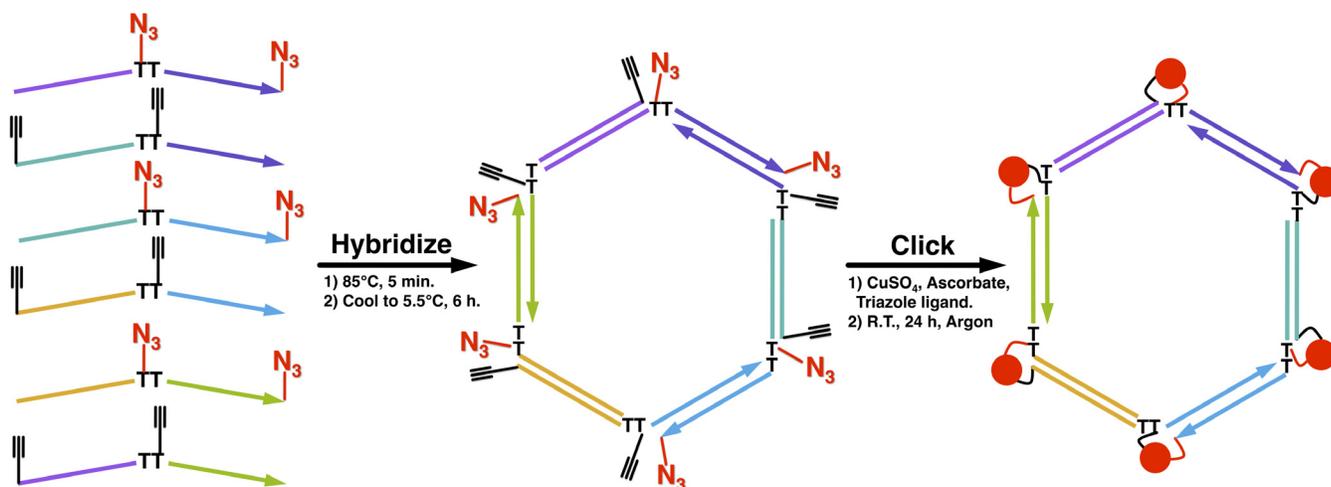
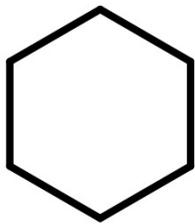


Figure 6. Reaction scheme of the site-specific click-fixation strategy of a hexagonal DNA nanostructure. Six alkyne and azide modified 22-mer oligonucleotides are hybridized to form a hexagonal construct with edge length 10 bases. The sequence of 22 nucleotides has two stretches of 10 nucleotides separated by two thymines. Each stretch of 10 bases is complementary to another sequence in the system, forming one side in the hexagon. The two thymines are unpaired and function as hinges. Upon formation of the hexagonal nanostructure, azide and alkyne modifications are situated to enable the six site-specific click reactions of the fixation step. Chemical structures of the triazole cross-links are shown in Figure S5, in SI.

When designing future modules it is vital to understand the fixation strategy, a system where six individual click reactions occur simultaneously in a one-pot reaction. As pointed out before, the precision of these reactions is governed by the exact positioning of the alkyne and azide functionalities intended to react. This creates a system of six identical reaction sites located on a nanoscale construct, *i.e.* a hexagon with a side length of 3.4 nm. It is possible to investigate if these reaction sites somehow interact with each other by looking at the resulting fractions of substructures formed after the click fixation process has occurred. Analogous to the case with yields of DNA nanostructures based on a one-

step approach considered above, the probability that a click reaction occurs is below unity. A result of this non-ideal behavior is that instead of one single end product, *i.e.* one totally covalently locked DNA hexagon, there will be a distribution of various substructures corresponding to species for which one, two or more click reactions have not taken place. The matter is further complicated by the fact that the range of substructures actually formed is not only due to the number of reactions, but also dependent on where in the structure these reactions occur. Table 1 summarizes all possibilities of different numbers of click reactions, where they occur on the hexagon, and the resulting substructures, starting from a hybridized DNA hexagon. Due to symmetry many possibilities are degenerate.

Table 1. Overview of possible combinations of click reactions (red dots) that can take place on a hexagonal starting material and the actual substructures resulting from a specific combination.

Start	No. of Click	Position	Degeneracy	Total No.	Substructures ^a	
	6		1	1	H(6)	
	5		6	6	L(6)	
	4		6	6	15	5, 1
			6	6		4, 2
			3	3		3, 3
	3		6	6	20	4, 1, 1
			6	6		3, 2, 1
			6	6		3, 2, 1
			2	2		2, 2, 2
2		6	6	15	3, 1, 1, 1	
		6	6		2, 2, 1, 1	
		3	3		2, 2, 1, 1	
1		6	6	6	2, 1, 1, 1, 1	
0		1	1	1	1, 1, 1, 1, 1, 1	

^aThe numbers correspond to the size of fragments a particular click combination gives rise to, *i.e.* the number of oligonucleotides involved. *e.g.* the first combination of 4 successful click reactions gives rise to a 5mer and a monomer substructure.

At this point of the analysis a model can help us understand the outcome of the process. First, each of the six click reactions constitutes a Bernoulli experiment, *i.e.* a binary option of success or failure. This means that if the reactions occur independently, then the resulting outcome of the six sequential, independent events should follow a binomial distribution (Eq. 2). The probability, $P(k)$, of a number of successful events, k , in a number of trials, n , is given by the following equation, where p is the probability of one successful event.

$$P(k) = \binom{n}{k} p^k (1-p)^{n-k} \quad (\text{Eq. 5})$$

Applied to click-fixation n is the total of six click reactions and k is the different number of successful reactions, starting with a fully hybridized DNA nano-hexagon. However, because the end result does not only depend on the number of successful click reactions but also on position (as illustrated in Table 1), the analysis is somewhat more complex. The next step is therefore to evaluate each substructure separately and sum up contributions from every specific number of successful reactions that contributes to one substructure, according to Table 1. The resulting binomial terms are displayed in Table 2. The contribution from a specific number of successful click reactions to a substructure depends on the degree of degeneracy, explaining the fractional coefficient of each term. Furthermore, to be able to compare substructures with each other they have to be of the same base, *i.e.* parts of six in this case, why there is also a factor correlated to the size of a substructure. *E.g.* for the hexagon there is only one, in a total of one, possibility of forming that structure and furthermore it is constructed of six parts out of six. This gives rise to the pre-factors of the $P(6)$ terms in Table 2 which, although not important in the present example, are displayed for clarity.

Table 2. Summary of the contribution of binomial terms to the different substructures.

Structure	Binomial Terms ^a	Values ^b
	$\frac{6}{6} \binom{6}{1} P(6)$	0.3122
	$\frac{6}{6} \binom{6}{6} P(5)$	0.4011
	$\frac{5}{6} \binom{6}{15} P(4)$	0.0716
	$\frac{4}{6} \left(\frac{6}{15} P(4) + \frac{6}{20} P(3) \right)$	0.0695
	$\frac{3}{6} \left(\frac{6}{15} P(4) + \frac{12}{20} P(3) + \frac{6}{15} P(2) \right)$	0.0633
	$\frac{2}{6} \left(\frac{6}{15} P(4) + \frac{18}{20} P(3) + \frac{18}{15} P(2) + \frac{6}{6} P(1) \right)$	0.0512
	$\frac{1}{6} \left(\frac{6}{15} P(4) + \frac{24}{20} P(3) + \frac{36}{15} P(2) + \frac{24}{6} P(1) + \frac{6}{1} P(0) \right)$	0.0311

^a The binomial terms are deduced by evaluating contributions from different origins (according to Table 1), with fractions correlated to the degeneracy of a substructure. The factors preceding the parentheses ensure every binomial term to be set to the same base, *i.e.* parts of six.

^b The values are the binomial terms for $p=0.82$, which applies to our experimental results for the hexagonal system.

An experimental parameter that has to be taken into account, if the binomial distribution model is to be used as justification for the behavior of the click-fixation system, is the presence of substructures due to incomplete hybridization (as indicated in ref. 42). The result of this situation, unless a purification step is applied, is that there are DNA structures of varying sizes already at the start of the click fixation step, and notably a decreased number of possible click-reaction sites (the linear 6-mer has 5 sites, 5-mer has 4 sites, 4-mer has 3 sites *etc.*). Consequently, each starting template gives rise to a different binomial distribution with a different pattern of substructures; analogous to the case with the hexagonal

template described in Tables 1 & 2 (corresponding tables for each starting template are given in Supporting Information). To generate an accurate representation of the system including this template spread, using the binomial model, all possible starting structures have to be taken into account. This gives an expression for each covalently cross-linked substructure in the form of a weighted sum of binomial terms, originating from every starting template structure with its corresponding distribution (as given for the hexagonal template in Table 2; for other structures see Supporting Information). The experimental yield data from the hybridization reaction⁴² are the coefficients used for the different substructures in the weighted sum. Starting with 27% of the hexagonal structure appearing from the hybridization reaction entails a factor of 0.27 to all the terms in Table 2 when summing the binomial distributions. Analyzing all substructures in a similar manner finally generates the binomial model for this system.

The result of this analysis is displayed in Figure 7, with the prediction of the binomial model given in black and the experimental data given in red (error bars correspond to the standard deviation of 14 identical experiments, data from ref. 42). The pattern of the model corresponds well with the experimental results, within experimental error in all but two cases (4-mer and monomer). There is only one variable in this model, which is the probability of a successful event (p), *i.e.* the probability of one click reaction occurring. This is deduced by evaluating the probability of achieving the percentage of cross-linked hexagon indicated by experimental data. Six click reactions need to occur and consequently the sixth root of the hexagonal experimental yield gives the probability of one of these events (taking the hybridization reaction into account). The experimental data of this study give a value of 82%, which is used in the presented distribution (Figure 7). Assessing this variable rather than a fitting procedure is motivated by a higher reliability for the experimental yield of the hexagonal structure compared to the monomer structure. On the polyacrylamide gels in this study, the hexagonal structure produces one distinct and easily quantified band, arising from the fact that there is only one possibility of forming this structure. The monomer structure, on the other hand, is actually six different oligonucleotides. As a consequence, the band on the gel broadens because of the sequential distribution of different monomers.

Furthermore, the monomer structure behaves differently on native and denaturing gel.⁴² These phenomena make the quantification of the monomers less precise, thus the yield data of the hexagon is more reliable in comparison. An analysis procedure where the variable p was fitted to the experimental data has been investigated (data not shown), however, without improvement in presentation or understanding.

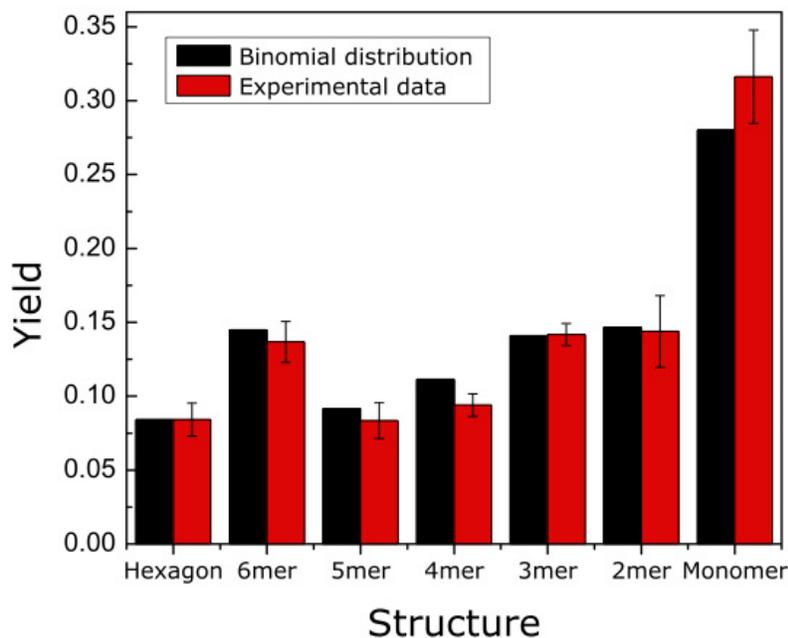


Figure 7. Yield data for click fixation of the DNA hexagon. Experimental distribution of possible substructures shown as red columns, error bars indicating standard deviation (14 samples). Experimental data taken from ref 42. Black columns show weighted sums of binomial distributions for each substructure of all possible starting materials. The probability, p (yield of one click reaction), is deduced to be 0.82 from the observed yield of totally cross-linked DNA hexagons.

The result presented in Figure 7 indicates that the binomial model can explain the yields for the click-fixation system studied in this paper. This analysis also gives us information about possible cooperativity. Because independency of events is a prerequisite for a binomial distribution, we may

conclude that the six click reactions occur without any interference with one another. Understanding and predicting the outcome of the click-fixation strategy could pave the way for this technology to be applied to systems of higher hierarchy using a modular approach for bottom-up nano-assembly.

Conclusions

We have demonstrated the assembly of a fully addressable nano-network, based on synthetic three-way branched DNA molecules. Even though the principle of total addressability relies on unique building blocks, we show that it is possible to assemble nanostructures of precise sizes containing a generic node. Furthermore, the four-ring DNA nanostructure is assembled from 28 building blocks in a one-step annealing process with a respectable yield of 28%. Importantly, we show that a simple statistical yield model can describe the limit of a one-step/one-pot assembly process for a non-periodic system like this. The yield of DNA nanostructures with controllable size decreases exponentially with increasing structure size (*i.e.* number of duplexes involved). Therefore, in order to open up a general route for nanofabrication based on DNA assembly, we envision an alternative future strategy that may solve the problem of decreasing yields when larger structures are desired: this is a modular build-up approach based on a click-fixation step, covalently cross-linking a DNA nanostructure and thereby creating a larger module, which is robust and importantly may be purified and used for further fabrication. Although the present yield of this module is low, less than 10%, the result defines a first step toward an alternative fabrication approach. More importantly, we show how the system of six simultaneous click reactions on a DNA nano-hexagon behaves in a way that can be described by a binomial model. Not only does this predict the outcome of the reaction but also, more importantly, tells us that the six reactions proceed independently from one another. Consequently this model allows us to use the binomial distribution function as a probe of cooperativity between chemical reactions on the confined template of the DNA nanostructure. Deviations from this model would indicate positive or negative chemical interference, a scientific path we will follow in future studies.

Methods

DNA sequence design. The system follows the same principals as in our previous work.^{35,36} The three-way oligonucleotides, serving as basis for the assembly, are synthesized using a 1,3,5-trisubstituted-benzene as node. The D_{3h} -symmetric node gives rise to the hexagonal geometry of the unit cell. However, a non-repetitive network demands unique sequences on all three positions, which can be achieved by incorporating suitably chosen orthogonal protection groups in the synthetic process. This enables not only the necessary freedom in base sequence design but also freedom in directionality of the strands. In the system of this study, all nodes are connected to the 5' end of one strand and the 3' end of the two other strands (Figure 1). The reason for this geometry is simplicity and optimization of the reaction yields in the solid-phase synthesis; this configuration is consistent with the use of conventional 5'-dimethoxytritylnucleoside-3'-phosphoramidite monomers in trigonal oligonucleotide assembly. Experimental details can be found in previous work.^{35,36} Each side in the system is a unique sequence of ten nucleotides and every 10-mer sequence is designed to be complementary to only one other sequence in the system and orthogonal to all others. The criterion of orthogonality between two sequences is that no more than four bases may match for any alignment of two strands. Furthermore, to enhance the thermodynamic stability of the system, the number of CG bases in any 10-mer sequence is set to be at least four. All three-way oligonucleotides in the system are unique apart from one (12 in figure 1), which is used twice in the nanoconstruct. This assesses the possibility of using generic nodes in the system and how this might influence the formation of the desired construct.

Self-assembly process. Every oligonucleotide was set to 2 μ M in phosphate buffer (pH 7.5; $[\text{Na}^+] = 200$ mM; $[\text{PO}_4^{3-}] = 109$ mM) using the DNA absorbance at 260 nm, with the extinction coefficients at 260 nm calculated by the nearest-neighbor approximation (NNA). To assemble a specific construct the oligonucleotides involved were mixed in equimolar amounts (typically 1.5 μ l per oligonucleotide) and hybridized by heating to 90°C for 5 min and allowing to cool to room temperature overnight (typically 15 h) by turning off the heating-block (Grant Inst.). The samples were then put on ice until further analysis.

Gel electrophoresis. All samples were analyzed using 4.5 % (w/w) MetaPhor® agarose (Cambrex) with phosphate buffer (pH 7.5; [Na⁺] = 200mM) as running buffer. Ficoll 400 (Sigma-Aldrich) was added to each sample as a loading agent, giving a concentration of 5% (w/w). The applied voltage was set to give a field strength of 3.5 V/cm, typically 67 V. The run time of electrophoresis was 4 h. Circulation of the buffer through a heat exchange system was performed to keep the temperature of the system below 4°C throughout the experiment.

Sample analysis. After electrophoresis the gels were visualized using a Typhoon 9410 (GE Healthcare) using appropriate laser excitation and emission filter for each fluorophore. Cy3 was excited with a 532 nm laser and emission collected through a 580 nm band pass filter. Yield estimations were done by quantification of emission intensities from each band correlated to the total intensity of the sample using ImageQuant TL software (GE Healthcare).

Acknowledgements

This research is funded by the European Research Council (ERC grant to B.N) and King Abdullah University of Science and Technology (KAUST awardee B.N.).

Supporting Information Available:

DNA sequences of oligonucleotides; analysis of the binomial distributions, starting from templates of different sizes and yield error analysis are available free of charge *via* the Internet at <http://pubs.acs.org>.

References

1. Seeman, N. C.; Kallenbach, N. R., Design of Immobile Nucleic Acid Junctions. *Biophys. J.* **1983**, *44*, 201-209.

2. Seeman, N. C., Nucleic Acid Junctions and Lattices. *Journal of Theoretical Biology* **1982**, *99*, 237-247.
3. Seeman, N. C., DNA in a Material World. *Nature* **2003**, *421*, 427-431.
4. Winfree, E.; Liu, F. R.; Wenzler, L. A.; Seeman, N. C., Design and Self-assembly of Two-dimensional DNA Crystals. *Nature* **1998**, *394*, 539-544.
5. Ding, B. Q.; Sha, R. J.; Seeman, N. C., Pseudo-hexagonal 2D DNA Crystals from Double Crossover Cohesion. *J. Am. Chem. Soc.* **2004**, *126*, 10230-10231.
6. Wang, X.; Seeman, N. C., Assembly and Characterization of 8-arm and 12-arm DNA Branched Junctions. *J. Am. Chem. Soc.* **2007**, *129*, 8169-8176.
7. Zhang, Y. W.; Seeman, N. C., Construction of a DNA-Truncated Octahedron. *J. Am. Chem. Soc.* **1994**, *116*, 1661-1669.
8. Kuzuya, A.; Wang, R. S.; Sha, R. J.; Seeman, N. C., Six-Helix and Eight-Helix DNA Nanotubes Assembled from Half-Tubes. *Nano Lett.* **2007**, *7*, 1757-1763.
9. Goodman, R. P.; Berry, R. M.; Turberfield, A. J., The Single-step Synthesis of a DNA Tetrahedron. *Chem. Commun.* **2004**, 1372-1373.
10. Erben, C. M.; Goodman, R. P.; Turberfield, A. J., A Self-assembled DNA Bipyramid. *J. Am. Chem. Soc.* **2007**, *129*, 6992-6993.
11. Zimmermann, J.; Cebulla, M. R. J.; Monninghoff, S.; von Kiedrowski, G., Self-assembly of a DNA Dodecahedron from 20 Trisoligonucleotides with C-3h Linkers. *Angew. Chem.-Int. Edit.* **2008**, *47*, 3626-3630.

12. Ke, Y. G.; Douglas, S. M.; Liu, M. H.; Sharma, J.; Cheng, A. C.; Leung, A.; Liu, Y.; Shih, W. M.; Yan, H., Multilayer DNA Origami Packed on a Square Lattice. *J. Am. Chem. Soc.* **2009**, *131*, 15903-15908.
13. Rothemund, P. W. K., Folding DNA to Create Nanoscale Shapes and Patterns. *Nature* **2006**, *440*, 297-302.
14. Andersen, E. S.; Dong, M. D.; Nielsen, M. M.; Jahn, K.; Lind-Thomsen, A.; Mamdouh, W.; Gothelf, K. V.; Besenbacher, F.; Kjems, J., DNA Origami Design of Dolphin-shaped Structures with Flexible Tails. *Acs Nano* **2008**, *2*, 1213-1218.
15. Zhao, Z.; Yan, H.; Liu, Y., A Route to Scale Up DNA Origami Using DNA Tiles as Folding Staples. *Angew. Chem.-Int. Edit.* **2010**, *49*, 1414-1417.
16. Kuzuya, A.; Komiyama, M., Design and Construction of a Box-Shaped 3D-DNA Origami. *Chem. Commun.* **2009**, 4182-4184.
17. Andersen, E. S.; Dong, M.; Nielsen, M. M.; Jahn, K.; Subramani, R.; Mamdouh, W.; Golas, M. M.; Sander, B.; Stark, H.; Oliveira, C. L. P., *et al.*, Self-Assembly of a Nanoscale DNA Box with a Controllable Lid. *Nature* **2009**, *459*, 73-76.
18. Ke, Y. G.; Sharma, J.; Liu, M. H.; Jahn, K.; Liu, Y.; Yan, H., Scaffolded DNA Origami of a DNA Tetrahedron Molecular Container. *Nano Lett.* **2009**, *9*, 2445-2447.
19. Dietz, H.; Douglas, S. M.; Shih, W. M., Folding DNA into Twisted and Curved Nanoscale Shapes. *Science* **2009**, *325*, 725-730.
20. Douglas, S. M.; Dietz, H.; Liedl, T.; Hogberg, B.; Graf, F.; Shih, W. M., Self-Assembly of DNA into Nanoscale Three-Dimensional Shapes. *Nature* **2009**, *459*, 414-418.

21. Han, D. R.; Pal, S.; Liu, Y.; Yan, H., Folding and Cutting DNA into Reconfigurable Topological Nanostructures. *Nat. Nanotechnol.* **2010**, *5*, 712-717.
22. Hogberg, B.; Liedl, T.; Shih, W. M., Folding DNA Origami from a Double-Stranded Source of Scaffold. *J. Am. Chem. Soc.* **2009**, *131*, 9154-9155.
23. Yurke, B.; Turberfield, A. J.; Mills, A. P.; Simmel, F. C.; Neumann, J. L., A DNA-Fuelled Molecular Machine Made of DNA. *Nature* **2000**, *406*, 605-608.
24. Bath, J.; Turberfield, A. J., DNA Nanomachines. *Nat. Nanotechnol.* **2007**, *2*, 275-284.
25. Mao, C. D.; Sun, W. Q.; Shen, Z. Y.; Seeman, N. C., A Nanomechanical Device Based on the B-Z Transition of DNA. *Nature* **1999**, *397*, 144-146.
26. Yan, H.; Zhang, X. P.; Shen, Z. Y.; Seeman, N. C., A Robust DNA Mechanical Device Controlled by Hybridization Topology. *Nature* **2002**, *415*, 62-65.
27. Turberfield, A. J.; Mitchell, J. C.; Yurke, B.; Mills, A. P.; Blakey, M. I.; Simmel, F. C., DNA Fuel for Free-Running Nanomachines. *Phys. Rev. Lett.* **2003**, *90*, 1-4.
28. Green, S. J.; Lubrich, D.; Turberfield, A. J., DNA Hairpins: Fuel for Autonomous DNA Devices. *Biophys. J.* **2006**, *91*, 2966-2975.
29. Sherman, W. B.; Seeman, N. C., A Precisely Controlled DNA Biped Walking Device. *Nano Lett.* **2004**, *4*, 1203-1207.
30. Shin, J. S.; Pierce, N. A., A Synthetic DNA Walker for Molecular Transport. *J. Am. Chem. Soc.* **2004**, *126*, 10834-10835.
31. Yin, P.; Yan, H.; Daniell, X. G.; Turberfield, A. J.; Reif, J. H., A Unidirectional DNA Walker that Moves Autonomously Along a Track. *Angew. Chem.-Int. Edit.* **2004**, *43*, 4906-4911.

32. Lund, K.; Liu, Y.; Lindsay, S.; Yan, H., Self-Assembling a Molecular Pegboard. *J. Am. Chem. Soc.* **2005**, *127*, 17606-17607.
33. Park, S. H.; Pistol, C.; Ahn, S. J.; Reif, J. H.; Lebeck, A. R.; Dwyer, C.; LaBean, T. H., Finite-Size, Fully Addressable DNA Tile Lattices Formed by Hierarchical Assembly Procedures. *Angew. Chem.-Int. Edit.* **2006**, *45*, 735-739.
34. Voigt, N. V.; Topping, T.; Rotaru, A.; Jacobsen, M. F.; Ravnsbaek, J. B.; Subramani, R.; Mamdouh, W.; Kjems, J.; Mokhir, A.; Besenbacher, F., *et al.*, Single-Molecule Chemical Reactions on DNA Origami. *Nat. Nanotechnol.* **2010**, *5*, 200-203.
35. Tumpane, J.; Sandin, P.; Kumar, R.; Powers, V. E. C.; Lundberg, E. P.; Gale, N.; Baglioni, P.; Lehn, J. M.; Albinsson, B.; Lincoln, P., *et al.*, Addressable High-Information-Density DNA Nanostructures. *Chem. Phys. Lett.* **2007**, *440*, 125-129.
36. Tumpane, J.; Kumar, R.; Lundberg, E. P.; Sandin, P.; Gale, N.; Nandhakumar, I. S.; Albinsson, B.; Lincoln, P.; Wilhelmsson, L. M.; Brown, T., *et al.*, Triplex Addressability as a Basis for Functional DNA Nanostructures. *Nano Lett.* **2007**, *7*, 3832-3839.
37. Mastroianni, A. J.; Claridge, S. A.; Alivisatos, A. P., Pyramidal and Chiral Groupings of Gold Nanocrystals Assembled Using DNA Scaffolds. *J. Am. Chem. Soc.* **2009**, *131*, 8455-8459.
38. Aldaye, F. A.; Sleiman, H. F., Dynamic DNA Templates for Discrete Gold Nanoparticle Assemblies: Control of Geometry, Modularity, Write/Erase and Structural Switching. *J. Am. Chem. Soc.* **2007**, *129*, 4130-+.
39. Pal, S.; Deng, Z. T.; Ding, B. Q.; Yan, H.; Liu, Y., DNA-Origami-Directed Self-Assembly of Discrete Silver-Nanoparticle Architectures. *Angew. Chem.-Int. Edit.* **2010**, *49*, 2700-2704.

40. Sharma, J.; Ke, Y. G.; Lin, C. X.; Chhabra, R.; Wang, Q. B.; Nangreave, J.; Liu, Y.; Yan, H., DNA-Tile-Directed Self-Assembly of Quantum Dots into Two-Dimensional Nanopatterns. *Angew. Chem.-Int. Edit.* **2008**, *47*, 5157-5159.
41. Wilner, O. I.; Weizmann, Y.; Gill, R.; Lioubashevski, O.; Freeman, R.; Willner, I., Enzyme Cascades Activated on Topologically Programmed DNA Scaffolds. *Nat. Nanotechnol.* **2009**, *4*, 249-254.
42. 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.
43. Chen, J. H.; Seeman, N. C., Synthesis from DNA of a Molecule with the Connectivity of a Cube. *Nature* **1991**, *350*, 631-633.
44. Goodman, R. P.; Schaap, I. A. T.; Tardin, C. F.; Erben, C. M.; Berry, R. M.; Schmidt, C. F.; Turberfield, A. J., Rapid Chiral Assembly of Rigid DNA Building Blocks for Molecular Nanofabrication. *Science* **2005**, *310*, 1661-1665.
45. Sacca, B.; Meyer, R.; Feldkamp, U.; Schroeder, H.; Niemeyer, C. M., High-Throughput, Real-Time Monitoring of the Self-Assembly of DNA Nanostructures by FRET Spectroscopy. *Angew. Chem.-Int. Edit.* **2008**, *47*, 2135-2137.
46. Nangreave, J.; Yan, H.; Liu, Y., Studies of Thermal Stability of Multivalent DNA Hybridization in a Nanostructured System. *Biophys. J.* **2009**, *97*, 563-571.
47. Sandin, P.; Lincoln, P.; Albinsson, B., Conformational Flexibility in DNA Nanoconstructs: A Time-resolved Fluorescence Resonance Energy Transfer Study. *Journal of Physical Chemistry C* **2008**, *112*, 13089-13094.

48. Sandin, P.; Tumpane, J.; Borjesson, K.; Wilhelmsson, L. M.; Brown, T.; Norden, B.; Albinsson, B.; Lincoln, P., Thermodynamic Aspects of DNA Nanoconstruct Stability and Design. *Journal of Physical Chemistry C* **2009**, *113*, 5941-5946.
49. Bloomfield, V. A.; Crothers, D. M.; Tinoco, I. J., *Nucleic Acids - Structures, Properties and Function*. University Science Books: Sausalito, CA, 2000.
50. Tornøe, C. W.; Christensen, C.; Meldal, M., Peptidotriazoles on Solid Phase: [1,2,3]-Triazoles by Regiospecific Copper(I)-Catalyzed 1,3-Dipolar Cycloadditions of Terminal Alkynes to Azides. *J. Org. Chem.* **2002**, *67*, 3057-3064.
51. Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B., A Stepwise Huisgen Cycloaddition Process: Copper(I)-Catalyzed Regioselective "Ligation" of Azides and Terminal Alkynes. *Angew. Chem.-Int. Edit.* **2002**, *41*, 2596-2599.

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-trisubstituted-benzene node¹ is positioned in the middle (N).

Additional evidence of formation of DNA-phenalene

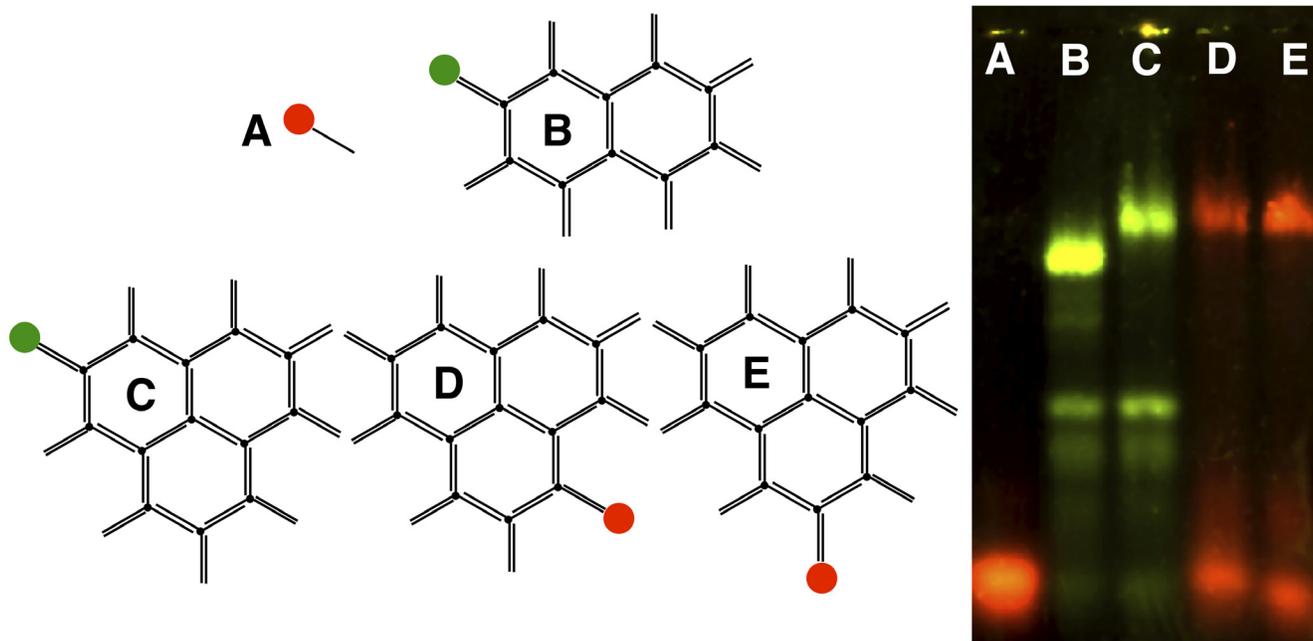


Figure S2. Three-cell construct DNA-phenalene with fluorophore label following the schematic to the left. Analysis was made using 4.5% MetaPhor agarose gel electrophoresis. Red and green dots indicate the position of the ROX and Cy3 label, respectively, used for visualization in gel scanner (Typhoon 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

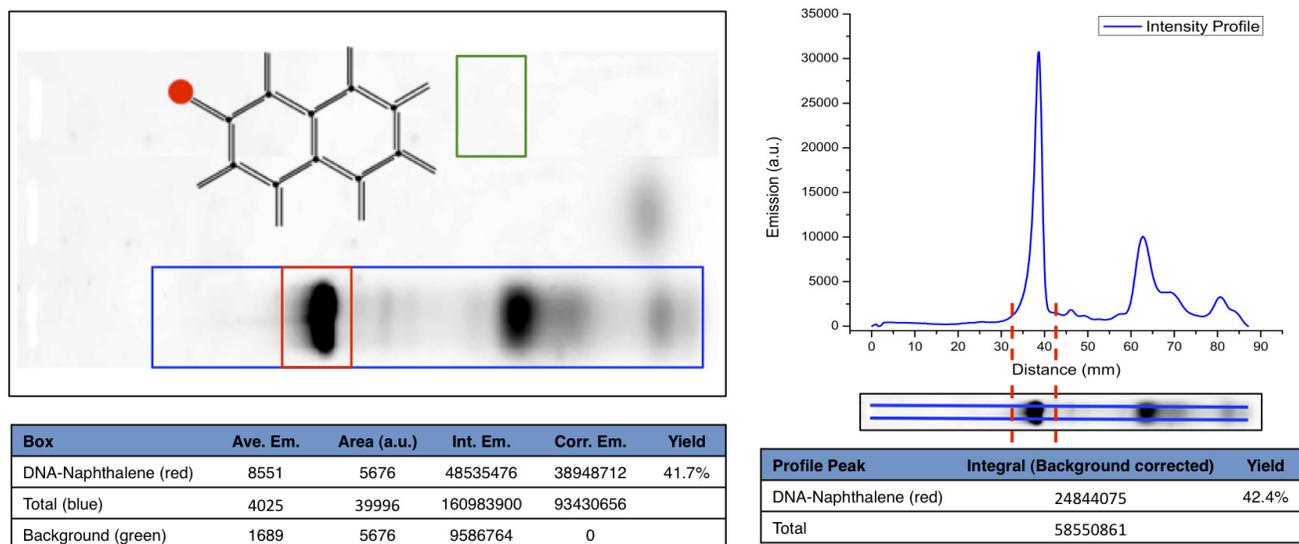


Figure S3. Illustration of yield analysis using emission data from gel electrophoresis results. Emission was collected using a Typhoon 9410 gelscanner (GE Healthcare). Cy3 (red dot) is the fluorophore used; excited at 532 nm (SYAG laser) and with a 580 nm bandpass filter (580 BP 30) to collect the emission. 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 than 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

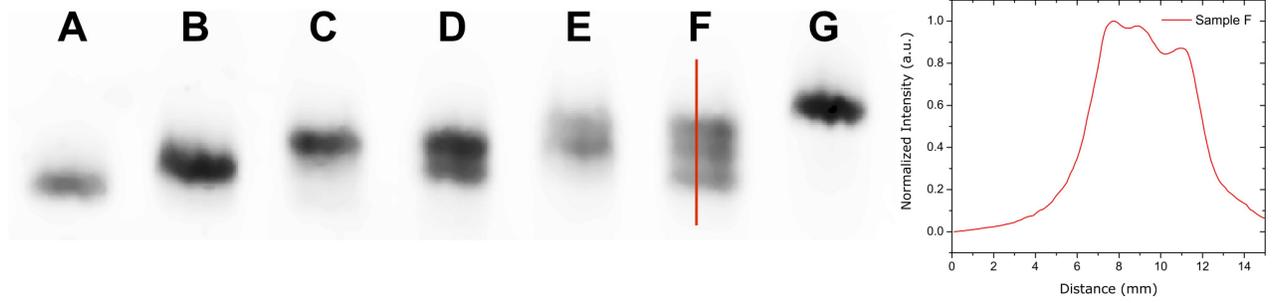


Figure S4. (Left) Zoom of bands of interest from Figure 2 in the main text. Red line gives the intensity profile of sample F (Right). Three strongly overlapping bands can be seen, corresponding to binding of one, two and three nodes to the two-cell construct.

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, \dots, W . The equilibrium of ring formation is:



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 :

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

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

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

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

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:

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

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

$$[x] = [-A -]K_a[-B -]K_b \cdots [-W -]K_w K_x \quad (S4)$$

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} :

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

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

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

if $a > b$, then

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

and if $a < b$, then

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

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

$$[x] = \frac{(a - [A - B])(b - [A - B])}{a} K_a \frac{(b - [B - C])(c - [B - C])}{b} K_b \dots \frac{(w - [W - A])(a - [W - A])}{w} K_w K_x \quad (\text{S8})$$

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

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

Let

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

where m is the average concentration in the sample, thus $\sum r_i = 0$. 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 ($Y = x/m$):

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

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:

$$\ln(1 - \alpha) - \ln(\kappa) = \frac{1}{\alpha} \sum (\min(r_i, r_{i+1}) - r_i) - \frac{1}{2\alpha^2} \sum (\min(r_i, r_{i+1})^2 - r_i^2) + \frac{1}{3\alpha^3} \sum (\min(r_i, r_{i+1})^3 - r_i^3) + \dots \quad (\text{S11})$$

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

$$\begin{aligned}
\sum \left(\min(r_i, r_{i+1}) - r_i \right) &= \sum \left(\min(r_i, r_{i+1}) - \frac{1}{2}(r_i + r_{i+1}) \right) = -\frac{1}{2} \sum |r_i - r_{i+1}| = -\frac{1}{2} n \langle |r_i - r_{i+1}| \rangle \\
\sum \left(\min(r_i, r_{i+1})^2 - r_i^2 \right) &= \sum \left(\min(r_i, r_{i+1})^2 - \frac{1}{2}(r_i^2 + r_{i+1}^2) \right) = 0 \\
\sum \left(\min(r_i, r_{i+1})^3 - r_i^3 \right) &= \sum \left(\min(r_i, r_{i+1})^3 - \frac{1}{2}(r_i^3 + r_{i+1}^3) \right) = -\frac{1}{2} \sum |r_i^3 - r_{i+1}^3| = -\frac{1}{2} n \langle |r_i^3 - r_{i+1}^3| \rangle \\
&\dots
\end{aligned} \tag{S12}$$

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 $\sin(u) > \cos(u)$:

$$\begin{aligned}
\langle |r_i^{2k+1} - r_j^{2k+1}| \rangle &= \frac{1}{2\sigma^2\pi} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |r_i^{2k+1} - r_j^{2k+1}| e^{-\frac{r_i^2+r_j^2}{2\sigma^2}} dr_i dr_j \\
&= 2 \int_{\pi/4}^{5\pi/4} (\sin^{2k+1}(u) - \cos^{2k+1}(u)) du \frac{1}{2\sigma^2\pi} \int_0^{\infty} t^{2k+2} e^{-\frac{t^2}{2\sigma^2}} dt
\end{aligned} \tag{S13}$$

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).

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

Taking κ to be unity, Eq.4 simplifies to Eq.2 for α large enough that the higher order terms can be neglected.

Finally, $k=0$ in S13, gives

$$\langle |r_i - r_j| \rangle = \langle \Delta \rangle = \frac{2\sigma}{\sqrt{\pi}} \quad (\text{S15})$$

as in Eq. 2 (in main text).

Analysis of Binomial Distributions

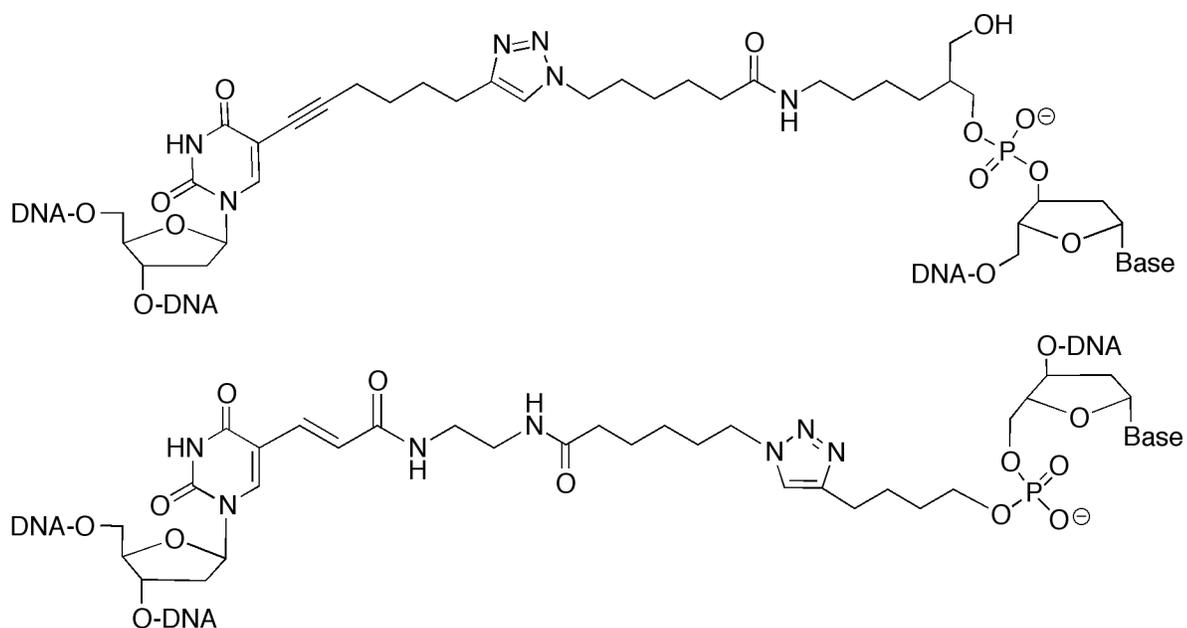


Figure S5. The two resulting triazole cross-links of the site-specific click reactions, from an unpaired thymine to either 3' or 5' phosphate of a complementary oligonucleotide. The relatively short linkers ensure reaction specificity.

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
		4		1	1
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
		3		1	1
	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
	2		1	1	3
	1		2	2	2, 1
	0		1	1	1, 1, 1
	1		1	1	2
	0		1	1	1, 1
	0		1	1	1

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

Structure	Binomial Terms	Values
	$\frac{6}{6} \binom{1}{1} Q(5)$	0.3791
	$\frac{5}{6} \binom{2}{5} Q(4)$	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
	$\frac{5}{5} \binom{1}{1} R(4)$	0.4602
	$\frac{4}{5} \binom{2}{4} R(4)$	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
	$\frac{4}{4} \binom{1}{1} S(3)$	0.5588
	$\frac{3}{4} \binom{2}{3} S(3)$	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
	$\frac{3}{3} \binom{1}{1} T(2)$	0.6784
	$\frac{2}{3} \binom{2}{2} T(1)$	0.1937
	$\frac{1}{3} \left(\frac{2}{2} T(1) + \frac{3}{1} T(0) \right)$	0.1279

Structure	Binomial Terms	Values
	$\frac{2}{2} \binom{1}{1} U(1)$	0.8237
	$\frac{1}{2} \binom{2}{2} U(0)$	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* 2010².

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.