DNA Knitting: robust and programmable fabrication of DNA nano-structures at sub-nanometer resolution

Technical Report

Ali A Atiia
School of Computer Science
McGill University

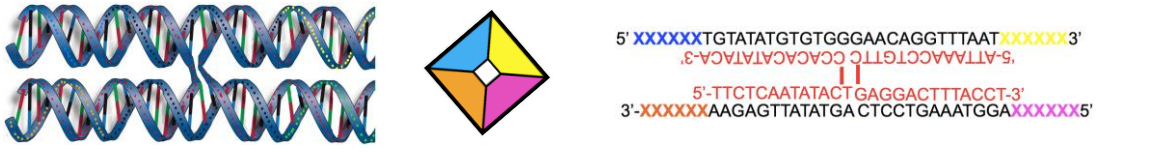
2017

1 Executive Summary

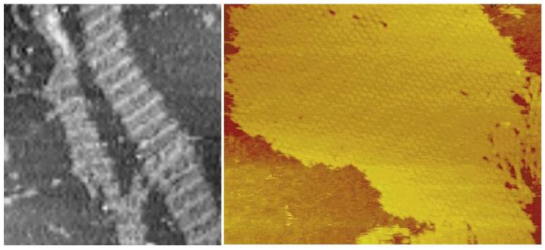
The current state-of-the-art method for creating DNA nanostructures, DNA Origami [1], involves the synthesis of hundreds of short strands (~200 strands, totaling 7249 bases, for a 100x100 nanometer square), driving the cost of of large-scale fabrication of DNA nano-structures to prohibitive levels. In addition, it involves the folding of a long DNA scaffolding strand which imposes a limit on the maximum achievable resolution (1.8 nm = 5.3 DNA bases). Since the long scaffold is viral DNA, it also limits researchers' access (since it's a regulated material) and applicable contexts (e.g. medical applications). The presented method, termed "DNA knitting", aims to overcome these limitations in terms of: resolution, availability, and cost. The aim is to achieve the ultimate fabrication resolution possible which is 1 base of DNA (~0.34 nm), use only synthetic DNA (no viral DNA involved), and reduce the cost of the each *de novo* fabrication to a negligible amount while keeping the programmability feature. The approach differs from "DNA origami" in that (1) there is no long (virus-based) scaffold, (2) programmability is achieved by polymerase-driven synthesis that faithfully follows a "program" in the form polymerase-guiding primers, and (3) chemical synthesis is kept to a minimum for each new design (namely a small number of primers that guide the construction process).

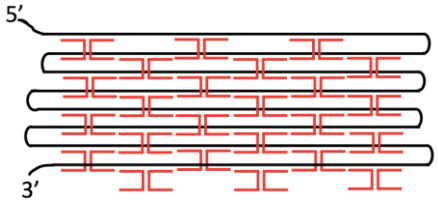
2 Background

Investigations into the potential of DNA structures that could be used for nanotechnological purposes had been pioneered by Ned Seeman, who showed as early as 1982 that the flexibility of DNA double-helix could enable creating DNA-based units for the fabrication of nanostructures [2]. The DNA double-crossover (DX) units (Figure 1a, left), which manifest in various conformations [3] and were initially inspired by naturally occurring Holliday junctions[4], have since then been the basis for many demonstrated 2-dimensional DNA nanotechnological constructions [5][1][6][7]. DNA DX units encoding a set of Wang tiles (Figure 1a, center) attach by virtue of "sticky ends" (ssDNA available for hybridizing to complementary sequence) of a DX complex (Figure 1a, right). In fact a single DX tile can grow in 2D, by virtue of its own sticky ends complementarity. In the example DX complex shown in Figure 1a (right), blue-to-purple and yellow-to-orange complementarity can result in a continuous aggregation of the same tile on the 2D plane. Figure 1b shows atomic force microscopy images of DX-based DNA lattices reported in [5] and [7], at 300 and 600 nanometer resolution, respectively. Winfree's model [8] of a Turing-universal DNA self-assembly model, which is founded on Wang's Turing-universal tiling system [9], also established a link between molecular computing and DNA nanotechnology whereby Wang tiles were implemented using DX units.

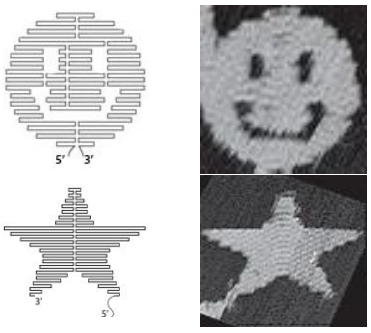


(a)





(b)



(c)

Figure 1: DNA as nano-fabrication material. **(a)** left: the DNA double-crossover (DX) unit; two double helices are stacked by virtue of complementary strands crossing over from one strand to continue complementing with another; center: an example Wang tile; right: a DX molecule representing the Wang tile. **(b)** Atomic force microscopy (AFM) images of 2D fabrication based on aggregating DX molecules reported in [5] (1998) and [7] (2005), at 300- and 600-nanometer resolution, respectively. **(c)** top: the DNA origami method; a long scaffold ssDNA (black) is folded into a specific shape by short ssDNA stapler strands (red); bottom: example DNA structures (smiley and star) fabricated using the origami method, the AFM images are at 165-nm resolution.

Despite the versatility of reported DNA lattices, control over their aggregation was not yet possible. In 2006, Rothemund [1] reported a breakthrough method, termed "DNA origami", which added the programmability feature. In this method, a long 7249-nt single stranded viral ssDNA (virus M13mp18) serves as a scaffold that twists and turns into a certain shape depending on what short ssDNA stapler strands are present in the mix. Stapler strands effectively weave DX units out of the scaffold strand. Figure 1c (top) shows an example of the mechanism, where a long scaffold strand (black) is folded into a rectangle by virtue of the short stapler strands (red). A computer-aided design can generate the set of stapler strands that can fold the viral DNA into a certain shape, and so different sets of stapler strands (~200) strands fold the same viral DNA into different shapes (Figure 1c, bottom) with impressive folding accuracy.

3 Method:

3.1 Overview

The groundbreaking "DNA origami" method [1] brought the programmability aspect into DNA nanotechnology [10] allowing precise control over the shape of DNA nano-structures. The method still fundamentally follows the geometrical principles of the double-crossover (DX) (section 2), but introduces the concept of "folding" the same long single-stranded DNA (ssDNA) into a different shapes depending on what set of short stapler strands ("program") are added in the mix. Figure 2a shows an example folding where

stapler strands (red) fold the "scaffold" strand (black) into a rectangular shape (the 5′ and 3′ ends of the long scaffold indicated). The scaffold used in DNA origami is a 7249-nt viral (M13mp18) ssDNA which when folded into a square area results in a DNA complex of ~100x100 nanometer (nm) area. Along its folding path, the viral scaffold must conform to the geometrical requirements of the DX, whereby crossover points must be at least 1 helical half-turn apart (a half-turn is the distance in bases that it takes for the strand to make 180 degrees rotation around the helical axis). This imposes a theoretical minimum of 1.8 nm = 5.3 DNA bases before the scaffold can crossover to another double-helix. The use of long viral DNA scaffold strand also limits researchers' access (since it's regulated material) and applicable contexts (e.g. medical applications). In terms of fabrication cost, every shape requires acquiring new samples of the virus (provided commercially, e.g. NEB) and a new set of stapler strands (~200 strands, totaling 7249 bases) that must be synthesized (the current rates of DNA synthesis ~0.13$ per base).

The presented method, "DNA knitting", aims to overcome these limitations: resolution, availability, and cost. The goal is to achieve the ultimate fabrication resolution possible which is 1 base of DNA (~0.34 nm), use only synthetic DNA, and reduce the cost of the each fabrication to a negligible amount. The approach differs from "DNA origami" in that (1) there is no long scaffold, (2) programmability is achieved by polymerase-driven synthesis that faithfully follows a "program" in the form polymerase-guiding primers, and (3) chemical synthesis is kept to a minimum. In what follows we describe the method in general terms first, then provide more details on the experimental protocol.

3.2 Implementation

3.2.1 Template construction:

In our DNA Knitting method, the idea of a long scaffold (Figure 2a) is replaced by a collection of synthetic DNA strands that are assembled through a process of splint ligation shown schematically in Figure 2 (b-d). A collection of "horizontal" strands (black in Figure 2b) are ligated using a set of splint strands (grey in Figure 2b) resulting in longer continuous strands. The ligation product (Figure 2c, ligation spot indicated by yellow dots) is used as a template in a polymerase chain reaction (PCR) which selectively and exponentially amplifies the complete ligation strands (and only those), resulting in fully double-stranded DNA (dsDNA) (Figure 2d). The resulting PCR product is mixed

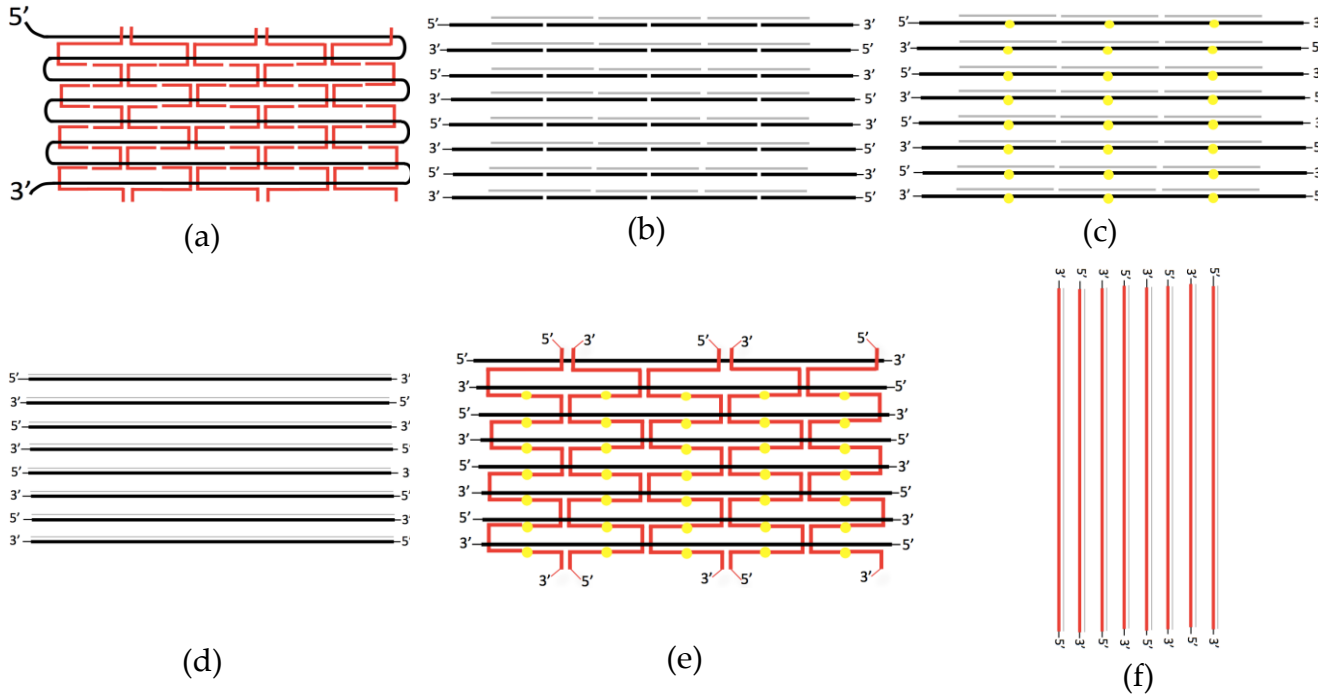


Figure 2: **DNA knitting vs. DNA origami method**. (a) The origami method showing the long scaffold viral strand (black) and the short stapler strands (red) folding it into a rectangular shape. (b-f) constructing a stock library horizontal and vertical dsDNA strands used in DNA knitting. Short chemically synthesized horizontal oligonucleotides (black) and splint strands (grey) are hybridized (b) and ligated (yellow dots in (c)); the ligation product serves as template in a PCR reaction (d) with primers at the 3′ extremities of each ligated horizontal strand resulting in a fully dsDNA. (e) dsDNA horizontal strands from (d) are hybridized with short vertical strands (red) and ligated (yellow dots). (f) The ligation product from (e) serves as template in a PCR reaction with primers at the 3′ extremities of each ligated vertical strand resulding in fully dsDNA.

with excess amount of stapler strands (red in Figure 2e), which complement the sense strands of the PCR product (black in Figure 2d) following the double-crossover principle (as in Figure 2(a), see also Figure 1a in Section 2). Stapler strands are ligated at their meeting points (yellow dots in Figure 2e), and the ligation product is used as template in a PCR reaction which amplifies the complete ligation product into fully dsDNA strands (Figure 2f). The "horizontal" and "vertical" dsDNA obtained in this process, Figure 2 (d) and (f), respectively, are the stock material which can themselves now be used as the template out of which a nano-structure of interest can be "knitted" .

3.2.2 Programmable sub-nanometer fabrication:

The horizontal and vertical template strands obtained in the previous phase, Figure 2 (d) and (f), respectively, are used as templates for the programmable fabrication of nano-structures of interest. The programmability

feature is achieved by virtue of primer-directed PCR amplification of horizontal or vertical strands at locations specified by one's choice of primers at the base level. In our context, the polymerase enzyme represents the "labor", dNTPs (A,G,C,T molecules) represent the "material", and the chosen primer oligonucleotides represent the blueprint (or the "program") guiding the enzyme to construct the nano-shape of interest. A set of primers (ssDNA synthesized commercially) are chosen such that they amplify each horizontal or vertical strand at a specific base location. Figure 3 shows a schematic representation where horizontal (black) and vertical (red) strands are represented as a grid, and a set of primers (blue dots) define where amplification starts and ends along a given strand. The region enclosed by a pair of primers is amplified to exponential amount, effectively excluding the region falling outside the primers' range (grey regions in Figure 3).

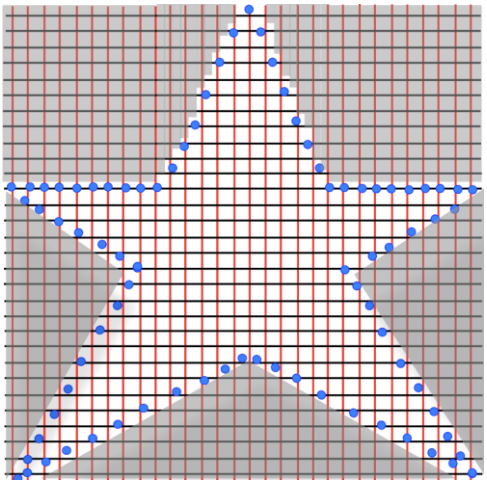


Figure 3: Programmability in DNA knitting method. The horizontal and vertical strands obtained in the template construction phase (setion 3.2.1) as templates for the programmable fabrication of nano-structures of interest. The choice of primer oligonucleotides (blue dots) determines the resulting nano-shape. Regions falling outside the range a given pair of primers are effectively amplified-out.

3.2.3 Experimental approach:

(1) 34 sets of DNA oligonucleotides, each set containing 10 oligonucleotides, and each nucleotide of length 32 bases, are software generated [11] to meet the following general combinatorial requirements:

I. Sequences have no more than 6 bases in common.
II. No self-complementarity of >= 6 bases.
III. Each sequence has 40% GC content.

Requirements I and II are chosen to maximize the hamming distance between sequences as much as computationally feasible (sequence design is a fundamentally hard problem [39]–[41]) while requirement III was inspired by the fact that the viral DNA sequences used in DNA origami has 40% GC.

34 sets × 10 oligonucleotides per set × 32 bases per oligonucleotides = 10880 bases in total have been chemically synthesized (Biocorp). The generated set represents black strands in Figure 2b. These strands are phosphorylated using polynucleotide kinase (NEB) in standard buffer conditions.

(2) 34 sets of DNA oligonucleotides, each set containing 9 oligonucleotides, and each nucleotide at length 32 bases, are software generated such that each set is complementary to the corresponding set in (1). These sets represent the splint strands (grey in Figure 2b).

34 sets × 9 oligonucleotides per set × 32 bases per oligonucleotides = 9792 bases in total have been chemically synthesized (Biocorp and IDT)

(3) The phosphorylated set from (1) and non-phosphorylated set from (2) are mixed at 1uM concentration of each strand in a ligation reaction using T4 DNA ligase (NEB) in standard buffer conditions (Figure 2c).

(4) 34 sets of horizontal primers are used to PCR-amplify the splint-ligated set from (3). Each set of primers contains two oligonucleotides: a forward and a backward primer that are software-generated to complement the 3′ extremities of each ligated horizontal strand in (3), and then chemically synthesized (Biocorp). The resulting dsDNA strands (Figure 2d) are each at 320-bp length.

(5) 20 sets of DNA oligonucleotides, each set containing 17 oligonucleotides, each nucleotide at length 32 bases, are software generated such that each set complements the horizontal strands (black in Figure 2e) in double-crossover fashion as shown (red in Figure 2e). These sets are phosphorylated using polynucleotide kinase (NEB) in standard buffer conditions.

20 sets × 17 oligonucleotides per set × 32 bases per oligonucleotides = 10880 bases in total have been chemically synthesized (Biocorp).

(6) The phosphorylated set from (5) and the dsDNA horizontal strands from (4) are mixed at 10:1 stoichiometry to favor the double-crossover assembly formation (Figure 2e) against the re-annealing of anti-sense (grey in Figure

2d) to sense strands in horizontal. The assembly is mixed in a ligation reaction using T4 DNA ligase (NEB) under standards conditions[1].

(7) 20 sets of vertical primers are used to PCR-amplify the ligation product from (6) (the splints being dsDNA PCR product in (4)). Each set of primers contains two oligonucleotides: a forward and a backward primer that are software-generated to complement the 3' extremities of each ligated vertical strands from (6), and then chemically synthesized (Biocorp). The resulting dsDNA strands (Figure 2f) are each 544-bp long.

The template DNA library is therefore composed of the resulting dsDNA in (4) and (7) corresponding to the horizontal and vertical scaffolds, respectively (Figure 2 (d) and (f)). Notice that: 34 horizontal strands × 320 base pair (bp) per strand = **10880 total bases** = 20 vertical strands × 544-bp per strand. The DNA lattice resulting for hybridizing the horizontal and vertical strands into the canvas structure (grid in Figure 3) corresponds to an area of 100×100 nanometer area.

4 Results and Progress:

Steps (1) to (4) (corresponding to Figure 2b-d) described above have been completed. Shown here are the ligation and PCR amplification results of the template's horizontal strands. Figure 4a shows the result of splint-ligation result describe in step (3) of the protocol (and shown schematically in Figure 2c). The ligation product is run on at 50 degrees on a 5% denaturing TBE-urea polyacrylamide gel. The lowest band at 32-nt shows non-ligated horizontal oligonucleotides and splint strands. The successive upper bands correspond to the expected lengths of 2×32=64, 3×32=96 ….. 10×32=320-nt. This ligation product is the template used in the subsequent PCR reaction generating the full-length 320-bp dsDNA horizontal strands described in Step (4) of the protocol (shown schematically in Figure 2d). The PCR result is shown in Figure 4b, where the full set (34 dsDNAs in total) of strands are run on 1.5% agarose gel run. The PCR product has been Sanger-sequenced, both forward and backward, at McGill University and Génome Québec Innovation Centre, confirming the integrity of

[1] In a preliminary trial experiment the success of ligation in double-crossover assemblies such as those shown in Figure 2e has been verified; data not shown here.

each band (ruling out erroneous ordering of splint ligation due to splint strands' mis-hybridization)[2].

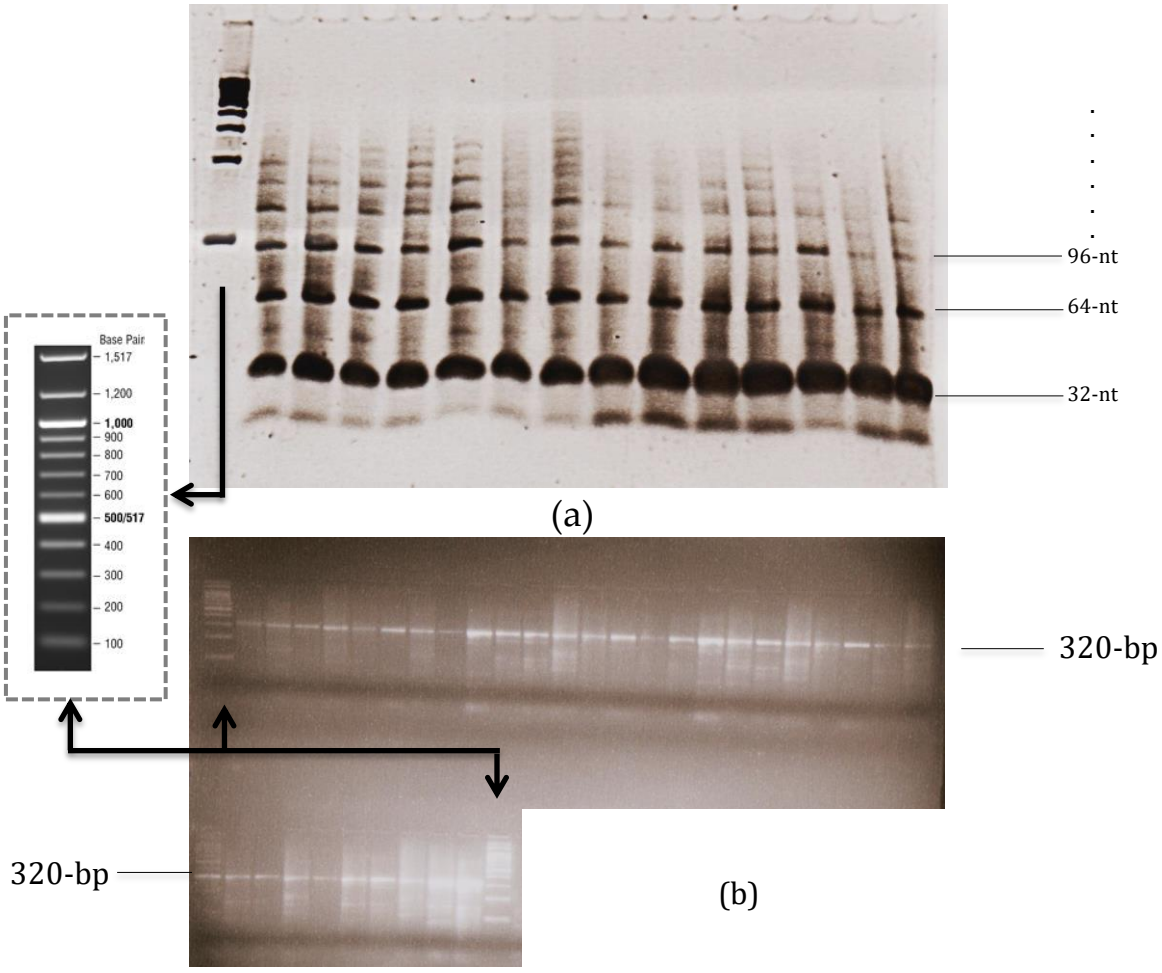


Figure 4: The ligation and PCR amplification results of horizontal library strands. Inset**:** a standard 100-bp marker. **(a)** Splint ligation results; left to right: well 1, marker; well 2-15, the ligation reaction of 14 out of 34 horizontal library strands. In each reaction, 10 32-nt horizontal oligonucleotides are mixed with 9 32-nt splint strands; the bands reflect the expected partial ligations corresponding at multiples of 32-nt: 2×32=64, 3×32=96 ….. 10×32=320-nt. Bands at 32-nt represent left over horizontal/splint strands. **(b)** PCR amplification of all using the ligation product as template; first and last wells: marker (inset), all others: the PCR product of dsDNA strands representing the full library of 34 horizontal library (represented schematically in the previous section in Figure 2(d)).

[2] The sequencing results and sequence alignment confirming 100% accuracy of all horizontal strands can be accessed thru: http://www.cs.mcgill.ca/~malsha17/Sequencing

References:

[1] P. W. K. Rothemund, "Folding DNA to create nanoscale shapes and patterns," *Nature*, vol. 440, no. 7082, pp. 297–302, Mar. 2006.

[2] N. C. Seeman, "Nucleic acid junctions and lattices," *J. Theor. Biol.*, vol. 99, no. 2, pp. 237–247, 1982.

[3] N. C. Seeman, H. Wang, X. Yang, F. Liu, C. Mao, W. Sun, L. Wenzler, Z. Shen, R. Sha, H. Yan, and others, "New motifs in DNA nanotechnology," *Nanotechnology*, vol. 9, no. 3, p. 257, 1998.

[4] Y. Liu and S. C. West, "Happy Hollidays: 40th anniversary of the Holliday junction," *Nat. Rev. Mol. Cell Biol.*, vol. 5, no. 11, pp. 937–944, 2004.

[5] E. Winfree, F. Liu, L. A. Wenzler, and N. C. Seeman, "Design and self-assembly of two-dimensional DNA crystals," *Nature*, vol. 394, no. 6693, pp. 539–544, 1998.

[6] C. Mao, W. Sun, and N. C. Seeman, "Designed two-dimensional DNA Holliday junction arrays visualized by atomic force microscopy," *J. Am. Chem. Soc.*, vol. 121, no. 23, pp. 5437–5443, 1999.

[7] D. Reishus, B. Shaw, Y. Brun, N. Chelyapov, and L. Adleman, "Self-assembly of DNA double-double crossover complexes into high-density, doubly connected, planar structures," *J. Am. Chem. Soc.*, vol. 127, no. 50, pp. 17590–17591, 2005.

[8] E. Winfree, "Algorithmic self-assembly of DNA," California Institute of Technology, 1998.

[9] H. Wang, "Proving theorems by pattern recognition—II," *Bell Syst. Tech. J.*, vol. 40, no. 1, pp. 1–41, 1961.

[10] N. C. Seeman, "DNA in a material world," *Nature*, vol. 421, no. 6921, pp. 427–431, 2003.

[11] A. Kick, M. Bönsch, and M. Mertig, "EGNAS: an exhaustive DNA sequence design algorithm," *BMC Bioinformatics*, vol. 13, no. 1, p. 138, 2012.