Biological Nanowires:

Integration of the silver(I) base pair into DNA with nanotechnological and synthetic biological applications

Simon Vecchioni

Submitted in partial fulfillment of the
requirements for the degree of
Doctor of Philosophy
in the Graduate School of Arts and Sciences

Columbia University

2019

© 2019
Simon Vecchioni
All rights reserved

Abstract

Biological Nanowires: Integration of the silver(I) base pair into DNA with nanotechnological and synthetic biological applications

Simon Vecchioni

Modern computing and mobile device technologies are now based on semiconductor technology with nanoscale components, i.e., *nanoelectronics*, and are used in an increasing variety of consumer, scientific, and space-based applications. This rise to global prevalence has been accompanied by a similarly precipitous rise in fabrication cost, toxicity, and technicality; and the vast majority of modern nanotechnology cannot be repaired in whole or in part. In combination with looming scaling limits, it is clear that there is a critical need for fabrication technologies that rely upon clean, inexpensive, and portable means; and the ideal nanoelectronics manufacturing facility would harness micro- and nanoscale fabrication and self-assembly techniques.

The field of molecular electronics has promised for the past two decades to fill fundamental gaps in modern, silicon-based, micro- and nanoelectronics; yet molecular electronic devices, in turn, have suffered from problems of size, dispersion and reproducibility. In parallel, advances in DNA nanotechnology over the past several decades have allowed for the design and assembly of nanoscale architectures with single-molecule precision, and indeed have been used as a basis for heteromaterial scaffolds, mechanically-active delivery mechanisms, and network assembly. The field has, however, suffered for lack of meaningful modularity in function: few designs to date interact with their surroundings in more than a mechanical manner.

As a material, DNA offers the promise of nanometer resolution, self-assembly, linear shape, and connectivity into branched architectures; while its biological origin offers information storage, enzyme-compatibility and the promise of biologically-inspired fabrication through synthetic biological means.

Recent advances in DNA chemistry have isolated and characterized an orthogonal DNA base pair using standard nucleobases: by bridging the gap between mismatched cytosine nucleotides, silver(I) ions can be selectively incorporated into the DNA helix with atomic resolution. *The goal of this thesis is to explore how this approach to "metallize" DNA can be combined with structural DNA nanotechnology as a step toward creating electronically-functional DNA networks.*

This work begins with a survey of applications for such a transformative technology, including nanoelectronic component fabrication for low-resource and space-based applications. We then investigate the assembly of *linear* Ag^+-functionalized DNA species using biochemical and structural analyses to gain an understanding of the kinetics, yield, morphology, and behavior of this orthogonal DNA base pair. After establishing a protocol for high yield assembly in the presence of varying Ag^+ functionalization, we investigate these linear DNA species using electrical means. First a method of coupling orthogonal DNA to single-walled carbon nanotubes (SWCNTs) is explored for self-assembly into nanopatterned transistor devices. Then we carry out scanning tunneling microscope (STM) break junction experiments on short polycytosine, polycationic DNA duplexes and find increased molecular conductance of at least an order of magnitude relative to the most conductive DNA analog.

With an understanding of linear species from both a biochemical and nanoelectronic perspective, we investigate the assembly of *nonlinear* Ag^+-functionalized DNA species. Using rational design principles gathered from the analysis of linear species, a *de novo* mathematical framework for understanding generalized DNA networks is developed. This provides the basis for a computational model built in Matlab that is able to design DNA networks and nanostructures using arbitrary base parity. In this way, DNA nanostructures are able to be designed using the dC:Ag^+:dC base pair, as well as any similar nucleobase or DNA-inspired system (dT:Hg^{2+}:dT, rA:rU, G4, XNA, LNA, PNA, etc.). With this foundation, three general classes of DNA tiles are designed with embedded nanowire elements: single crossover Holliday junction (HJ) tiles, T-junction (TJ) units, and double crossover (DX) tile pairs and structures. A

library of orthogonal chemistry DNA nanotechnology is described, and future applications to nanomaterials and circuit architectures are discussed.

Contents

List of Figures

Acknowledgements

My deepest gratitude goes out to Professor Shalom Wind for opening his doors to an unconventional candidate from a non-engineering background and giving me the opportunity to become a doctoral student in his group. His hands-off approach let me embark on creative escapades in odd directions, but excellent mentorship and patience kept me pointed the right way. At every turn, he gave me the opportunity to explore and grow, and I've never met a kinder scientist. I thank Dr. Lynn Rothschild for her support and vision from the early beginnings of a group project idea, and for fighting off the demons of practicality and near-sightedness. She helped me follow my childhood dream of working at NASA. I would never have made it through this process without the continual support and encouragement of my research mentors.

I thank Professor Gary Wessel for lending his creative whetstone in a time of need, and for his advice to follow my nose. A huge thank you to my iGEM teammates and mentors for listening to and supporting my strange ideas. Most notably, to Emily Toomey, a sincere thank you for pragmatism and perseverance in equal measure, and for staying on the ship when things weren't at their best. I thank Dr. Mark Capece for incredible patience and believing in collaboration. I owe it to the former members of the Wind group for sharing their deep expertise: thank you, Erika, Diego, Risheng, Haogang and Jinyu (l'abbiamo fatto!). To Dr. Dan Paley and the SMCL team, thanks for letting me discover through play. I wish to thank Professor Ned Seeman and Dr. Ruojie Shah for lending their time and considerable expertise to this project in its last year. Thanks to Professor Henry Hess for signing onto this not-quite-standard BME project, and for supporting our last year.

This project was funded first and foremost by a National Space Technology Research Fellowship, and my thanks goes out to the STMD team for their constant support and allowing me to represent them in Washington, to Claudia and Michael for reading all my reports, and to Ingrid for her open-door welcome

at Ames. This project was also supported by the Rhode Island Space Grant, by Stanford VPUE, the Brown Science Foundation, Brown UTRA, the Universities Space Research Association, and the NSF.

Most fundamentally, I want to thank my parents for being with me every step of the way, for giving all of themselves to keep me moving forward—no matter the challenge. I would not have gotten through the chaotic engine of graduate school without their help, even if all it took sometimes was an ice cream delivery. Finally, thank you Adaora. You kept me sane and focused, and reminded me that there's more to life than writing. I wouldn't have made it without you.

Preface: Structure of this Thesis

Chapter 1: Introduction: Significance of DNA nanowire and nanostructure design

We begin with a discussion of DNA in the context of nanotechnology, synthetic biology and space technology. DNA is an attractive candidate for integration into nanoelectronics as a biological nanowire due to its linear geometry, definable base sequence, easy, inexpensive and non-toxic replication and self-assembling properties. DNA has been developed as a powerful building block for a diverse set of technologies since the early 1980s, and developments in DNA conductivity might allow for new applications in earth- and space-based applications. We discuss the context and implications of the Ag^+ base pair in DNA and set the stage for experimentation on this system.

Chapter 2: Biochemical Analysis of DNA Nanowires: A foundation for nanotechnology

To use Ag^+-functionalized DNA as a nanoscale building block, it was first necessary to understand the behavior of DNA with inherently orthogonal chemistry. To map the functionality and biostability of this system, we built and characterized internally-functionalized DNA nanowires through non-canonical, Ag^+-mediated base pairing in duplexes containing cytosine-cytosine mismatches. We utilize a variety of biochemical analyses to characterize the structure, kinetics, yield, and stability of the dC:Ag^+:dC bond. Reaction conditions are identified that control product size and reliability, while sequence design constraints are explored to understand the effect of C:Ag^+:C bonds on neighboring base pair assembly. We demonstrated continuous ion chain formation in oligonucleotides of 11-50 nucleotides (nt), and enzyme ligation of mixed strands up to six times that length. This construction is feasible without detectable silver nanocluster contaminants. We conclude that the C-Ag^+-C bond forms DNA duplexes with replicable geometry, predictable thermodynamics, and tunable length. An optimized protocol for the assembly of Ag^+-functionalized DNA duplexes with minimal defects is developed. Finally, functional gene parts for the synthesis of DNA- and RNA-based, C-Ag^+-C duplexes in a cell-free system are

constructed in an *Escherichia coli* expression plasmid and added to the open-source BioBrick Registry, suggesting a path toward inexpensive industrial production.

Chapter 3: Finding the Spark: Electrical assay of silver(I) DNA nanowires

Linear DNA assemblies are shown to be thermostable and free of contaminants, and in this chapter we explore the interface of these linear species with nanoscale nanoelectronic arrays. First we discuss the diverse history of DNA conductivity and the current state-of-the-art. We then build on previous studies to link orthogonal DNA duplexes to single-walled carbon nanotubes (SWCNTs). The production of nanoscale heterostructures is carried out at high yield and assayed via atomic force microscopy (AFM). This chemical coupling is designed for integration into nanoscale transistors. The design, construction and assay of SWCNT field-effect transistors is outlined, and charge transport is measured across SWCNT bundles. Finally, molecular conductance in polycytosine, Ag^+-functionalized DNA nanowires is measured using a scanning tunneling microscope (STM) break junction technique. It is found that 11 bp duplexes composed of dC:Ag^+:dC bonds are at least an order of magnitude higher than the most conductive Watson-Crick (WC) duplex of similar length; and our results are compared to studies in the literature on short duplexes with similar baseline conductance values. We conclude that the dC:Ag^+:dC bond confers increased molecular conductance on DNA, rendering metal paired duplexes and heterostructures a functional nanomaterial for electronic architectures.

Chapter 4: DNA by Design: *De novo* computational framework for DNA sequence design and nanotechnology

Chemical analysis of metalized DNA made it quite clear that traditional models of DNA thermodynamics are insufficient to predict and control the self-assembly of orthogonally-paired nucleotides. The ability to add or remove environmental parity between nucleobases alters the fundamental language of

dogmatic DNA assembly. This change in behavior necessitates an accompanying shift in computational design. We begin by exploring the state-of-the-art in DNA modeling, and include both sequence analysis and sequence design practices. We then start from first principles and establish a mathematical basis for heterostructure and 'nmer' analysis in connected DNA networks that operates without assumptions about nucleobase parity. A generalized search algorithm is then constructed in Matlab and implemented using evolutionary techniques. We then discuss DNA nanostructure design criteria, operation efficiency in differentially-connected networks, and the application of computationally-aided sequence design for nanotechnological applications. In sum, we present a novel computational tool for geometry-informed optimization of DNA networks. This tool is meant to enable design of both linear (duplexed, hairpin or otherwise) and nonlinear polynucleotide assemblies with inherent modularity for parity, metalation, or more exotic nucleotide substitutions that may arise from advances in synthetic biology, nanomaterials and nanomedicine.

Chapter 5: Silver(I) nanotechnology: DNA nanostructures with Ag^+-mediated, conductive base pairing for self-assembling electronic arrays

The ability to incorporate metal ions into DNA without chemical modification, nucleobase substitution, or significant perturbation to the B-form helix was a significant impetus for the development of orthogonal DNA nanotechnology that takes advantage of the electrical and chemical advantages of these ionic components. First we review the history and design of structural DNA nanotechnology to date, and suggest avenues of modification for Ag^+ (and other orthogonal element) incorporation. We then present three broad classes of DNA tiles; we design and build increasingly exotic Ag^+- and other heterostructure-functionalized prototypes; and we compare them to their historical, Watson Crick analogs. Single-crossover Holliday junction (HJ) tiles are fractured and reconnected with orthogonal components, while other designs involve the fusion of strand oligos into a secondary layer for stacked

nanowire assembly. With these HJ lattices, the assembly of guanine tetraplexes into DNA nanostructures is observed for the first time. Sticky end / kissing loop coordinating T-junction (TJ) tiles are designed into windowed lattices and rings with Ag^+-functionalized connections, and flat sheets are attained larger than ten microns in size. Finally, double crossover (DX) units are designed with current and former members of the Seeman Group at NYU, and we demonstrate computationally-optimized DNA nanostructures of novel rotational and chemical behavior over long range. In sum, we present a library of DNA nanostructures based on a variety of tile sizes and integration schemes to serve as a foundation for future work on the electrical functionalization of DNA (and DNA-inspired) networks.

Chapter 1

Introduction: Significance of DNA nanowire and nanostructure design

"OR: Off-label uses of DNA (with apologies to Ned Seeman)"

-Professor Shalom Wind

1.1. Mesoscale technologies for Earth- and space-based applications

The origin of nanotechnology as a field of study is often attributed to Richard Feynman's 1959 talk at Caltech entitled *There's Plenty of Room at the Bottom*, where he first described the 'inversion' of the electron microscope, an attention to biological systems, and the ordering of single atoms for use in future technologies.[1] Since Feynman's talk, various advances introduced seminal changes in the understanding of matter at the mesoscale, such as the development of the scanning tunneling microscope (STM) in 1979,[2] and the subsequent invention of the atomic force microscope (AFM) in 1982;[3] but the tandem development of pattern transfer technologies key to nanolithography, paired with the ever-increasing market forces for semiconductor device fabrication, have generated some of the most ordered systems in the known universe. In this way, nanoelectronics has become a transformative technology for both industrial and personal computing applications, yet the limitations of traditional manufacturing methods make current prototypes expensive and difficult to mass produce.[4] Nanoelectronics—and, by extension, nanowires—produced by common lithographic techniques suffer from high resource cost, population heterogeneity, lack of structural precision, and difficulty of construction.[5-11] Furthermore, and most fundamentally, the design of ever denser arrays of circuit components require full nanofabrication technology overhaul on a regular basis: by design, nanoelectronics and the machines that build them are obsolete almost as soon as they reach the market. Consequently, conventional nanowire synthesis often requires prohibitively complex and expensive manufacturing infrastructure.

Beyond consumer electronics, there are myriad applications for nanoelectronics that require repair, fabrication and even design without the luxury of traditional, state-of-the-art manufacturing complexes. Space technology applications in particular would be vastly improved with *in situ* electronics manipulation. In its latest call for advanced space technologies, the National Aeronautics and Space

Administration (NASA) dedicates an entire chapter to nanotechnology and electronic systems (TA10-Nanotechnology[12]), and has called for the miniaturization of nanoelectronic system features to less than 10 nm by 2025. The secondary goal of this call was for the liberation of nanoscale technologies from the overwhelmingly macroscale fabrication toolkit employed in the manufacture of modern devices. At greater than $10,000 per pound of upmass, the prospect of traditional nanofabrication in low-Earth orbit or long-range missions is impractical, if not impossible. Furthermore, there exists to date no reliable technology to fabricate sub-5 nm features at high volume; nor is there a manufacturing process to produce nanoelectronic features of any size with pitch (spacing) below 25 nm, though plans for future semiconducting device technology development exist.[13] In the scope of space technology, such technologies will be unreliable and require exorbitantly heavy machinery to produce and maintain. As a result, unanticipated failures or altered device demands encountered during a mission would be unaddressable, offsetting the impact of current and proposed nanoscale systems.

This paradigmatic limitation has been a driving motivation for the development of throwaway, short term satellites powered by mobile phone technologies. These “cubesats” are built small, are relatively disposable, and maintained only within the operational lifetimes of the smart devices contained inside them. Anecdotally, the Rothschild lab has experienced extended delays in satellite launch: despite a resounding call for novel miniaturization and synthetic biological technologies, these missions are subject to postponement, technical failure, and change in mission scope as a result of satellite hardware, upmass regulation and flight windows. Though the Rothschild lab at NASA’s Ames Research Center has experienced recent success in payload launch in sister projects to this one, it remains an iron rule that all non-personnel technologies are subject to—and designed around—failure. The ability to manufacture, repair or redesign nanosystems outside the confines of a gravity well would represent a significant lexical shift in space technology capabilities, leaving aside the benefits to consumer electronics. In

keeping with this philosophy, NASA invested a full training grant in the present study for the development of orthogonal, synthetic-biology-based nanoelectronics technologies.

1.2. DNA as a building block

In light of miniaturization, upmass, and technology-development obstacles, the use of DNA as an alternative nanowire appears very attractive; with its linear configuration, alterable binding sites and sequence-defined geometry, the structure of DNA can be reliably controlled and replicated at low relative cost with error rates approaching ~10^{-6},[14] and can be dynamically manipulated in multiple ways, including through chemical linkers,[15-17] fluorescent markers,[18,19] ligand-protein,[20,21] and nanoparticle/nanotube attachment.[22-24] Careful sequence design in self-pairing oligos can even produce stranger structures such as self-cleaving hammerhead ribozymes[25] and viral pseudoknots.[26,27] In laboratories, investigators have been able to stabilize several structural variants such as left-handed helicies (Z-form DNA),[28] low-pH cytosine tetraplexes (*i-motif*),[29] and nonhelical arrays of guanine tetraplexes (G4).[30,31] The affinity of oligonucleotides for hydrogen-bond-driven heterostructures has led to the development of small molecule detection with molecule-specific oligonucleotides called aptamers, [32-35] leading to a variety of fluorescence-driven and redox-powered applications in sensing and antigen capture.[36,37]

In particular, DNA has been used as a structural polymer for organizing complex crystals, or *nanostructures*, by design since the 1980s.[38] Small molecules have been folded together to make larger assemblies, while conversely large packets of DNA have been folded down to form precisely-shaped, origami-like features.[39-42] A variety of medical technologies have been explored for these constructs, predominantly covering static or mechanically-active drug delivery to exploit the preferential shuttling of large biopolymers into oncogenic areas of the body.[43,44] For the most part, these nanostructures exist as mechanical or templating objects, contributing to their environments as inert structural assemblies.

Recent work has suggested a means for assembling DNA heterostructures using crystallographic assemblies of semiconductors inside DNA cages, but a reliable technology has yet to emerge.[45]

While other naturally-conducting biological nanowires and hybrids have been developed, including bacterial pili, coated nanoparticles, metalated protein filaments,[46-51] double-stranded DNA (dsDNA) remains at least an order of magnitude smaller at a diameter of 2 nm—a size scale appropriate for future semiconductor device integration at sub-10 nm production nodes.[13,52] Relative to inorganic nanowire production, the annealing temperatures of DNA are significantly lower, the reagents to construct them far less toxic, and the technical requirements greatly reduced. As in any biosynthetic system, there is the potential for rapid, cheap, and environmentally-safe mass production from cellular components that can be easily tuned for different purposes. Furthermore, this system does not require the heavy infrastructure utilized for inorganic production, and thus it has the potential for distributed manufacturing in low-resource environments, including in space, the moon or Mars.

1.3. Metallo-DNA: electrical potential in DNA

Despite these advantages, native DNA has yet to demonstrate robust functionality as a molecular wire. Measurements of DNA conductivity have shown behavior ranging from insulating[53,54] to induced-superconducting.[55] This disparity in behavior is likely due to technical difficulties in performing electrical measurements on individual molecules.[53,56-62] Still, efforts to enhance the conductivity of DNA are ongoing through various metalization schemes, including the non-specific exchange of imino protons for metal ions;[63] gold,[64,65] palladium,[66] and cobalt[67] nanocluster attachment using azide-alkyne interactions or reduction-based schemes; nanoparticle-catalyzed formation of E-DNA (eccentric DNA) in GC-dominated duplexes;[68] nanosphere assembly from polycytosine i-motif oligomers;[69] and site-specific thiol functionalization in rolling circle amplification,[70] and DNA origami.[22,71,72] Modification of natural

DNA with site-specificity has been sparse, and it ultimately began with the chemical study of ionic interactions with a subset of nucleotides.

Metal-nucleobase interactions have been studied since the 1960s when an affinity was discovered between the mercury cation (Hg^{2+}) and thymine-enriched DNA polynucleotides,[73] but several decades passed before the notion of introducing a metal ion directly into a DNA duplex was first described in detail by Tanaka and coworkers in 2002.[74] This was followed by work that demonstrated the replacement of Watson-Crick G-C base pairs with mismatched cytosine pairs capable of forming a coordinating bond with Ag^+.[75,76] Due to the smaller size of pyrimidines bases relative to their purine cognates, the mispairing of two cytosines introduces a gap in the helical structure which selectively incorporates Ag^+ (Figure 1.1) in a reaction energetically driven by the dehydration of the cation. [77,78] Consequently, the introduction of cytosine-mismatch repeats into oligonucleotide sequences allows functionalization of the helical core with a chain of single metal ions that is surrounded by π-stacking nucleobase rings and stabilized by the sugar-phosphate backbone. Recent studies suggest that in ion-coordinating pyrimidine systems, a metallophilic bond between axially-adjacent cations may cause a compression in the inter-pair bond length of B-form DNA from 3.4 Å to 3.3 Å (Figure 1.2).[79,80] This potential attraction in the Ag^+ system suggests an interaction between the valences of the two ions and an open pathway for electron mobility. Long coordinating ion chains in both linear duplexes and ring-like species have been developed with great promise by several groups for their attractiveness as a nanowire.[81-84]

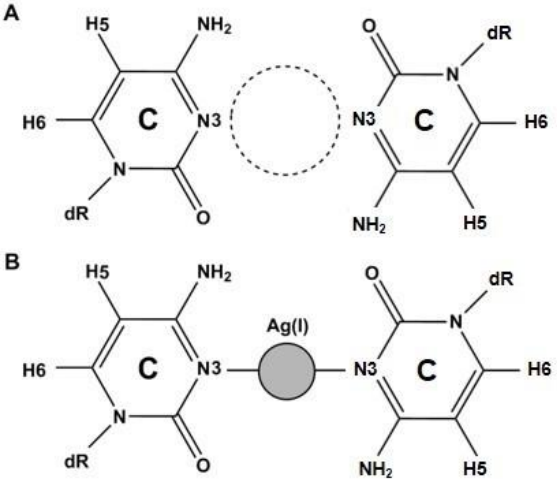


Figure 1.1. Structure of the C-C mismatch predicted from NMR studies:[57] A) unbound with gap; B) bound to Ag^+ at N3 position to form the dC:Ag^+:dC metal base pair.

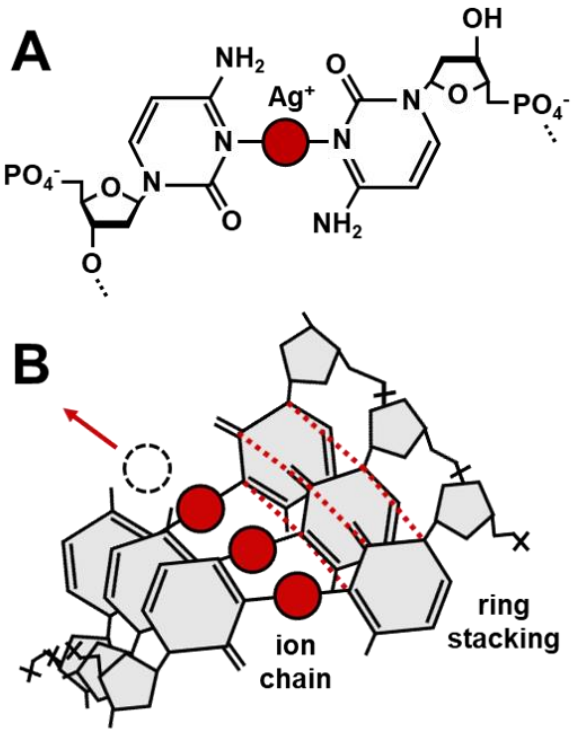


Figure 1.2: Bond and ion coordination chain. A) The C:Ag^+:C base pair coordinates a silver ion between opposing nitrogen atoms (N3 ring position). B) Homobase oligomers exhibit stacking of coordinated rings; while stacked ion-pairs are expected to exhibit metallophilic compression through interactions between the ions, which suggests a novel conduction pathway.

The ionic radius of elemental silver matches the hole size both between radially-opposing cytosines and axially-separated base pairs at 172 pm. The specificity of CC mismatches for Ag^+ uptake is likely driven by this Van der Waals interaction distance, which is also repeated in bulk silver crystals. Other metals may be able to bind single cytosine-cytosine pairs in solution, but the b-form helix may be perturbed by inappropriately-sized metals. Platinum, with an ionic radius of 175 pm, may be well suited to CC mispair binding; whereas cadmium (158 pm) may be a suitable substitute for mercury (155 pm) in TT mismatches. Other metal-nucleobase interactions have been described, including a salicylic aldehyde copper pair (dS–Cu^{2+}–dS) mediated by a covalent ethylene diamine bridge;[85] dC–Ag^+–dT, 5-methylisocytosine (m^5iC)–Ag^+–dT, m^5iC–Ag^+–dC, and dC–Hg^{2+}–dT base pairs with variable stability;[76,86]

and thiopyrimidine–thiopyrimidine pairs binding two Ag^+ ions per mismatch;[87]. Within the dC–Ag^+–dC system, other studies have demonstrated parallel orientations of dC–Ag^+–dC-containing oligomers[76] and the assembly of large, self-assembling ring species [69,84] based on i-motif or silver-paired components. Metal-mediated base pairing is a dynamic platform for noncanonical helical assemblies, and future studies may identify XNA:M^+:XNA base pairs,[88,89] hachimoji (eight-letter) metal assemblies,[90] or other exotic binding schemes.

Given the great potential of Ag^+-functionalized polynucleotides in both biological and nanoscale devices, we propose a viable route to fabricating stable DNA nanowires for nanoelectronic applications in cell-free and synthetic biological systems. To this end, this thesis has two fundamental objectives: the integration of orthogonal, conductive, nucleobase chemistry into the DNA nanotechnology lexicon; and furthermore, the development of computational tools to predict and inform the behavior of modified-parity, nucleic acid nanostructures. To do this, we assay first linear DNA constructs based on orthogonal chemistry in both chemical and electrical probe settings. Then we develop computational tools for the design and prediction of behavior in nonlinear assemblies based on the findings in linear species. Finally, we explore the new language of design for orthogonal DNA tiles and nanostructures; and we comment on the usefulness of these and future designs in synthetic biological and nanotechnological applications.

Chapter 2

Biochemical Analysis of DNA Nanowires:

A foundation for nanotechnology

"I am, as I said, inspired by the biological phenomena in which chemical forces are used in repetitious fashion to produce all kinds of weird effects (one of which is the author)."

-Richard Feynman, There's Plenty of Room at the Bottom[1]

2.1. Chemical analysis: a framework

DNA has attracted attention as a potential electronic material due to its linear geometry, definable base sequence, easy, inexpensive and non-toxic replication via biological processes and self-assembling properties. To assess its appropriateness for technology integration, it was critical to begin with chemical and thermodynamic studies. The approach in this work has been to create conductive DNA nanowires by incorporating metal ions (specifically Ag^+) into the double helix. For this to work, we needed to determine the following: 1) *whether* Ag^+ integration into DNA could be performed reliably; 2) what factors would promote reliable integration into DNA; 3) what effect this integration would have on the overall stability and behavior of the modified helix; and 4) whether these nanowire candidates could be subsequently integrated into biological processes.

To this effect, a battery of chemical tests and characterizations was employed comparing dsDNA of canonical and orthogonal (dC:Ag^+:dC-pairing) base chemistry. A variety of sequence designs, buffers, and reaction stoichiometries were employed, and the products were tested isothermally, at high temperature, and in the presence of enzymes, varying pH, mixed counterion species, and various staining dyes. By the end, a picture of a new DNA system emerged with specific and oft-changing reaction requirements, superior aqueous stability, potential for bio-integration, and a need for more study, new math, and new applications. What began with “does this work” became “how to make it work,” and ultimately “what else can we make work.” In sum, chemical analysis was the foundation for more complex designs, computational modeling, and a great deal of future study.

To map the functionality and biostability of this system, we built and characterized internally-functionalized DNA nanowires through non-canonical, Ag^+-mediated base pairing in duplexes containing cytosine-cytosine mismatches. We assessed the thermal and chemical stability of ion-coordinated duplexes in aqueous solutions by ultraviolet-visible spectrophotometry (UV-Vis), polyacrylamide gel

electrophoresis (PAGE), nuclear magnetic resonance spectroscopy (NMR) and zero mode waveguide (ZMW); and we conclude that the C-Ag^+-C bond forms DNA duplexes with replicable geometry, predictable thermodynamics, and tunable length. We demonstrate continuous ion chain formation in oligonucleotides of 11-50 nucleotides (nt), and enzyme ligation of mixed strands up to six times that length. This construction is feasible without detectable silver nanocluster contaminants.

Building on enzyme compatibility, functional gene parts for the synthesis of DNA- and RNA-based, C-Ag^+-C duplexes in a cell-free system are constructed in an *Escherichia coli* expression plasmid and added to the open-source BioBrick Registry, paving the way to realizing the promise of inexpensive industrial production. With appropriate design constraints, this conductive variant of DNA demonstrates promise for use in synthetic biological constructs as a dynamic nucleic acid component and contributes molecular electronic functionality to DNA that is not already found in nature. We propose a viable route to fabricating stable DNA nanowires in cell-free and synthetic biological systems for the production of self-assembling nanoelectronic architectures.

2.2. Characterization of Ag+ intercalation

The following tests utilize a variety of DNA oligomers with a variety of modifications. Sequences were chosen to illustrate various ion incorporation motifs or distributions, and the final oligo sequences were prepared by hand or using sequence design algorithms (see *Chapter 4*, below). They were uniformly annealed using a standard protocol (detailed below), unless otherwise indicated. Some reactions, explicitly noted in their experimental setup, were annealed at different temperatures or in different buffers. All unmodified and fluorescently-labeled DNA oligonucleotides were synthesized by Elim Biopharmaceuticals (Hayward, CA, USA) and Integrated DNA Technologies (Coralville, IA, USA), respectively (Table 2.1). Strand alignment and ion intercalation sites are shown diagrammatically in Figure 2.1. Oligonucleotides were suspended at 100 µM in deionized water as a working solution.

Table 2.1. Oligonucleotide sequences used for chemical analysis

C11. 11 nt polycytosine	
Template	5'-CCCCCCCCCCC-3'
C20. 20 nt polycytosine	
Template	5'-CCCCCCCCCCCCCCCCCCCC-3'
C30. 30 nt polycytosine	
Template	5'-CCCCCCCCCCCCCCCCCCCCCCCCCCCCCC-3'
T30. 30 nt polythymine	
Template	5'-TTTTTTTTTTTTTTTTTTTTTTTTTTTTTT-3'
Oligo A. 32bp10CC duplex	
Template	5'-TTATATTTACCACCTCCTCCACCTTTTAGATT-3'
Complement	5'- AATCTAAAACCTCCACCACCTCCTAAATATAA-3'
Oligo B. 50bp6CC duplex	
Template	5'-TAAACCACTCATACCACAACAACTCTCTACTCCTACACATCATCCATCTC-3'
Complement	5'-GAGATGGATCATGTCTAGGACTAGAGAGTTCTTGTGGTATCAGTCGTTTA-3'
Oligo C. 25bp1CC duplex palindrome	
Template	5'-TATTAAATAAAACTTTTATTTAATA-3'
Oligo D. 26bp0CC canonical duplex	
Template	5'-CATTAATGCTATGCAGAAAATCTTAG-3'
Complement	5'- CTAAGATTTTCTGCATAGCATTAATG-3'
Hairpin A. 19bp8CC hairpin	
Template	5'-biotin-TTTGTGT(Cy3)CTCCTTCATCCCTCACTTTGTATGCAAACTCACCCATCAACCA GACA-Cy5-3'

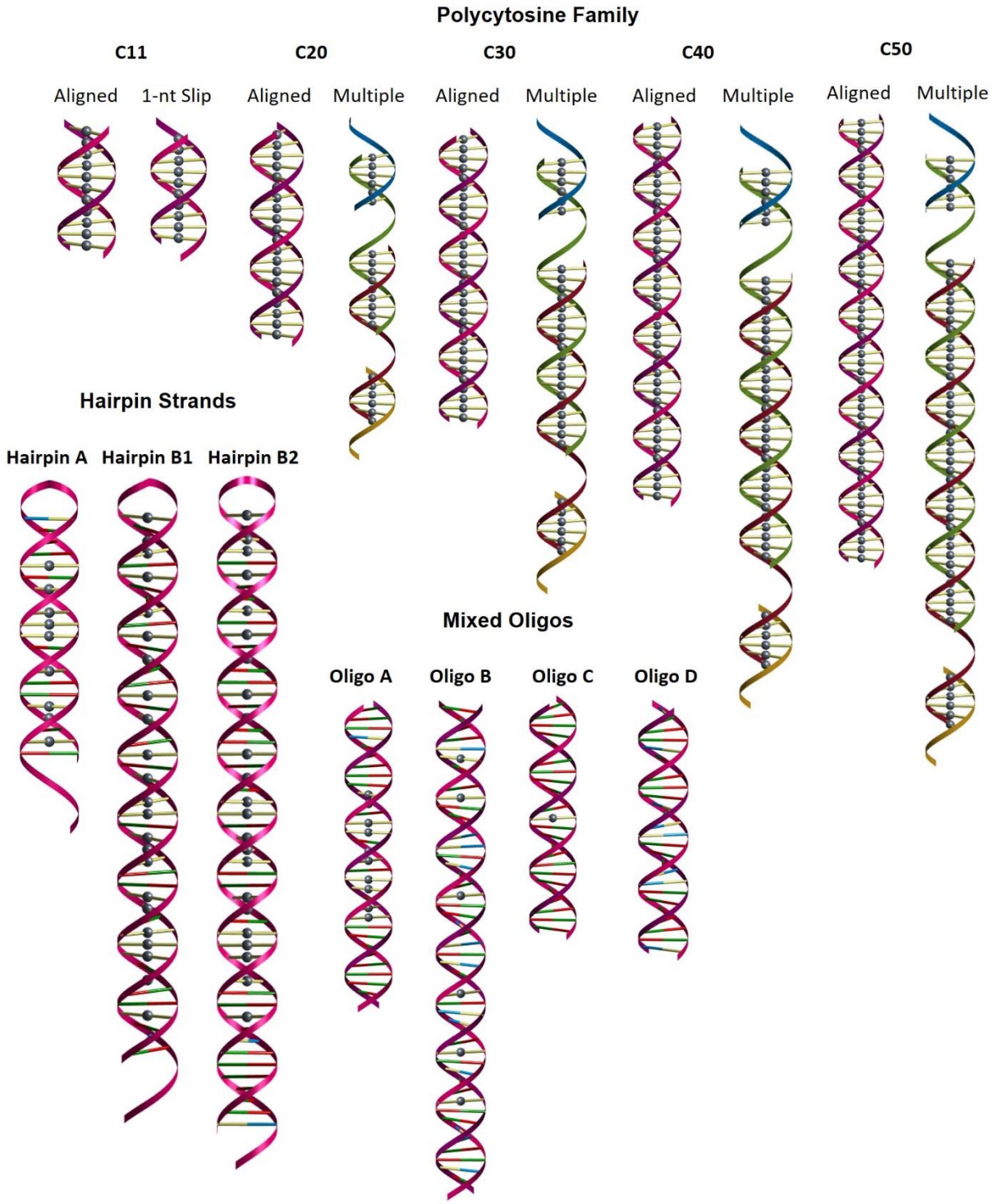


Figure 2.1. Rendering of the oligonucleotides in this chapter (see Table 2.1 and supplementary figures in 2019 manuscript[91]). Bases are color coded: guanine (blue), cytosine (yellow), adenine (green), thymine

(red). Polycytosine structures include the aligned annealing structure as well as a multiple-strand, non-unitary duplex. Models were implemented in Matlab (Mathworks, Natick, MA, USA).

Oligonucleotides were annealed at final concentrations of 2 µM in 10 mM MOPS ($C_7H_{15}NO_4S$) [pH 7.2] and 100 mM $NaNO_3$ with various concentrations of $AgNO_3$ specified in the text for no more than 5 min at 95 °C, followed by cooling to 25 °C, and subsequently 4 °C, over 2 hr.[92]

2.2.1. Analysis by molecular weight: reaction kinetics, sequence design and enzyme compatibility

A set of parallel experiments was employed to analyze the reaction stoichiometry, Ag^+ and CC dependence, and enzyme compatibility of silver(I) DNA nanowires. This was done through titration of Ag^+ into a highly-mismatched sequence (Oligo A, Figure 2.1), as well as into molar excess with a long polycytosine (C30). Annealing of this reaction in ionic excess was compared to polythymine (T30) at various temperatures. Finally, mixed-sequence Oligo B was annealed with Ag^+ and subjected to enzymatic end-ligation in order to assess whether changes in DNA binding chemistry inhibit the overall compatibility of these nanowires with other biological processes.

a. Methods of phosphorylation, enzymatic ligation and electrophoresis

To perform enzymatic ds-ligation, DNA oligomers were first phosphorylated to allow efficient coupling. For 3' phosphorylation, 5 µL each of 2 µM template and complement strands for Oligo D were annealed with a saturating amount of Ag^+ and mixed with 3 µL of MilliQ-purified water, 1 µL (10 U) of T4 phosphonucleokinase (PNK) (Life Technologies, Carlsbad, CA, USA, Cat. #EK0031), and 1 µL (1 unit) of 10X T4 DNA Ligase Buffer (New England Biolabs, Ipswitch, MA, USA) and incubated for 60 min at 37 °C. End-ligation was achieved by subsequently adding 1 µL of T4 DNA Ligase (New England Biolabs) and incubating for 2 hr at 25 °C. Nondenaturing PAGE gels were prepared via standard techniques.[93] Staining was performed either using SYBR Gold (Life Technologies), silver stain kit (Thermo Fisher), toluidine blue

(Thermo Fisher), or strand functionalization with FAM fluorophores (Elim Biopharamceuticals); and gels were visualized on a Typhoon Trio gel scanner (Amersham Biosciences, Amersham, UK).

b. Results: characterization of Ag^+ intercalation into DNA duplexes

Thermal annealing of cytosine-mismatched DNA duplexes from cytosine-enriched DNA oligomers subjected to varying salt conditions was readily visualized by mass distribution using polyacrylamide gel electrophoresis (PAGE)**(**Figure 2.2**)**. Using two complementary 32 bp ssDNA sequences with ten cytosine mismatch points (Oligo A) as a model, we found that strand annealing is directly proportional to the concentration of Ag^+ ions in the reaction (Figure 2.2A). The sequence of Oligo A was designed with a $(CCN)_x$ pattern (where N is any canonically-pairing DNA nucleotide) to produce a duplex with a repeating motif of two C-Ag^+-C pairs followed by one Watson-Crick standard pair; thus, the duplex is tailored to accommodate exactly 10 Ag^+ ions (1 Ag^+ ion per CC mismatch [Figure 2.1]). Nine Watson-Crick terminal pairs on each end served as a clamp to force alignment of the central mismatch motif. Oligo A duplexes were first observed at a Ag^+:CC molar ratio of 0.75, and a solid duplex with a fixed molecular weight is seen at molar unity (Figure 2.2A). Apparent band shift from 0.05 to 0.75 Ag^+ occupancy is likely a gel artifact and is also observed in the tilt of the 10 bp ladder band at that position. Similar band shifts from the left to the right can be seen in other uncropped polyacrylamide gels run in the same gel box (data not shown).

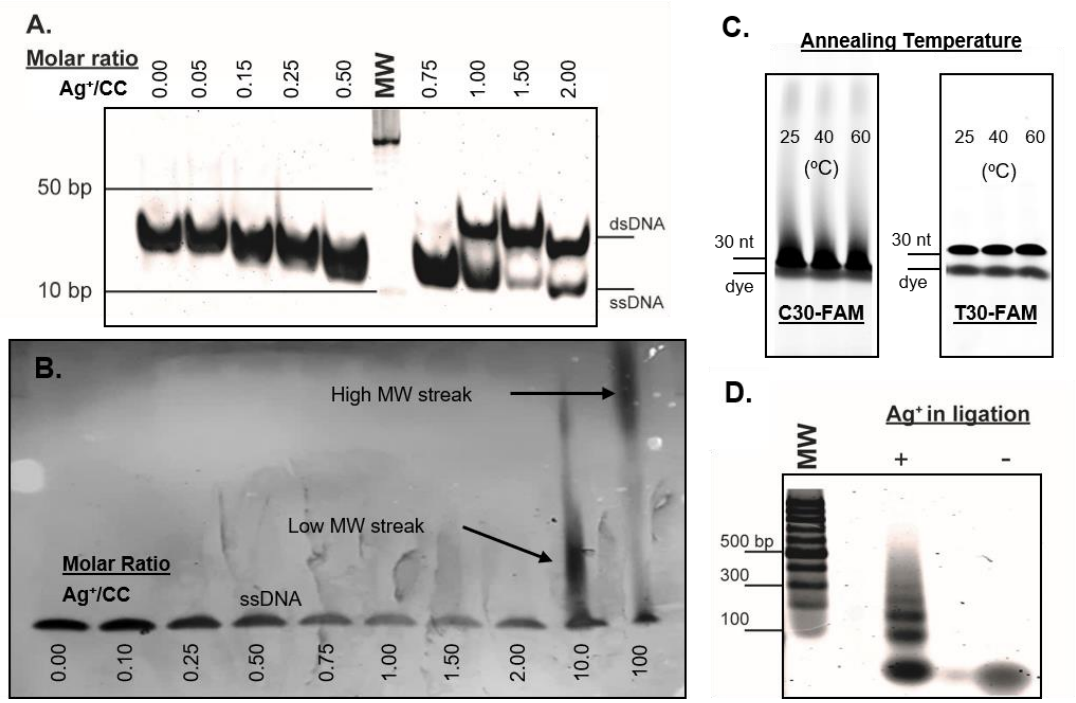


Figure 2.2: Polyacrylamide gels showing formation of nanowires of tunable lengths. A) Equimolar dependence of Ag^+ and mismatches is shown in end-clamped, fixed-annealing-frame, mismatched Oligo A. B) Polycytosines are generally ill-behaved; here ion excess promotes formation of sticky end regions and long-chain polymerization into high-MW regions for C30. C) Ten-fold ion excess produces multimeric polymerization in FAM-labeled C30 at low temperatures of 25 °C, 40 °C and 60 °C, while FAM-labeled T30 controls show no such effect. D) Successful phosphorylation and end-ligation of 50-bp Oligo B is exhibited of up to 300 bp. Uncropped gels for this figure can be found in the supplemental information for the larger manuscript.[91]

When a similar analysis of molar ratio annealing was performed on the C30 oligonucleotide, a 30-nt polycytosine DNA sequence, a markedly different result is achieved. Unlike Oligo A, which featured 9 bp Watson-Crick “clamps” at both ends, C30 lacks a defined pairing frame and may anneal in a range of thermodynamically-promiscuous configurations (Figure 2.1). As a result, instead of producing a unitary duplex, C30 strands interlock into chains whereby overhanging nucleotides from one duplex can pair with the overhangs from another. The size of these chains increases with the Ag^+ ion concentration

(Figure 2.2B). In these experiments, annealing is not observed until $AgNO_3$ is added to a 10:1 Ag^+:CC ratio, producing an ill-defined, multi-band streak representing a distribution of chained C30 strands. An even higher molecular weight streak is apparent at a 100:1 Ag^+:CC ratio, suggesting that C30 chain interlocking is driven by Ag^+ ion availability. This effect was also reproduced in longer polycytosines (data not shown). Multimeric polycytosine streaking can be achieved preferentially with excess environmental Ag^+, even at temperatures far below 100 °C, if given sufficient reaction time (Figure 2.2C). Though the single-stranded bands are more prominent than those produced by high-temperature conditions in Figure 2.2B, streaking into arbitrarily high molecular weight regions of the gel can be observed even when annealed at room temperature for times over 48 h.

2.2.2. NMR

To confirm that the bond geometry in the nanowires matched those observed in dC-Ag^+-dC pairing by other groups,[77,94] a collaboration was carried out with Dr. Mark C. Capece at the Stanford University Magnetic Resonance Laboratory. To remain consistent with previous studies demonstrating the effectiveness of 2D nuclear magnetic resonance spectroscopy (NMR) at observing the secondary effects of spin-neutral Ag^+ intercalation, a model palindromic sequence was annealed and analyzed. In this manner, our collaborators were able to confirm that Ag^+-intercalation occurs at the N3 ring position on opposing cytosines. Further analysis was done to assess the strength of that bond in the presence of precipitating agents.

a. Methods of NMR

Palindromic 25 nt sequence, Oligo C, was selected for its single cytosine nucleotide and absence of confounding guanines. To closely examine the energy of ring hydrogens without signal contamination, standard annealing reactions were carried out in 99.99% D_2O with 10 mM K_3PO_4 and 75 mM $NaNO_3$—none of which contain elemental hydrogen. In doing so, the coupling energy between H5-H6 atoms

become the most prolific ^{1}H-^{1}H coupling in the environment, allowing for accurate correlation spectroscopy (COSY) NMR. Data acquisition and analysis was carried out by collaborators using a 600 MHz Varian UNITY INOVA spectrometer (Agilent Technologies, Santa Clara, CA, USA) and 5 mm $^{1}H\{^{13}C, ^{15}N\}$ Z-PFG conventional probe at 25 °C. The covalent binding of Ag^{+} to the cytosine ring creates an observable shift in hydrogen coupling energy that can be observed and analyzed. Oligo C was subjected to COSY NMR prior to Ag^{+} intercalation, after reannealing with Ag^{+}, and finally after precipitation of solvent-accessible Ag^{+} as AgCl with the addition of 100 mM NaCl and threefold buffer exchange.

b. Results of structural analysis

Following the observations of Torigoe and colleagues,[77] the COSY spectrum for $^{3}J_{HH}$ coupling of cytosine ring H5 and H6 was analyzed in the range of 5.60 ppm and 7.50 ppm, respectively. Palindromic Oligo C allows the data to collapse into a single observable peak, which is expected to shift upon disruption by an electron-rich neighbor like Ag^{+}. When annealed in the absence of Ag^{+}, a single peak is observed (black, Figure 2.3A) in the expected region. Addition of Ag^{+} in annealing shifts this peak upwards in the manner observed by previous studies (red peak). This suggests direct perturbation of ring coupling energies and the binding of Ag^{+} by cytosine ring N3. Precipitation and buffer exchange do not cause reversion of the field shift, indicating that helical sterics or bond strength render the dC:Ag^{+}:dC-participating ion solvent inaccessible. By contrast, $^{3}J_{HH}$-coupled hydrogens in the phosphate backbone are shifted and then reverted after dialysis, suggesting that casual counterion stabilization by Ag^{+} is both reversible and non-covalent.

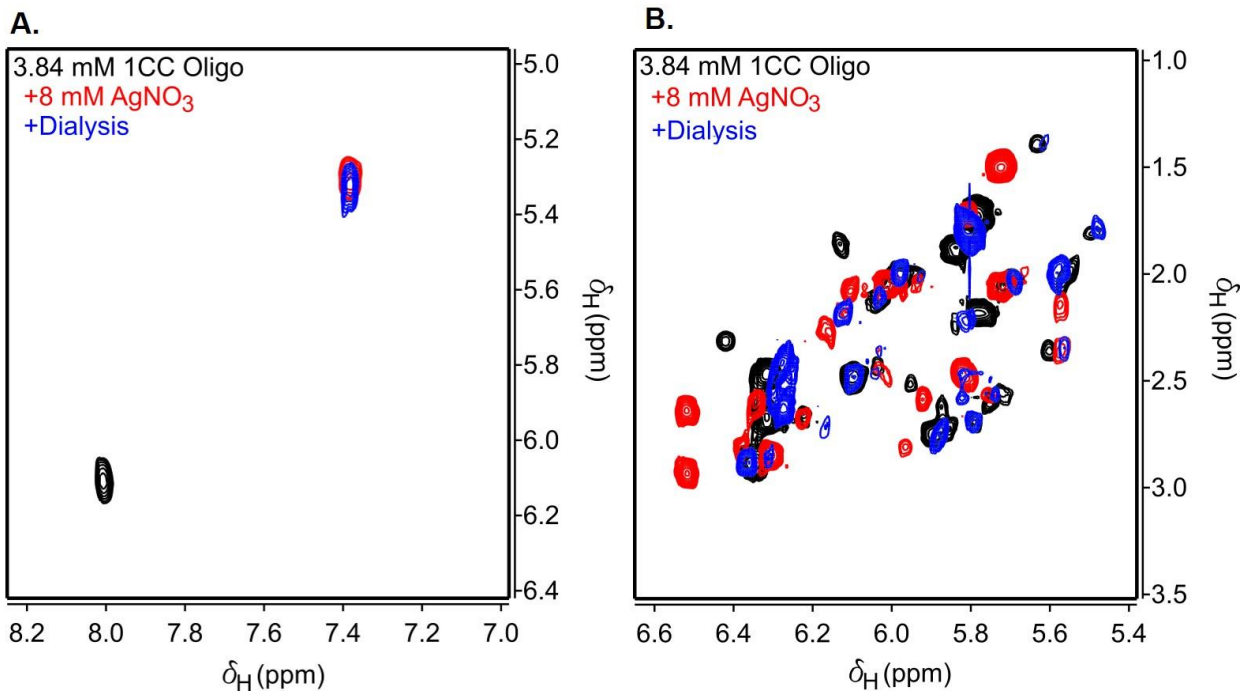


Figure 2.3: COSY NMR spectra of Oligo C duplexes in varying ionic environments: without Ag^+ (black), with Ag^+ (red), and after removal of solvent-accessible Ag^+ as AgCl post-annealing (blue). **A)** Spectra in the range of the $^3J_{HH}$-coupled H5-H6 crosspeak show a baseline black peak in the expected range that is perturbed by Ag^+ intercalation into the cytosine ring at the N3 position. Subsequent Ag^+ precipitation and buffer exchange produced a (blue) peak in the same place as the Ag^+ anneal, suggesting that the ring structure was not perturbed by precipitation. **B)** By contrast, other coupled hydrogens in the phosphosugar backbone shift upon Ag^+ addition, suggesting counterionic stabilization, but are for the most part collapsed to their original positions upon buffer exchange. This suggests that solvent-accessible, backbone-bound Ag^+ is less thermostable than ring-intercalated ions.

2.2.3. Single-molecule FRET via zero-mode waveguide (ZMW)

Single molecule kinetics can be used to understand folding times and Ag^+-dependencies in DNA nanowires. Here we collaborated with Dr. Capece of the Stanford Magnetic Resonance Lab to capture the real-time dynamics of single molecule fluorescence using a zero-mode waveguide.

a. Methods of FRET and ZMW

Förster resonance energy transfer (FRET) between Cy3 and Cy5 was measured on DNA oligo Hairpin A. Designed with 9 CC mismatches and an otherwise C-G-depleted structure, the oligo forms a hairpin through its gTATGc tetraloop in the presence of Ag^+ only. While studies have shown that gold-sulfide-immobilized DNA oligomers form temporary pseudocoils,[95] the strong Ag^+-dependent FRET due to helical alignment should occur only after the addition of intercalating ions, and should be a much stronger, specific interaction between the fluors.

The zero-mode waveguide employed in this experiment was a converted DNA sequencer used for fluorescence analysis in DNA duplex formation[96] and translation kinetics[97] for single-molecule studies. Neutravidin-coated ZMW chips with 150,000 wells are bound with 25 µM biotinylated Hairpin A in assay buffer (10 mM Tris [pH 8.25], 2.5 mM 3,4-dihydroxybenzoic acid, 250 nM protocatechuate dioxygenase,[98] and 2.5 mM TSY triplet-state quencher solution [Pacific Biosciences, Menlo Park, CA, USA]). It is estimated that only 13% of wells have a bound oligomer, in this way preferentially promoting single-molecule occupancy. Either 1 mM $NaNO_3$ or $AgNO_3$ in the assay buffer was mechanically delivered to the sample chip while the sample was illuminated by only the green (532 nm) laser at a power of 0.48 $\mu W/\mu m^2$. Emission measurements for both Cy3 and Cy5 fluorophores were recorded for each waveguide at 100 ms intervals for a total observation period of 10 min. The detection of Cy5 emission indicated that energy was transferred from excited Cy3 to ground state Cy5 through the process of FRET.

Data were filtered down to 200 representative traces in which FRET was recorded after the start of the experiment, either with the infusion of control or Ag^+ buffer. Fit of the data to either single or double exponentials allowed for the extraction of folding kinetics. Curation of the data was performed in Matlab (Mathworks).

b. Results of SM kinetics

Hairpin A was assayed via ZMW for FRET with either infusion of $NaNO_3$ or $AgNO_3$. In control experiments, two broad and non-specific Cy3 quenching efficiencies were observed (Figure 2.4A), with low (0.227±0.119) and medium (0.587±0.173) FRET efficiencies that were, by and large, independent of buffer infusion, suggesting that a pseudocoil or random conformation is formed by the oligo. These wide Gaussian peaks suggest structural heterogeneity in the conformations, indicating lack of strong duplex formation. When silver ions are mechanically infused, one low-FRET Gaussian at 0.457±0.234 and one high-FRET Gaussian at 0.738±0.049 is recorded (Figure 2.4B). The high-FRET state has a much greater density, and is only observed to form from the ground state—no high-FRET is observed in which a medium-FRET molecule upconverts to high-FRET. This indicates that the complexity of Ag^+-intercalation via duplex formation requires the full unfolding of a pseudocoil prior to subsequent hairpin formation. This matches the expectation that Ag^+ is bound in a two-stage reaction, adhering to one cytosine before forming a bond with the opposing mismatched cytosine complement.

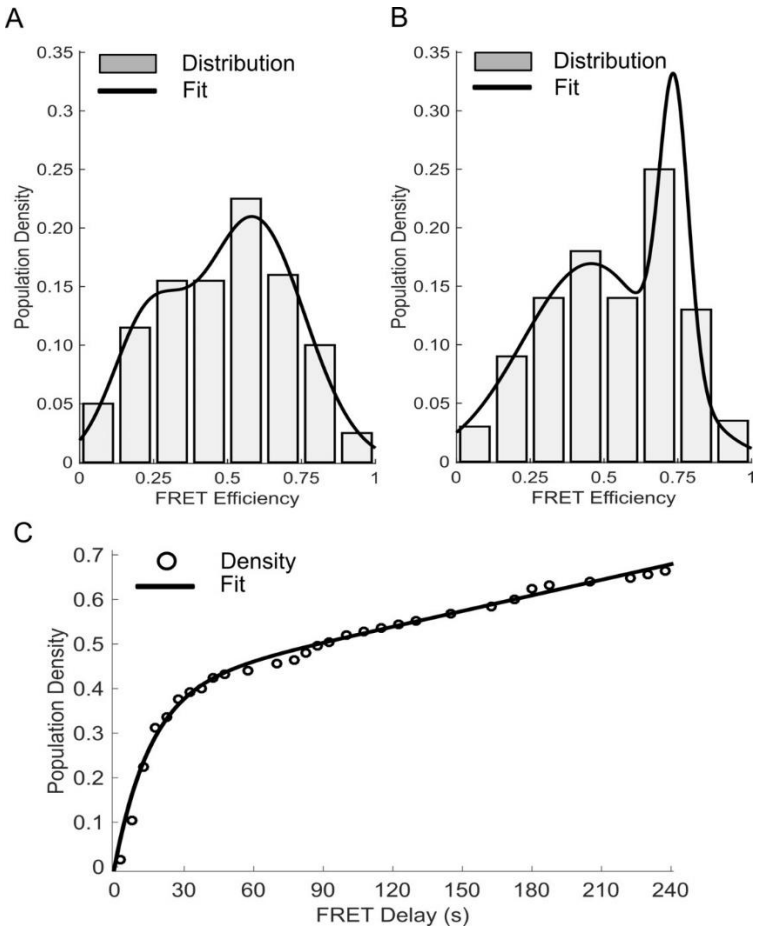


Figure 2.4. A) Histogram of FRET intensity values for the $NaNO_3$ control of Hairpin A fit to a sum of two Gaussian functions centered at 0.227 and 0.587 intensities. **B)** Histogram of FRET intensities for the $AgNO_3$ delivery to Hairpin A with a double-Gaussian fit of peaks 0.457 and 0.738. **C)** Extraction of the FRET delay parameter (time until FRET) for only the AgNO3-delivered high FRET traces. The folding rate for the Ag-dependent structure is approximated from the exponential fit to be 0.065 s^{-1}.

An exponential fit of high-FRET molecules is performed, and the folding rate is determined to be $6.5\pm0.7\times10^{-2}$ s^{-1} for Ag^+-dependent hairpin formation. Returning to the raw data and isolating the medium-FRET state, a folding rate of approximately 15 s^{-1} for Ag^+-independent pseudocoils is obtained, which is typical for conformational rearrangement of DNA molecules.[99] These result underscore a

relatively slow, strong bond formation between opposing mismatched cytosines in an Ag^+-dependent reaction, with significant and specific alignment of helical fluorophores in designed conformations.

2.3. Stability of Ag+ intercalation

In tandem with kinetic and stoichiometric analysis of the C-Ag^+-C bond formation, tests were performed to investigate the thermostability and purity of the DNA nanowire product. UV spectrophotometry was combined with thermal denaturation to assess the strength and reversibility of binding in different sequence and buffer conditions, while a full UV-Vis spectrum was investigated under various conditions to understand the purity and tendency (or lack thereof) to reduce Ag^+ into silver nanocluster contaminants.

2.3.1. Thermal denaturation and UV-Vis

In keeping with previous thermodynamics studies, UV spectrophotometry combined with temperature ramping was employed to analyze duplex stability.[94,100] This technique was employed to assess three different reaction conditions: Ag^+ availability; C:C mismatch availability over increasing sequence length; and Cl^--dependent Ag^+ precipitation.

a. Methods of thermal denaturation

An annealed volume of 20 µL of 2 µM double-stranded DNA oligonucleotides was diluted into 1 mL MOPS buffer in a crystal cuvette on a Perkin-Elmer Lambda 950 UV spectrophotometer (PerkinElmer, Inc., Shelton, CT, USA) with Peltier temperature control. Absorbance relative to the MOPS buffer blank was measured at 280 nm. Measurements were carried out using a D_2 lamp with 2-nm spectral bandwidth and 1-cm path length. Readings were collected across a temperature range of 25-100 ^oC with 60-s pauses at each temperature point to achieve thermal equilibrium. Five measurements in 1-s intervals were recorded at each point and averaged. Melt curves were constructed in GraphPad Prism (Origin Labs, Northampton, MA, USA); and the melting temperatures were extracted from a generalized

Hill equation, or five-parameter logistic (5PL) regressions. Equation 1 describes the normalized absorbance (Y) as a function of temperature (independent variable T), which was fit using the absorbance extremes (Y_{min} and Y_{max}), the Hill slope (fit parameter H), the temperature at half-maximum absorbance (melting temperature T_m), and an asymmetry coefficient (fit parameter S) to account for clipping of the curves near 100 °C. Equation 2 is provided to simplify the exponent in the final expression for Y(T). Control samples without injection of Ag^+ were normalized, in identical buffer and concentration, to the absorbance start point (fit parameter Y_{min}) of the Ag^+ experimental condition, such that the relative change was correlated to that of the experimental sample. This analysis yields a relative change of 0.1% in the ion-free control, with an F-value of non-zero fit of 0.0108.

$$Y(T) = (Y_{max} - Y_{min})/(1 + 10^{H(Z-T)})^S \quad (1)$$

$$Z = T_m + \frac{1}{H}\log(2^{1/S} - 1) \quad (2)$$

b. Results of thermal denaturation

Results of thermal melts can be seen in Figure 2.5. We see from Figure 2.5A that the C30 (30 nt polycytosine DNA) sequence, when annealed at a 10:1 Ag^+:CC molar ratio, yields a full melting profile with a T_m of 91 °C. The melting profile of the C30 sequence with Ag^+ is far more articulated than the same C30 sequence without Ag^+, as the latter lacks any ability to base pair without Ag^+. Similarly, the melting profiles of C11 (11 nt polycytosine DNA) and C20 (20 nt polycytosine DNA) also produced melting temperatures of 90 °C and 93 °C, respectively (Figure 2.5B). When compared with calculated melting temperatures of their canonical dC:dG duplex counterparts (62.5 °C, 83.3 °C, 90.7 °C, respectively),[101,102] the Ag^+-incorporated polycytosine chains are significantly more thermostable, especially with the shorter C11. It seems that the Ag^+-intercalated polycytosine chains are resistant to temperatures approaching the boiling point of water. Further analysis was performed on the metalated

C30 sample in the presence of excess NaCl, which precipitated uncoordinated Ag^+ ions as AgCl (Figure 2.5C). In agreement with the NMR experiments, exchanging the backbone-bound Ag^+ cations with Na^+ in this manner resulted in a small drop in melting temperature of <5 °C, but did not revert the melting profile to that of ssDNA. Ag^+ thus may have a secondary role as a monovalent cation backbone stabilizer for duplex formation, similar to the well-documented effect of other cations, such as Na^+ and Mg^{2+}.[103,104]

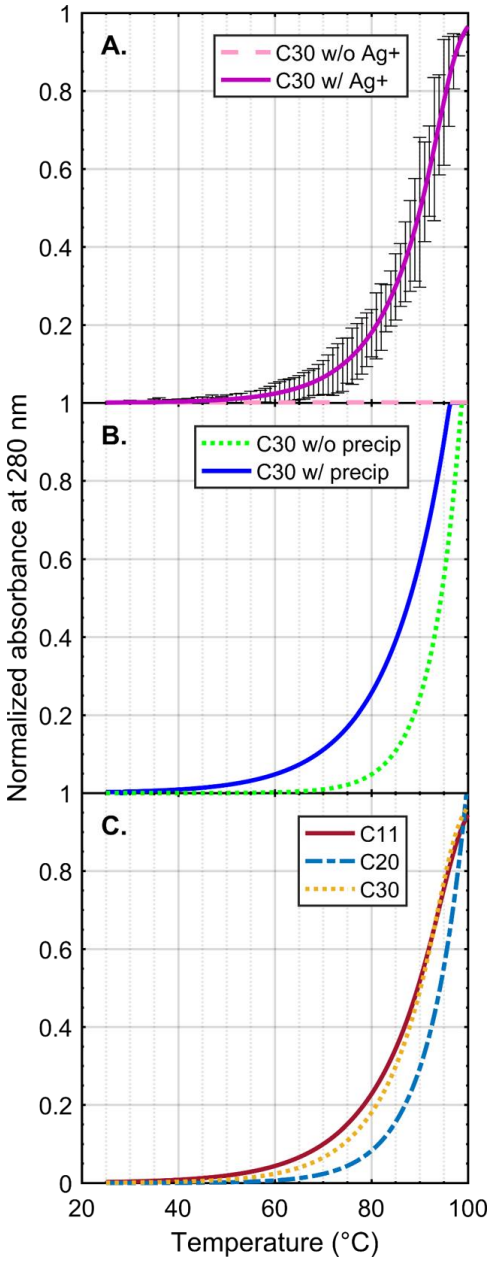


Figure 2.5: Melt curves of polycytosine oligomers. **A)** The ion requirement for successful C30 melting is shown with standard deviation (n=3). Without Ag^+, no change in absorbance is shown over temperature, while the experimental condition shows full melting behavior characteristic of DNA duplexes. **B)** Representative curves of C30 before and after precipitation of backbone-bound Ag^+ as AgCl suggest a measurable difference in melting behavior, but that the primary contributor to duplex stability is the pyrimidine-coordinated cation chain. **C)** Polycytosine oligomers of varying lengths show melting temperatures at or above 90 °C, indicating a powerful stabilizing effect mediated by the internal ion chains. The similarity between these melting temperatures likely reflects the maximum capacity of the instrument to assess thermal stability in water-based buffers.

2.3.2. Cluster analysis

Concerns of Ag^+ reduction into nanocluster contaminants were raised upon early demonstration of elevated charge transport in 11 nt polycytosine nanowires, and as a result it was necessary to rule out conduction through catalyzed nanoclusters that are known to form around polycytosines under specific reaction conditions.[105,106] It has been shown that reduction through the aggressive agent BH_4^- can be carried out using cytosine nucleobases as a catalyst, perhaps through electron exchange near the N3 amine, which is used to stabilize mismatched duplexes with the same ion. To replicate these results, a variety of conditions are assayed, with canonically- and orthogonally-paired sequences, with and without Ag^+, with and without AgCl precipitation, and with or without BH_4^-. The only difference from previous studies lies in the lack of reduction agent infusion at high temperatures, as this does not mimic the conditions found in the DNA nanowire system in this work. Characteristic peaks in the absorbance between 400-550 nm are investigated as a spectral footprint of cluster contamination, in agreement with the literature.

a. Methods of cluster analysis

UV absorption spectra of variously-processed oligomers were collected in buffer conditions in 2 nm intervals over a range of 250-700 nm on a Lambda 950 UV spectrophotometer (PerkinElmer, Inc.) at room temperature. Measurements were carried out using a D_2 lamp with spectral bandwidth of 0.05-5.00 nm and a path length of 1 cm. Sample absorbance relative to the empty buffer was collected in wavelength intervals of 1 nm across the spectrum. DNA sequences annealed with and without Ag^+ were compared in cytosine-enriched sequence C11 and in canonically-paired sequence Oligo D.

To strip Ag^+ ions from 2 µM DNA samples after annealing, NaCl was added to a concentration of 50 mM and was allowed to react at room temperature for 15 min before pelleting the AgCl precipitate with a tabletop centrifuge (12,000 g, 12 min). To force reduction of solvent accessible Ag^+ ions, $NaBH_4$ was added to 2 µM duplexes to achieve a final reductant concentration of 1 mM. Samples were then shaken vigorously for 2 min at 6 °C and left to react at room temperature overnight. Spectra were collected both before and after precipitation of AgCl and reduction by BH_4^-.

b. Results of nanocluster detection assay

A comparison between the C11 sequence and the Watson-Crick paired 28-bp DNA duplex Oligo D with and without $AgNO_3$ showed no significant difference in the UV-visible absorption of the nucleotide solution (Figure 2.6). Addition of BH_4^- and precipitation of uncoordinated Ag^+ ions did not affect the absorption spectra for either sample. In all tests, two peaks were observed in the critical range (450 nm and 483 nm), and this profile was conserved for a large variety of sequences, regardless of processing (C20, CG20, C30, Hairpin A; data not shown). These results demonstrate that Ag^+-intercalated polycytosine DNA duplexes exhibit high thermal stability without the drawback of undesirable nanocluster formation in common aqueous solutions.

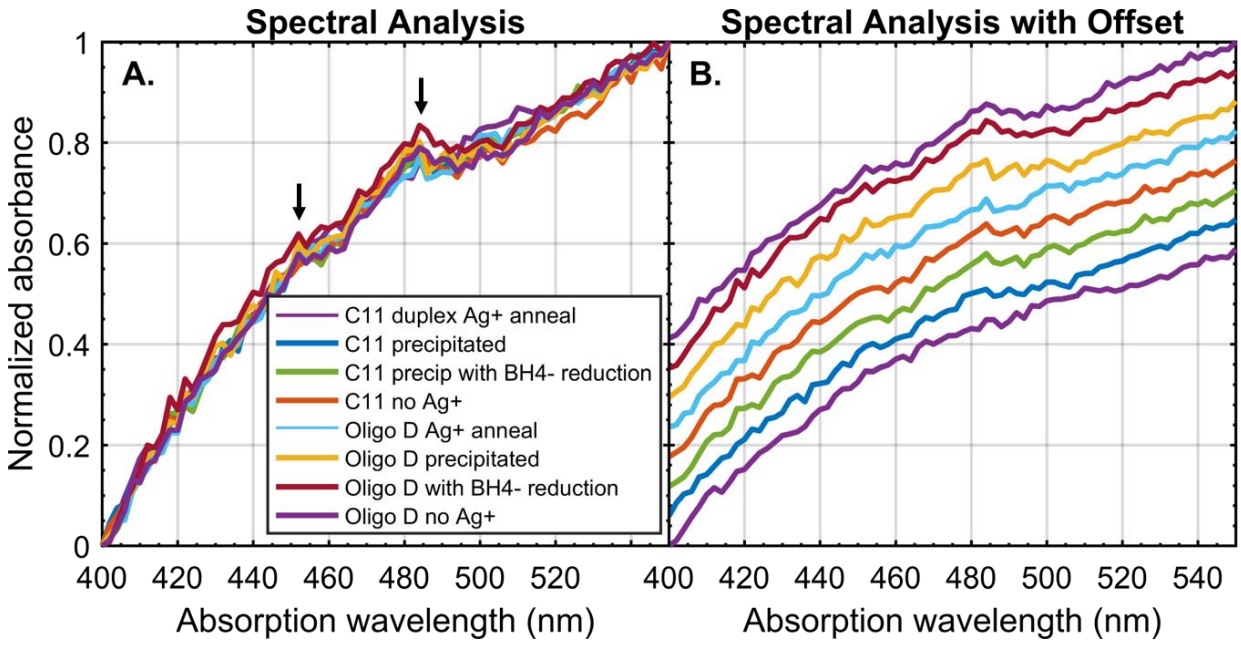


Figure 2.6: Normalized absorbance spectra of various annealed oligos from 400-550 nm. **A)** Spectral overlay of C11 and Oligo D subjected to minus-Ag^+ and positive-Ag^+ conditions, precipitation of AgCl, and reduction by BH_4^-. Peaks are observed at 483 nm and 450 nm and appear to be universal across all conditions (black arrows). No significant perturbation is observed to suggest nanocluster formation. **B)** The same data are offset visually for readability.

2.4. Best annealing protocol for Ag+ intercalation, devised by chemical analysis

Based on the biochemical analysis carried out in this chapter, and driven by an attempt to streamline, optimize, and control successful annealing of DNA nanowires based on the dC:Ag^+:dC system, a thorough investigation of buffer and counterion contribution was carried out. In doing so, a better picture of nanowire annealing emerged. It is *possible* to produce products in a variety of buffers (see above for Tris, MOPS, K_2PO_4, etc.), but it was necessary to develop a high-yield process prior to building more complex structures. To perform this analysis, FRET quenching over nearly 1000 trials was investigated for a wide array of buffer compositions, and a picture emerged which was both surprising

and important for future work. With these results, we write a best-practice annealing protocol for Ag^+ DNA nanowires, and discuss some of the degrees of freedom in this assembly.

2.4.1. Buffer analysis using FRET quenching

a. General considerations: Ag^+ compatibility

The choice of buffer will determine the favorability of ion uptake. For the most part, standard annealing buffers are appropriate for Ag^+-functionalized DNA nanowires [94], with several caveats. Most importantly, there should be minimal presence of chloride ions, as silver chloride is highly insoluble. At high temperature, ^1H-^1H COSY NMR suggests that cytosine-mismatched oligonucleotides can outcompete anions for Ag^+ when already incorporated into the DNA (Figure 2.3), but it is best to avoid the confounding effect of precipitation. This means that one should avoid adjusting buffer pH with HCl, as this can leave residual Cl^- of relatively high molarity. Furthermore, there should be no chelating buffer agents. Thus, ethylenediaminetetraacetic acid (EDTA) and other Mg^{2+} intercalators should be avoided, rendering buffers such as Tris-acetate-EDTA (TAE) and Tris-borate-EDTA (TBE) inappropriate for annealing. These buffers are still acceptable for downstream applications such as gel electrophoresis after Ag^+ is already snugly bound by a cytosine-mismatched duplex.

Several other considerations are less obvious. To elucidate the effects of buffer component species, we screened 33 buffers with different buffer salt, pH, counterion species and Ag^+ availability. The experiment was performed by mixing MOPS, potassium phosphate, Tris, and sodium acetate buffers at standard reaction concentrations with either 100 mM Na^+, 12.5 mM Mg^{2+}, or both counterions. Oligonucleotides C30 (5'-FAM-CCCCCCCCCCCCCCCCCCCCCCCCCCCCCC-BHQ1-3') and AT24 (5'FAM-AACAATACATACTAAAACCAAAAT-BHQ1-3', 5'-ATTTTGGTTTTAGTATGTATTGTT-BHQ1-3') (Elim Biopharmaceuticals) were labeled with 5’ fluorescein (FAM) and 3’ Black Hole Quencher 1 (BHQ1), which, when in close proximity, will cause FRET, whereby BHQ1 will absorb and suppress the signal from

FAM, to cause an overall quenching in the fluorescence. Data were collected by measuring fluorescence before and after annealing of labeled oligos (2 μM per strand, or 4 μM for palindromic C30), either with or without 10x Ag^+ (600 μM). FRET efficiency was calculated by normalizing to the minimum FRET efficiency in deionized water, and the maximum efficiency (100 % quenching). FRET quenching is never 100% due to the promiscuity of the C30 sequence—it may anneal out of frame to form a daisy-chain type structure and thus prevent full quenching by annealed duplexes. The results can be seen in Figure 2.7.

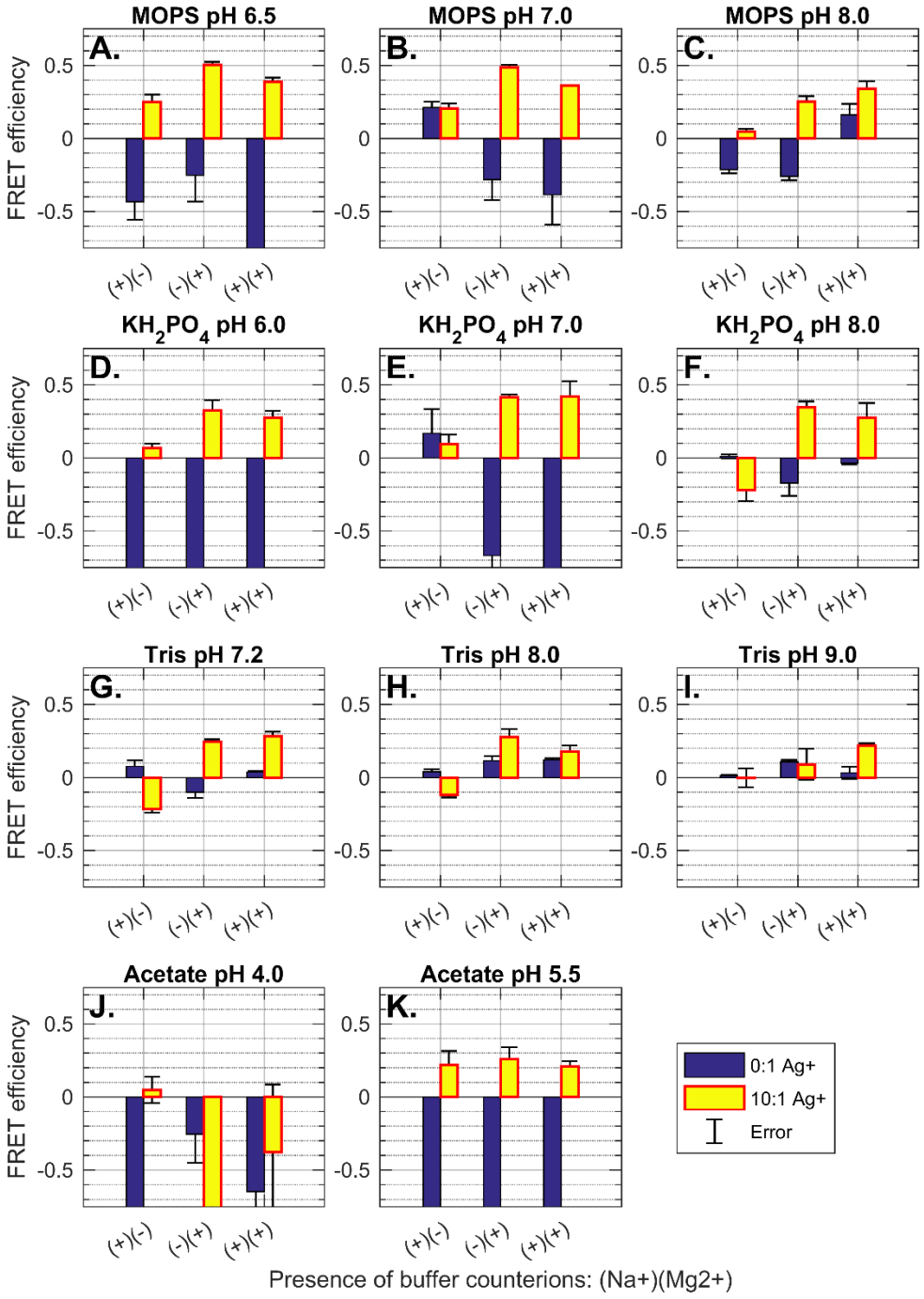


Figure 2.7: Comparison of FRET efficiency (normalized fluorescence quenching) of palindromic 30 nt polycytosine "C30"—5'-FAM-CCCCCCCCCCCCCCCCCCCCCCCCCCCCCC-BHQ1-3' in buffers with varying buffer salts, counterions, pH and Ag^+ availability. Buffer and pH are indicated in plot titles. Counterion species are indicated on the X-axes, where (+)(-) denotes 100 mM Na^+; (-)(+) denotes 12.5 mM Mg^{2+}, and

(+)(+) indicates 100 mM Na^+ *and* 12.5 mM Mg^{2+}. Error bars show standard deviation for FRET efficiency (n=3). Presence of 0:1 Ag^+ molarity compared to sequence mismatches (blue) and 10:1 Ag^+ molarity to sequence mismatches (yellow) is indicated by bar color (see legend). For both MOPS buffers and potassium phosphate buffers, the presence of Mg^{2+} greatly increases FRET efficiency and strand annealing: **A)** MOPS pH 6.5, **B)** MOPS pH 7.0, **C)** MOPS pH 8.0; and **D)** KH_2PO_4 pH 6.0, **E)** KH_2PO_4 pH 7.0, **F)** KH_2PO_4 pH 8.0. The same trend is observed for Tris buffers, but with greatly reduced efficiency throughout: **G)** Tris pH 7.2, **H)** Tris 8.0, and **I)** Tris 9.0. Low pH sodium acetate buffer does not promote duplex formation: **J)** NaAc pH 4.0; while higher pH acetate buffer promotes i-motif formation across all counterion conditions: **K)** NaAc pH 5.5. This conclusion is supported by the lack of FRET quenching in control strand AT24 in similar buffer (data not shown; see Figure S1[107]). In all buffers, acidic pH (≤ pH 7.0) promotes fluorescence enhancement (negative FRET efficiency) after annealing, while basic pH (> pH 7.0) dampened the fluorescence signal overall, both fluorescence and quenching.

b. Counterion species: Na^+ and Mg^{2+}

Thermodynamic simulations implemented in Matlab based on the best-available nearest neighbor and buffer correction models [101,102] suggest that in Watson-Crick pairing sequences, [C30:G30] will have weak stabilization by Mg^{2+}, but will experience a strong increase in thermostability due to Na^+ association with the phosphosugar backbone (Figure 2.8). The presence of both counterions should have an intermediary effect, increasing the melting temperature and subsequent strength of duplexing relative to deionized conditions, but to a lesser degree than Na^+ alone. Control sequence AT24 is predicted by a similar analysis to be stabilized equally by buffers containing Na^+, Mg^{2+}, or both. Experimentally we find results for C30, when stabilized by Ag^+, that are not in agreement with this model, suggesting that the binding mechanism does not adhere to traditional Watson-Crick thermodynamics. To account for these differences, future studies may inform a more accurate thermodynamic model to accommodate orthogonal ion pairing. Preliminary insight on how to do this is

provided by Torigoe et al. [11], but more work is needed to match the level of analysis available for canonical base pairs.

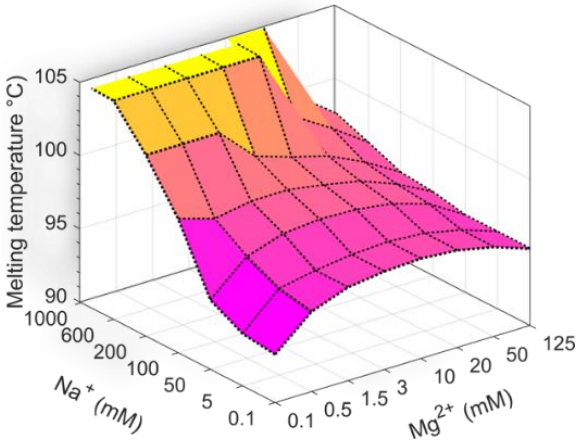


Figure 2.8: Thermodynamic simulation of buffer counterion effects on [C30:G30]. Nearest neighbor models based on Peyret et al. [101] and buffer correction equations from Owczarzy et al. at IDT DNA [102] allow for melting temperature modelling of oligonucleotides based on sequence and buffer composition. We implement these methods in Matlab to investigate the predicted effects of Na^+ and Mg^{2+} on the thermostability of Watson Crick pairing of C30 (5'-CCCCCCCCCCCCCCCCCCCCCCCCCCCCCC-3') to G30 (5'-GGGGGGGGGGGGGGGGGGGGGGGGGGGGGG-3'). The canonical model suggests a strong stabilizing effect by Na^+, but little to no contribution by Mg^{2+}. Experimental data on ion paired C30 suggests that the opposite is true, and that a new computational model is needed to account for orthogonal metal pairs.

In intermediary-pH MOPS and phosphate buffers, the C30 sequence demonstrates minimal fluorescence quenching in the presence of monovalent cation Na^+ alone (Figure 2.7 A-F). The addition of Mg^{2+} causes maximal FRET efficiency of ~0.5. The presence of both Na^+ and Mg^{2+} together still promotes FRET, but in no case does the combined counterion solution outperform the buffer containing only divalent cations by any statistically significant amount. There are two potential reasons for the success of Mg^{2+} buffers. It may be the case that axial compression due to metallophilic attraction between stacked silver ions may

cause crowding effects on monovalent counterions such as Na^+, creating a need for the higher charge density offered by divalent cations such as Mg^{2+} in order to stabilize duplex formation in highly-cytosinated oligos. It may also be the case that divalent cations are better at sterically hindering Ag^+ during the annealing process, causing intermediary, misaligned states to have a lower energy availability. This may force strand rearrangement to better-aligned duplexes which exhibit a higher degree of FRET quenching due to the better terminal fluorophore-quencher proximity. In either case, it is clear that Mg^{2+} provides a significant boost to FRET efficiency, and thus to duplex formation in cytosine-mismatched nanowires.

c. Buffer pH

It is known that buffer pH at or below 6.5 promotes the formation of cytosine-quadruplex (i-motif) DNA in the presence of oligos with high cytosine counts [29]. This structure forms non-B-form DNA that, to our knowledge, has not been assayed for electrical conductivity. Experimentation on this effect shows that low pH sodium acetate (pH 4.0) inhibits formation of any coherent structure (Figure 2.7J). Evidence of i-motif formation can be seen in medium-acidity acetate buffer (Figure 2.7K) by the observation of successful quenching only in excess Ag^+ as well as a general insensitivity to counterion species. By contrast, AT24 does not cause FRET quenching either with or without Ag^+, highlighting the requirement of cytosine bases in formation of acid-mediated structures like a cytosine quadruplex (see supporting information[107]). We do not see a similar insensitivity to counterion salts in MOPS (pH 6.5, Figure 2.7A) or phosphate (pH 6.0, Figure 2.7D) buffers, suggesting that i-motif formation may not be supported in these buffers at slightly-acidic pH.

High pH Tris buffer (pH 9.0, Figure 2.7I) shows reduced FRET efficiency compared to intermediary pH, and slightly-basic compositions of pH 8.0 MOPS (Figure 2.7C), phosphate (Figure 2.7F) and Tris (Figure 2.7H) show slightly reduced efficiency. Overall, it seems that successful duplexing can occur at pH 6.5

and pH 8.0; however, neutral buffers near pH 7.0 provide the highest overall FRET quenching, suggesting optimal strand alignment and duplex strength.

d. Final buffer composition

At neutral pH, MOPS and potassium phosphate buffers (Figure 2.7B, E) outperform Tris buffer (Figure 2.7G) for overall reaction favorability. MOPS buffer can be used for many molecular biology and molecular electronics applications, while potassium phosphate in 99.99% D_2O can be used for ^1H-^1H COSY NMR and other analytical applications. For use in single-molecule conductivity experiments, we suggest the use of pH 7.0, 10 mM MOPS buffer with 100 mM $NaNO_3$ and 12.5 mM $MgSO_4$. Annealing may be done at a 10x Ag^+ availability with respect to sequence mismatches.

2.4.2. Protocol for high-yield DNA nanowire annealing and processing

a. Reagent preparation: buffers and silver salts

The following protocol is used to prepare the necessary chemical solutions for nanowire annealing. Keep in mind that Ag^+ is easily reduced into silver nanoclusters, and therefore constituent solutions should be reformulated for each use, and, if necessary, stored away from light. A measure of solution purity can be inferred from the amount of smudging on the tube walls after vigorous vortexing or shaking. At 1 M, the solution will appear cloudy, but no accumulation on the tube walls should be apparent.

1. Add 10.0 mL nuclease-free water to a 15 mL tube
2. Using tabletop scale and weigh boats, carefully measure out 209 mg (1 mmol) MOPS, 850 mg (10 mmol) $NaNO_3$ and 308 mg (1.25 mmol) $MgSO_4 \cdot 7H_2O$. Add to nuclease-free water and mix vigorously. Allow to mix evenly at least 1 min.
3. Optional: consider autoclaving. This is not an absolute necessity, but it can improve buffer cleanliness. Perform if integrating with any enzyme, RNA, or molecular biology application.

4. Adjust pH to 7.2 using 10 % H_2SO_4 and 10 % NaOH. Label "10x MOPS" and indicate pH and date. Avoid using chloride salt to adjust pH as it will precipitate Ag^+ in later steps, preventing nanowire synthesis.
5. Make a 10-fold dilution to produce 1x MOPS by adding 9 parts deionized water and one part 10x MOPS in a 10 mL tube. Label one "1x MOPS pH 7.2" and indicate the date.
6. Pipette 1.00 mL 1x MOPS into a 1 mL centrifuge tube. Add 169 mg $AgNO_3$ and vortex vigorously until Ag^+ is fully in solution. Label "1 M Ag^+ in 1x MOPS."
7. Perform serial dilutions to produce 10 mM and 1 mM solutions of Ag^+: first add 10.0 µL 1 M Ag^+ and 990 µL deionized water and vortex. Label "10 mM Ag^+." Mix 100 µL 10 mM Ag^+ and 900 µL deionized water and vortex. Label "1 mM Ag^+."

b. Annealing

Polycytosine sequences are palindromic and thus will bind to themselves in the presence of Ag^+ ions. As a result, the concentration of ssDNA added will be double the concentration of the final duplex in the presence of these types of sequences, but not for mixed-base sequences. For this reason, the amount of oligo added to experimental conditions should be adjusted accordingly. It should be noted that the presence of Ag^+ may cause the formation of a dC:Ag^+:dC duplex instead of dC:dG in positive controls, though experimentally this population seems to be quite small. This may be a result of slowed kinetics in the dC:Ag^+:dC bond formation process from the dehydration energy penalties on the Ag^+ ion, as well as the two-step, three-body coordination between opposing cytosines and the cation, compared with the relatively-straightforward dC:dG electrostatic bond.

When there is system equimolarity between sequence mismatches and coordinated Ag^+, individual ions may interact with single-stranded polycytosines before forming a complex inside a double-stranded helix [105]. It is well-known that a thermodynamic penalty is paid to initiate duplex formation [108], and by

adding molar excess to the system, we reduce the amount of ion rearrangement that must occur within the duplex while strand initiation is occurring. We find that a ratio of 10 ions per CC mismatch is a good annealing ratio, and that excess ions can be removed in later steps. Annealing is performed in the following manner:

1. Heat the water bath or hot block to 95 °C. An impromptu water bath can be made by using a 1 L glass beaker inside a Styrofoam box. The tubes can float on a weigh boat with holes in the bottom. In this setup, the tubes should be sitting in the water but should not be submerged below their caps. The ideal heat source is a hot block in which the block itself can be removed to slow cool.
2. Anneal in 1x MOPS in 100 µL total volume. Add all reagents to a 100 µL microcentrifuge tube with a pipette, label, and use parafilm to seal the caps in order to prevent solution concentration and acidification by evaporation of water during annealing.
3. Vortex or mix until solution clears (do not vortex after annealing).
4. Consider spinning for 3-5 s on tabletop microcentrifuge to collect fluid at the bottom of the tube. Alternatively, tap several times.
5. Place in water bath/hot block/PCR programmed to 95 °C. Leave at that temperature 3-5 minutes.
6. Remove block/water from heat and allow to cool to room temperature. This should take ~1 h.
7. Place at 4 °C to store. Leave at 4 °C at least one hour.

c. Precipitation

As ionic excess is used in annealing, there will be phosphate-bound Ag^+ ions in solution. In order to separate the effect of these environmental ions on molecular conductivity from the effect of pyrimidine-coordinated ions, chloride precipitation is used to remove all solvent-accessible Ag^+ from the solution.

NMR data show this leaves the duplex unperturbed (Figure 2.3). Purification can be performed in the following manner:

1. Add NaCl to 50 mM.
2. Wait 15 min.
3. Use tabletop microcentrifuge to form a pellet. This may take ~30 s or slightly longer. The pellet may or may not be visible to the naked eye.
4. Pipette away the top of the solution, being careful not to remove the pellet. Approximately 60-80% of the solution can be drawn out and placed in a new microfuge tube. Not all the liquid need be removed.
5. If it appears that some of the precipitate has been removed by the pipetting, repeat precipitation protocol until clean extraction is achieved.
6. Each sample should last 3-4 weeks. Store at 4 ˚C.

2.5. Synthetic biological integration of Ag^+ intercalation

One of the primary advantages of nanotechnology and nanomaterials based on DNA is their ability to be produced, modified and integrated into biological processes. The promise of biological fabrication and/or repair of nanowire-integrated devices lies at the horizon, but as a proof of concept we investigate the ability of core-functionalized DNA nanowires to be modified by enzymatic processes; and we further design model gene parts that fabricate low-level nanowires as placeholders for more complex designs or processes. In doing so, we attempt to reintegrate the biology into the biochemical processes that drive Ag^+-dependent nanowire formation.

2.5.1. Enzyme compatibility

DNA polymerases and biased nucleotide pools can be used to synthesize nanowires in the presence of the mediating ion *in vitro*.[85,109,110] In the present study, we tested whether longer C-Ag^+-C DNA duplexes

can be constructed by traditional molecular biology methods in standard media. To this end, we carried out blunt-end ligation reactions on medium-length, double-stranded oligonucleotides. The 50-bp Oligo B duplex with six CC mismatches distributed throughout the sequence, using sections of Watson-Crick pairing to promote strand alignment and homogeneity between the products, thereby avoiding the chaining effect previously observed with the C20 and C30 sequences (Figure 2.2B, as well as SI in longer manuscript[91]). In the absence of Ag^+, a duplex product was formed but did not undergo double-stranded end-ligation, likely due to "puckering" at the mismatch points from unbonded nucleotides, which may cause bad strand alignment and inhibit ligase binding. When annealed in the presence of excess Ag^+, 100, 150, 200, 250, and 300 bp dsDNA, bands were observed as a result of successful end-ligation (Figure 2.2D). Ag^+ intercalation between the mismatched cytosine pairs repaired the Oligo D duplex and enabled ligase recognition and activity that can extend the length of a DNA nanowire length more than six-fold.

2.5.2. Synthetic gene parts

To further expand the compatibility of core-functionalized DNA nanowires with living cells, we constructed two novel BioBricks for the synthesis of cytosine-mismatched, Ag^+-binding nanowire templates in *E. coli* plasmids. BioBricks are a standard way to craft gene parts for general use through the open-source gene repository at the iGEM Registry. BioBricks are designed for modularity and compatibility with other gene parts, and are documented and curated by the BioBricks Foundation for ordering and use by registered labs.[111] PCR amplification and restriction digestion of part BBa_K1219026 produces both 32bp10CC Oligo A strands, which can be annealed into a duplex with a 10:1 Ag^+:CC molar ratio. Similarly, PCR amplification of part BBa_K1218022 produces a DNA hairpin sequence with 24 cytosine mismatches (Figure 2.9). Restriction digestion and T7 transcription of this template synthesizes a 54bp24CC RNA hairpin. These gene parts are the first templates for the synthesis of cytosine-enriched

duplexes and hairpins in cell-free transcription systems with the use of PCR, and both parts and their respective implementations are documented in the iGEM registry for use by other groups.

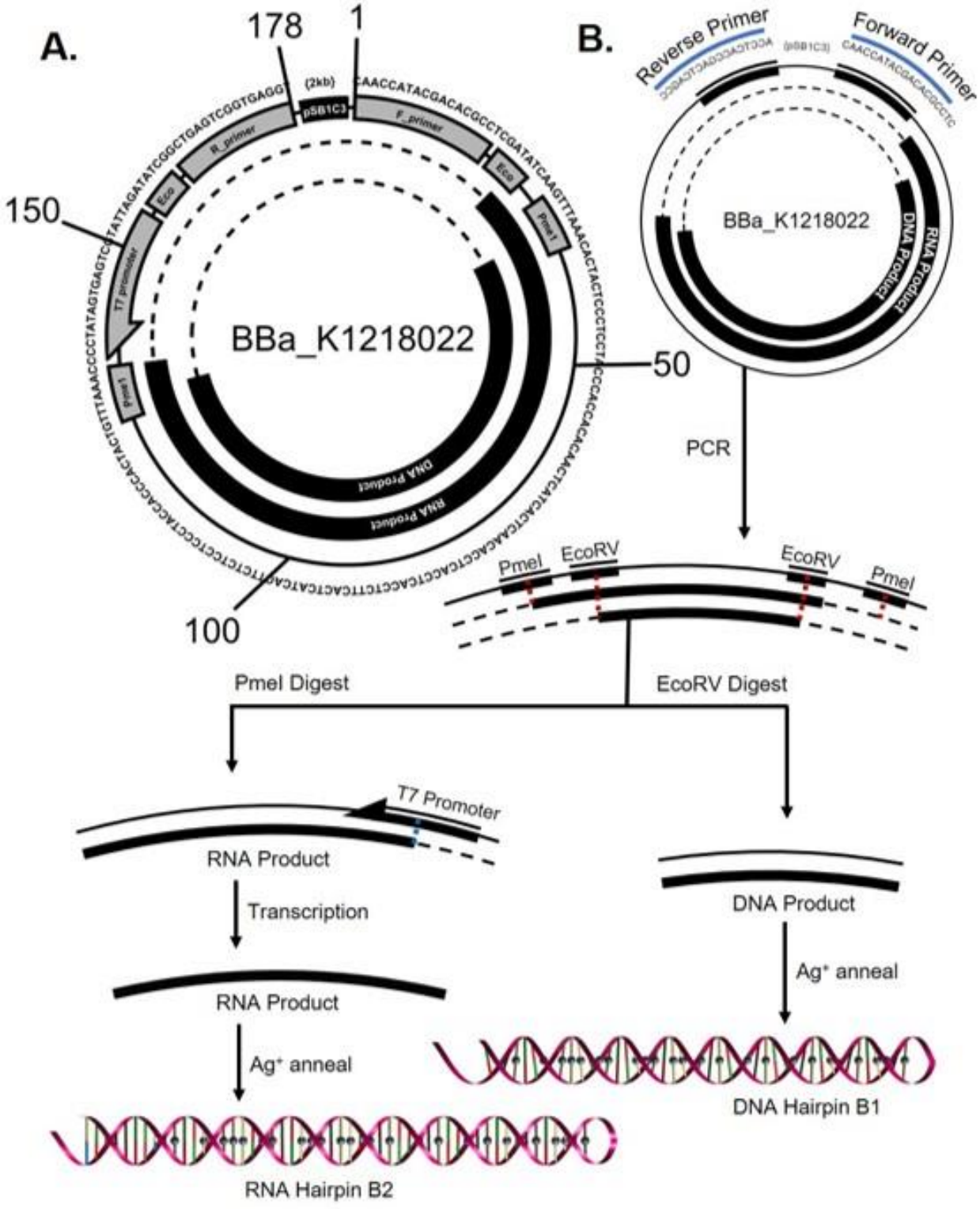


Figure 2.9: Part BBa_K1218022 serves as the template for DNA nanowire Hairpin B1 and RNA nanowire Hairpin B2. **A)** Scheme is shown with sequence data, excluding the BioBrick prefix, suffix and plasmid backbone, collectively marked as "pSB1C3," the expression plasmid backbone. Diagram includes the 178-bp functional region with forward and reverse primer binding sequences, EcoRV and Pme1 restriction sites, a T7 promoter and the two nanowire products. Concentric circles are employed to illustrate the different functions of gene part components. **B)** To utilize this BioBrick, PCR amplification of the functional region should be performed using the identified promoters. The DNA nanowire product may be subsequently isolated by PmeI excision and annealed as described above. The template for RNA Hairpin B2 is excised by EcoRV and is designed to produce an RNA nanowire hairpin in a cell-free transcription system as described above.

The functional region of gene part BBa_K1218026, including the oligonucleotide sequences of interest, the BioBrick prefix and suffix, and the expression primers, was synthesized by Elim Biopharmaceuticals and cloned into plasmid backbone pSB1C3 (BioBricks Foundation, San Francisco, CA, USA) using an EcoRI-PstI (New England Biolabs, #E0546S) cloning protocol for BioBrick assembly.[112] Part BBa_K1218022 was synthesized by GeneBlocks (Integrated DNA Technologies) and cloned into pSB1C3 using the same protocol. Part BBa_K1218022 served as the template for PCR amplification of DNA and RNA nanowires Hairpin B1 and B2 (Table 1.1). This BioBrick was utilized through PCR amplification of the functional region using forward and reverse primers 5'-CAACCATACGACACGCCTC-3' and 5'-ACCTCACCGACTCAGCC-3', respectively. The DNA nanowire product was subsequently isolated by PmeI excision (R0560S, New England Biolabs), annealed in the manner described above in the presence of 24-fold excess $AgNO_3$ with respect to oligonucleotide molarity, and purified by PAGE. The template for RNA Hairpin B2 was instead excised by digestion with EcoRV (R0195S, New England Biolabs), and the RNA nanowire itself was synthesized in a cell-free system by transcription with T7 RNA polymerase (E2040S, New England Biolabs). The RNA nanowire was annealed in the presence of 24-fold excess $AgNO_3$ and purified by

PAGE. BioBrick BBa_K1218026 was designed to allow PCR amplification using forward and reverse primers 5'-CCAAGCACGCCCACCT-3' and 5'-TGGTAGGTGGCGGTGC-3', respectively, in order to generate DNA nanowire 32bp10CC Oligo A. This product was annealed with a 100:1 molar ratio between $AgNO_3$ and oligo in the manner described above. Additional documentation may be found on the BioBrick registry.[113,114]

2.6. Discussion of Ag+ intercalation

In order to drive a better understanding of the fundamentals of DNA nanowire behavior during synthesis and processing, we employed a variety of techniques toward the study of the structural, thermodynamic, and kinetic properties of Ag^+-intercalated DNA duplexes. We develop a fuller understanding of the Ag^+-binding DNA system, and develop a high yield protocol for nanowire assembly. These results suggest a foundation for nanotechnology applications, and we further present two methods for synthesizing DNA nanowires with standard synthetic biology techniques.

In DNA strands with alternating cytosine mismatches and Watson-Crick base pairing, we have found that strands are annealed when Ag^+ ions are equimolar with the number of cytosine mismatch sites. Ag^+ ion incorporation between cytosine bases was visualized by 2D NMR, and the kinetics of incorporation was calculated from single-molecule fluorescence of hairpin formation to be 230-times slower than non-metalated hairpins. For polycytosine sequences, however, Ag^+ ion concentration must be in at least 10-fold excess and the lack of Watson-Crick base pairing results in a loss of base pair "frame," allowing polycytosine chains to develop randomly; these chains of interlocking strands can extend up to many times their individual strand length, in proportion to the environmental concentration of Ag^+ ions. Metalated polycytosine chains have melting temperatures greater than their Watson-Crick duplex counterparts, even after accounting for counterion stabilization of the phosphate backbone. Notably,

free Ag^+ ions were not found to reduce to elemental Ag in the presence of DNA duplexes and form nanoclusters, even when the strong reducing agent BH_4^- was added to the solutions.

The Ag^+-intercalated DNA duplexes were assayed by a variety of means, and reporters of both chemical and spectral nature were employed to visualize their behavior. Differences were observed between nanowire and duplex behavior, but the successful application of standard assay techniques allowed for direct comparison of the different systems. Silver, ethidium bromide, and toluidine blue staining protocols were used successfully to visualize Ag^+-driven strand annealing by PAGE; NMR and melting curve profiles were entirely consistent with dsDNA; and Ag^+-intercalated duplexes were recognized as substrates and successfully extended by T4 ligase. Ligation is especially noteworthy, indicating that multiple pyrimidine-ion pairs leave the overall helical structure relatively unperturbed, which agrees with reported structures of Ag^+-intercalated DNA duplexes.[78] Predefined, medium-length Ag^+-intercalated DNA duplexes can be synthesized by PCR amplification from standardized BioBrick plasmids and then elongated several-fold by enzymatic ligation. This result opens the door for downstream molecular biology reactions, and demonstrates that long nanowire synthesis can be achieved with definable length in this way through band excision or gel chromatography.

The primary difference between cytosine-mismatched and canonical DNA duplexes lies in the multimeric chaining behavior of polycytosine sequences. For these strands, high molecular weight streaks were discovered by PAGE at great Ag^+ ion excess. This disparity suggests that ion overload causes long chain polymerization, perhaps due to local energy minima attained when all cytosines are ion-coordinated, even in a single-stranded state during high-temperature states in annealing. With lower ion availability, lower energy states may be more readily achieved by improving the annealing frame or strand alignment; with ion excess, these states may be unavailable or disfavored due to enthalpic penalties incurred from dislodging partial coordinating bonds or entropic penalties associated with ion rehydration.

NMR analysis of the dC:Ag^+:dC bond showed the irreversible uptake of Ag^+ into the CC mismatch, as visualized by $^3J_{HH}$ coupling of cytosine H5 and H6. The spectral shift observed by Ag^+ intercalation was dependent upon ion infusion, and was not able to be perturbed by solvent exchange or counterion precipitation. Furthermore, Ag^+ that did not intercalate into the duplex, but rather adhered to the phosphosugar backbone in a stabilizing formation, was shown to be stripped by downstream processing without disrupting the ion coordination between opposing cytosines. NMR study suggests that the dC:Ag^+:dC bond is strong and pseudo-covalent, rather than hydrostatic, in nature.

The melting curves of Ag^+-intercalated polycytosine chains also were mostly unaffected by the precipitation of excess Ag^+, indicating that the vast majority of the stabilizing effect occurs as a result of nucleobase-ion coordination and not solvent-based electrostatic stabilization. These results suggest that Ag^+-intercalated DNA nanowires could be assembled by simply incubating strands of DNA with soluble Ag salts and then washing off the unbound charges; the nanowire product would associate tightly with the Ag^+ ions incorporated into the C-C mismatches, while the unbound ions would be removed. This is a highly desirable property for conducting nanowires to maintain an almost-linear conductive interior (Ag^+ ion chain) while the exterior dielectric (solvent) remains resistive.

Single-molecule FRET analysis shows distinct folding pathways for mismatched molecules: transient, low-FRET conformations occur in the absence of Ag^+, and, to a lesser degree, in the Ag^+ population. These molecules fold 230x faster than true duplexes, and represent pseudocoil states observed by other studies of bound redox-catalyzing hairpins in MEMS studies.[95] By contrast, high-FRET molecules are present only in Ag^+ conditions. This high-FRET state occurs, on average, at a rate of 0.065 s^{-1}, and represents a real-time kinetics solution to the isothermal folding of DNA nanowires.

In order to utilize the Ag^+ DNA system in nanomaterials applications, it is critical that the nanowires be easily synthesized with sequence specificity and length control. Most notably, studies of the behavior of

these wires need to exclude the possibility that results are dependent not on DNA duplexes with orthogonal chemistry, but rather polycytosine-catalyzed nucleation of silver nanoclusters.[79-80] *The spectroscopic peaks in the 450-500 nm range that represent silver nanoclusters are not observed in this study*. This is a departure from those studies on polycytosine-induced cluster nucleation which use exclusively isothermal reactions in saturating levels of reducing reagent. Our results indicate that the annealing conditions and processing experienced by duplexing DNA in the presence of Ag^+ are not sufficient to cause nanocluster formation and that the ion competition and energetic stability of the metal-mediated, double-stranded structure can prevent reduction of internally coordinated ion chains by aqueous species. To further support this conclusion, electro-spray ionization mass spectroscopy (ESI-MS) has been used in the past to analyze the distribution of $(Ag^+)_N$-DNA products for homo-base oligomers, including C11.[115] Those results showed that the dominant product was, in fact, C_{11}-$(Ag^+)_{11}$-C_{11}, indicating that the silver ions participate primarily in a 1:1 base pair binding interaction. This supports the conclusion that the Ag^+ ions are not reduced to nanoclusters along the chain, but rather assist in the duplex formation and contribute to base-pairing stability, ultimately suggesting that reduction contaminants are not responsible for the increase in molecular conductivity for precipitation-cleaned C11-amine duplexes reported previously.[62]

BioBricks, with standardized plasmids, can provide a convenient and inexpensive path toward the synthesis and scaled production of DNA nanowires. The BioBrick parts we have designed for this work require multiple processing steps, including enzymatic digestion, PCR, and Ag^+ infusion. There are, however, no fundamental barriers preventing future designs from engineering around these steps. With the addition of mechanisms like Ag^+ transporters, nanowire export proteins and tags, or co-transcriptional/reverse transcriptional assembly in cells,[116,117] future designs can be envisioned to produce useful constructs without direct intervention. Nonetheless, some obstacles still need to be overcome to ensure successful replication of more complex dC:Ag^+:dC nanowires *in vivo*. Long template

sequences with cytosine repeats will have complementary guanine-rich areas, which have been shown to assume a quadruplex conformation that may interfere with plasmid transcription and replication.[118,119]

Furthermore, any long nanowires or nanostructures assembled will be subject to a greater degree of base-pair promiscuity in the presence of Ag^+, and will necessarily require careful sequence design, which may necessitate the development of new modeling tools. Some of these issues may be addressed by incorporating different metal base pairs together in a single sequence, including metalated dT:Hg^{2+}:dT and dC:Ag^+:dC pairs,[75,86] as well as other orthogonal pairs such as the imidazole-Ag^+-imidazole or the Cu^{2+} complex bond.[120,121]

This issue also complicates the formation of nonlinear assemblies containing Ag^+ base pairs, for which it will be necessary to minimize heterostructures while maximizing cytosine occupancy: this could also introduce additional challenges in WC pairing regions as a result of the emergent parity asymmetry between standard nucleotides. Standard origami scaffold M13mp18[39] has 169,452 heterostructures greater than 3 bp in the absence of Ag^+, the largest of which comprise five 12 bp dimers; conversely, when dC:dC bonds are facilitated by Ag^+, the scaffold would have 295,755 heterostructures greater than 3 bp, including 21 size 12 bp dimers, two 13mers, one 14mer and two 16mers (analysis performed in Matlab, Table 2.2). The calculated free energy of the largest homodimers are -20 kcal/mol and -31 kcal/mol, respectively, without adjustment for the free energy of Ag^+ intercalation.[101] As such, it may be impractical to use native ssDNA plasmids without extensive mutagenesis to tailor the sequence for orthogonal chemical environments. Future studies may focus first on the assembly of short-oligonucleotide nanostructures[122,123] using Ag^+ pairing, rather than origami/staple interactions. Such studies will likely begin with extremely short oligomers with few heterostructures and work toward longer, more complex designs for plasmid integration.

Table 2.2: Heterostructure analysis of M13mp18 ssDNA origami scaffold

Self-dimers in WC pairing environment													
Size	**4 bp**	**5 bp**	**6 bp**	**7 bp**	**8 bp**	**9 bp**	**10 bp**	**11 bp**	**12 bp**	**13 bp**	**14 bp**	**15 bp**	**16 bp**
#	125630	32200	8561	2270	581	152	35	18	5	0	0	0	0
Self-dimers in dC:Ag^{+}:dC pairing environment													
Size	**4 bp**	**5 bp**	**6 bp**	**7 bp**	**8 bp**	**9 bp**	**10 bp**	**11 bp**	**12 bp**	**13 bp**	**14 bp**	**15 bp**	**16 bp**
#	205527	62423	19489	5740	1775	567	140	68	21	2	1	0	2

The BioBricks presented here are an important first step toward microbial genes designed to build electrically-active nanowires from nucleic acids, and as such, demonstrate a preliminary approach to implementing a genetic system using orthogonal base chemistry in DNA with tools found naturally in and around *E. coli.* Bacterial chasses can provide highly reliable and homogenous production capabilities, and by harnessing their onboard manufacturing systems, significant progress can be made toward the directed integration of living biological systems into nanofabrication. Oligonucleotide production may take place in cell-free BioBrick systems, or with other synthetic biological approaches such as ribozyme cleavage through rolling circle amplification[124] or enzymatic extension via reverse transcriptase.[117] Scaling of such a biologically-derived nanowire system would require only those materials involved in cell culture and DNA synthesis, as well as small amounts of ionic silver—thereby avoiding many of the resource- and reagent-associated barriers associated with silicon device technology. With this synthetic biological platform, we believe that the field of nanotechnology is capable of harnessing the dynamic properties of DNA to build effective, conductive, biological wires for use in self-assembling nanoelectronic architectures at reduced cost and increased scale.

Chapter 3

Finding the Spark:

Electrical assay of silver(I) DNA nanowires

"All that glitters is not gold"

-Common aphorism (repeated frequently by Emily Toomey in the early years)

3.1. Assay of DNA conductivity

While interest in the electrical properties of DNA has been around since the early 2000's, the study of electrical conductivity in DNA has produced a broad diversity of results, largely as a consequence of differences in experimental design. Early studies on viral DNA demonstrated resistive behavior,[53] while microbial DNA was suggested to have an RFID-like signal in bulk solution.[125] Subsequent attempts to probe the current carrying properties of DNA have followed many approaches, also with a variety of results: small bundles of DNA were found to be excellent insulators;[54] cooled DNA displayed proximity-induced superconductivity;[55] cAFM across guanine tetraplexes demonstrated semiconducting behavior;[30] and, recently, scanning tunneling microscope (STM) studies have reliably shown weak conductivity in stacked dG:dC pairs.[126]

The atomic scale break junction platform, based on the scanning tunneling microscope (STM), has become a mainstay of the field of molecular electronics, and it has proven to be a reliable and reproducible method for the assay of electron tunneling across short single molecules.[127,128] This result relies upon a reversible covalent linkage between a gold STM tip and one end of a molecule in tandem with a linkage between the other end of the molecule and a metal – generally an atomically-flat gold substrate. By oscillating the tip elevation in order to sequentially form and break contact with single molecules, it is possible to measure charge transfer across the molecule in the presence of a voltage bias and to construct height-dependent current plots that represent the ability of electrons to tunnel across the gap, that is, through the molecule itself. Because this can be repeated rapidly – up to thousands of times within a short measurement period – statistical methods can be applied to determine the minimal step size between measured currents, thereby allowing a determination of the resistance of the molecule itself. This has proven a powerful method for the study of conductance in molecules in units of the quantum conductance, since each measurement can be calibrated to a gold-to-gold single atomic-chain junction. This technique has shown repeatable differences in conductance

between DNA molecules of different lengths[61] and sequence composition,[126] and, in this work, we use it to show a significant difference between the electrical behavior of canonical Watson-Crick strands and DNA duplexes with orthogonal binding chemistry.

With experimental evidence that charge transport in native DNA decays over any meaningful distance,[53,61] efforts have focused on approaches to increase DNA conductance by design. In particular, the development of metal binding schemes either through ion-coordinating pyrimidine-mismatch systems,[75,77,120] or proposed schemes built on orthogonal nucleobases,[82,88] have generated a great deal of interest in the role of electrically-functional nucleic acid nanotechnology. It has long been hypothesized that the discovery of metal-mediated base pairs would allow for the construction of electrically-active DNA elements, but prior to the development of the break junction technique (and other single-molecule device architectures), the lack of molecular resolution in conductivity assays made it difficult to study these systems. Until recently, chemical and thermodynamic characterization of orthogonal DNA chemistries prevailed. Structural studies on these systems[121,129] and improved tight-binding (TB) and density functional theory (DFT) simulations of canonical DNA[130,131] suggested that the reduction of inter-pair distances, the hopping of electrons along bases with similar energy levels, and the depletion of dA:dT pairs could improve results in electron mobility. A key study showing the electrical functionality of guanine tetraplexes as a wide-bandgap-semiconducting material opened the door to direct electrical studies in altered-chemistry nucleic acids.[30]

Building upon advances in metal-mediated pairing [11, 20], we insert silver (Ag^+) ions into the helical gaps between mismatched cytosine bases to form molecular wires using the C-Ag^+-C base pair. We utilize two complementary methods of single-molecule (SM) electrical assay: nanofabrication of electrode arrays and STM break junction analysis. Fabrication of molecular-scale nanoelectrodes is based largely upon lab-specific expertise, and builds on a long line of studies showing assembly and testing of nanoscale heterostructures [22,132-134]. Complementary work by other groups has shown the

usefulness of blind and/or guided surface assembly on nanopatterned electrodes to drive electrical studies of molecular devices.[30,135,136] In this work, we initially sought to build single-molecule devices based on the coupling of single-walled carbon nanotubes (SWCNTs) with orthogonal DNA nanowires, which could be assembled on "larger-scale" (~100-200 nm) electrodes for the assay of these heterostructures.

In parallel, a collaboration with Toomey, Xu and colleagues at Brown University produced an STM break junction system for the assay of electron tunneling through polycytosine nanowires. As described below, this was used to demonstrate significant difference between the electrical behavior of canonical Watson-Crick strands and DNA duplexes with orthogonal binding chemistry,[62] obviating the need for the planar electrodes as a means to electrical characterization. Here we describe in full the process and results of the two methods.

3.2. Nanofabricated molecular transistors

3.2.1. CNT linking chemistry—a route to SM leads

Palma, Penzo and Wind demonstrated the ability to open carboxyl groups on the terminal end of single-walled carbon nanotubes for the linkage and assembly of diverse nanoscale heterostructures.[22,137] This opened the door to both assemblies of DNA nanostructures with CNT-functionalized regions as well as bi- and trivalent CNT-molecule-CNT assemblies. Penzo demonstrated that functionalization of CNT pairs with a dsDNA molecule was attainable with proper reaction stoichiometry.[138] In this study, we replicate the linkage of SWCNTs with dsDNA, and further attempt to combine this linkage with C:Ag^+:C-containing DNA duplexes. The ultimate goal of this chemical assembly is to use carbon nanotubes as leads to access the internal rings or other axial conductive pathways in DNA without contact resistance conferred by resistive backbone states.

In order to perform this assembly, insoluble SWCNTs are functionalized by collaborators with ssDNA,[134,139] which displays sequence-dependent adhesion to tubes of different chirality. In this way, DNA has been used to sort and purify SWCNTs and prepare them for aqueous reactions. A potential disadvantage of this approach is that transistors formed with DNA-wrapped SWCNTs as the channel have thus far exhibited low transconductance.[134] This may be a result of defects formed in the nanotubes as a result of the ultrasonication used in the process or possibly the small diameter of the tubes themselves.

Three linking chemistries are tested for CNT adhesion: terminal C6-amine functionalization ('C6-amm'); 1 nt 5' guanine overhang for internal amine access; and unfunctionalized DNA, whose terminal nucleotides also possess amine groups that, when not participating in hydrogen bonds, may be accessible for kinetic reactions. Two DNA oligos were tested: first, the sequence used by Penzo, 'EP26,' (5'-CATTAATGCTATGCAGAAAATCTTAG-3', 5'-CTAAGATTTTCTGCATAGCATTAATG-3'); and the second 11 nt C11, the principal focus of STM break junction analysis. Linking amines were either supplied by 3' amine linkers, overhanging purines, or without any modification to allow access to the terminal ring amine.

A two-step reaction[138] is employed in which 20 µL (1 nM) metallic, DNA-wrapped SWCNTs[139] is mixed with 20 µL MES buffer (200 mM), EDC (4 mM) and sulfo-NHS (10 mM) for 30 min to functionalize the terminal end of the CNTs. dsDNA is annealed in 10 mM MOPS buffer and 100 mM $NaNO_3$, with 5 min at 95 ^{o}C, followed by cooling to 25 ^{o}C over 60 min and subsequent cooling to 4 ^{o}C over 60 min. C11 samples are annealed with 2x Ag^{+}, introduced as $AgNO_3$.

The activated tubes are mixed with 20 µL 500 nM dsDNA and left to react for 24 hr. Unpaired dsDNA is removed through centrifugation in 100K Millipore Amicon filters (Sigma) at 10,000 RPM for 10 min. Activation of a second batch of 20 µL 1 nM SWCNTs in MES, EDC and sulfo-NHS is repeated for 30 min,

and the products are mixed with the DNA-bound tubes. The resulting molecules have been shown to be preferentially bivalent, with two SWCNTs and one dsDNA molecule per junction. Results of the hybridization are assayed via AFM on Si/SiO_2 freshly cleaned with oxygen plasma (18W, Harrick) prior to drop casting in a humidified chamber. Representative 10 µm x 10 µm AFM images are collected of the surface and analyzed in ImageJ using the 'Analyze Particles' feature and internal circularity of < 0.4. The resulting data are plotted in histogram format in Figure 3.1**,** and average lengths and distribution can be found in Table 3.1.

Table 3.1	**Metallic**	**EPbare**	**EPamm**	**EP-GG**	**C11amm**	**C11-AA**
Mean length (nm)	170	325	184	453	305	273
Median length (nm)	138	281	158	351	227	231
St. Dev. (nm)	108	185	115	306	209	145

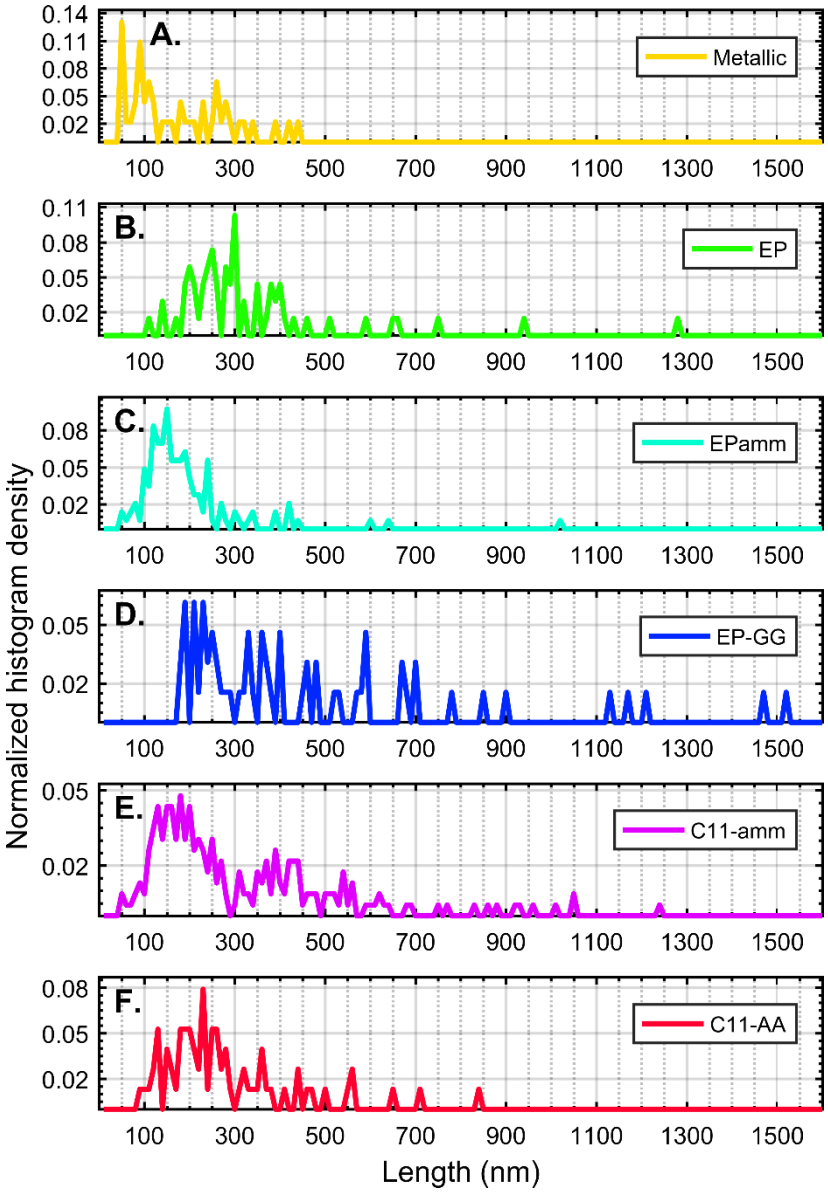


Figure 3.1: Coupling of SWCNTs and dsDNA variants, as assayed by AFM. **A)** Uncoupled metallic SWCNTs with no activation or DNA present. No tubes longer than 450 nm are observed. The average SWCNT measures 170 nm ± 108 nm. **B)** dsDNA sequence used by Dr. Penzo[138] is employed without an amine linker to harness the amines of the terminal base rings. The average length of the SWCNT conjugate was 325 nm ± 185 nm, with almost no tubes less than 200 nm in length. In addition, higher length conjugates corresponding to bi- or trivalent linking were also observed. **C)** Use of C6-amine linker increased the average length to 184 nm ± 115 nm, and eliminated all tubes less than 100 nm in length. Some longer

conjugates were observed. **D)** Use of a 3' guanine overhang with the EP26 sequence increased the average length to 453 nm ± 306 nm, suggesting highly successful coupling between SWCNTs and the dsDNA. Many long conjugates observed, and the population of tubes shorter than 200 nm is severely depleted. **E)** The C11 sequence with a C6-amine linker saw tube length grow to 305 nm ± 209 nm with few tubes less than 100 nm. Long conjugates were observed. **F)** Use of a 3' adenine overhang produced a population of tubes with an average length of 273 nm ± 145 nm, with a similar distribution. In sum, a variety of linkers were observed to successfully bridge the gap between SWCNT pairs.

Overall, the shortest population of nanotubes observed was the unmodified, metallic SWCNTs, in keeping with results attained by Dr. Penzo.[138] Whereas that study demonstrated an average value of 147 nm ± 92 nm with the same tubes, differences in experimental and measurement protocol generated an average value of 170 nm ± 108 nm. Shared among all linking chemistries was the tendency to deplete the micrographs of short nanotubes, with global tube populations rising near or above 200 nm. This suggests that, for the most part, all SWCNTs are participating in a junction with another tube. In the metallic tubes, no long conjugates greater than 500 nm in length were observed. By contrast, all the linkers were able to produce some tube conjugates of greater than 500 nm lengths, and in the dramatic case of EP-GG linking, several tubes were measured at longer than 1000 nm. It is clear overall that the addition of a dsDNA linker is able to increase the length of SWCNTs by coupling the terminal carboxyl group with amines available on the DNA helix. It is notable that the amine need not be provided by an industrial linker—instead, purine base overhangs or even the terminal nucleobase amine groups seem sufficient to form a junction, though a more thorough investigation of the bond geometry is needed in this case. These results underscore the ability of natural amines on the DNA helix to form covalent linkages with high-energy functional groups, and they further suggest that a more efficient STM junction could be assembled without the use of a six-carbon linker.

Finally, it is likely that the measurement algorithm in ImageJ underestimates the length of the carbon nanotubes. AFM does not always produce feature maps with uniform height, and there are often breaks in the image that do not correspond to the end of a nanotube on the surface. Furthermore, the ability to measure internal diameter algorithmically breaks down when a tube makes a turn of high angle. In this case, many bivalent species may be under- or uncounted by the algorithm. A manual measurement of two representative images is shown in Figure 3.2, where metallic tubes (A) are compared with C11-AA junctions (B). The overall lengths are 242 nm ± 140 nm and 419 nm ± 255 nm, respectively, when sampled randomly across the image. In this counting, the length of both populations increases, but the number of > 500 nm nanotube junctions greatly increased in the presence of a linker. Better counting algorithms and the use of high-resolution AFM may resolve these problems in the future.

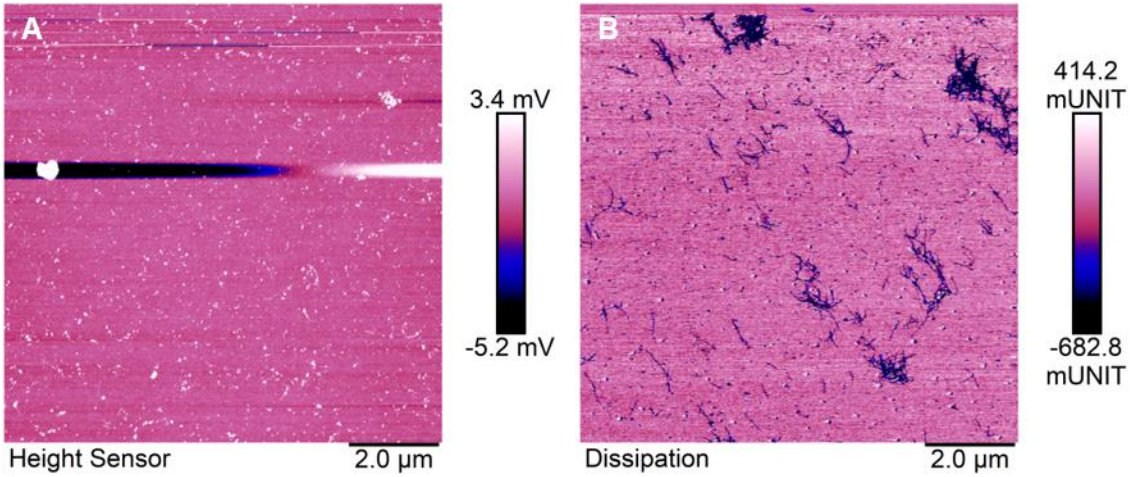


Figure 3.2: Nanotube junction AFM, 10 µm x 10 µm. **A)** Unmodified metallic tubes shown with height data. Tubes are white lines, averaging 170 nm. **B)** SWCNTs with C11-AA dsDNA linkers shown in dissipation map to enhance contrast. Many tubes can be observed of > 500 nm length that display sharp bending or branching behavior indicative of dsDNA linkers. Average length from software is 273 nm for this population, but likely an underestimate due to branching phenotype. Many tubes longer than 500

nm and some longer than 1000 nm are observed. Manual measurements of A and B show lengths of 242 nm ± 140 nm and 419 nm ± 255 nm, respectively.

In sum, dsDNA with both canonical and orthogonal base chemistry can be used to link SWCNTs in chains of at least two molecules, and furthermore that linkage may take place with amine groups supplied by industrial synthesis, base overhang, or terminal base amines typically participating in helical hydrogen bonding. This chemical linkage provides a solid basis for electrical probing of the dsDNA molecule with metallic SWCNT leads.

3.2.2. Device nanofabrication

To promote the assembly of single molecules or groups of molecules onto planar electrodes, a device fabrication is carried out in four phases: 1) alignment mark fabrication and dicing; 2) electrode fabrication using electron beam nanolithography (EBL); 3) surface passivation and subsequent patterning; 4) molecular deposition and electrical assay. Alignment marks are fabricated in order to allow subsequent patterning over existing features—the ability to place features with nanometer precision requires exact positioning and calibration of existing device features. Indeed, the primary barrier to directed self-assembly into high-energy areas has been the stability of alignment marks under aggressive cleaning protocols, limiting downstream nanofabrication.[138] Gold electrodes are subsequently fabricated with EBL using the alignment marks to precisely define the placement of large and small features. A gap of 100 nm between electrodes allows the assembly of one or several molecules to create a transistor. When directed surface assembly (DSA) is employed, the surface is passivated with polyethylene glycol (PEG) and a 2 nm-wide high-energy trench is subsequently etched into the nanogap to allow directed SWCNT placement. If the transistor employs several molecules, this step is omitted. The primary failure of most devices of this nature has been the precision of the trench relative to the

nanogap, owing to alignment mark blurring during fabrication. Finally, the molecules are deposited and measured electrically.

a. Alignment mark fabrication on SiO_2 gate oxide

Alignment marks are 50 nm tall gold squares spanning 20 μm. A chromium underlayer of 2 nm is employed to ensure adhesion of the gold to the substrate. The substrates are oxidized Si wafers, which can serve as a gate electrode (the oxide is the gate dielectric) in a final device structure. A bilayer resist is employed to allow effective deposition of metal features after lithographic patterning.

A silicon (100) wafer with 300 nm, undoped, N-type silicon dioxide layer, 3 in x 0.50 mm (SI-SO-Ua76D05C1-300nm, MTI Corporation, Richmond, CA, USA) is cleaned in aged piranha (one part hydrogen peroxide to three parts concentrated sulfuric acid, cooled at RT for 90 min) for 5 minutes. The wafer is carefully removed from piranha and washed briefly in distilled water, then in ethanol. The surface is subsequently blown dry in an inert gas such as H_2, N_2 or Ar. The surface is coated in e-beam resist EL5 (Microchem, Westborough, MA USA) with a spin coater (4000 RPM, 2000 ACC) and baked for 5 min at 180 ^oC on a hot plate. A second layer of PMMA (495K A2, Microchem) is spun (2500 RPM, 1000 ACC) and baked for 10 min at 180 ^oC.

Alignment mark patterns are then written using an electron beam lithography system (Nanobeam nB4, Cambridge, UK) at 37 nA using a dose of 12 C/m^2. Development of the pattern is carried out for 120 s in cold IPA/MIBK 3:1 with gentle sonic agitation. In a high-vacuum e-beam deposition chamber (Angstrom EvoVac, Angstrom Scientific, Ramsey, NJ, USA), 2 nm Cr and 50 nm Au are deposited at a rate of 0.1 Å/s and 0.5 Å/s, respectively. Liftoff of the remaining PMMA is carried out in remover PG overnight.

b. Electrode fabrication

Electrodes are fabricated on 1 cm Si/ SiO_2 chips with gold alignment marks. Again, a two-layer resist is used for optimal feature development and liftoff, and due to the disparity in feature size between the

pads (micrometers) and contacts (nanometers), the e-beam write is carried out in two stages. The following protocol was adapted from studies on CNT field-effect transistors by Dr. Erika Penzo.[134,138]

To protect the surface of the wafer during dicing, a layer of PMMA 495K A2 is spun (2500 RPM, 1000 ACC). The wafer is then diced into 1 cm chips using a cleanroom auto-dicing saw under distilled water stream (DAD3220, Disco Corporation, Santa Clara, CA USA). When the chip is ready for electrode fabrication, liftoff of the resist is performed in remover PG overnight. The substrate is then washed in acetone (10 s), isopropyl alcohol (IPA, 30 s), blown dry with N_2, and exposed to oxygen plasma for 5 min (18W PDC-32G, Harrick Plasma Inc., Ithaca, NY USA). A layer of EL copolymer (Microchem) is spun 100 nm thick (4000 RPM, 2000 ACC) and baked at 155 ^{o}C for one hour. A subsequent 50 nm coat of PMMA 495K A2 (Microchem) is spun (2500 RPM, 1000 ACC) and baked for 1 hr at 170 ^{o}C. Electrodes are then written in two stages using an electron beam (Nanobeam), with large features writing at 37 nA, and the small features subsequently being written at 1 nA. The resist is developed in 3:1 IPA:DIW for 1 min at room temperature, exposed to oxygen plasma for 15 s (18 W, Harrick), and then deposition of 1 nm Ti and 50 nm Au is carried out in an e-beam deposition chamber (Angstrom Scientific) at a rate of 0.5 Å/s. Liftoff of the resist is carried out in remover PG overnight.

c. CNT Deposition

Two distinct methods of DNA-wrapped SWCNT deposition onto patterned electrodes are employed, depending on the application. The first method is a drop casting of CNTs randomly onto the surface to create a multi-tube CNT transistor. The second method, more broad-ranging in scope, involves creating high surface energy trenches in PEG-passivated substrates to direct single CNT assembly into desired areas. This directed self-assembly, developed by Penzo et al,[134,140] offers the promise of heterogeneous single-molecule nanomaterial transistors for direct probing of molecular behavior. This method has suffered from high resistance in low-quality tubes, as well as feature loss during surface cleaning in an

attempt to reduce contact resistance. As successful devices in this work were only constructed with non-directed assembly, we describe here only a random deposition protocol. Details of a similar DSA can be found in studies by Dr. Penzo.[134,138]

A 40 µL droplet of (5,5) and (6,6) metallic SWCNTs is prepared in 0.1x TAE buffer, 0.25x DPBS buffer, and 12.5 mM $MgCl_2$. After liftoff, the chips are treated with 1 min oxygen plasma , and the CNT droplet is placed at the center of the substrate. A humidified chamber is prepared by wetting a Kimwipe and placing in a dish near the chip, sealing the dish, and allowing evaporation to occur slowly for at least 3 hr. The chip is then washed in 1:1 EtOH:DIW for 10 s, and then incubated in 9:1 EtOH:DIW for 1 hr. Samples are then blown dry in N_2 gas and stored in an airtight container until assay.

3.2.3. Electrical assay, results, limitations

Devices are subjected to voltage sweep using an Autoprobe system with Matlab script. Data are analyzed and devices are subsequently imaged by AFM to determine results of nanofabrication. A successful set of devices using randomly deposited SWCNTs reports conductance in the range of 100 nA, while a similar device with no tubes and a gold-bridged nanogap demonstrated conductance of 1 µA (Figure 3.3). A device fabricated by Dr. Penzo with single-molecule assembly demonstrated conductance in the range of 1 nA,[134] corroborating data showing that successful devices produced here were bridged by many nanotubes. The feature height of the contacts made AFM assay of tube occupancy in the trench impractical, but a close look at Figure 3.3I shows the 'ciliated' phenotype of drop cast CNTs, suggesting that the 100-200x recorded conductance is, in fact, bulk nanotubes. By contrast, the device has a maximum conductance on the order of 1 µA at ±500 mV, as seen by bulk gold contact in Figure 3.3H,K.

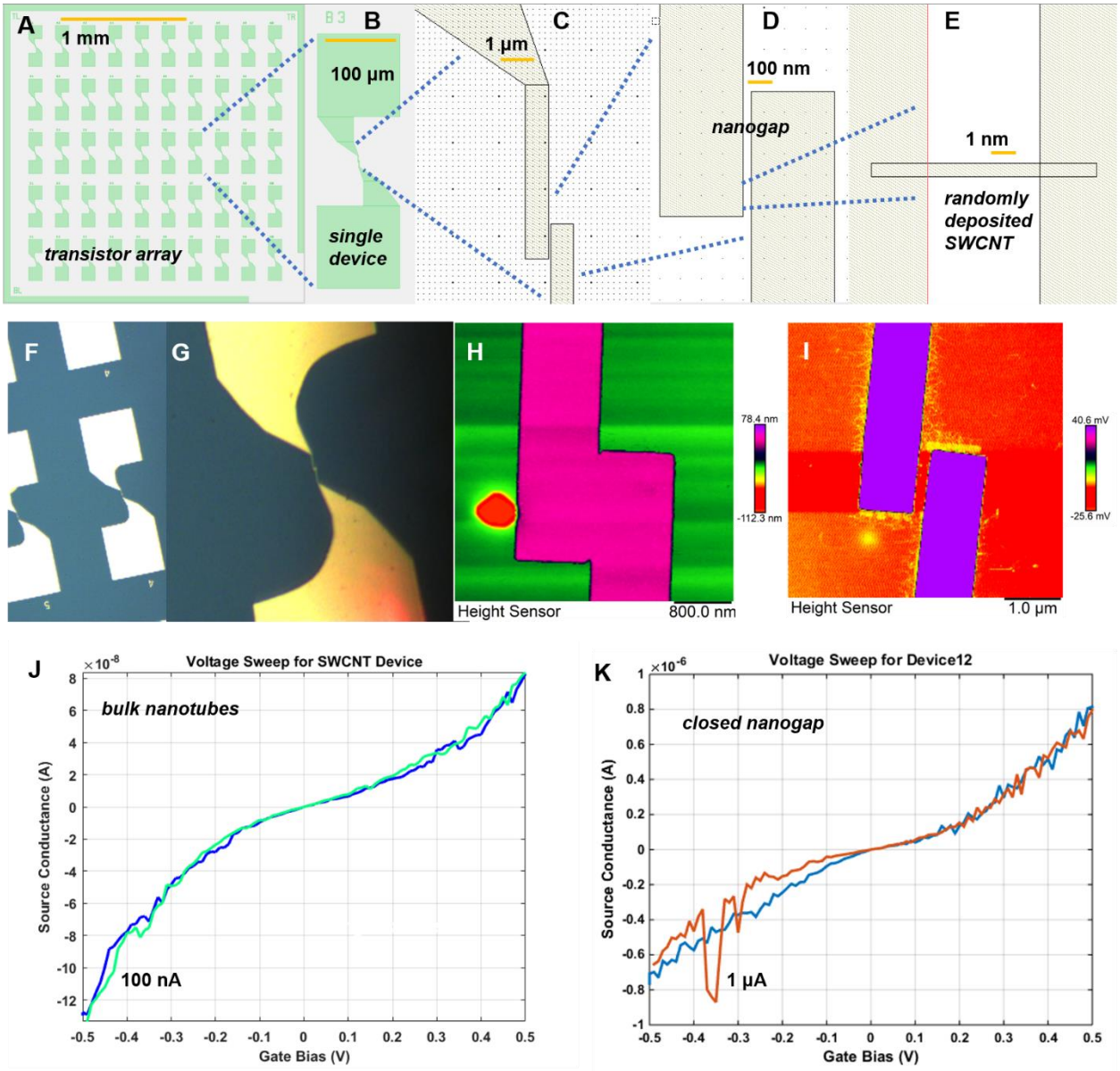


Figure 3.3: Nanofabrication and electrical assay of SWCNT transistors. **A)** A 10x5 array of devices is designed in L-Edit (Mentor Graphics, WIlsonbille, OR, USA) in a symmetrical grid, with 100 µm spacing on all sides between features. **B)** A single device consists of 100 µm gold pads with a tapered end. **C)** The terminal portion of the pads consists of 1 µm gold lines. **D)** The ends of the pads are separated by a 100 nm gap over a distance of 1 µm. **E)** The gap is fabricated to accommodate one or many SWCNTs, depending on subsequent surface preparation. All gold features are 100 nm in height. **F)** After fabrication, arrays of devices can be seen via optical microscopy. **G)** A single device seen in detail shows

the overlapping region where the nanogap lies. **H)** Some devices are overexposed during electron beam writing, causing the pads to fuse across the gap. In the AFM micrograph shown, the contacts have uniform height. No nanotubes were added to this device. **I)** In a correctly-fabricated device, a clear gap is developed between the pads, demonstrating an effective width of 270 nm at its widest and 210 nm at its narrowest, deviating from the original design. Self-assembly of carbon nanotubes onto the surfaces is done without passivation layering, and as a result, nanotubes are ubiquitous at the edge of the contacts (yellow in scale) to produce a ciliated phenotype. It is clear that in this device, many nanotubes bridge the nanogap, though the AFM tip did not resolve to layout inside the gap. **J)** Voltage sweep of ciliated device in I shows metallic behavior across the nanotubes, with a conductance in the range of 100 nA. **K)** Fused electrodes (H) demonstrate similarly metallic curves with a magnitude of ~1 µA. Plots were generated in Matlab.

As a platform, the ability to probe several molecules in a nanoscale transistor is quite powerful. With further development of the DSA protocol, better nanotubes, and some attention to the CNT-DNA heterostructure assembly, it may be possible to assay directly the conductive behavior of single C:Ag^{+}:C DNA nanowires. This approach is time-intensive, and was put on hold after obtaining reasonable measures of conductivity from STM, as described below.

3.3. STM break junction experiments: bridging the gap

The STM break junction technique offers SM resolution with thousands of iterations within a span of a few minutes, generating statistical data on the performance of a molecular population. Comparison of conductance between molecules is typically done by measuring the behavior of gold-gold junctions, which typically exhibit conductance of one quantum unit: $G_0 = \frac{2e^2}{h} \approx 77\ \mu S$.

In order to facilitate effective STM experimental design, several important factors need to be addressed, namely: noise reduction in instrumentation to allow pA sensitivity; surface preparation to

promote atomic flatness and feedback reduction during measurement; and molecular homogeneity both within the wires and in bond formation with the scanning. Strategies for overcoming these difficulties are addressed below, and suggestions for improvement in future iterations are considered.

3.3.1. Device design

The STM break junction used in this work[141] used an ultrasharp gold tip made from sonication of a gold wire and subsequent washing with acetone and ethanol. Piezoelectric stage control was carried out using a trianglular waveform generator regulated by a Microdrive controlled and linear actuator. Voltage was applied through gold substrate at 300 mV. Noise reduction was carried out by a high gain amplifier to allow signal enhancement above background levels. The tool was set up to run using Matlab scripts and manual surface identification, as no imaging mode was available through the tool. Calibrated using gold-gold junctions, the most common conductance step was found to be G_0.

3.3.2. Linking chemistry

The sample sequence used in this work is an 11 nucleotide (nt) polycytosine which can either form a silver-coordinating duplex with 11 Ag^+ ions and two C11 oligos, or a native duplex with C11 and its Watson-Crick reverse complementary sequence, G11 (Table 3.2). This sequence is selected because it provides an excellent template and is the basis of previous molecular electronics work.[62] While polycytosine oligomers suffer badly from alignment issues during annealing, the C11 oligo is short enough that it preferentially forms a duplex without an overhang.[142] Longer polycytosines begin to form sticky-end-like duplexes that anneal into long chains—also interesting as electrical wires, but their length and inhomogeneity makes them unsuitable for straightforward electrical characterization. Furthermore, the C11 molecule demonstrates a very high thermal stability and shows a lack of nanocluster formation after annealing with Ag^+ (Figure 2.6).

Table 3.2: Oligonucleotide sequences for SPM assay

Code	Sequence Name	Sequence with IDT DNA mod codes
[1A]	C11 thiol	CCC CCC CCC CC /3ThioMC3-D/
[1B]	G11 thiol	GGG GGG GGG GG /3ThioMC3-D/
[2A]	C11 amine	/5AmMC6/ CCC CCC CCC CC
[2B]	G11 amine	/5AmMC6/ GGG GGG GGG GG
[3A]	C11 overhang	CCC CCC CCC CCA
[3B]	G11 overhang	GGG GGG GGG GGA

In order to perform direct electrical measurements or other operations on DNA duplexes, a chemical linker must be used to access the ring system or ion coordination chain inside the double helix (Figure 1.2B, Figure 3.4). Every linker system has its advantages, but none offers the perfect electrical contact. Amines provide weaker coupling than thiols and are known to chelate silver ions, while thiols are unstable electrically and add noise into the measurement.[143] To reduce the effects of chelation on bond availability, Ag^+ not coordinating cytosine mismatches is precipitated out as AgCl (Figure 3.5). Standard linkers that are obtained commercially often come with three- or six-carbon spacers (C_3, C_6) between the DNA oligo and the functional group, providing steric accessibility but generating a good deal of contact resistance and reducing the measured conductivity of the molecule (see [126] vs. [61]). Commercially available linkers are furthermore added to the phosphate backbone of these molecules and are consequently in direct contact with the least conductive part of the DNA molecule. Here we describe a single nucleobase overhang in which the native C6-site amine of adenine is used as a substitute for longer linkers. This allows an electrical contact directly into the ring system of the DNA and is at most 4-6 Å from the ion chain in a terminal dC:Ag^+:dC bond. When assayed via PAGE and thermal denaturation, the annealing profile of the molecules appears standard across the linker variants, suggesting that the bond chemistry does not disrupt ionic bond formation (data not shown). The STM studies utilized amine

linkers,[62] but future studies may expect to measure increased molecular conductance by at least one order of magnitude through the use of natural ring amines.

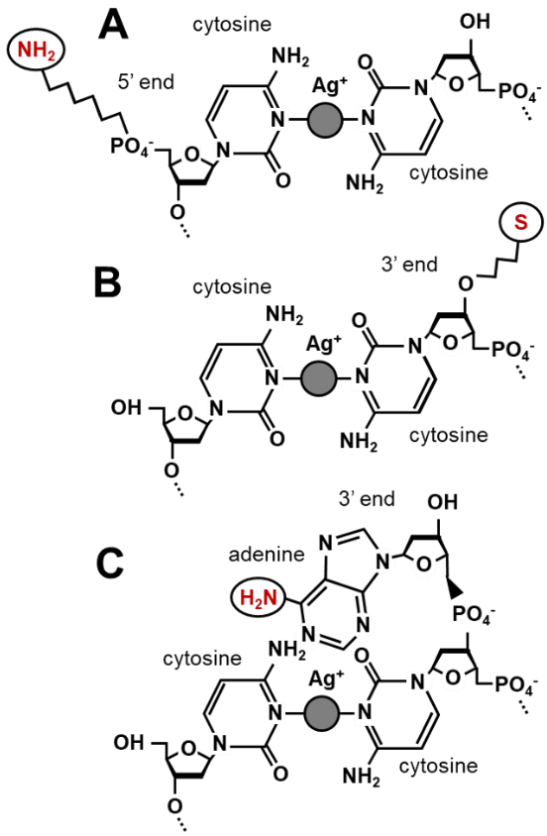


Figure 3.4: Linker chemistries. **A)** 5' C_6 amine provides a stable link with higher resistance. **B)** 3' C_3 thiol provides a more volatile link with lower resistance. **C)** Overhanging terminal adenines present primary amines that can be used to directly access the DNA ring structure.

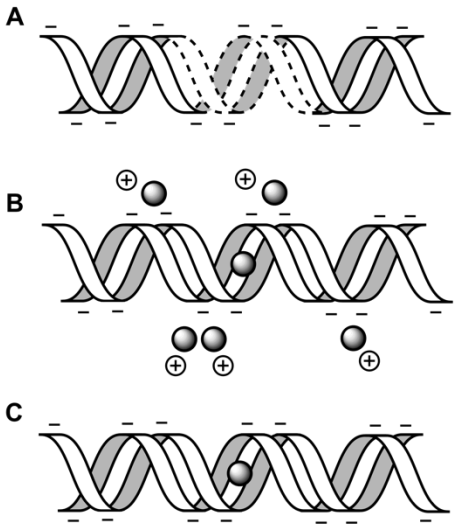


Figure 3.5: Duplex annealing and ion localization. A) DNA strand with mismatches in the absence of Ag^+. B) After annealing with Ag^+, ions are coordinated between cytosine mismatches and also form weaker electrostatic interactions with phosphate groups on the backbone. C) Precipitation of AgCl does not perturb cytosine-coordinated ions but does strip Ag^+ from the backbone.

The DNA oligos themselves can be ordered from a variety of commercial sources, such as Integrated DNA Technologies (IDT, Coralville, IA, USA) or Elim Biopharmceuticals (Hayward, CA, USA). For convenience, the modification codes for IDT ordering are added to sequences in Table 3.2. To ensure the purity of the oligonucleotides and exclude any outlying sequences, the strands are either PAGE- and/or high-performance liquid chromatography (HPLC)-purified. Unannealed oligos are resuspended at a final concentration of 100 µM in nuclease-free distilled water and stored at most one year at -20 ^{o}C.

3.3.3. Atomically-flat gold surface preparation

This section presents a protocol for sample surface preparation for scanning probe microscopy (SPM) measurements, focusing specifically on electrical assay of single-molecule conductivity through the application of a voltage bias between a metal SPM tip and a gold surface. These methods can include non-contact atomic force microscopy (AFM) modes such as: electrical force microscopy (EFM); static

charge (Kelvin probe) force microscopy (KPFM); contact-mode conductive AFM (cAFM); scanning tunneling microscopy (STM) modes such as imaging STM and break junction experiments; and hybrid techniques such as tapping-mode PeakForce tunneling AFM (PF-TUNA) (proprietary mode: Bruker Corporation, Billerica, MA, USA), which gathers quantum nanomechanical and electron tunneling data in tandem. What these scanning probe measurements have in common is the necessity for smooth, conductive gold surfaces for sample bias application and data acquisition.

While AFM imaging may be reliably carried out on mica, any form of electrical measurement must be carried out on a conductive substrate such as gold. The degree of smoothness depends on the desired resolution and feedback-sensitivity of the SPM mode. As such, processing of the gold film is unnecessary in low-resolution modes such as KPFM and EFM; similarly, non-imaging STM such as break junction experiments can be reliably carried out without post-deposition processing due to the statistical nature of data analysis. Conversely, SPM modes requiring high-resolution, feedback-enabled imaging (e.g., STM and PF-TUNA) are unyielding and require atomically-flat gold.

There are several methods for creating ultra-flat gold surfaces for SPM measurements; these generally involve deposition of gold on cleaved mica and the subsequent use of hydrogen flame annealing under nitrogen atmosphere at varying temperatures.[144,145] These methods are effective at producing locally-flat gold but suffer from low throughput. Deposition of gold usually carried out in a vacuum-sealed chamber in which individual samples need to be affixed to a deposition chuck. Mica offers true atomic flatness due to its surface grain, but it is impractical to cleave mica in chips larger than one centimeter in size. This restricts the number of samples that can undergo deposition at one time, which is frequently a time- and resource-intensive process. To circumvent the problem of throughput, we demonstrate here a protocol for forming ultra-flat gold films on three-inch silicon/ silicon dioxide (Si/SiO_2) wafers that can be diced into 12 to 16 1-cm^2 chips and subsequently processed. Depending on the size of the deposition

chamber, up to four wafers may be prepared at once, greatly reducing the time and cost of preparing imaging surfaces.

a. Analysis of various thin film processes: assay of surface roughness using AFM

In order to determine an effective thin film process for achieving atomically-flat or nearly-atomically-flat gold, we assayed the surface roughness of a wide variety of cleaning and deposition techniques via AFM (Table 3.3 and extended data in JSAME manuscript).[107] Gold films were deposited using an electron beam evaporation chamber (Angstrom Scientific) (100-200 nm at 0.5 Å/s) on either: cleaved mica, Si/SiO_2 chips cut with a diamond knife, or wafer-scale Si/SiO_2. Films were tested both with and without a 1 nm titanium adhesion layer for atomic lattice alignment. After deposition, surfaces were subjected to a variety of processing steps, including sonication, acid bath, oxygen plasma, flame annealing and controlled-temperature furnace annealing. Samples were then assayed by AFM (Dimension Icon, Bruker) using 2 nm nominal radius silicon nitride tips (spring constant 0.4 N/m, ScanAsyst-Air, Bruker) using a peak-force tapping mode (PF-QNM in Air). Representative 2 µm x 2 µm scans were analyzed for surface roughness: surface area, percent difference in surface area from 4 μm^2; and line roughness: feature average height (R_a), root mean squared feature height (R_q), and maximum feature height (R_{max}). These results can be visualized in Figure 3.6. Detailed process notes and AFM micrographs for each sample can be found in the supporting information for publication at JSAME.[107]

Table 3.3: Processing of Gold Films

Methods					Surface roughness (2 μm x 2 μm scan)					Notes
Substrate	Cleaning	Dicing Method	Deposition	Processing	SA (μm^2)	SA % diff from 4 μm^2	Ra (nm)	Rq (nm)	Rmax (nm)	
Mica disc adhered to metal	Fresh layer cleaved with tape prior to deposition	none	1 nm Ti (0.5 Å/s); 100 nm Au (0.5 Å/s)	none	4.16	4.08	1.47	1.88	19.7	*Surface bubbly:* nodes are ~70 nm wide and ~10 nm tall
Si/SiO2 (300 nm)	Aged piranha (5 min); distilled water (30 s); ethanol (30 s); dry with nitrogen stream; oxygen plasma (5 min)	Diamond knife; manual cut prior to deposition	1 nm Ti (0.5 Å/s); 100 nm Au (0.5 Å/s)	none	4.26	6.49	1.29	1.63	13.3	*Surface bubbly:* nodes are ~70 nm wide and ~10 nm tall
Si/SiO2 (300 nm)	Aged piranha (5 min); distilled water (30 s); ethanol (30 s); dry with nitrogen stream; oxygen plasma (5 min)	Auto-dicing saw with distilled water stream	1 nm Ti (0.5 Å/s); 200 nm Au (0.5 Å/s)	none	4.14	3.39	2.45	3.08	25.3	*Surface bubbly:* nodes are ~70 nm wide and ~10 nm tall
Si/SiO2 (300 nm)	Aged piranha (5 min); distilled water (30 s); ethanol (30 s); dry with nitrogen stream; oxygen plasma (5 min)	Auto-dicing saw with distilled water stream	1 nm Ti (0.5 Å/s); 200 nm Au (0.5 Å/s)	Oxygen plasma 18 W (5 min)	4.32	8.1	2.42	3.42	49.3	Nodes are *grain-like* and flat; nodes are ~100 nm wide with ~15 nm trenches
Si/SiO2 (300 nm)	Aged piranha (5 min); distilled water (30 s); ethanol (30 s); dry with nitrogen stream; oxygen plasma (5 min)	Auto-dicing saw with distilled water stream	1 nm Ti (0.5 Å/s); 200 nm Au (0.5 Å/s)	Benchtop bunson burner: high heat (150 s); low heat (150 s)	4.47	11.8	3.98	6.55	103	*Rolling surface:* grain size ~300 nm with boundary peaks ~20 nm tall
Si/SiO2 (300 nm)	Aged piranha (5 min); distilled water (30 s); ethanol (30 s); dry with nitrogen stream; oxygen plasma (5 min)	Auto-dicing saw with distilled water stream	1 nm Ti (0.5 Å/s); 200 nm Au (0.5 Å/s)	Benchtop bunson burner: high heat (150 s); low heat (150 s); constant nitrogen gas	4.27	6.67	2.66	4.21	55.5	*Rolling surface:* grain size ~500 nm with boundary peaks ~20 nm tall
Si/SiO2 (300 nm)	Aged piranha (5 min); distilled water (30 s); ethanol (30 s); dry with nitrogen stream; oxygen plasma (5 min)	Auto-dicing saw with distilled water stream	1 nm Ti (0.5 Å/s); 200 nm Au (0.5 Å/s)	Anneal in forming gas at 300 ^{o}C (21 h)	4.07	1.86	1.04	1.62	31.7	*Rolling surface:* grain size ~300 nm with boundary peaks ~10 nm tall
Si/SiO2 (300 nm)	Aged piranha (5 min); distilled water (30 s); ethanol (30 s); dry with nitrogen stream; oxygen plasma (5 min)	Auto-dicing saw with distilled water stream	1 nm Ti (0.5 Å/s); 200 nm Au (0.5 Å/s)	Oxygen plasma (5min); anneal in forming gas at 300 ^{o}C (21 h); oxygen plasma (5 min)	4.04	0.919	0.78	1.31	27.1	*Rolling surface:* grain size ~400 nm with boundary peaks ~5 nm tall

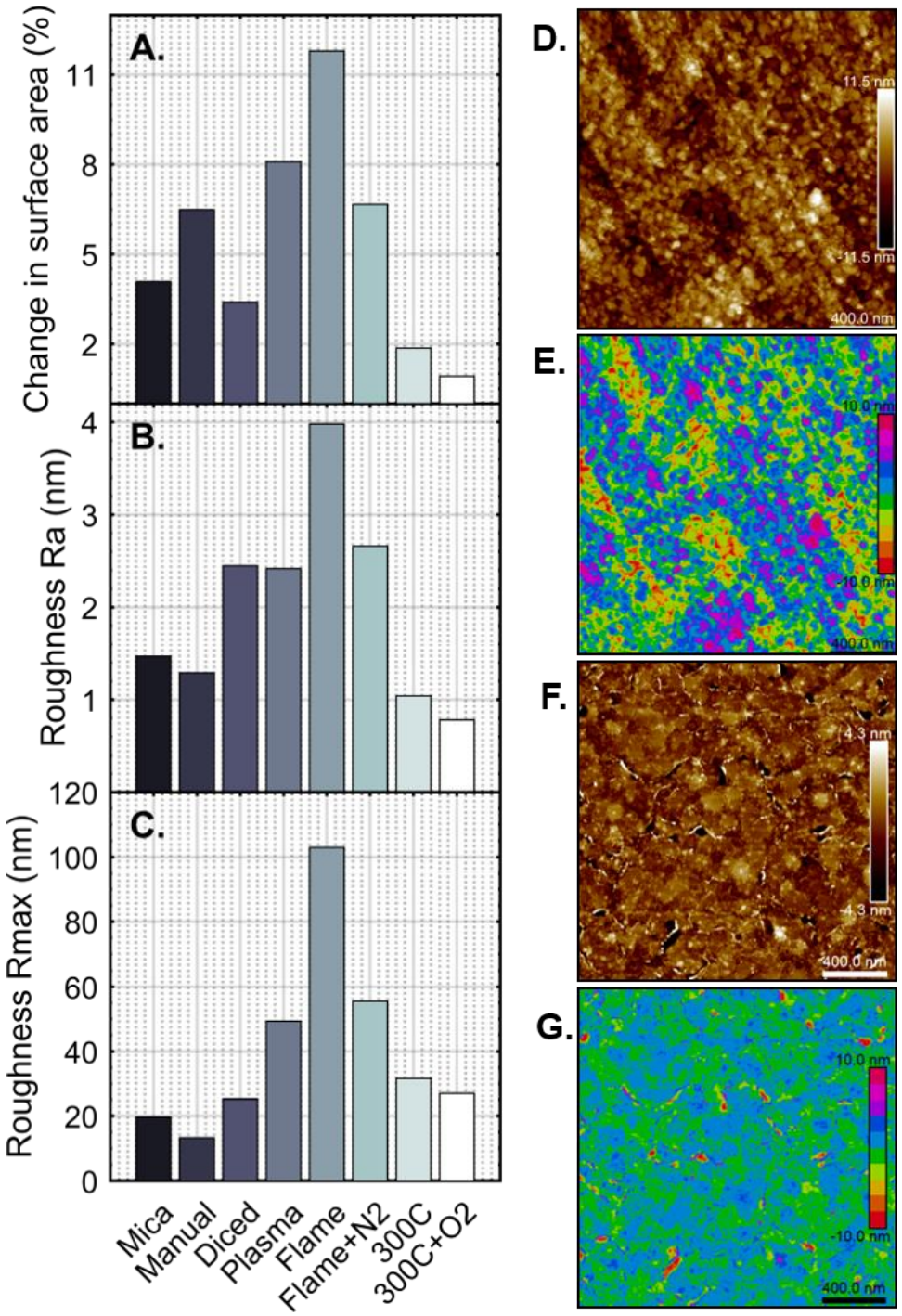


Figure 3.6: Surface roughness of processed gold films. **A)** Percent difference in surface area of 2 µm x 2 µm AFM scan from 4 μm^2 on gold films with differential processing. **B)** Quantitative roughness measurement of height deviation average from flattened image (R_a) in the same AFM images. **C)** Maximum feature height in flattened AFM images (R_{max}). Samples shown are (from left to right): cleaved

mica; manually-cut Si/SiO_2; diced Si/SiO_2 wafer; diced Si/SiO_2 with 5 min oxygen plasma; diced Si/SiO2 annealed in a gas flame; diced Si/SiO_2 annealed in gas flame with nitrogen gas stream; diced Si/SiO_2 annealed at 300 ˚C in forming gas for 21 h; diced Si/SiO_2 annealed similarly with 5 min oxygen plasma both before and after annealing. Detailed process notes can be found in Table 3.3. D-G) 2 µm x 2 µm AFM scans of representative surfaces. **D)** Titanium (1 nm) and gold (200 nm) are deposited on a clean Si/SiO_2 wafer, diced using a dicing saw, and imaged directly after ("Diced" in A-C). Height scale ±11.5 nm. **E)** Threshold-processed micrograph reproduction of D with 2.5 nm height steps to illustrate surface levels. **F)** Titanium (1 nm) and gold (200 nm) are deposited on a clean Si/SiO_2 wafer, diced using a dicing saw, annealed in forming gas at 300 ˚C for 21 h, and cleaned afterward for 5 min in a tabletop oxygen plasma tool. Samples were imaged directly after ("300 ˚C + O_2" in A-C). Height scale ±4.3 nm. **G)** Threshold-processed micrograph reproduction of F with 2.5 nm height steps to illustrate surface levels. AFM images taken on a Dimension Icon AFM (Bruker). Data analyzed using NanoScope Analysis software (Bruker).

b. Effects of dicing method on surface roughness

To validate a high-throughput, batch deposition process, we compare the roughness of films grown on cleaved mica, Si/SiO_2 wafers cut manually with a diamond knife, and Si/SiO_2 wafers cut using an auto-dicing saw found in many cleanrooms (e.g., DAD3220, Disco Corporation, Santa Clara, CA USA). The use of a dicing saw enables highly precise sample cutting, but the process itself involves the spray of deionized water, possible ionization of that water during dicing, and ultimately the deposition of silicon dust onto the gold surface. The samples themselves show more visible dust in a light microscope, but AFM imaging demonstrates only slight increase in surface roughness (2.45 nm to 1.47 nm) and comparable surface area (4.14 μm^2 to 4.16 μm^2) when contrasted with samples deposited on cleaved mica (Figure 3.6A-C, Table 3.3). Surface features have similar phenotype with nodes ~70 nm in diameter and ~10 nm in height. Like thin films on cleaved mica, silicon wafers diced using a dicing saw may be

appropriately flat for KPFM and break junction experiments but should be subjected to downstream processing for SPM modes that require gain and feedback modulation.

c. Effect of surface processing after dicing

After deposition of gold on whole wafers and use of the dicing saw, samples can be subjected to additional processing to increase the overall surface homogeneity of the gold film. The primary objectives of downstream processing are to produce grain-like surface features, rather than the rounded nodes produced during deposition, and to decrease the quantitative surface roughness to produce the best imaging substrate. Flame annealing, as suggested by various sources,[144,145] proved to be an effective method for producing grain-like surface features with locally flat areas up to 500 nm (Table 3.3). Unfortunately, these surfaces demonstrated tall features at the grain boundaries of up to 100 nm without nitrogen gas, or 60 nm when annealed in a two-stage benchtop flame process with a stream of nitrogen. Repeating this process five times did not improve overall flatness (see Vecchioni, 2018,[107] Table S3).

Other procedures such as acid cleaning and sonication were tested and did little to improve the imaging surface (Vecchioni, 2018,[107] Table S2). We did, however, find that oxygen plasma in a tabletop plasma cleaner (18 W, Harrick) has the effect of smoothing out surface features—in the case of the rounded nodes produced by deposition, the surface of the nodes were made relatively flat, though the size and overall shape was not changed (Figure 3.6A-C, JSAME manuscript Table S2).[107] This suggests that oxygen plasma smooths existing surface features but does not change the topography of these features. As such, oxygen plasma may be used to erode any tall features produced during surface processing.

We tested the effects of controlled temperature annealing on the quality of gold films in a controlled furnace with constant application of forming gas (Ar/H_2). While gold has a bulk melting temperature of approximately 1000 $^{\circ}$C, a thin film will aggregate and cease to act as a surface at much lower

temperatures. We found that for a 200 nm gold film with a 1 nm titanium adhesion layer on Si/SiO_2, the denaturation temperature is roughly 400 ˚C (Table S1).[107] To test the effect of low temperature annealing, gold films were subjected to a temperature of 300 ˚C in forming gas and annealed for 21 h. The results demonstrate that surface flatness is lowest in samples annealed at 300 ˚C and cleaned with oxygen plasma both before and after annealing. Compared with unprocessed gold films, these samples exhibited a 70% overall reduction in area roughness (4.04 μm^2 to 4.16 μm^2) and line roughness (R_a 0.78 nm to 2.45 nm), and comparable maximum feature height (27.1 nm to 25.3 nm) (Table 3.3, Figure 3.6A-C). Surface phenotypes assayed with AFM show ~400 nm grains with ~5 nm boundary features in annealed samples (Figure 3.6XF-G), compared with ~70 nm wide nodes with ~10 nm edge features on unprocessed gold (Figure 3.6XD-E). In sum, the slow anneal at 300 ˚C appears to provide enough energy for gold atoms to migrate but insufficient energy for the formation of large clusters and surface degradation. Slow annealing has the overall effect of producing flat surface grains that are ideal for SPM imaging.

d. Final protocol: atomically-flat imaging substrates

Surface Cleaning

The purpose of surface cleaning on Si/SiO_2 wafers is to etch away any dust that may have been deposited on the surface in order to reduce the substrate to an atomically-flat crystal. The use of piranha solution allows for a thorough cleaning of surface dust, but it is also hazardous to the researcher. Pre-deposition oxygen plasma is a recommended but optional step, depending on whether a wafer can fit inside the given plasma cleaner. In general, oxygen plasma will remove organic residues, and, at relatively low power (18 W), cleaning should occur for at least three minutes, but can continue for any reasonable amount of time.

1. Prepare piranha solution in an acid-safe fume hood by slowly adding one part (10 mL) hydrogen peroxide (H_2O_2) to three parts (30 mL) concentrated sulfuric acid (H_2SO_4) in a glass dish that can fit your silicon wafer. Cover with perforated aluminum foil and label with caution markers and the time of mixing. Use extreme caution as the dish will be quite hot.
2. Allow the piranha solution to age for 90 min. The glassware should be warm but no longer hot.
3. While the piranha is aging, prepare two glass dishes, one with 40 mL deionized water, the other with 40 mL ethanol.
4. Using acid-safe tweezers, place the silicon wafer in the piranha solution and leave it for 5 min to clean.
5. Remove the wafer carefully from piranha and wash briefly in distilled water, then in ethanol.
6. Blow wafer dry in an inert gas such as H_2, N_2 or Ar, letting the gas stream blow toward the tweezers. Store in an airtight container.
7. Safely dispose of the piranha solution in an appropriately-labeled, acid-safe waste container.
8. Clean wafer in oxygen plasma for 5 min (optional).
9. Deposit 1 nm titanium (0.5 Å/s) followed by 200 nm gold (0.5 Å/s) using a cleanroom deposition chamber.

Surface processing

The purpose of surface processing is to reduce the surface roughness of the gold film that is acquired during deposition, during dicing, and by passive exposure to air, as discussed above.

1. Cut the wafer into 1 cm^2 squares using a cleanroom dicing saw. Store in airtight container until individual surfaces are needed.
2. Remove individual samples from diced wafer using tweezers in cleanroom environment.
3. Clean samples in tabletop plasma sterilizer (18 W) for 5 min.

4. Anneal samples in forming gas (Ar/ H_2) for 21 h at 300 ˚C.
5. Again, clean samples in tabletop plasma sterilizer (18 W) for 5 min.
6. Store in an airtight container and use the surface for imaging within 48-72 h.

Depositing the DNA nanowires

In order to control for the effects of buffer salts on single-molecule conductivity, DNA samples may be filtered into deionized water, as in Toomey, 2016.[62] STM break junction experiments on DNA have been performed on surfaces with both in buffer salt fluid as well as deionized water, and it is left to the discretion of the researcher whether this is a necessary step. DNA is not fully stable in salt-free solutions; therefore, after performing a buffer exchange, the sample should be deposited on an imaging surface as soon as possible to avoid degradation. In this protocol, we do not perform buffer exchange, but rather wash the imaging surface in deionized water after deposition of the DNA in order to remove residual salt. Samples producing images will need to have the solution concentration tailored to the image results—if messy, aggregated DNA is observed on the imaging substrate, consider dilution up to 100x in nuclease-free water reduce surface density of DNA. An initial sample may use undiluted DNA or 2-5x dilution as a first guess.

1. Clean the surface in tabletop plasma sterilizer (18 W) for 3-5 min in order to remove any dust and increase adhesiveness of the surface.
2. Within 30 s of removing from plasma cleaner, deposit 5 µL DNA sample in the center of the surface. Cover and set aside on the bench to allow the water (5-10 min).
3. Wash surface by pipetting 10 µL nuclease-free water onto the surface. Wait 30-60 s and slowly wick the water off the edge of the imaging surface with a sterile cloth or wipe.
4. Wash up to twice more, depending on surface salt density, as seen in the light microscope in your SPM. The surface should appear less dirty to the naked eye or under an optical microscope.

3.3.4. STM break junction assay and results

The results obtained from the STM break junction measurements benefitted from neither an optimized binding chemistry nor an atomically-flat gold protocol. The use of C_6 terminal amines and gold-on-glass evaporation was sufficient to generate data, but the curves were subject to noise and possibly reduced conductance as a result. However, comparison of canonical $[dC:dG]_{11}$ and orthogonal $[dC:Ag^{+}:dC]_{11}$ was carried out with significant results (Figure 3.7). The canonical DNA oligomer demonstrated a molecular conductance in the range of 1×10^{-4} G_0, which matches the range identified by other studies.[61] In contrast, the polycytosine duplex showed two distinct levels, one at 1×10^{-3} G_0, and the other at 0.5×10^{-3} G_0.

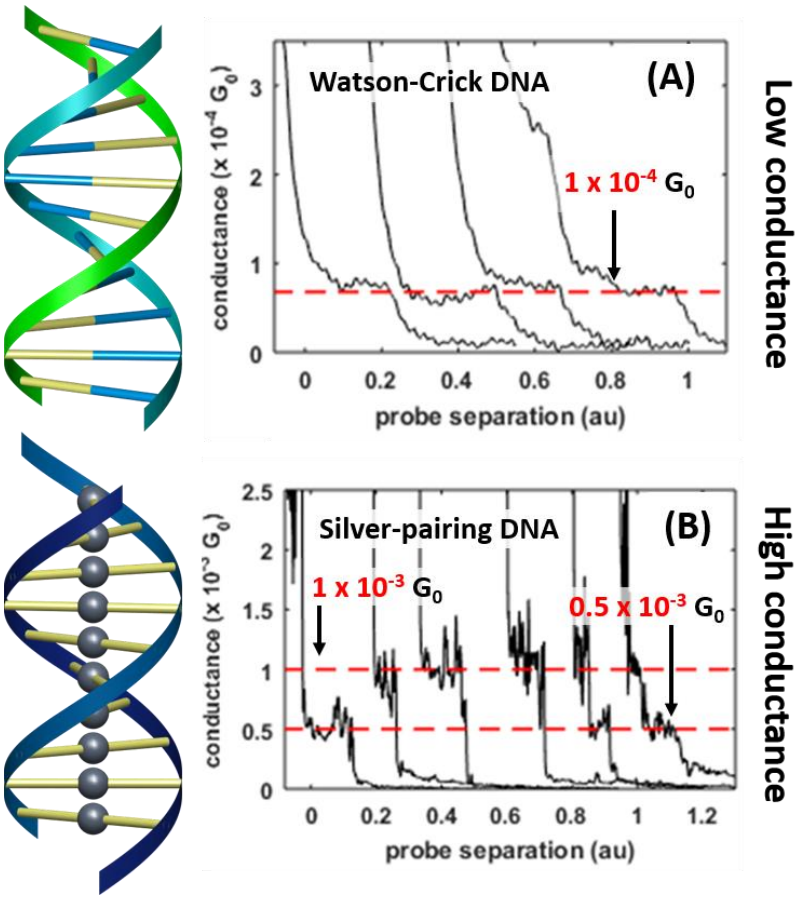


Figure 3.7: STM break junction comparison of amine-linked canonical and Ag^{+}-mediated oligonucleotide duplexes. **A)** Canonical, relatively-highly-conducting DNA sequence $[dC:dG]_{11}$ (design shown inset) was

measured to have a conductance step of approximately 1 x 10^{-4} G_0, or ~7 nS. **B)** Metal-pairing $[dC:Ag^{+}:dC]_{11}$ demonstrated conductance steps of both 1 x 10^{-3} G_0 and 0.5 x 10^{-3} G_0, or ~70 nS and ~40 nS, respectively. Measurements taken by Emily Toomey while at Brown University.

Xu et al perform similar measurements on native DNA and find that conductance decays with sequence length, and the 10bp sequence measured by their group conducts at 7.5 x$10^{-5}G_0$.[61] Despite differences in experimental setup, it is notable that the 11bp control duplex presented here has a conductance of the same order at 6.8 x$10^{-5}G_0$ (Figure 3.7A). Charge transport is likely slightly lower in the Xu work, possibly due to the presence of two A-T pairs at the center. It has been shown that stacked guanine bases have a lower resistance than other geometries, and it is thus reasonable that an 11bp sequence comprised of consecutive G and C bases would perform similarly to a shorter strand with a conductance bottleneck. Though the tests performed by Xu et al are done through an aqueous layer of buffer salts, the similarity in results between our two groups suggests that the control sequence studied here provides a reliable baseline for DNA conductance. This supports the conclusion that the enhanced charge transport in silver-paired strands—behavior characteristic of much shorter strands—is not a system artifact; rather it represents a significant finding specific to core-functionalized DNA.

There remains, however, a core measurement feature that requires further experimental analysis: there exist two distinct peaks for polycytosine wires. This could be the result of several factors. One possibility is that repetitive measurement of the same molecule may introduce stretch and or breakage into the DNA helices over iterative timescales. This may also be related to ion displacement by the tip, in which the scanning probe first makes contact with the terminal end of the nanowire, displaces and binds an ion, and subsequently forms a junction with the C_{11}-Ag^{+}_{10}-C_{11} conjugate, demonstrating an effectively reduced conductance. Furthermore, a study by Swasey et al[142] analyzes the yield of C11 nanowires via electrospray ionization mass spectrometry (ESI-MS) and determine that the resulting oligomers are far from uniform: in the same way that C20, C30, C40 and C50 demonstrate promiscuous annealing

properties owing to their unconstrained annealing frames, C11 annealed with equimolar Ag^+ formed a duplex with only 10 Ag^+ ions in ~20% of all duplex molecules, and ~9 Ag^+ ions in 2% of duplexes, while C_{11}-$[Ag^+]_{11}$-C_{11} accounted for only 40% of all measured DNA oligos, the rest presumably comprising hairpins, daisy chains, monomeric strands and broken material. While the protocol did not benefit from Mg^{2+} or buffer alterations used here, the trend of poly-phenotypic annealing is telling. Indeed, whether introduced in annealing or subsequently by the forceful removal by a scanning probe, it is likely that defects account for the two conductance modes seen in Figure 3.7. In sum, it is probable that the higher conductance value of 1×10^{-3} G_0 corresponds to a defect-free, 11 bp DNA nanowire, while the reduced value of 0.5×10^{-3} G_0 described a defective state of the 11 bp DNA nanowire with one chemically-, electrically-, or kinetically-liberated silver ion. Experimental corroboration of this hypothesis could be carried out through ESI-MS, per Swasey, or through analysis of the C11 oligomer at each stage of the process for defect introduction, much as was done for nanocluster analysis.

To conclude, it is apparent that the introduction of Ag^+-cytosine bonds increases the molecular conductance by introducing resonant states or pathways unavailable to canonical, ring-delocalizing, DNA oligonucleotides. Further study is warranted to elucidate the exact conducting behavior and the effects of sequence heterogeneity, but this represents a proof-of-concept for DNA nanowire technology.

Chapter 4

DNA by Design: *De novo* computational framework for DNA sequence design and nanotechnology

"To succeed, planning alone is insufficient. One must improvise as well."

-Isaac Asimov, Foundation

4.1. Use of Computational Modeling in DNA Sequence Design

4.1.1 DNA hybridization by design

DNA hybridization is an emergent property that is predicated upon the base complementarity of two or more oligonucleotides. The formation of a double helix or other nucleic acid heterostructure requires the association of typically more than three complementary nucleotides via hydrogen bonding. Qualitative analysis shows that in Watson Crick (WC) pairing environments, deoxyadenosine (dA, or A) will form two hydrogen bonds with deoxythymidine (dT, or T), while deoxyguanosine (dG, or G) will form three hydrogen bonds with deoxycytidine (dC, or C). The introduction of the metal ions Ag^{+} and Hg^{2+} will mediate the formation of homonucleotide bonds between opposing C and T bases, respectively.[75,94] Formation of consecutive antiparallel (5'-3': 3'-5') base pairs will allow the assembly of a B-form double helix, which has a periodicity of 10.4 base pairs (bp), or 3.57 nm in neutral pH.[146]

Alternative structures can be formed, by either forcing parallel base pairing (5'-3': 5'-3') to form a left-handed Z-form helix;[28] by dehydration to form an A-form helix;[147] or through the introduction of homonucleotide repeats, such as polycytosine assumption of an i-motif structure at acidic pH,[29] or polyguanosine assumption of a tetraplex helix over long distances in the presence of divalent counterions.[31] More fundamental to the self-assembly of macromolecular nanotechnology, proper design of multi-oligo systems can allow for helical branching, an idea first proposed in 1982 by Nadrian Seeman, and which has been expounded upon many times over in subsequent decades.[38,123,148] Complex structures, called DNA origami, have been subsequently created using large ssDNA viral plasmids and hundreds of short (20 nt) staple oligos to pinch the template into desired conformations, a technology invented by computer scientist Paul Rothemund in 2006.[39,116]

With only four nucleotides to work with, and only two base pairs, the design and construction of DNA oligos with long-range fidelity is a significant challenge. In the same way that single amino acid

mutations can inhibit proper protein folding to generate disease states, imprecise DNA sequence design can promote the formation of thermodynamic mush, rather than meticulously prepared lattices or networks.

Each DNA nanotechnology group has their own method for sequence design, driven by a need to promote hybridization over tens of nanometers with ever-increasing geometric complexities to form vast arrays of two- and three-dimensional structures, including rings, coils, crystallographic arrays, and semiconducting cubes.[40,43,45,71,122,149] The Seeman lab utilizes a legacy sequence design algorithm written in Fortran, while Rothemund has utilized a variety of computational tools to design non-interfering oligos for their origami structures, first using unpublished Matlab code, and then switching to origami tool caDNAno.[150] A field-unifying, modular sequence design tool has yet to emerge that covers the diverse needs of researchers, though there are many examples of software packages that address different stages of the design process. [151-154]

Comprehensive oligonucleotide design parameters were driven largely by the expanding needs of molecular biologists and the researchers at the forefront of the Human Genome Project: it was necessary to design DNA and RNA primers to amplify particular fragments of DNA.[155] The design of effective polymerase chain reaction (PCR) primers required the application two critical pieces of information: 1) the oligo dissociation temperature or melting temperature and subsequent matching within +/- 2 ^{o}C of the melting temperatures for the forward and reverse primers; and 2) the exact size, and subsequent minimization of, unwanted primer heterostructures, or 'primer dimers.'[151] Modern software packages for molecular biology include primer design tools that execute iterative optimization to fix these two parameters.

The requirement of accurate, sequence-specific melting temperatures produced a strong drive to fully elucidate nucleic acid thermodynamics, while the need to design DNA sequences with ever-increasing

genetic databases covering millions of nucleotides spurred the development of machine learning and iterative optimization approaches to multi-objective algorithmic design.

4.1.2 DNA thermodynamics

It was realized early on that the individual nucleotides in long oligo chains were not linearly-independent; rather, there are adjacency effects along the chain unrelated to the contribution of hydrogen bond energy in base pairs. The best example of this resulted from analysis of dG:dT mismatches, which are considered a wobble base pair generally recognized and excised by DNA repair enzymes.[108] The confounding effect of GT mismatches on overall strand formation served as a model for eventual analysis of adjacency, where sets of three consecutive nucleotides were analyzed in a frame, using the 3-nt codon/anti-codon set found in the ribosome as a template.

It was through painstaking thermal denaturation analysis performed via UV-Vis spectrophotometry in the SantaLucia laboratory in the late 1990s that a nearest-neighbor model of DNA assembly was elucidated, using sets of two—not three—adjacent nucleotides summed axially along the oligonucleotide.[101,108] In this approach, each pair of adjacent bases contributes to the energy profile of the oligomer, such that the energy contribution of each nucleotide will be accounted for twice: once for each neighboring DNA base. The terminal bases, left out of this accounting regime, instead contribute strand initiation and termination energies, namely the energy required to form the first base pair and break the last base pair in a DNA duplex. Extensive analysis showed that each set of nearest neighbors is linearly independent, providing a unique enthalpic, entropic and energetic contribution. Studies soon followed to add corrections for ribonucleic acid (RNA), locked nucleic acid (LNA) and mixed, non-specific bases.[156,157] In this way, the free energy of any duplex or primer dimer could be calculated to a high degree of accuracy.

Unfortunately, a reasonable extension of the relatively simple free energy calculation to duplex melting temperature was not achievable. Simplistic formulas for melting temperature prediction were utilized, one of the most popular using the energetics and the concentration:[158]

$$T_{m_{simple}} = \frac{\Delta H}{\Delta S + R_{gas}\log[oligo]} \quad (1)$$

In (1), *ΔH* represents the nearest-neighbor sequence enthalpy, *ΔS* represents the nearest-neighbor entropy, R_{gas} is 1.987 cal/Kmol, and [*oligo*] is the concentration of the oligomer in moles/L. The final temperature in degrees Kelvin is independent of any contributions of salt stabilization and pH, and is therefore only useful as a reference number in very rough calculations. Other forms of this equations began to surface, using the GC concentration, the Na^+ concentration, and other factors, but it was the DNA sequence design companies, in an effort to provide reliable, competitive products, who funded and carried out the experiments necessary to identify a comprehensive formula for buffer-specific melting temperature in arbitrary oligonucleotides.

Working with oligonucleotide synthesis company IDT-DNA to predict primer melting behavior for customers, Owczarzy and colleagues provided the most complete melting analysis of nucleic acids to date.[102] This model corrects for buffer counterions at intermediate concentrations by separating samples into three regimes based on the ratio of monovalent to divalent cation buffer species (constant R, in their literature). Different constants of varying complexity are applied in Na^+-heavy, Mg^{2+}-heavy, and balanced situations, relying on the oligo concentration, salt concentration, pH, and GC content (sample data shown in Figure 4.1).

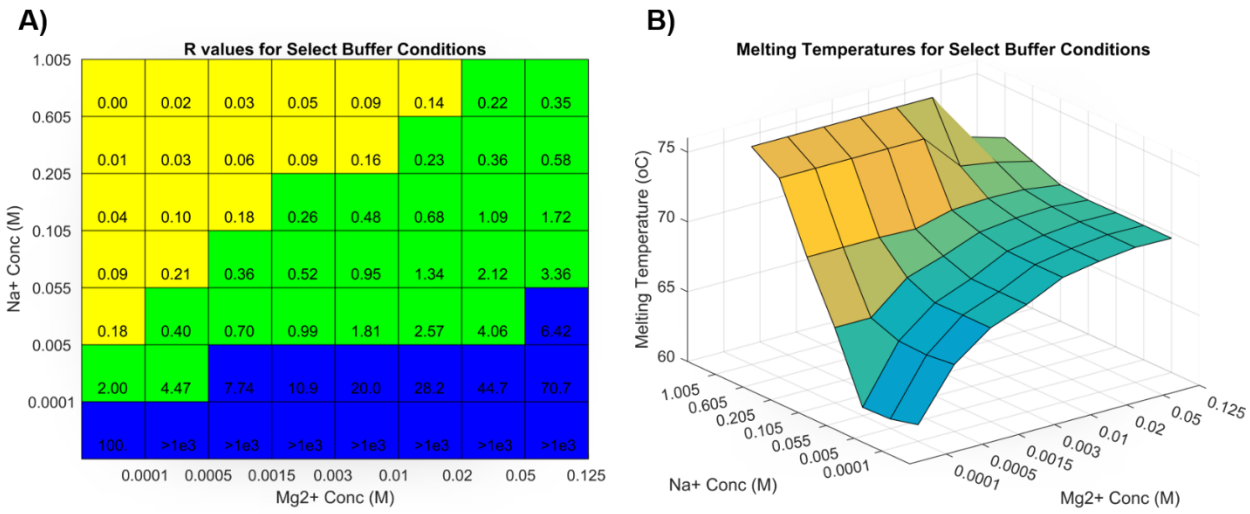


Figure 4.1: Reproduction of Figure 8 in Owczarzy, 2008[102] in Matlab, where oligonucleotide 5'-TGGTCTGGATCTGAGAACTTA-3' is analyzed. **A)** Ion fraction R is calculated using the formula: $\sqrt{[Mg^{2+}]}/[Na^{+}]$, and three parameter regimes are shown in blue, green and yellow. **B)** The counterion-dependent melting temperature is calculated.

The behavior profile was elucidated using nonlinear fits and correction parameters that are not intuitive in the way of Allawi and SantaLucia's nearest-neighbor model of free energy. Though highly specific, these formulas do not work for all oligos in all aqueous environments: homobase repeats, low concentrations of counterions, or the presence of other stabilizing and destabilizing molecules, modifications and substrates do not perform according to this model.

The application of both nearest-neighbor and melting temperature analyses does not extend to orthogonal base pairs: a full nearest-neighbor energy profile of the dC:Ag^{+}:dC pair has not been elucidated, leading researchers to simply correct the predicted melting temperature of mismatched oligonucleotides by adding 1-2 ^{o}C per metal base pair. This practice does not fully take into account metallophilic attraction between silver ions,[78] which would contribute adjacency effects outside the two-body problem of nearest-neighbor calculations, though initial studies suggest that reaction

favorability of a second ion is drastically reduced after a nearby ion is intercalated.[77] These factors suggests that a nearest neighbor model may require knowledge of the opposing oligomer to allow for base pair and intercalant corrections, greatly increasing the overall number of variables and the complexity of the calculation.

4.1.3 Sequence optimization via genetic algorithms

Genetic algorithms (GAs) were first described by Dr. John Holland in the 1960s as an optimization tool based on observations of evolutionary biology.[159] The premise of GAs, and most optimization algorithms, involves assessing the strength, or fitness, of a particular solution to a problem. Like elementary guess-and-check methods, slight variation, or mutation, is applied to a solution. The relative fitness of the new solution is assessed, and over successive iterations, the fitness landscape is mapped to find local or global fitness maxima—the optimization algorithm seeks *better* answers through iteration. The primary difference between guess-and-check methods and GAs is the number of parallel solutions. Like in population biology, a set of solutions is initialized, fitness analyzed, and tournaments carried out to produce new generations of the population. In order to promote solution diversity, successful tournament winners exchange information to allow independent assortment of solution characteristics.

The linear DNA sequences in primer and nanostructure design make the analogy to living systems highly useful. Use of mutation and crossover *in silico* directly corresponds to natural DNA design and optimization. As such, GA tools are particularly well-suited to evolutionary optimization. The challenge to developing *in silico* models of DNA optimization lies in defining a fitness function that is both simple to calculate and well-suited to the design objectives. Unlike in living systems, the computational researcher does not have the luxury of hour- or yearlong generations over which to assess DNA sequence fitness. The fitness of a particular solution, or set of DNA sequences, is necessarily subject to

more than one criterion, and can include: melting temperature, heterodimer size, guanine repeats, etc. There are several strategies for taking a multi-objective fitness function and producing a simple, comparison-based algorithm to assess relative fitness. The most common heuristic involves condensing the many criteria into a single numerical score based on the relative importance of the different design criteria.[160] In doing so, the various objectives are assigned variable weights, and sorted according to numerical size. A GA iterator may be designed to either maximize or minimize this fitness score, depending on the nature of the criterion.

The primary drawback of GA tools lies in this definition of this multi-objective fitness function. If there were only one criterion and a continuous fitness landscape, as in the problem: 'minimize the distance from the square root of 200,' there would be little trouble in attaining a highly precise *near-best* solution. The use of a multi-objective fitness score with weights makes the fitness landscape discontinuous: once a change occurs to increase the highest-weight criterion, solutions are boxed into this corner of the fitness landscape, preventing any true global optimization of lower-weight criteria. The fitness hill-climbing function becomes a stair-climbing function, where it is improbably that the iterator will climb back down to get around a block in the landscape. In this way, multi-objective GAs frequently find only local, rather than global fitness maxima.

To combat the intractability of local fitness maxima, evolutionary algorithms turn to population dynamics, introducing diversity artificially. This takes the form of randomization or hypermutation, niche penalties based on population similarity, and, most importantly, gene flow. Randomization involves the formation of a small number of new solutions in a population with a much higher mutation rate than average to introduce diversity. While potentially quite useful, large tournament sizes will forcibly exclude these solutions from passing on information. Niche penalties involve increasing the global mutation rate when the overall similarity reaches a critical value. In the case of oligonucleotides, this may be assessed via the mean squared distance from the average sequence value at a particular

position. Depending on the size of the steps in the fitness score, this approach may introduce diversity, or simple nudge the average solution only temporarily away from the local fitness maximum, only to return in several optimization iterations.

The two most successful approaches to diversity include elitism and gene flow. Elitism is a method that allows for greater mutation rates by saving a small number of the fittest solutions in each generation without any modification. In this way, the best solution is not lost while the general population can be subjected to higher rates of change to promote diversity. There are several methods for applying elitism, but the simplest in the context of DNA design is simply copying without mutation.[160] As in population biology, diversity can be introduced into an isolated population through the introduction of new, medium-fitness, individuals. In this way, gene flow is used to combat genetic drift, or niche creation. There are several established methods of information exchange, but all rely primarily on optimizing several independent populations and exchanging information either at distinct intervals, or after the simulation in a subsequent 'F2 cross.'[160] Crossing the best solutions in several populations in a follow-up simulation pits solutions clustered around distinct fitness maxima against one another, allowing for a more granular blending of characteristics at different levels in the fitness function. While there is no approach that completely fixes the diversity problem, a pointed combination of these strategies may lead to the identification of *near-best* solutions, given sufficient computational resources.

Several tools for primer design by GA already exist,[150,151,161] and they rely on a multi-objective fitness function that weighs melting temperature, relative length, and dimer size between forward and reverse primer pairs. This relatively simple fitness approximation relies the primary assumption that any ~20 nt primer will have only one target site in a prokaryotic or eukaryotic genome, and the calculation restricts heterodimer analysis to the primer oligonucleotides only, *excluding effects of off-target dimerization on fitness calculation.* This significantly speeds computation, allowing for a swift identification of primers in gene targets in the context of thousands to millions of nucleotides. Conventional DNA synthesis

techniques at companies like IDT DNA are ~99% efficient at adding a correct nucleotide to a growing oligo, meaning that for each nucleotide letter in an oligonucleotide sequence, there is a compounding 1% chance of an improper base. In 2019, synthesis of 25 nmol of a 20 nt primer cost ~$7. Subsequent purification of the sequences via gel electrophoresis (PAGE) or liquid chromatography (HPLC) at the same company cost an additional $60. Due to the relaxed specificity needs of PCR, primers are generally not purified—99% sequence specificity over 20 nt leaves ~80% primer accuracy to perform PCR ($0.99^{20} = 0.82$). Over successive amplification, defective primers remain inert. Assuming a reasonable concentration of correctly-synthesized primer, the reaction will run without oligo purification. As a result, the low cost of 20mer oligonucleotides allows for occasional off-target dimerization, and primer failure, meaning that a computationally-efficient fitness approximation is good enough for the problem statement in PCR.

By contrast, in the case of DNA nanostructure design, any off-target dimerization will cause the failure of structural assembly. The critical regions for crossover between two adjacent helices is 4 bp, and highly conserved between different laboratories (almost always CAGG:CCTG).[20,148,162] Any deviation from this sequence and the crossover will not occur, necessitating purification of oligos after synthesis: a lattice built with 30mers will be unable to polymerize without removing the 27% failed product. The time and cost of purification thus makes the penalty of using a relaxed fitness calculation far more prohibitive. To mitigate this much tighter problem definition, a more robust fitness calculation and accompanying genetic algorithm toolbox are required.

There exist a variety of tools for DNA nanotechnology design, often with an emphasis on aiding definition of sequence geometry (e.g. caDNAno,[150] industry standard). The sequence generation tools are quite relaxed, and there does not yet exist a convincing tool for specifying conserved regions of DNA for the purposes of crossover sequences, hairpin formation, or importantly, the incorporation of orthogonal base pairing.

In this chapter, we present a new computational model with the goal of bridging the gap between structure topology and the sequences that exist within that framework. This model seeks to ultimately provide a generalizable, geometry-informed platform for sequence generation with diverse design criteria and base pairing environments for nucleic acid nanotechnology applications.

4.2. Computational Analysis of Nanostructure Composition

A *de novo* computational platform for DNA sequence analysis is derived and described from first principles. This platform is built from the perspective of modular nanostructure design, though it can be used in both branched and linear applications. Formulas are written to be independent of Watson-Crick parity and make note of any base pairing assumptions that are made in their derivation.

4.2.1. Nanostructures, nodes and sequences

Let M be a DNA nanostructure composed of a finite number of oligonucleotide sequences (S) where $S \in \mathbb{N}^*$ for $\mathbb{N}^* = \{1, 2, \ldots\}$, which bind together in topological units called nodes (N) for $N \in \mathbb{N}^*$. M is a structure with L total base pairs, where $L \in \mathbb{N}^*$, and the nodes in this network have lengths L_i in base pairs (bp) for $i \in [1, N]$. A network with one node is considered an unbranched duplex. By convention, a node is considered to be double stranded (ds), where each oligonucleotide (n_i) is bound to its logical complement ($\bar{n}_i$) in a base pair (b_i), where $b_i = [n_i{:}\bar{n}_i]$, though some nodes may be single-stranded (ss) by design.

$$\begin{array}{l} 5'-n_1\ n_2\ n_3 \ldots n_i-3' \\ \quad\ |\ \ |\ \ |\ \ \ |\ \\ 3'-\bar{n}_1\ \bar{n}_2\ \bar{n}_3 \ldots \bar{n}_i-5' \end{array} \qquad (1)$$

DNA network M can be drawn conventionally using its continuous oligonucleotide sequences (Figure 4.2A), but can be more robustly analyzed using a nodal model (Figure 4.2B), where a node contains a

simplex or duplex with internally consistent pairing rules, chemistry, geometry, and has connections to other nodes from its template and complement strands, and its 5' and 3' ends:

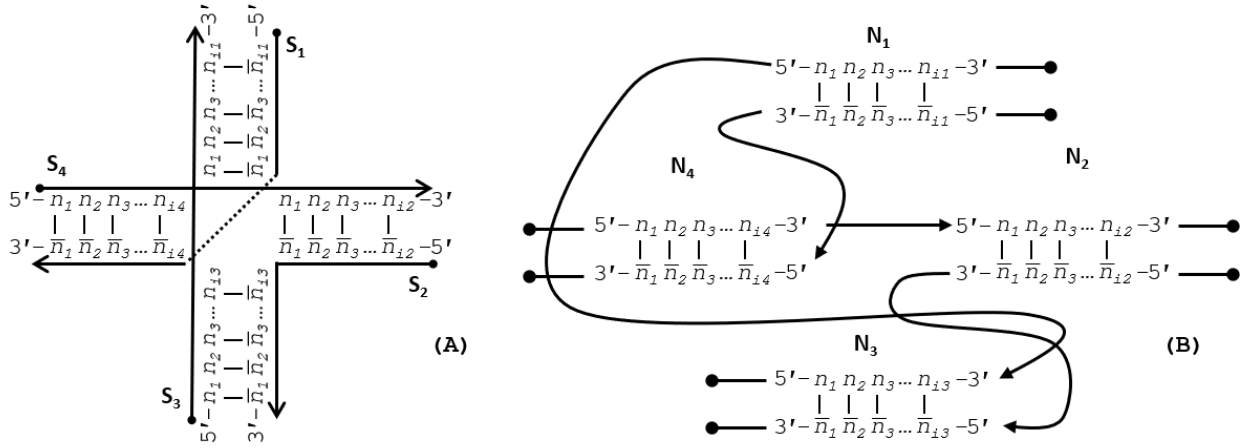


Figure 4.2: Representations of a Holliday junction, **A)** in sequence format with four sequences S_i for for $i \in [1, 4]$, and **B)** in node format with four nodes N_j for for $j \in [1, 4]$. The primary difference between these modes of representation is that the geometry and complementarity is carried on individual nucleotides in sequence format; whereas the nature of node-based representations omits the need for complementarity indicies and just tracks the connections, or geometry, at the edges of the nodes.

Complementarity in nucleobases typically follows the binding of one purine (R) to one pyrimidine (Y) in the following set of Watson-Crick (WC) rules: { [dA:dT], [dG:dC], [rA:rU], [rG:rC] }, where dn_i is a deoxyribonucleotide, and rn_i is a ribonucleotide. Orthogonal base pairs can break the symmetry of WC pairing, as in the case of metal base pairs { [dC:Ag^+:dC], [dT:Hg^{2+}:dT] }. The difference in nucleobase correspondence between canonical DNA and a silver metal pairing system (CC) can be seen in the following matrix:

$$\begin{matrix} \text{WC} \\ \begin{matrix} dA \\ dT \\ dG \\ dC \end{matrix} \begin{bmatrix} dT \\ dA \\ dC \\ dG \end{bmatrix} \end{matrix} \qquad \begin{matrix} \text{CC} \\ \begin{matrix} dA \\ dT \\ dG \\ dC \end{matrix} \begin{bmatrix} dT \\ dA \\ dC \\ dG, dC \end{bmatrix} \end{matrix} \qquad (2)$$

4.2.2. Heterostructures

In order to self-assemble *N* nodes from *S* oligos, analysis of nucleobase sequence must be performed to bias the energy landscape toward the desired geometry. This requires analyzing *M* for its heterostructures (*d*), or unwanted binding configurations, with length L_d. Analysis of homo- and heterodimers is carried out by aligning component strands of *S* or *N* and comparing each base in varying frames. An alignment frame (*k*) can be defined as the juxtaposition of two oligonucleotides, one in the 5′-3′ orientation and the other in 3′-5′ orientation, with fixed base correspondence, where *k* will be a frame of k_{size} potential base pairs. Two complementary oligonucleotides of length L_N will form a perfect dimer of size $L_d = L_N$ with perfect alignment by design (see (1) above).

A homodimer is a kinetic trap for an oligonucleotide in which it is able to bind to *itself* in a misaligned frame, as with single-stranded oligo S_a of length L_a = 11 nt, k_{size} = 9 bp, L_d = 3 bp:

```
5′– CATTTGCGAAA –3′
     |||    |||        (3)
  3′– AAAGCGTTTAC –5′
```

A heterodimer can be defined as the inappropriate binding of two disparate oligonucleotides, as with single-stranded oligos S_b and S_c of lengths L_b = 11 nt, L_c = 14 nt, k_{size} = 11 bp, and L_d = 7 bp:

```
   5′– ATATATCGCCG –3′
       |||||||            (4)
3′– CTTATATAGGTAGA –5′
```

4.2.3. Oligo slip

Testing for inappropriate heterostructures requires comparison of nucleotides with different alignment frames, in which 'slip' can be defined as the size of the 5′ or 3′ overhang of the first sequence. Slip can be visualized in the following manner:

```
   5′–|ATATATCGCCG–3′
      |||||||                (5)
3′–CT|TATATAGGTAGA–5′
```

```
5′–TCA|TTTGCGAAA–3′
      |||   |||              (6)
   3′–|AAAGCGTTTAC–5′
```

Scheme (5) shows the top strand slipping to the right two nucleotides relative to alignment, while (6) the top strand slipping to the left three nucleotides. These two slip conditions can be called forward and backward slip (s_f and s_b), respectively. The case where two strands are aligned in the correct frame can be called a special case of forward slip, where $s_f = 0$. To avoid repeat comparison, the case where $s_b = 0$ is disallowed. A general formula for slip can be derived from the following example, where two sequences (S_a, S_b) with unequal lengths (L_a = 5 nt, L_b = 3 nt) are aligned in all possible frames:

$s_f = 0$

$$\begin{array}{l} 5'\text{–}\,n_1\,n_2\,n_3\,n_4\,n_5\text{–}3' \\ \quad\;|\;\;|\;\;| \\ 3'\text{–}\,n_6\,n_7\,n_8\text{–}5' \end{array}$$

$s_f = 1$

$$\begin{array}{l} 5'\text{–}\,|n_1\,n_2\,n_3\,n_4\,n_5\text{–}3' \\ \quad\;|\;\;| \\ 3'\text{–}\,\mathbf{n_6}\,|n_7\,n_8\text{–}5' \end{array}$$

$s_f = 2$

$$\begin{array}{l} 5'\text{–}\,|n_1\,n_2\,n_3\,n_4\,n_5\text{–}3' \\ \quad\;| \\ 3'\text{–}\,\mathbf{n_6\,n_7}\,|n_8\text{–}5' \end{array}$$

(7)

$s_b = 1$

$$\begin{array}{l} 5'\text{–}\,\mathbf{n_1}\,|n_2\,n_3\,n_4\,n_5\text{–}3' \\ \quad\;|\;\;|\;\;| \\ 3'\text{–}\,|n_6\,n_7\,n_8\text{–}5' \end{array}$$

$s_b = 2$

$$\begin{array}{l} 5'\text{–}\,\mathbf{n_1\,n_2}\,|n_3\,n_4\,n_5\text{–}3' \\ \quad\;|\;\;|\;\;| \\ 3'\text{–}\,|n_6\,n_7\,n_8\text{–}5' \end{array}$$

$s_b = 3$

$$\begin{array}{l} 5'\text{–}\,\mathbf{n_1\,n_2\,n_3}\,|n_4\,n_5\text{–}3' \\ \quad\;|\;\;| \\ 3'\text{–}\,|n_6\,n_7\,n_8\text{–}5' \end{array}$$

$s_b = 4$

$$\begin{array}{l} 5'\text{–}\,\mathbf{n_1\,n_2\,n_3\,n_4}\,|n_5\text{–}3' \\ \quad\;| \\ 3'\text{–}\,|n_6\,n_7\,n_8\text{–}5' \end{array}$$

Forward slip begins with frame alignment and proceeds until the first nucleotide in S_a (5′-3′) corresponds to the last nucleotide in S_b. Slip values can be tallied: s_f = [0,1,2], while s_b = [1,2,3,4]. Frame sizes can be written as an array: k_{size} = [3,2,1] for forward slip; and k_{size} = [3,3,2,1] for backslip. The relationship between these arrays can be formalized as follows. The maximum forward slip value is one less than the length of the second sequence:

$$s_f \in [0,\ L_b - 1] \qquad (8)$$

Backward slip follows a similar convention, avoiding perfect frame alignment:

$$s_b \in [1,\ L_a - 1] \tag{9}$$

4.2.4. Frame alignment

There is one alignment frame per slip condition. With (8) number of alignment frames (k_{range}) in forward slip can be expressed as the total number of nucleobases on the second strand:

$$k_{range} = L_b \tag{10}$$

The number of backward alignment frames corresponds with the number of nucleobases in the first strand, disallowing a slip of zero:

$$k_{range} = L_a - 1 \tag{11}$$

The alignment frames for comparison will have varying sizes, depending on whether slip is forward or backward, and which sequence is longer, and can be defined in the following manner:

$$k_{size} = \left[k_1 \dots k_j \dots k_n\right]; \quad n = k_{range} \tag{12}$$

For **forward slip** conditions where $L_b > L_a$ for $k_j \in k_{size}$, k_j may be written as an array with a formula in two halves. The first part of k_{size} is simply the length of the smaller sequence, written out for the length differential between the two sequences:

$$\left(k_j\right)_{j \in [1,\, n - L_a + 1]} = L_a \tag{13}$$

The second part is the decay of the previous length to the minimum frame size, 1:

$$\left(k_j\right)_{j \in [n - L_a + 2,\, n]} = L_a - \left(j - (n - L_a + 1)\right)$$

$$\left(k_j\right)_{j \in [n - L_a + 2,\, n]} = n - j + 1 \tag{14}$$

When $L_b \leq L_a$, there is one expression, as the frame decays from the length of the shorter sequence (L_b) to one:

$$\left(k_j\right)_{j\in[1,\,n]} = L_b - (j-1)$$

$$\left(k_j\right)_{j\in[1,\,n]} = L_b - j + 1 \qquad (15)$$

For **backward slip** alignment with $L_a > L_b$, k_{size} may again be written as an array in two parts. The first section is again the length of the shorter sequence written for the length differential between the two strands:

$$\left(k_j\right)_{j\in[1,\,n-L_b+1]} = L_b \qquad (16)$$

The second part of the array contains frames of decaying size from L_b to 1:

$$\left(k_j\right)_{j\in[n-L_b+2,\,n]} = L_b - \left(j - (n - L_b + 1)\right)$$

$$\left(k_j\right)_{j\in[n-L_b+2,\,n]} = n - j + 1 \qquad (17)$$

Whereas when $L_a \leq L_b$, we have a single expression for k_j in k_{size}:

$$\left(k_j\right)_{j\in[1,\,n]} = L_a - j \qquad (18)$$

The relationship between length, slip and alignment is summarized in Tables 4.1-4.2 below.

Table 4.1: Forward slip alignment formulations

	Variable name	Definition	Number in text
Slip values	s_f	$s_f \in [0,\ L_b - 1]$	(8)
Number of frames	k_{range}, n	$k_{range} = L_b$	(10)

Frame sizes	k_{size}, k_j	$k_j \begin{cases} (k_j)_{j\in[1,\ n]} = L_b - j + 1 & L_a \geq L_b \\ (k_j)_{j\in[1,\ n-L_a+1]} = L_a, \quad (k_j)_{j\in[n-L_a+2,\ n]} = n - j + 1 & L_a < L_b \end{cases}$	(15) (13,14)

Table 4.2: Backward slip alignment formulations

	Variable name	Definition	Number in text
Slip values	s_b	$s_b \in [1,\ L_a - 1]$	(9)
Number of frames	k_{range}, n	$k_{range} = L_a - 1$	(11)
Frame sizes	k_{size}, k_j	$k_j \begin{cases} (k_j)_{j\in[1,\ n-L_b+1]} = L_b, \quad (k_j)_{j\in[n-L_b+2,\ n]} = n - j + 1 & L_a > L_b \\ (k_j)_{j\in[1,\ n]} = L_a - j & L_a \leq L_b \end{cases}$	(16, 17) (18)

4.2.5. Nucleobase comparison indices

With these definitions in hand, exact formulas can be written for nucleobase comparison indices at prescribed slip conditions. The nucleobases to compare can be written in a visually simple manner where the first base index of S_A is at the 5' end, and base indices for S_B start at the 3' tail—a process akin to flipping the 'complement' sequence for the reader (see Figure 4.3). Implementing this formula *in silico* is, however, impractical: the cost of transposing a sequence array of size 11 (one helical turn) in the Matlab environment is approximately 1 µs. We will investigate the number of operations below, but for any iterative algorithm, this operation should be avoided. As such, we write formulas for comparison of two oligonucleotides with both sequences written in the 5' to 3' direction.

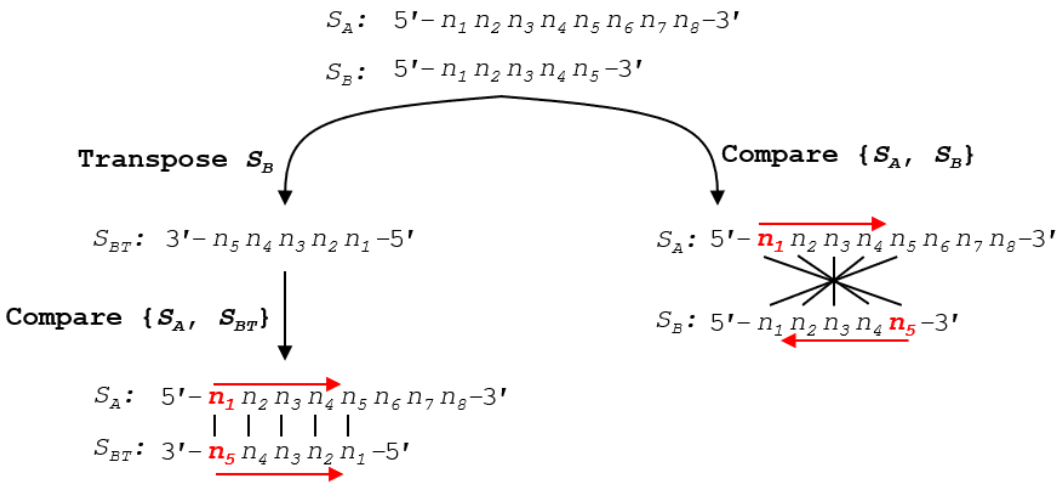


Figure 4.3: Using a simplified formula to compare two sequences requires the transpose of the second sequence into the 3'-5' orientation. This transpose costs ~N operations per slip frame, and at worst N^2 total operation for all slip conditions. A more efficient algorithm can be written without conversion to S_{BT}, where the formula for indexing in S_B begins from the end of the sequence, rather than the start. Performance analysis of this reverse indexing algorithm is carried out in Section *4.2.9.d*.

We define the variable m as the position within the alignment frame, or the number of nucleobases from the leftmost (5') base in sequence S_A. According to Figure 4.3, the comparison position in S_A, a_m, will proceed from the 5' end, while the comparison index in S_B, b_m, will proceed from its 3' end. The nucleobase indices for comparison in S_A and S_B (a_m, b_m) with slip s_j and alignment frame size k_j for $j \in [1, k_{range}]$, and an alignment frame index $m \in [1, k_j]$ can be written for **forward slip**:

$$a_m = m \tag{19}$$

$$b_m = L_B - s_j - m + 1 \tag{20}$$

A similar index pair can be written for **backward slip**:

$$a_m = m + s_j \tag{21}$$

$$b_m = L_B - m + 1 \quad (22)$$

In this way, for all m in a given alignment frame, comparison can be carried out between $S_A(a_m)$ and $S_B(b_m)$.

4.2.6. Dimer indices

This also means that a potential base pair in the given alignment frame can be written as $[S_A(a_m): S_B(b_m)]$. For a dimer *d* of length L_d that terminates at frame index *m* in alignment window k_j with slip s_j, the indices of the base sequence of *d* in sequences S_A and S_B (nucleobase sequence indices d_a and d_b, respectively) can be written with the following formulas:

$$d_a \in [a_m + 1 - L_d : a_m] \quad (23)$$

$$d_b \in [b_m - 1 + L_d : b_m] \quad (24)$$

Plugging the definitions of a_m (19) and b_m (20) for **forward slip** conditions into (23) and (24), we achieve the following generalized expression for dimer index:

$$d_a \in [m + 1 - L_d : m] \quad (25)$$

$$d_b \in [L_B - s_j - m + L_d : L_B - s_j - m + 1] \quad (26)$$

The same process can be carried out for **backslip** conditions, plugging (21) and (22) into (23) and (24):

$$d_a \in [m + s_j + 1 - L_d : m + s_j] \quad (27)$$

$$d_b \in [L_B - m + L_d : L_B - m + 1] \quad (28)$$

These results are summarized in Tables 4.3-4.4, and can be used to directly track and index dimers in a programming environment.

Table 4.3: Forward slip comparison indices			
	Variable name	Definition	Number in text
Strand 1 index	a_m	$a_m = m$	(19)
Strand 2 index	b_m	$b_m = L_B - s_j - m + 1$	(20)
Dimer position strand 1	d_a	$d_a \in [m + 1 - L_d : m]$	(25)
Dimer position strand 2	d_b	$d_b \in [L_B - s_j - m + L_d : L_B - s_j - m + 1]$	(26)

Table 4.4: Backward slip comparison indices			
	Variable name	Definition	Number in text
Strand 1 index	a_m	$a_m = m + s_j$	(21)
Strand 2 index	b_m	$b_m = L_B - m + 1$	(22)
Dimer position strand 1	d_a	$d_a \in [m + s_j + 1 - L_d : m + s_j]$	(27)
Dimer position strand 2	d_b	$d_b \in [L_B - m + L_d : L_B - m + 1]$	(28)

4.2.7. Operation count for dimer analysis

In order to design an efficient algorithm for *in silico* dimer analysis, it is necessary to elucidate the number of operations required to analyze *M* for heterostructures. A generalized expression can be derived by first looking at a one-node structure, or duplex.

a. One-node structure

Comparison of any set of sequences for dimers requires checking each strand against itself and against each other strand for all allowable slip and alignment conditions. In a single-node network, or DNA duplex, we can see the operations in Figure 4.4:

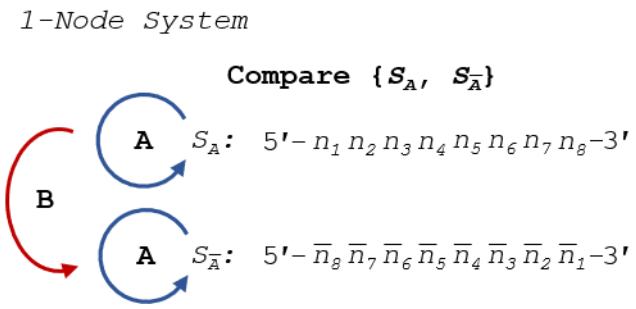


Figure 4.4: Comparison of S_A and its complement for heterostructures requires three sets of operations: two self-comparisons (A), and one complement comparison (B).

We can introduce two different types of operations, *A* and *B*: operation A involves the comparison of a single-stranded oligo against itself; whereas operation B involves the comparison of an oligo and its direct complement, ignoring the case where they are properly aligned, which is precluded by their complementarity. Formulation of these comparisons requires, by definition, that the lengths of the two oligos be identical. The number of comparisons (*N*) directly scales with the length of the sequences, and involve the summation of the allowable alignment frames, or k_{size}. To do so, we introduce the summorial operator *$*, which is the additive cousin of the factorial operator:

$$x! = \prod_{i=1}^{x} i$$

$$x\$ = \sum_{i=1}^{x} i \tag{29}$$

The summorial operation can be decomposed into the following formula:

$$x\$ = \frac{n(n+1)}{2} \tag{30}$$

a.i. Homodimer operation counts (Operation A):

The number of comparisons in A can be calculated by summing all components (k_j) of forward and backward frames, k_{size}, where the number of slip cases, n_{sf} and n_{sb}, are equivalent to forward and backward k_{range}, respectively:

$$N_A = \sum_{j}^{n_{sf}} k_j + \sum_{j}^{n_{sb}} k_j \tag{31}$$

Substituting (10) for n and (15) for k_j in the left sigma:

$$\sum_{j}^{n_{sf}} k_j = \sum_{j=1}^{L_a} (L_a - j + 1)$$

$$= (L_a)\$$

$$= \frac{(L_a)(L_a + 1)}{2}$$

Again substituting (11) and (18) in the right sigma:

$$\sum_{j}^{n_{sb}} k_j = \sum_{j=1}^{L_a - 1} (L_a - j)$$

$$= (L_a - 1)\$$

$$= (L_a)\$ - L_a$$

$$= \frac{(L_a)(L_a + 1)}{2} - L_a$$

Collecting terms, we can simplify to obtain the operation count for A:

$$N_A = \frac{(L_a)(L_a+1)}{2} + \frac{(L_a)(L_a+1)}{2} - L_a$$

$$N_A = {L_a}^2 \tag{32}$$

We can see a practical demonstration of this operation in Figure 4.5. In this example, sequence S_A has a length of 8 nt. Summation of the comparisons across each annealing frame is shown.

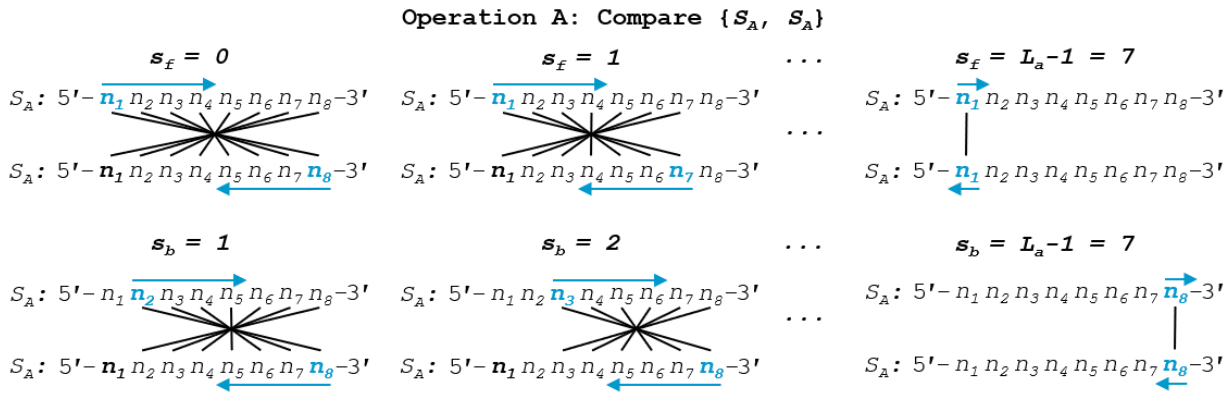


Figure 4.5: Self-dimerization (Operation A) comparison of generalized single-stranded, 8 nt oligonucleotide S_A is shown in both forward and backward slip conditions. Note that, by convention, backslip starts with a value of 1, while forward slip begins at perfect alignment, or a slip value of 0. To speed computation, formulas for the comparison of S_A without 5' and 3' realignment are shown. This involves the reverse indexing formulas found in Tables 4.3-4.4.

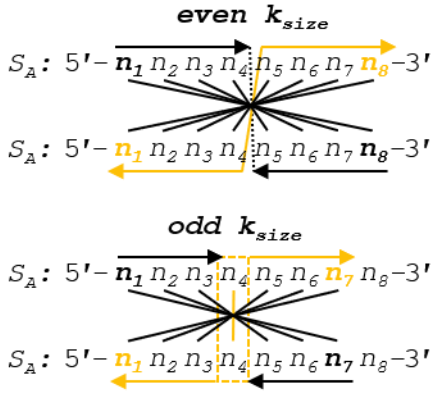


Figure 4.6: Two different alignment frames are shown for oligonucleotide S_A, one of even ksize (8 nt) and the other with odd ksize (7 nt). A line of symmetry can be drawn down the middle for the even case, and around the pivot nucleotide for the odd case. Because the nucleotides are identical in the opposing strands, the comparisons across the symmetry lines will also be identical: (n_1 == n_8) = (n_8 == n_1).

Operation A is efficient at order $O(n^2)$ for *n* nucleotides. In the specific case of Operation A, we can draw lines of symmetry at the halfway point of the alignment frame in the case of even k_{size}, or around the center nucleotide in the case of odd k_{size}, and note that the comparisons will be identical across this line of symmetry (Figure 4.6). With this in mind, we can reduce the number of comparisons by tracking the even- or oddness of the frame size and performing operations up to, but not past, the line of symmetry. Any dimer that exists once the center of the alignment frame is reached will be doubled in length after omitting odd-length pivot nucleotides. The tracking of frame size modulo 2 requires several floating point operations, and is only practical in high-performance situations for large *L*. A more efficient Operation A can be seen in Figure 4.7. This reduced dimerization search operates at roughly $O(\frac{1}{2}n^2)$.

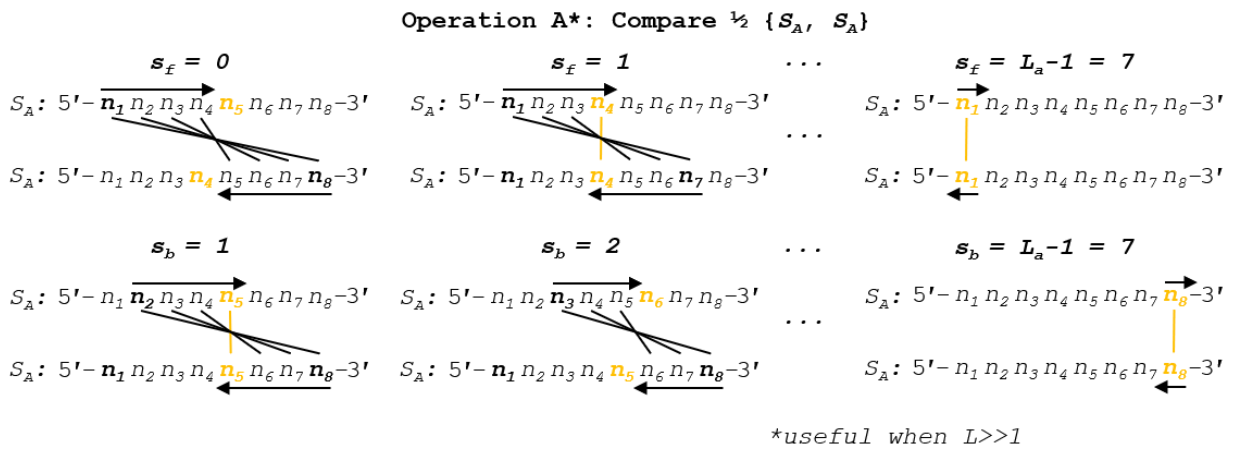


Figure 4.7: A high-performance algorithm for Operation A is shown, wherein the frame size and evenness are tracked. Comparisons are not continued past the line of symmetry, and any extant dimers that occur

at this line are doubled (subtracting 1 nt for odd frame sizes). Note reverse indexing formulas found in Tables 4.3-4.4 to account for two 5'-3' aligned sequences.

a.ii. Heterodimer counting with logical complements (Operation B):

We can sum k_{size} for operation B in a similar way, making sure to exclude $s_f = 0$ to avoid counting logical complements as dimers, and we arrive at a general expression. Setting up the problem, we generate an expression identical to (31):

$$N_B = \sum_{j}^{n_{sf}} k_j + \sum_{j}^{n_{sb}} k_j \tag{33}$$

By disallowing $s_f = 0$, we generate identical backslip conditions for both sigmas, which condenses with the application of definitions (11) and (18):

$$\sum_{j}^{n_{sf}} k_j = \sum_{j}^{n_{sb}} k_j = \sum_{j=1}^{L_a-1} L_a - j$$

$$= (L_a - 1)\$$

$$= \frac{(L_a - 1)(L_a)}{2}$$

$$N_B = \frac{(L_a - 1)(L_a)}{2} + \frac{(L_a - 1)(L_a)}{2}$$

$$N_B = {L_a}^2 - L_a \tag{34}$$

We can see a practical demonstration of this operation in Figure 4.8. In this example, sequence S_A has a length of 8 nt. Summation of the comparisons across each annealing frame is shown.

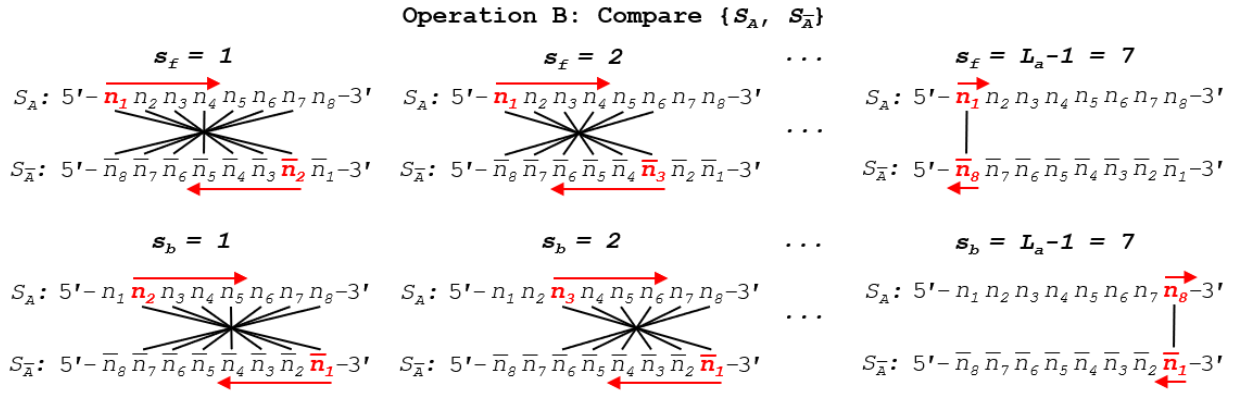


Figure 4.8: Complement dimerization (Operation B) comparison of generalized single-stranded, 8 nt oligonucleotide S_A and its logical complement $S_{\bar{A}}$ is shown in both forward and backward slip conditions. Note that both forward and backward slip start with values of 1 in order to avoid counting complementarity-by-design as a heterodimer. In order to speed computation, formulas for comparison without 5' and 3' realignment are shown, which involves the reverse indexing formulas found in Tables 4.3-4.4.

Operation B is efficient at order $O(n^2 - n)$ for *n* nucleotides. The number of overall comparisons shown in Figure 4.8 will include redundant operations when the base pairing rules are symmetric (see Figure 4.9). This is only true when each nucleobase has one and only one logical complement, which indicates that a full oligonucleotide or dimer substring will have exactly one complement as well. In the case of ion pairing, or any scheme where a nucleobase has multiple complements, the full operation, shown

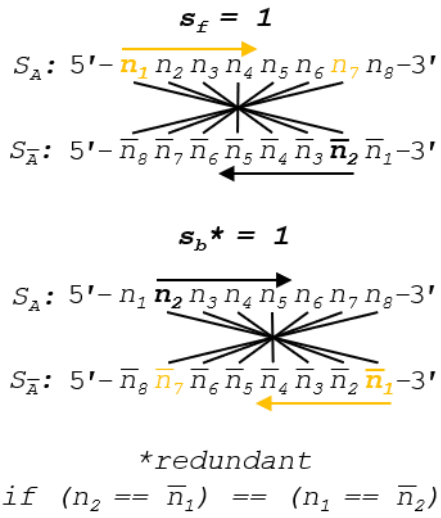


Figure 4.9: Slip value of 1 is shown for forward and backward slip. We can see that when there is Watson Crick complementarity, these two comparison conditions are identical. A higher performance algorithm in canonical pairing environments can be enacted by omitting backslip, as $(n_2 == n_1) = (n_1 == n_2)$.

in Figure 4.8, is necessary. If true DNA Watson Crick rules are in effect, Operation B may be halved in size by omitting backslip, as shown in Figure 4.10. The updated WC algorithm for Operation B is efficient at order $O(\frac{1}{2}(n^2 - n))$.

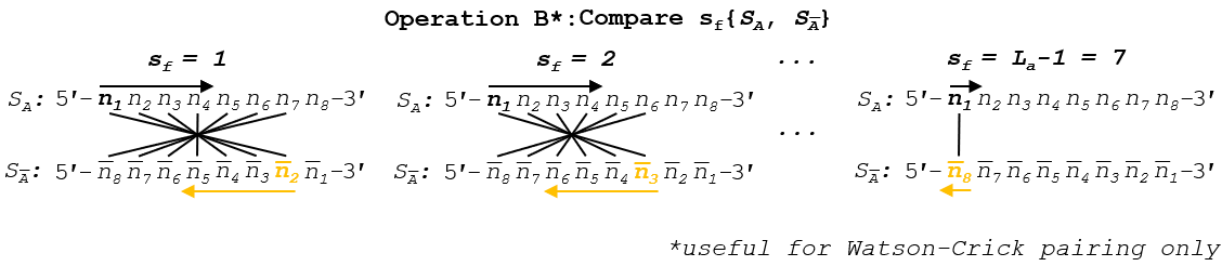


Figure 4.10: When Watson-Crick or other unitary complement pairing rule system is in effect, backslip may be omitted for symmetry reasons. Tracking the pairing regime does not require significant computation time, and thus tracking the usefulness of this updated algorithm can be considered essential for dimer identification as it will double comparison speed. Reverse indexing formulas for nucleotide position tracking are found in Table 4.3.

a.iii. Application to a one-node structure (DNA duplex)

For any DNA duplex consisting of two complementary oligonucleotides of equal length (*L*), we can express the number of comparisons (*N*) by adding together the operations above. Operation A occurs twice, as each strand must be compared to itself, while Operation B occurs once for the union of the two strands in non-complementary configurations.

$$N_{duplex} = 2N_A + N_B$$

$$N_{duplex} = 3L^2 - L \qquad (35)$$

b. Multi-node structure

A DNA nanostructure, primer-target complex, or other mixture of oligonucleotide species can be represented as a multi-node network such as the two-node structure shown in Figure 4.11. As before, the comparisons can be broken down into distinct procedures. Operations A and B return in the same form, and we introduce operation C, which we define as the alignment and comparison of sequences from disparate nodes. In this operation, we dispense with the requirement that $L_a = L_b$.

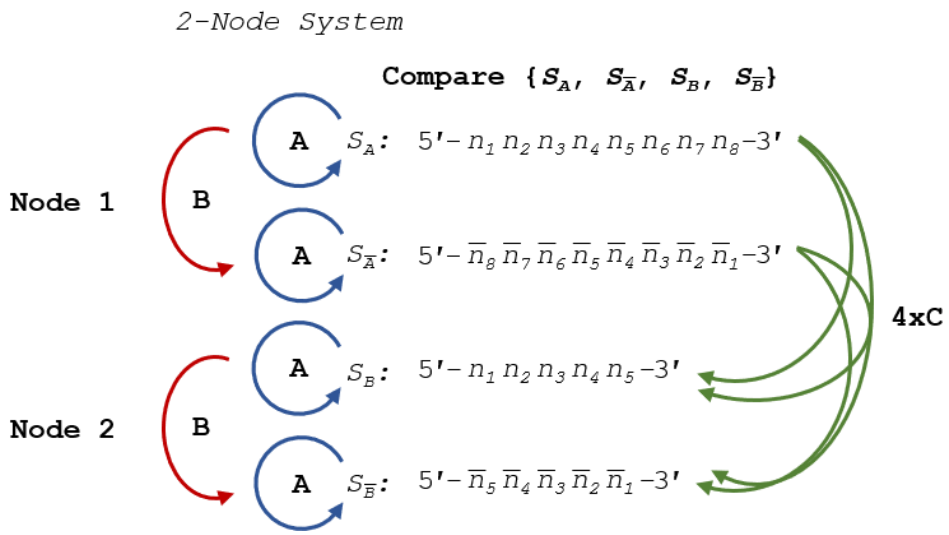


Figure 4.11: Comparison of nucleotides for dimer search in a two-node structure with three distinct operations, A, B and C, representing self-comparison, complement comparison, and other sequence comparison, respectively.

b.i. Heterodimer comparison in disparate nodes (Operation C)

To formulate the number of comparisons, we sum forward and backward ksize vectors with the allowance of $s_f = 0$. Recall that the formulas for k_{size} are dependent upon the relative lengths of the sequences being compared. As a first assumption, we require that the network be sorted in descending length order, where $L_1 \geq L_2 \ldots \geq L_i \geq \ldots L_n$. This requirement comes from the length-dependent derivation

of k_{size} (Tables 4.1, 4.2). The final form of the formula will be independent of assortment, and we can dispense with this assumption after derivation.

When $L_a \geq L_b$, we can write the following sum, where n_{sf} and n_{sb} are the number of slip conditions (k_{range}) forward and backward, respectively, and k_j is the corresponding frame size, similar to (31) and (33):

$$N_C = \sum_{j}^{n_{sf}} k_j + \sum_{j}^{n_{sb}} k_j \tag{36}$$

The left sigma can be rewritten by again substituting (10) and (15):

$$\sum_{j}^{n_{sf}} k_j = \sum_{j=1}^{L_b} (L_b - j + 1)$$

$$= (L_b)\$$

$$= \frac{(L_b)(L_b + 1)}{2} \tag{37}$$

The righthand sigma can be rewritten by breaking it into two vectors, substituting (11) and (17). To address the case where $L_a = L_b$, it can be noted that (17) and (18) are equivalent over the region of interest (substitute (11) into (17) to check):

$$\sum_{j}^{n_{sb}} k_j = \sum_{j=1}^{L_a - L_b} L_b + \sum_{j=1}^{L_b} (L_b - j)$$

$$= L_b(L_a - L_b) + (L_b - 1)\$ \tag{38}$$

Collecting terms from (37) and (38), we have:

$$N_C = L_b(L_a - L_b + 1) + 2(L_b - 1)\$$$

Simplifying, we arrive at an expression independent of the relative size of L_a and L_b:

$$N_C = L_b(L_a - L_b + 1) + (L_b - 1)(L_b)$$

$$N_C = L_a L_b \tag{39}$$

We can see a practical demonstration of Operation C in Figure 4.12. In this example, sequence S_A has a length of 8 nt, while S_B has a length of 5 nt. Note that, by convention, $L_A \geq L_B$. Operation C is efficient at order of approximately $O(n^2)$ for *n* nucleotides, or exactly $O(n^2)$ where $L_A == L_B$. An analysis of the effect of the relative sizes of L_A and L_B on the order of convergence can be found below in Section *4.2.7.b.iv.*

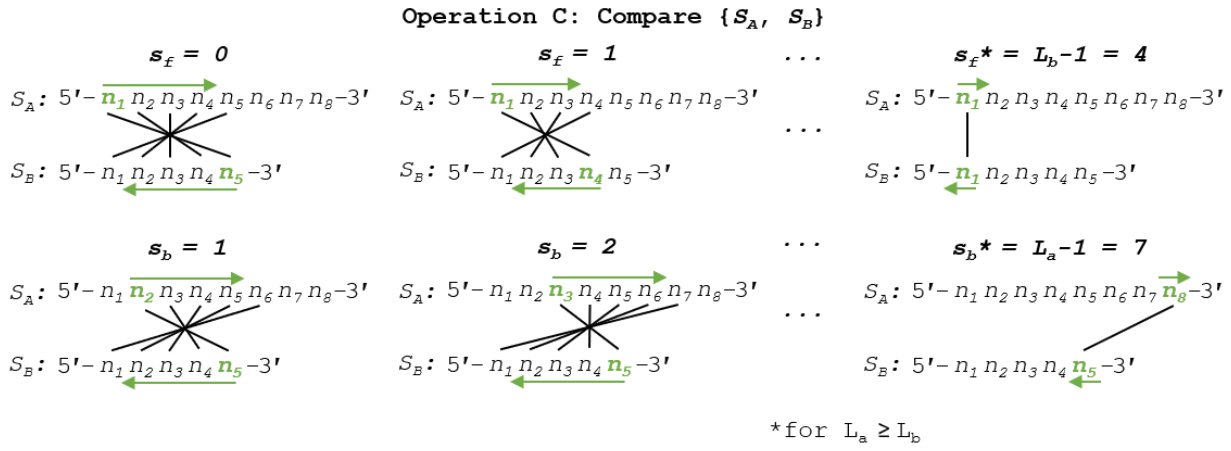


Figure 4.12: Nucleotide comparison for heterodimer search between disparate oligos (Operation C) is shown. By convention, backward slip starts at a value of 1 to avoid identical comparison, and $L_A \geq L_B$ to maintain consistency of k_{size} formulas (Tables 4.3-4.4). In practice, this need not be the case, and the relative oligo lengths can be tracked without significant computational overhead.

b.ii. General expression for any network

Returning to Figure 4.11, we can see that for each node in the system, self-comparison contributes $2A(L_a)+B(L_b)$ operations. Comparison between nodes, for node number D greater than one, we contribute $4C(L_a,L_b)$ comparisons. We can generalize this expression for total comparisons (N) in the following manner:

$$N = 2\sum_{i=1}^{D} A(i) + \sum_{i=1}^{D} B(i) + 4\sum_{j>i}^{D} C(i,j) \tag{40}$$

Substituting in (32), (34), and (39), we arrive at a general expression for the number of comparisons in any network with D double-stranded nodes:

$$N = 2\sum_{i=1}^{D} {L_i}^2 + \sum_{i=1}^{D}\left({L_i}^2 - L_i\right) + 4\sum_{i}^{D}\sum_{j>i}^{D} C(i,j)$$

$$\boldsymbol{N = \sum_{i}^{D}\left(3{L_i}^2 - L_i\right) + 4\sum_{j>i}^{D} L_i L_j} \tag{41}$$

b.iii. Comparison of fragmented and single-node networks

An important relationship in computational network analysis is whether the comparison efficiency converges faster in a duplex or a fragmented network. Specifically, will an increased number of nodes for the same number of total base pairs speed or slow computation? To investigate, let us define a network M_1 of D_1 nodes where $D > 1$. To perform this analysis, we require that the network M_1 be possessed of nodes with equal lengths, L_n. We can count the number of operations as before:

$$N_1 = \sum_{i}^{D_1} \left(3{L_i}^2 - L_i\right) + 4\sum_{j>i}^{D_1} L_i L_j \tag{42}$$

We can further count the total number of base pairs (B_1) in the network by adding together the lengths of all the nodes (L_i) in the network:

$$B_1 = \sum_{i}^{D_1} L_i \tag{43}$$

We can then create a M_2 linear structure for comparison containing one node ($D_2 = 1$) and whose total length (L_q) is identical to the length of all the nodes in M_1:

$$B_2 = \sum_{1}^{D_2} L_i = L_q \tag{44}$$

$$B_2 = B_1 \tag{45}$$

We can then rewrite the number of operations (N_2) in terms of L_q and substitute in (43):

$$N_2 = \sum_{i}^{1} \left(3{L_i}^2 - L_i\right)$$

$$N_2 = 3{L_q}^2 - L_q \tag{46}$$

$$N_2 = 3\left(\sum_{i}^{D_1} L_i\right)^2 - \sum_{i}^{D_1} L_i$$

Invoking the requirement that all nodes in M_1 have equal length, L_n, we can write:

$$N_2 = 3D_1^2 L_n^2 - D_1 L_n \tag{47}$$

Returning to N_1 we can simplify the number of operations, again requiring that all $L_i = L_j = L_n$. We rewrite (42) using L_n:

$$N_1 = \sum_{i}^{D_1} \left(3{L_n}^2 - L_n\right) + 4 \sum_{i}^{D_1} \sum_{j>i}^{D_1} L_n^2$$

$$N_1 = D_1\left(3{L_n}^2 - L_n\right) + 4(D_1 - 1)\$\, L_n^2$$

$$N_1 = D_1\left(3{L_n}^2 - L_n\right) + \frac{4(D_1^2 - 4D_1)}{2}\, L_n^2$$

$$N_1 = \left(2{D_1}^2 + D_1\right)L_n^2 + D_1 L_n \tag{48}$$

We can then take the difference between the two operations (47) and (48):

$$N_2 - N_1 = \ (D^2 - D)L_n^2 \tag{49}$$

Normalizing to the sequence length (B_1) of the network, we obtain the following relationship:

$$\frac{N_2 - N_1}{B_1} = \frac{(D_1^2 - D_1)L_n^2}{D_1 L_n}$$

$$\frac{N_2 - N_1}{B} = D(L_n - 1) \tag{50}$$

This formulation suggests that a fragmented network is more efficient to analyze for heterostructures than a one-node network of equal size. This relationship can be seen visualized in Figure 4.13, where 100 random networks of sizes $D \in [1, 30]$ are analyzed, averaged, and compared to (48) and (49).

Ultimately, the speed that is gained by fragmenting a network is lost in the tracking of that geometry, but the trade-off has important implications for algorithm design.

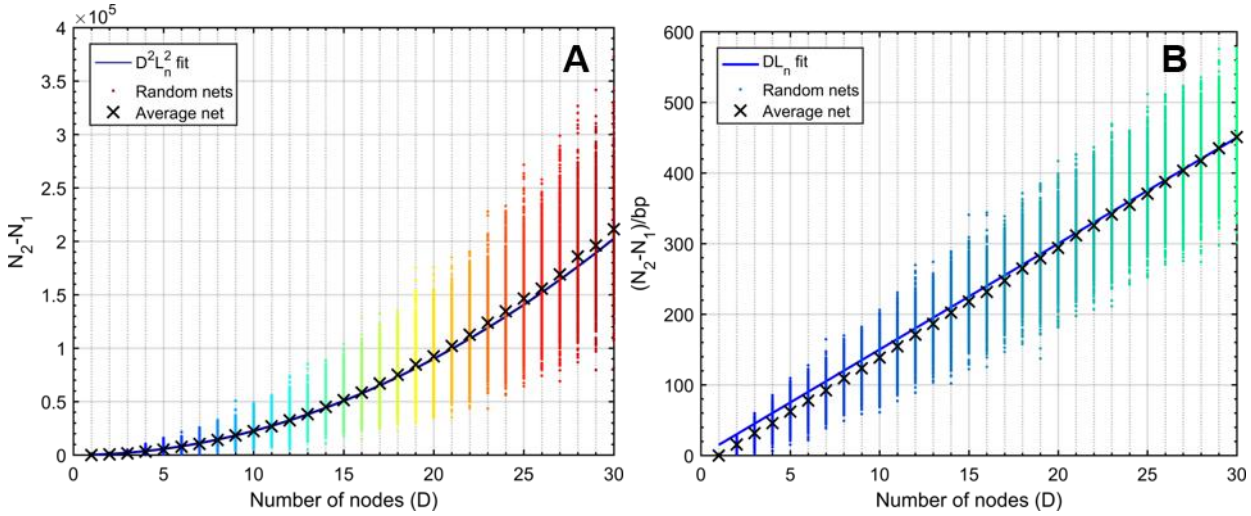


Figure 4.13: Comparison operation counts in multi-node networks and single-node networks of equal size. Random networks are initialized, analyzed, and averaged for the number of nodes, with 100 trials at each value of *D*. **A)** Operation counts for random networks over increasing number of nodes. Overlaying (49) as a fit line for the average behavior serves as a reasonable approximation, where L_n is given as the average node size, or $L_{max}/2$, 15 bp. **B)** Normalizing over the number of base pairs achieves a linear relationship between the difference and the number of nodes. Here, equation (50) serves as a reasonable fit line for the average data at each value of *D*.

b.iv. Comparison of symmetric and random networks

It is clear that fragmented networks are more efficient to analyze as the number of nodes increases. The analysis thus far has considered symmetric networks where each node has the same length as every other node, a requirement that was imposed to streamline the comparison. How then does the symmetry component affect the computation? We can calculate the difference in operation count between any random network M_0 with total base pairs B_0 and a symmetric network M_1, possessed of

the same total length, $B_1 = B_0$, and the same number of nodes $D_1 = D_0 = D$. For network M_1, we know the operation count N_1 in terms of average node L_n and D (48):

$$N_1 = (2D^2 + D\,)L_n^2 + DL_n$$

We have an equation for the operation count N_0 for a generalized network of unspecified node lengths (41):

$$N_0 = \sum_{i=1}^{D}\left(3{L_i}^2 - L_i\,\right) + 4\sum_{i}^{D}\sum_{j>i}^{D} L_i L_j$$

Recall that in a symmetric network, we know the number of base pairs in terms of L_n and D:

$$B_1 = \sum_{1}^{D} L_n = DL_n \tag{51}$$

And we know that L_n is an average of random node sizes:

$$L_n = \left(\sum_{i=1}^{D} L_i\right)\Big/ D \tag{52}$$

We can then substitute (52) into the definition of N_1 and simplify:

$$N_1 = (2D^2 + D\,)\frac{1}{D^2}\left(\sum_{i=1}^{D} L_i\right)^2 + \left(\sum_{i=1}^{D} L_i\right) \tag{53}$$

We then subtract N_1 from N_0:

$$N_0 - N_1 = 3\sum_{i=1}^{D}{L_i}^2 + 4\sum_{i}^{D}\sum_{j>i}^{D} L_i L_j - \left(\frac{2D^2 + D}{D^2}\right)\left(\sum_{i=1}^{D} L_i\right)^2 \tag{54}$$

The sums in this sequence must be rearranged to cancel, as $(\sum x)^2 \neq \sum x^2$. To simplify, we separate the sum squared term into two separate sums:

$$\left(\sum_{i=1}^{D} L_i\right)^2 = \left(\sum_{i=1}^{D} L_i\right)\left(\sum_{j=1}^{D} L_j\right)$$

$$= \left(\sum_{i=1}^{D} L_i\right)\left(L_i + \sum_{j\neq i}^{D} L_j\right)$$

$$= \sum_{i=1}^{D} {L_i}^2 + \sum_{i=1}^{D}\sum_{j\neq i}^{D} L_i L_j$$

Substituting into (54), we can rearrange:

$$N_0 - N_1 = \left(\frac{D^2 - D}{D^2}\right)\sum_{i=1}^{D} {L_i}^2 + 4\sum_{i}^{D}\sum_{j>i}^{D} L_i L_j - \left(\frac{2D^2 + D}{D^2}\right)\sum_{i=1}^{D}\sum_{j\neq i}^{D} L_i L_j \quad (55)$$

Now we can split the last term into two sums and combine like terms across the expression:

$$N_0 - N_1 = \left(\frac{D^2 - D}{D^2}\right)\sum_{i=1}^{D} {L_i}^2 + 4\sum_{i}^{D}\sum_{j>i}^{D} L_i L_j - \left(\frac{2D^2 + D}{D^2}\right)\sum_{i=1}^{D}\sum_{j<i}^{D} L_i L_j - \left(\frac{2D^2 + D}{D^2}\right)\sum_{i=1}^{D}\sum_{j>i}^{D} L_i L_j$$

$$N_0 - N_1 = \left(\frac{D^2 - D}{D^2}\right)\sum_{i=1}^{D} {L_i}^2 + \left(\frac{2D^2 - D}{D^2}\right)\sum_{i}^{D}\sum_{j>i}^{D} L_i L_j + \left(\frac{-2D^2 - D}{D^2}\right)\sum_{i=1}^{D}\sum_{j<i}^{D} L_i L_j$$

We then simplify this expression by introducing the constants α_m and β_m and splitting into three terms, (56-58)*:*

$$N_0 - N_1 = \sum_{i,j}^{D} \alpha_m L_i L_j \tag{56}$$

$$\alpha_m = \frac{\beta_m D^2 - D}{D^2} \tag{57}$$

$$\beta_m = \begin{cases} -2 & i < j \\ 1 & i = j \\ 2 & i > j \end{cases} \tag{58}$$

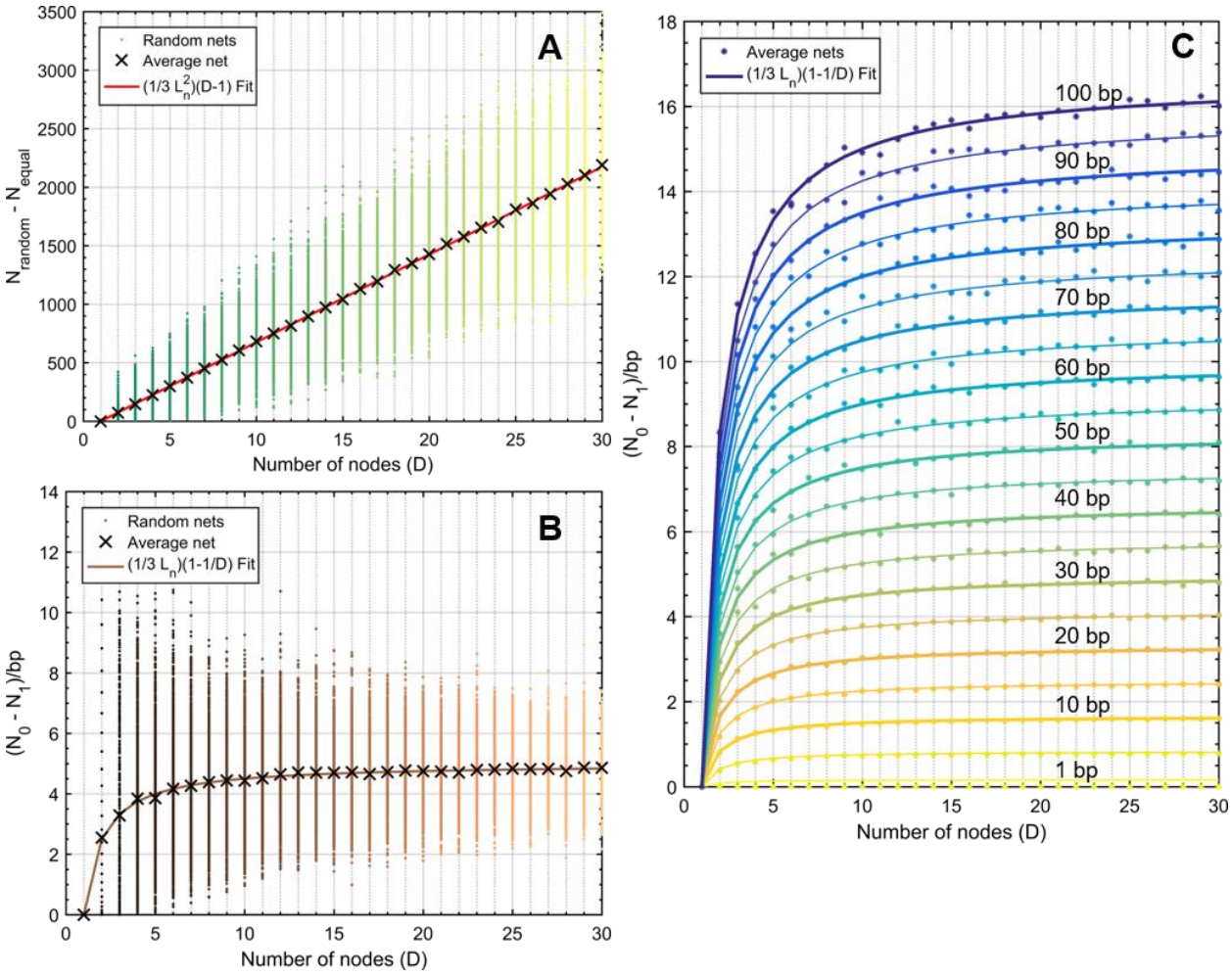


Figure 4.14: Difference in operation counts between random networks and their symmetric counterparts ($N_0 - N_1$). Symmetric networks were obtained by summing averaging the size of all nodes in a given random net, allowing for fractional base pairs. At node numbers $D \in [1,30]$, random lengths were

assigned between 1 and 30 bp. At each D value, 100 trials were run. **A)** The average behavior at each D value for $L_{max} = 30\ bp$ (X marks) was subjected to polynomial fit and determined to correspond to $\frac{1}{3}L_n^2(D-1)$, where L_n is $L_{max}/2$, or 15 bp. Data points were obtained using (56) and are identical to data generated without the use of a simplified expression [(41) – (48)]. **B)** The same experiment was performed and normalized to the total number of base pairs according to (51), where $B = \sum_i^D L_i \sim D L_n$. Here the fit line also behaves with $\frac{1}{3}L_n\left(1-\frac{1}{D}\right)$ relationship, which converges to $\frac{1}{3}L_n$ for $D \gg 1$. **C)** Experiments for values of $L_{max} \in [1,100]$ were carried out in the same manner as B. Only the average data are shown with their corresponding fit lines. The value of L_{max} is inset on lines every 10 bp.

If we reassume that all nodes in M_0 are equally sized, we can see that, on average, N_0 is greater than N_1 by approximately L_n^2, though the meaningfulness of that assertion is questionable. On the whole, (56) will describe the difference in performance between any random network and a symmetric network of equal node number and base pairs. The relationship between random networks and their symmetric counterparts can be seen in Figure 4.14.

4.2.8. Other sequence design criteria

In addition to dimer size minimization, DNA nanostructure design requires the minimization of other types of criteria as a result of thermodynamic or design-related criteria. Thermodynamic sequence design requirements were identified in discussion with Dr. Ruojie Shah in the group of Dr. Nadrian Seeman. All of the following analyses can be carried out using the single strand indexing rules found in Tables 4.3-4.4. Because they pertain to the presence of consecutive sequence elements, the rules described do not need to be subjected to repetition through slip and frame alignment, as with homo- and heterodimer search algorithms.

a. Gap analysis: 'gapN'

To utilize the emergent electrical characteristics of nucleotide pairing, defining, e.g. (based on generally accepted results in the literature, as discussed in Chapter 3), a resistor (WC pairing), a conductor ($dC:Ag^{+}:dC$ pairing), or a semiconductor (guanine tetraplex formation), it is most likely necessary to maximize density of a particular pairing regime across a polymerized oligonucleotide. Criteria such as gate analysis, which sets a percent occupancy for a particular nucleotide, may lead to unintended clustering of the preferred nucleotide at one end of the sequence, as with 75% guanine sequence: 5'-GGGGGGGGGATA-3'. While guanine may have the lowest base ionization energy of the four WC nucleotides,[163] an AT 3mer will interfere with the proposed conductance pathway. Gate criteria are thus useful for applications such as primer design, where GC occupancy is analyzed in percentages. For nanowire design, we instead introduce gap criteria, in which the maximum distance between a target *base pair*, not nucleobase, is tracked and minimized. Take for instance the sequence in (59) below where AT gap analysis is tracked and set to a maximum of 1 base pair.

```
5′–ATATATCAAGT–3′
   ||||||||||||        (59)
3′–TATATAGTTCA–5′
```

Gap analysis on this sequence shows the presence of 2 gaps in the AT occupancy, each of size 1 bp. Note the equvalence between dA:dT and dT:dA. The sequence shown in (60) instead fails the 'gap1' criterion for the AT base pair:

```
5′–ATATACCAAGT–3′
   ||||||||||||        (60)
3′–TATATGGTTCA–5′
```

Two gaps are identified, the first [dCdC:dGdG], and the second [dG:dC]. Highlighted in red is the size 2 bp gap that violates the prescribed 'gap1' rule. We can again perform the same 'gap1' analysis for CC pairing (dC:Ag^+:dC):

```
5′– CCACCCCACCT –3′
    ••|•|••|••|          (61)
3′– CCTCGCCTCCA–5′
```

In (61), the sequence presents 3 size 1 bp gaps in the CC paired nanowire, passing analysis for 'gap1' consideration. Note that the cytosine base in [dC:dG] does not evade detection as it is not paired to an opposing cytosine nucleobase, and it therefore counts toward a CC sequence gap. In contrast, the following sequence will fail 'gap1' analysis:

```
5′– CCACACCACCT –3′
    ••|•||•|••|          (62)
3′– CCTCTGCTCCA–5′
```

In (62), the presence of the [dAdC:dGdT] subsequence acts as a size 2 bp CC gap, failing 'gap1' analysis, but passing 'gap2' analysis. To design dC:Ag^+:dC nanowires with reasonable molecular conductance, we apply a 'gap1' CC rule to force any region that is conductive by design to have at most one non-metallic base pair in a row.

b. Purines and pyrimidines: 'R4' and 'Y4'

The overabundance of consecutive purines or pyrimidines in single oligonucleotides may affect the twist of the double helix. To maintain predictable rotational dynamics within a structure, allowing the formation of predictable geometry, it is conventional to restrict the total number of consecutive purines ('R' = A,G) and pyrimidines ('Y' = C,T) to four. Note that unlike gapping, this tracking occurs on single nucleotides, not base pairs, generating parallel analyses for each half of the node. An example of 'Y4' analysis is shown in (63):

```
5′– CCACTCCTCCT –3′
    ••|•|||•|••|          (63)
3′– CCTCAGCACCA–5′
```

Here the top sequence presents a pyrimidine 8mer as well as a 2mer. The bottom strand presents a 4mer, a 2mer and a 1mer. The top sequence will fail 'Y4' analysis. In contrast, the following duplex shows 'R4' analysis:

```
5′– CCACTCCTCCT –3′
    ••|•|||||•|          (64)
3′– CCTCAGGAGCA–5′
```

Here, the top sequence passes purine analysis (though it has 8 consecutive pyrimidines), while the bottom strand presents 5 consecutive purines [dAdGdGdAdG] and summarily fails the 'Y4' criterion. Because of CC pairing, neither (63) nor (64) have symmetrical purine and pyrimidine analysis. In WC conditions, all purine repeats will pair with pyrimidine repeats of equal size within the same node.

c. GC/AT: 'S4'/ 'W4'

For similar reasons, it is customary to minimize the number of consecutive CG and AT pairs to 4 bp. By convention, CG pair are abbreviated as 'strong' (S) owing to their three hydrogen bonds, while AT bonds are abbreviated as 'weak' (W) for containing two hydrogen bonds. Because currently known metal base pairs are homobase dimers, both strands in a duplexed node will achieve the same fitness score:

```
5′– ATATACCAAGT –3′
    |•|||||||||          (65)
3′– TTTATGGTTCA–5′
```

In (65), the sequence presents 5mer, 2mer and 1mer weak pair repeats, which fails the 'W4' design criterion, even with a dT:Hg^{2+}:dT base pair. By contrast, the CC bonding sequence in (66) will symmetrically fail 'S4' analysis:

```
5′– CCACACCGCCT –3′
    ••|•||•|••|        (66)
3′– CCTCTGCCCCA –5′
```

Here, both the top and bottom strands present strong base pair repeats of sizes 5,2 and 1 nt, regardless of ion distribution. The symmetry of this fitness criterion is immune to homobase orthogonal base pairs.

In practice, the W4 criterion may be relaxed to accommodate other design parameters.

d. Guanine repeats: 'G3'

The overexpression of guanine in an oligonucleotide may lead to the formation of a guanine tetraplex. To avoid kinetic traps or unintended geometries associated with tetraplex formation, we restrict the number of consecutive guanines to 3 nt, applying a 'G3' cap on single stranded sequences. This rule is not symmetrical, and the fitness score does not carry over from template to complement.

```
5′– GGCTCCCCAGT –3′
    ||||||||||||        (67)
3′– CCGAGGGGTCA –5′
```

The duplex shown in (67) will fail 'G3' analysis, as the bottom sequence has a guanine 4mer. The largest guanine repeat in the top sequence is 2 nt. To fix a sequence with this design, we can consider switching one [dC:dG] pair for a [dG:dC] pair.

4.2.9. Performance considerations

In order to speed computation, it is considered best practice to 'vectorize' operations in algorithmic design. In the case of oligonucleotide analysis, this involves avoiding time-intensive character vector comparison. We can see this spelled out in the following two prominent cases.

a. Float conversion

The storage of an oligonucleotide is typically done as a character array, for example: 'GGACTAG', where the sequence is read left-to-right to represent the 5'-3' orientation. Unfortunately, no mathematical operations can be performed on a character array. By contrast, a string converted to a number, either an integer or float data type, can carry out comparisons through subtraction, allowing for a vectorized analysis of sequence composition. Through experimentation (see Section *4.2.9.c* below), double-precision floating point numbers, 'double' in Matlab or 'float' in Python, demonstrate the shortest nucleotide comparison times. To correspond with one-indexed programming languages like Matlab, ['ATGC'] can be converted to [1, 2, 3, 4]. In a zero-indexed programming environment, we convert to [0, 1, 2, 3] to promote efficient vectorization. To account for all nucleobase possibilities, we utilize the following conversion: [dA, rA = 1.0]; [dT = 2.0]; [dG, rG = 3.0]; [dC, rC = 4.0]; [dC/dG, rC/rG, S = 5.0]; [dA/dT, rA/rT, W = 6.0], [rU = 7.0].

b. Comparison by vector subtraction

To compare substrings to character elements, a for loop iterates through the array and uses a comparison method such as strcmp(). To identify the complement of a nucleotide substring in a larger oligonucleotide array, we utilize the following comparison matrix for WC and CC pairing conditions:

$$\begin{matrix} & \text{WC} \\ \begin{matrix} 'A' \\ 'T' \\ 'G' \\ 'C' \end{matrix} & \begin{bmatrix} 'T' \\ 'A' \\ 'C' \\ 'G' \end{bmatrix} \end{matrix} \qquad \begin{matrix} & \text{CC} \\ \begin{matrix} 'A' \\ 'T' \\ 'G' \\ 'C' \end{matrix} & \begin{bmatrix} 'T' \\ 'A' \\ 'C' \\ 'G', 'C' \end{bmatrix} \end{matrix} \qquad (68)$$

This comparison algorithm uses two for loops, one to iterate through the oligo array, and the second to iterate through the rule matrix to identify the identity of the base and its possible complements using string comparison. A second string comparison is then carried out to compare complementary

characters to the opposing oligonucleotide array. In this way, dimers can be identified within an alignment frame.

A more efficient algorithm stores the oligonucleotide sequence as double-precision floating point number (float64, or double) and subtracts each 'nucleotide' from the nth row of float64 rule matrix, where ['A','T','G','C'] is equivalent to [1,2,3,4]:

$$\begin{matrix} \text{WC} \\ 1 \\ 2 \\ 3 \\ 4 \end{matrix}\begin{bmatrix} 2 \\ 1 \\ 4 \\ 3 \end{bmatrix} \qquad \begin{matrix} \text{CC} \\ 1 \\ 2 \\ 3 \\ 4 \end{matrix}\begin{bmatrix} 2 \\ 1 \\ 4 \\ 4,3 \end{bmatrix} \qquad (69)$$

One for loop is used to iterate through the oligonucleotide array. At each base, the row corresponding to the nucleotide identity is subtracted from the complement nucleotide. Where a value of 0 is reached, complementarity is obtained. This eliminates a for loop and both expensive character comparisons, speeding computation by two orders of magnitude (see Section *4.2.9.c* below).

In Matlab, this operation utilizes a one-indexed adenine, whereby the first row of the rule matrix in (69) is considered *row 1*. In a zero-indexed coding environment such as Python, adenine will be converted to 0, and subtraction of *row 0* for nucleobase comparison will occur.

c. Performance analysis

Analysis was performed in Matlab to identify the most efficient method of nucleobase comparison. The time to perform a single comparison was averaged across 10^8 iterations. Comparison of char vector 'G' with the second element in 'TTTATG' cost an average of 471 ns.

By contrast, subtraction of row 2 from element 2 of the floating point representation of the same sequence took on average 3.65 ns. Though utilizing less disk space, storing the sequence as a single precision floating point number (float32, single) took longer, clocking at 3.89 ns. Finally, even though the

rows and columns of a matrix have integer values, storage as an integer (int8, int) took 6.63 ns. These data can be seen in Table 4.5:

Table 4.5: Comparison efficiency of various Matlab data structures

Data type	*Sequence representation*	*Complementarity rule matrix*	*Average operation time*	*(xSlower) timeN/timeDouble*
double	[2, 2, 2, 1, 2, 3]	[2, 1, 4, 3]	3.65 ns	-
single	[2, 2, 2, 1, 2, 3]	[2, 1, 4, 3]	3.89 ns	1.06
int	[2, 2, 2, 1, 2, 3]	[2, 1, 4, 3]	6.63 ns	1.81
char	['T','T','T','A','T','G']	['T', 'A', 'C', 'G']	472 ns	129

There are tens of thousands of nucleotide comparisons to analyze a single DNA nanostructure (see 1.7.b.ii). Any sequence optimization through iterative algorithms will involve hundreds of millions to billions of comparisons. It is therefore clear that a conversion to double-precision floating point arrays is a critical feature for dimer search, complement generation, and nucleobase identification algorithms, and can enable high-performance optimization of DNA sequences for nanotechnology.

d. Reverse indexing

The tracking of base index inside an alignment frame is an operation requiring realtively few floating point operations. When aligning two sequences to test for complementarity, the usual practice on paper is to take a 5'-3' oligo and place it adjacent to a 3'-5' sequence. Though this makes conceptual sense, it requires that the user perform a flip operation on one oligo in an environment where all oligos are stored in the 5'-3' orientation. In a programming environment, this requires matrix transposition, which uses a relatively large number of floating point operations. An analysis of the time to transpose a floating point 'oligo' array of varying length was carried out over 10^6 iterations, and was determined to cost ~1 μs per fliplr() operation (Figure 4.15). A single nanostructure can have thousands of alignment

frames; over an iterative time scale, these transpositions will introduce significant delays in sequence design.

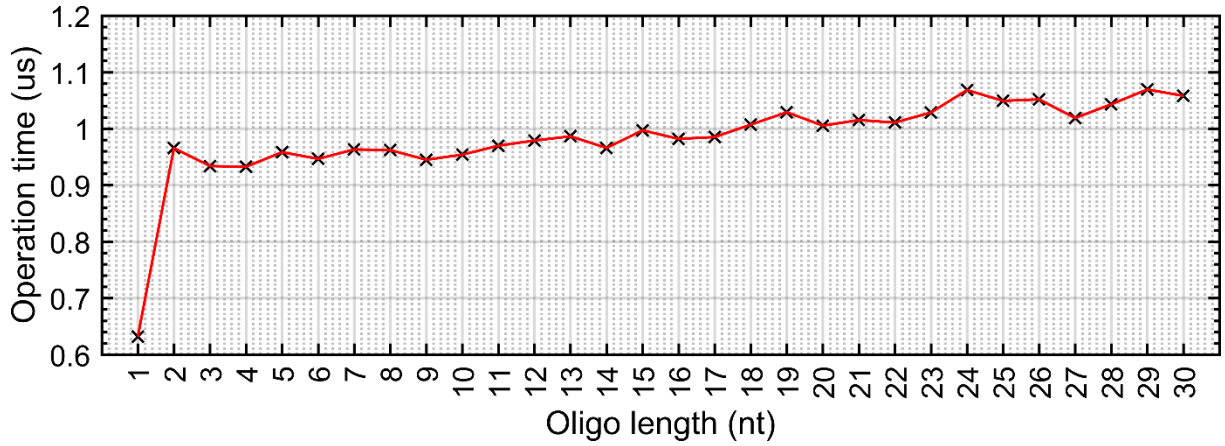


Figure 4.15: Average time to take a 5'-3' sequence and use fliplr() to store it in 3'-5' orientation in a Matlab floating point array. Operation carried out over 10^6 trials. For all sequence lengths >1 nt, transposing the array took approximately 1 µs. Omitting the first data point, the linear fit equation is approximately: 45 ns/nt + 920 ns (R^2 = 0.87).

To avoid gratuitous transposition of DNA arrays, we introduce reverse indexing, whereby base position in complement strands can be tracked going backwards from the end of the alignment frame. This allows the two oligos of interest to remain in the 5'-3' orientation. Indexing formulas of this type can be found in (25)-(28) and are summarized in Tables 4.3-4.4.

4.3. Computational Design of Nanostructures using a Genetic Algorithm

Using the computational framework established above, we outline the design and construction of a genetic algorithm toolbox in Matlab for branched, orthogonal DNA sequence design.

4.3.1. Generic workflow

The algorithm follows a general GA workflow with initialization and iteration phases. Design of topology and conserved regions occurs on paper and in other software packages. Within this toolbox there are

three distinct phases of operation. In the first phase, optimization and iteration criteria are defined precisely through graphical user interface (GUI) and .m file entry. In the second phase, the model is then initialized, and data structures are pre-allocated based on user specifications. Iteration begins in separate population files, and a subsequent F2 cross between best solutions is carried out. In the final phase, runtime analysis is performed, sequences data store, and graphical data is exported for the user. A diagrammatic explanation of this information can be found in Figure 4.16:

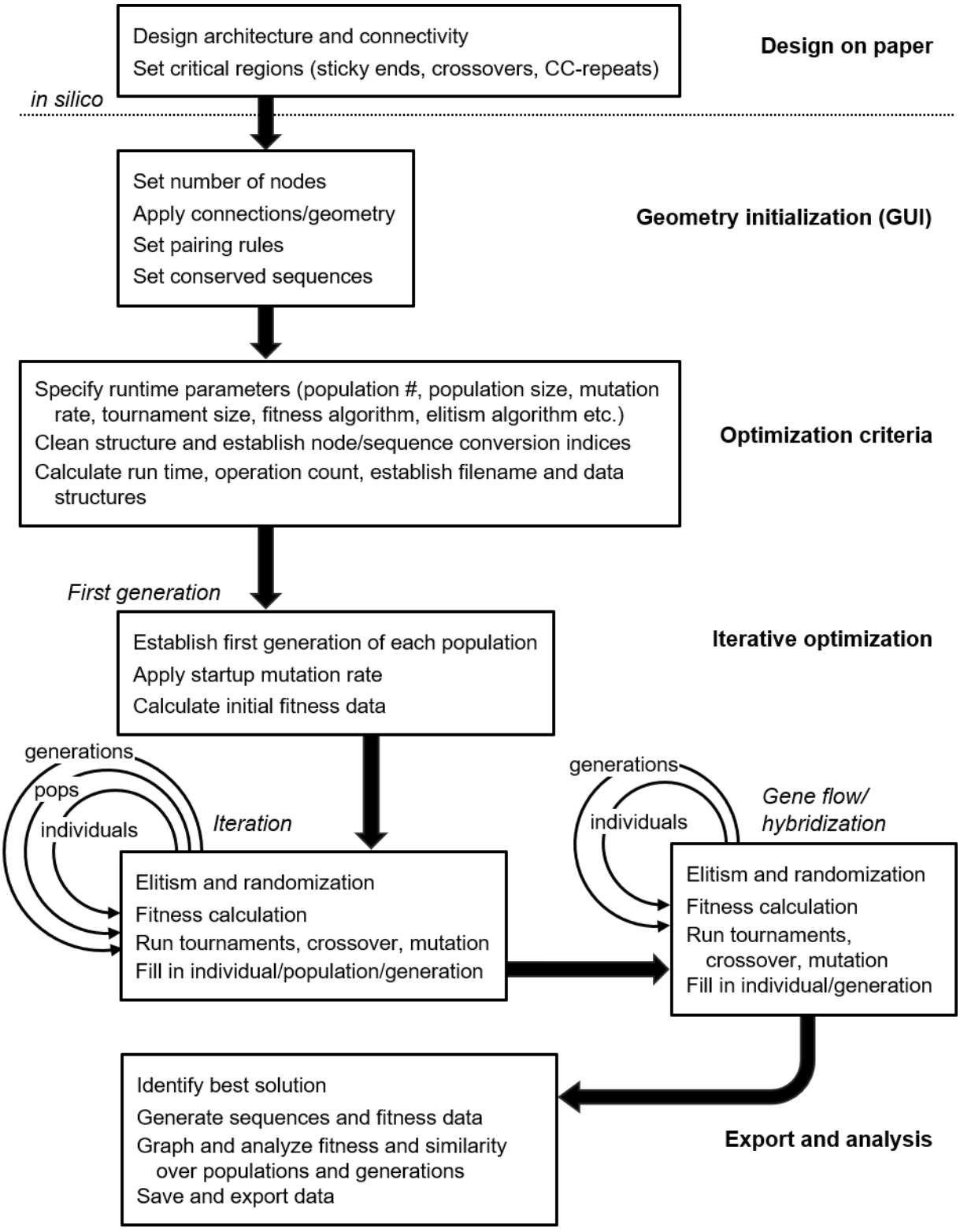


Figure 4.16: Process workflow for nanostructure sequence design algorithm.

4.3.2. UI component

A nanostructure is a struct() data type with N nodes. Each node is also a data struct(), with the following properties:

Table 4.6: Node properties

Property Name	**Data type**	**Overview**
nodeLength	integer	Number of nucleotides in the node
hasTemp	boolean	Indicates presence/absence of template sequence
hasComp	boolean	Indicates presence/absence of complement sequence
tempSeq	double	1D array of floating point nucleotide values for template (5'-3')
compSeq	double	1D array of floating point nucleotide values for complement (5'-3')
sequenceEditable	boolean	Determines whether sequence may be modified by mutation/crossover
ruleMatrix	cell	Vectorized parity rules, for mutation, not for assessing dimers, see (69) for representation of ruleMatrix
tempInNode	integer	Node number for template 5' connection, -1 for no connection
tempOutNode	integer	Node number for template 3' connection, -1 for no connection
compInNode	integer	Node number for complement 5' connection, -1 for no connection
compOutNode	integer	Node number for complement 3' connection, -1 for no connection
tempInDir	boolean	5' (1) or 3' (0) connection to template in node
tempOutDir	boolean	5' (1) or 3' (0) connection to template out node
compInDir	boolean	5' (1) or 3' (0) connection to complement in node
compOutDir	boolean	5' (1) or 3' (0) connection to complement out node

These parameters are entered using a GUI, in which the geometry and sequence of each node is manually specified, as well as its ability to be edited and internal pairing regime. The specification of node-by-node rule matrices does not at this stage govern which pairs are allowed during dimerization analysis, but rather governs the types of base pairs that are able to generate during mutation and crossover. At this stage, certain regions can be restricted to WC or CC parity; while other nodes can be specified as sticky ends by having no complement, no incoming node connections, and uneditable sequences. At this stage, the user has quite a bit of flexibility over the types of sequences and the ways that they connect, allowing for entry of complex nanostructures with heterogeneous local design rules.

4.3.3. Optimization criteria setup

After the user specifies the geometric frame that will be optimized, the model proceeds to the setup stage, in which a GA iterator is initialized and run with specific criteria, outlined below.

a. Population and runtime parameters

The runtime environment is initialized with the following user-defined parameters:

Table 4.7: Runtime parameters

Property Name	**Data type**	**Overview**	**Default Value**
popNum	integer	Number of independent populations in the simulation	*5*
popSize	integer	Number of unique solutions within a single population	*40*
generations	integer	Number of iterations for the genetic algorithm	*500*
tournamentSize	integer	Size, in number of solutions, of fitness tournaments, with a default one winner	*3*
crossoverRate	double	Probability [$0 \leq x \leq 1$] of exchanging information between two successful tournament winners at a given base position	*0.3*

mutationRate	double	Probability [0 ≤ *x* ≤ 1] of generating a random nucleotide pair in the allowable ruleMatrix of the component node	*0.05*
firstMutationRate	double	Probability [0 ≤ *x* ≤ 1] of generating a random nucleotide pair in the allowable ruleMatrix of the component node *in the first generation*, allowing for greater initial diversity	*0.8*
dimerWeight	double	Fitness score weight corresponding to the dimer size determined to be equivalent in gravity to gap2, G3, S4, R4, and Y4	*10*
numberOfElites	integer	Number of best solutions copied without mutation to the next generation	*5*
numberOfRandom	integer	Number of solutions with randomly generated nucleotides in editable regions in each generation	*2*
savePopData	boolean	Indicates whether longitudinal population data is stored throughout the simulation, and governs the preallocation of the accompanying cell matrix	*1*
saveFitnessData	boolean	Indicates whether longitudinal fitness data is stored throughout the simulation, and governs the preallocation of the accompanying cell matrix	*1*

b. Environmental parameters and polymorphic structures

The model will also be subject to several environmental parameters that affect performance. Firstly, the user will specify what base pairs are globally available for dimer analysis, in essence specifying the presence or absence of environmental factors such as silver and mercury cations, RNA bases, and orthogonal nucleotides. A global *ruleMatrix* is established (69), which governs fitness function

evaluation across all nodes. The nanostructure *ruleMartrix* differs from the nodal *ruleMatrix* in that it does not govern which nucleotides are generated during randomization and complement formation, but rather the penalties applied to existing nucleotides, regardless of regional design criteria.

A patch has also been developed to allow for certain regions to be specified as polymorphic. It is sometimes the case that sequences for nanostructures should be optimized to allow interaction with sticky ends in other structures. It is also the case that optimizing one nanostructure to operate with multiple versions of a single region may reduce synthesis costs and allow for greater diversity of behaviors from swapping out a single oligo. This may have usefulness in developing molecular motors and other nanomachines built from extremely similar nucleotide sequence swapping. The design tool allows for additional versions of certain nodes to be pre-specified, and are governed by the boolean operator *swapBool* and accompanying cell matrix *SEswap*. When *swapBool* is set to 1, the fitness function will be evaluated twice, once for each version of the nanostructure, and the fitness scores added together for a composite score. Scores will be consequently elevated, and optimization will be carried out on both versions of the structure, with weight distributed equally between phenotypes.

4.3.4. Iterative optimization

The runtime parameters for model setup are found in Table 4.7. Each population is initialized, then carried through the full simulation and stored. The best answers from each population are then collected and pitted against one another in a subsequent 'F2 cross' simulation.

a. One-time setup

At the start of the simulation, prior to the creation of any populations, analysis of the user-input nanostructure geometry is performed, and appropriate runtime data structures are generated. Most prominently, the GA iterator will generate a map and corresponding mapping function for converting dsDNA nodes into ssDNA sequences. The downstream fitness function will track the node connections

across sequences, allowing for node-based parameter specification while iterating through real oligonucleotides. The map (*linearSequenceIndicies*) is generated only one time at a cost of ~320 µs, and the mapping function (*linearSequenceExtraction*) is applied at a cost of 50 µs to generate ssDNA sequences (*linearSequences*) on each individual solution prior to fitness calculation. As some fitness criteria are applied on base pairs while others are applied to nucleotides (heterodimers vs. guanine repeats), it is necessary to track both node and sequence indices with this mapping function. Node objects and containers are subsequently cleaned and prepared for simulation, and each node is assigned a *ruleMatrix* (i.e. [69]), or set of allowable base pairs for mutation and randomization.

b. Population initialization

Populations are built from collections of mutually-interacting solutions that are separated from other populations. The number of total solutions tested is the product of *popSize* and *popNum*. The model runs for a set number of iterations, or *generations*, using the gen number as a stop criterion. At the start of each population simulation, longitudinal data structures can be set up to hold fitness, similarity and nucleotide data for each solution across the whole simulation (*fitnessScoresAllGens*, *dimerScoresAllGens*, *ATScoresAllGens*, etc.), at the cost of runtime speed.

A population P_j, or *pop*, is a cell data structure with dimensions (1, *popSize*). The contents are *nanostructure* objects, containing *Node* objects, a *ruleMatrix*, and other parameters specified above. To generate a *pop*, the basic *nanostructure* frame is subjected to an initial, elevated mutation rate (*firstMutationRate*), which causes all editable nodes to be subjected to near-randomization of base pairs allowed in their corresponding *ruleMatrix*. Setting a low initial mutation rate may be beneficial when continuing to optimize an already-modeled set of sequences. After initializing the data structures and the first generation *pop*, the model is ready for subsequent iteration.

c. Iteration by generation:

The GA iterator performs the same operations for each cycle, namely sequence generation, fitness calculation, and pop creation. At the end of the simulation, the best solutions are saved in a subsequent population (*F2pop*) for secondary optimization.

c.i. fitness score calculation

The fitness score is a weighted composite penalty comprised of several criteria, namely: guanine repeats ('G_N'), GC repeats ('S_N'), AT repeats ('W_N'), purine repeats ('R_N'), pyrimidine repeats ('Y_n'), nucleobase gaps ('gap_n'), and heterostructure dimer size ('D_N') (see Section *4.2.8*). The type of solution will be influenced by the relative weight of these parameters. To allow for user input, the variable *dimerWeight* is introduced, which corresponds to the stringency of dimer size requirements. This parameter corresponds to the dimer size that is equal in penalty to G_3, S_4, R_4, and Y_4. In the presence of a gap criterion, a gap size exceeding the user-specified maximum is also weighted similarly with dimers of size *dimerWeight (dW)*. The fitness penalty of the following parameters is identical: D_{dW}, G_3, S_4, W_4, R_4, gap_2. The Y_N criterion is ignored.

For each solution, histograms of all dimers, guanine repeats, purine repeats, nucleobase gaps, etc. are generated using the indexing and analytical framework established in Section *4.2*. From these histograms, GC, purine and pyrimidine repeats of size less than 4, guanine repeats and dimers less than size 3, and allowable gaps are all pared from these histograms. The remaining histogram bins are then subjected to a series of exponential functions :

$$F = \sum_{i\geq 3}^{n_D} c_{D,i}10^{(i-2-dW)} + \sum_{i\geq 4}^{n_R} c_{R,i}10^{(i-2)} + \sum_{i\geq 4}^{n_W} c_{W,i}10^{(i-2)} + \sum_{i\geq 4}^{n_S} c_{S,i}10^{(i-2)} + \sum_{i\geq 3}^{n_G} c_{G,i}10^{(i-1)} + \sum_{i}^{n_{gap}} c_{gap,i}10^{i} \quad (70)$$

In (70), summation across each histogram bin occurs for repeats of size *i*, starting at empirically-identified sizes and finishing at the maximum repeat, or n_X, for each parameter. The number of repeats at each size, c_i, corresponds to the histogram bin value at *i.* This number is multiplied by a decimal exponent containing *i* and an adjustment value to match the weights of criteria with different critical *i* values. A solution with better (fewer) repeats will have a smaller fitness value *F.* As such, the global optimization problem is the minimization of *F*. Pyrimidine repeats are omitted from the fitness function for redundancy, and GC repeats may be omitted for CC pairing environments to allow polycytosine repeats (this may affect the helical angles and assembly of the constituent nanostructure). Gaps are only included where specified by the user.

Dimer fitness is adjusted by *dW* or *dimerWeight* as dimers are of lesser relative importance compared with other types of repeats: a nanostructure may assemble with a size 7 bp heterodimer, while it will not be able to form a B-form duplex with a size 7 nt guanine repeat. The exponential is applied in order to clearly differentiate between the steps in the fitness hill: 5 dimers of size 8 bp are of lesser importance than 2 dimers of size 9 bp. By contrast, 10 dimers of size *i* will be equal in penalty to 1 dimer of size *i*+1. In this way, each level of the fitness hill is clearly segregated by repeat size, while the repeat count (*c*) is weighed in a granular fashion within each step (Figure 4.17).

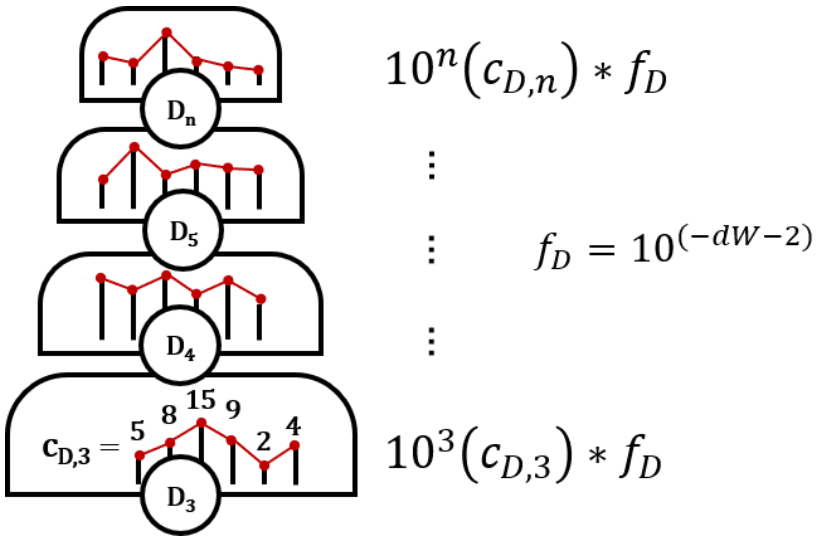


Figure 4.17: Graphical representation of dimer fitness hill. Each level is of size 10^i for a dimer of size *i*. Within each grade, the number of dimers of each size—the histogram bin count $c_{D,i}$— is the exponent constant. Here, *f* represents the multiplication factor, or relative weight, of the given fitness characteristic. Different criteria will have different weights, seen in (70).

The full fitness function can be described in a compact sum, shown in Figure 4.18. This graphical representation of the different fitness criteria demonstrates the relationship between fitness criteria and their weights. It can be seen that as other criteria pass their minimum penalties, the fitness hill becomes less granular, as the contribution of heterodimers will predominate. Selection of parameter *dW* by the user will determine the size of heterodimer where this solution characteristic becomes more heavily penalized. A good starting value for dW may be between 8 and 10.

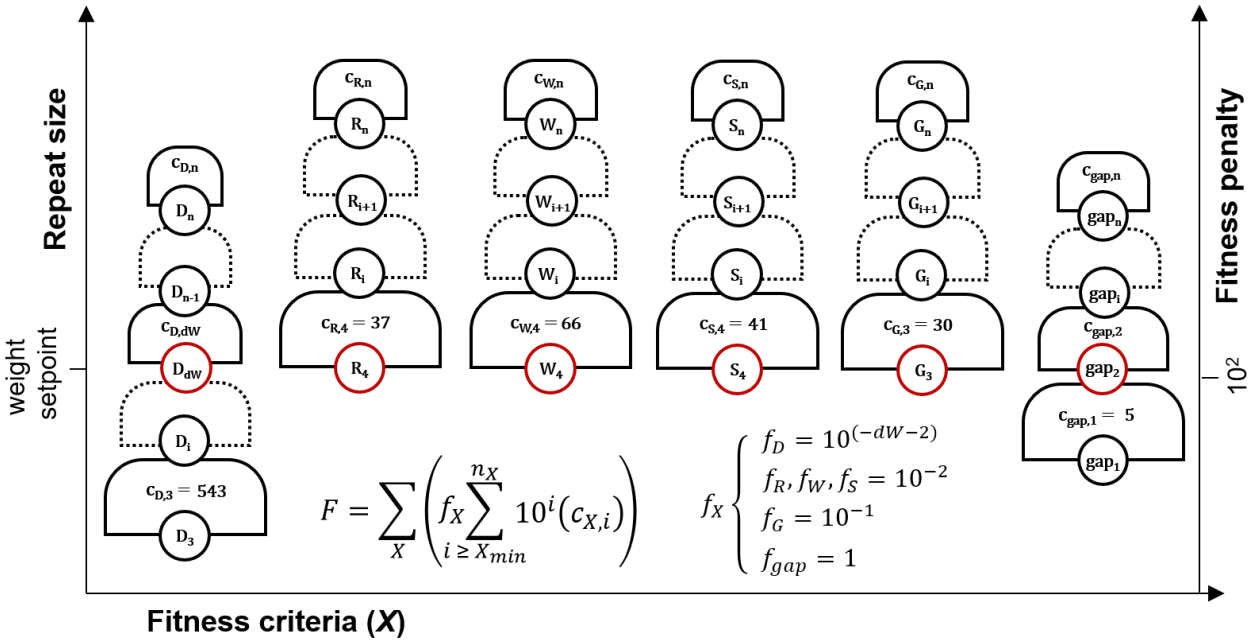


Figure 4.18: Schematic representation of fitness function (70). The fitness (*F*) of a solution is comprised of several sequence design criteria (*X*): dimers (*D*), purine repeats (*R*), AT pair repeats (*W*), GC pair repeats (*S*), guanine repeats (*G*), and nucleobase gaps (*gap*). Within each criterion, there is a graded

fitness hill with levels segregated by the exponential function $c*10^i$ for repeat size *i* and histogram bin count $c_{X,i}$. Levels marked in red are weighted equally. The weights are controlled by adjustment factor f_X. Each repeat type has a maximum recorded value n_X, and a minimum size X_{min}, under which histogram bins are not counted.

The fitness function *F* for fitness criteria *X* = [(***D***)imer, pu(***R***)ine, *(**W**)eak pair, (**S**)trong pair, (**G**)uanine,* nucleobase (***gap***)], repeat size *i*, repeat count $c_{X,i}$, minimum repeat size X_{min}, maximum repeat size n_X, and adjustment factor *f* may be expressed in compact sum form, as seen in Figure 4.18:

$$F = \sum_X \left(f_X \sum_{i \geq X_{min}}^{n_X} 10^i \left(c_{X,i} \right) \right) \qquad f_X \begin{cases} f_D = 10^{(-dW-2)} \\ f_R, f_W, f_S = 10^{-2} \\ f_G = 10^{-1} \\ f_{gap} = 1 \end{cases} \tag{71}$$

c.ii. create population for next generation

Once the fitness scores for a current generation are calculated, these rankings are used to populate the next generation. First, an elitism algorithm is executed in which a specified number of solutions (*numberOfElites*) are copied without alteration from the top of the fitness list. If the number of elites meets or exceeds the population size, no optimization will occur. Once elites are copied, a specified number of random solutions are generated (*numberOfRandom*) where editable nodes are subject to a pseudorandom number generator.

The remaining slots in the next generation pop are filled using tournament style competition. Solutions are selected at random to fill a tournament of specified size (*tournamentSize*), and the fittest individual is chosen. Ties are broken by choosing the first of the solutions selected during tournament filling. A second tournament is carried out, and the two best solutions are subject to a crossover algorithm, in which editable sequence information is exchanged with a specified probability (*crossoverRate*). A random number is rolled *r* = [0,1), and if *crossoverRate* exceeds *r*, the nucleotides in template and

complement strands at that position are swapped using forward and reverse index values. The two modified solutions are then subjected to nucleobase randomization at a specified rate (*mutationRate*) where editable nucleotides are subjected to a random number roll r = [0,1). When the mutation rate exceeds the roll, a random base is selected from that node's *ruleMatrix*, and if a complement sequence is available, a complementary base is selected at random from the allowable complementary nucleotides in the same row of the *ruleMatrix*. In this way, two new solutions are generated with information mostly conserved from previous answers based on fitness calculation and tournaments. When there is only one available slot left in the next generation pop, the first new solution to be generated will be selected. Once the population is filled, the generation number is advanced.

c.iii. repeat for F2 cross

When the last generation is reached for a given population P_j, fitness scores are calculated and the best solutions (specified by *numberOfElites*) are copied without alteration to a separate population, F_2. The best solutions from all P_j are collected and subsequently optimized in the same manner. The F_2 population is subjected to the same runtime parameters as each of the feeder populations. The fittest solution in the final F_2 generation is designated as the end result of the genetic algorithm. This type of forced migration, or gene flow, serves to overcome fitness niches created through the steepness of the fitness landscape by combining disparate solutions.

4.3.5. Data analysis and output

a. Save data

If the user specified *savePopData,* the final *F2pop* is saved to a Matlab data file with a timestamp and pertinent parameters marked. If *saveFitnessData* is marked, longitudinal arrays are saved to files with the same filestring and '*fitData*.' Any fitness graphs generated from these data are marked 'DIM', 'SIM' and 'CMP' for dimerization, similarity and repeat analysis, respectively.

b. Graphical analysis

When *saveFitnessData* is specified, data analysis across the lifetime of the simulation can be analyzed. In particular, the fitness scores and various metrics can be plotted over iterations and compared with the final *F2pop*. The parameters displayed are modular and may be chosen with some alteration to plot code. The graphical interface is shown without alteration in Figures 4.19-4.22. Figure 4.19 displays logarithmic fitness data over the lifetime of the simulation. The *F2pop* (bold) combines the best solutions from the feeder populations and is subject to intense initial competition, followed by slight incremental change over the lifetime of the final simulation. Figures 4.20-4.21 show the magnitude and frequency of repetitive segments in the nanostructures, with fitness criteria (*X*) from (70)-(71). Repeat sizes are subjected to a weight function to illustrate both the maximum repeat size X_n and the abundance of that repeat $c_{X,n}$. This function can be expressed as repeat size Xn with an additional weight w = (0,0.5]:

$$X_n + 0.5(1 - e^{c_{X,n}+1}). \tag{72}$$

Figure 4.22 shows the nucleobase similarity for all bases and all positions between all members of a nanostructure population. The similarity calculation (*S*) for nanostructures (*M*) in a population (P_j) with total number (*L*) of nucleotides (b_i) at position *M(i)* can be expressed in the following way:

$$S = \frac{1}{L}\sum_{u}^{N}\sum_{v>u}^{N}\sum_{i}^{L} g(M_u(i), M_v(i)) \qquad g(b_1, b_2)\begin{cases}1 & if\ b_1 - b_2 = 0\\ 0 & otherwise\end{cases} \tag{73}$$

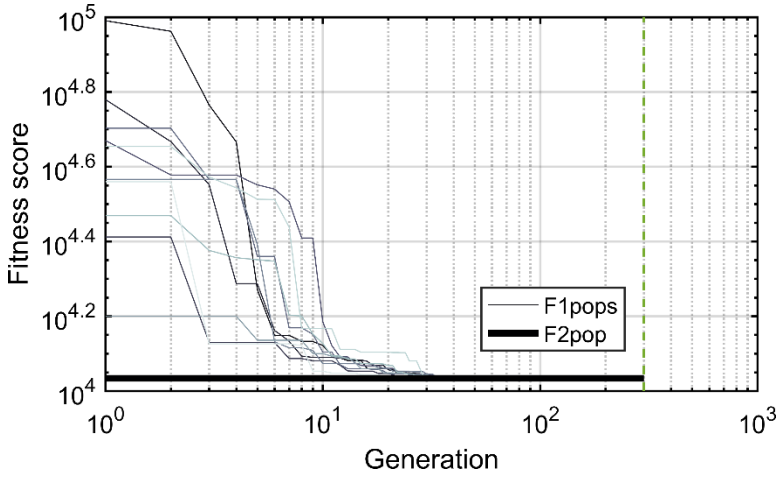


Figure 4.19: Longitudinal fitness data, plotted on logarithmic axis. The *F2pop* (marked bold) combines the best solutions of all other populations at the end of their respective simulations. Subsequent optimization selects for the best traits of each of these populations and may arrive at a further-optimized answer.

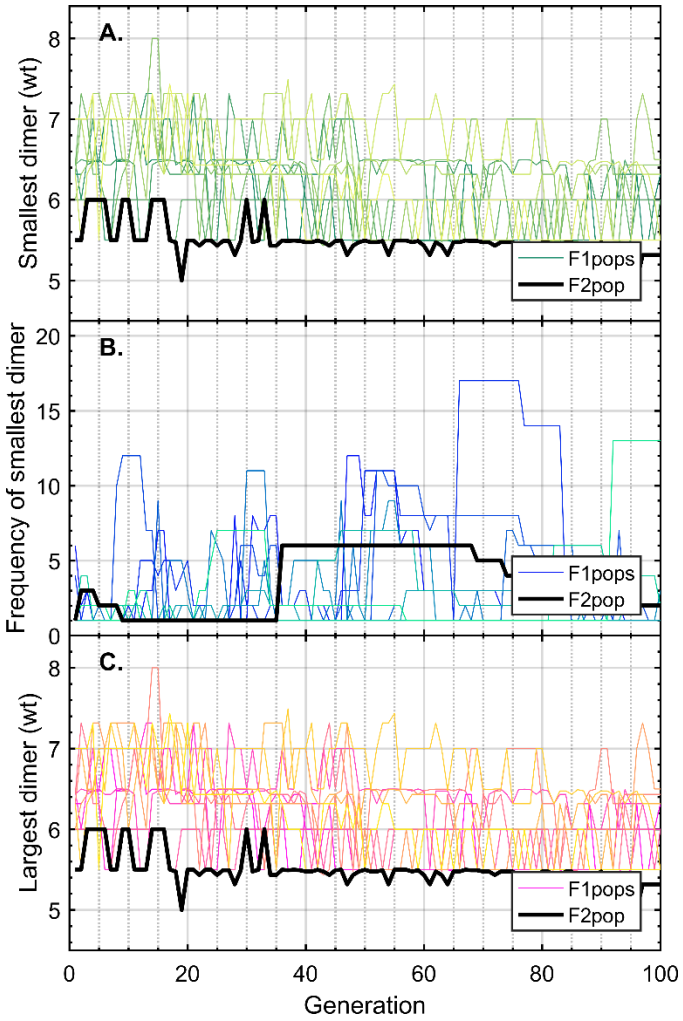


Figure 4.20: Longitudinal dimer data plotted for F1pops P_j (colored lines) and *F2pop* (bold). **A)** The largest dimer (D_n) of the fittest solution in each generation for all pops is shown. In order to show more granular data, the frequency of the dimer $c_{D,n}$ is included in this smallest dimer with the weight function (72): $D_n + 0.5(1 - e^{c_{D,n}+1})$. This function caps the weight of the dimer to $D_n + 0.5$. **B)** The frequency of the largest dimer ($c_{D,n}$) from A within a single nanostructure is shown. As a result of the fitness function, dimer count $c_{D,i}$ for dimer size D_i greater than 10 is counted as equal in weight to a dimer of size D_{i+1}, which causes most frequencies to rapidly fall under 10. **C)** The least fit solution for each generation is plotted in a similar fashion to A. It can be seen from these graphs, which are automatically displayed at the end of simulation, that the F2pop far outperforms the other populations, and that where feeder

populations generally reduce the max dimer size D_n, the F2pop will primarily reduce the histogram bin count $c_{D,n}$ for D_n.

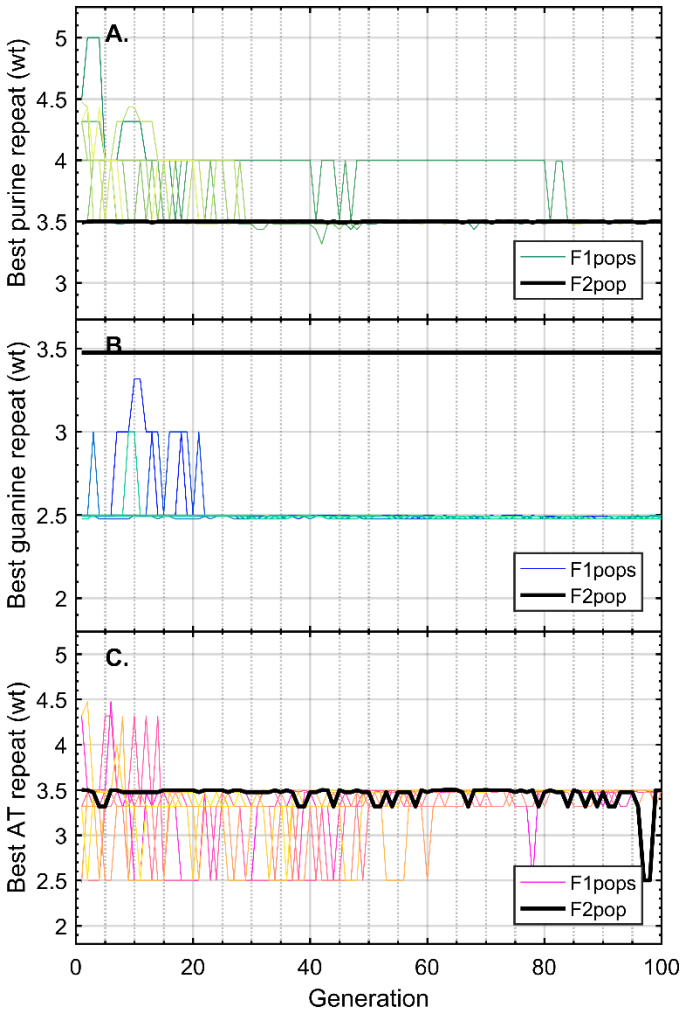


Figure 4.21: Plots of maximum repeats (X_n) for varying criteria (X) over all generations for all pops with weight function (72): $X_n + 0.5(1 - e^{c_{X,n}+1})$. As in Figure 4.20, the maximum display value for a repeat of size X_n is $X_n + 0.5$. Criteria shown: **A)** single-strand purine nucleotide (A,G) repeats (R_n); **B)** single-strand guanine nucleotide repeats (G_n); and C) weak base pair (A:T) repeats (W_n). Due to the nature of the fitness function, these repeats rapidly drop to the minimum recorded values and experience little change across the lifetime of the simulation.

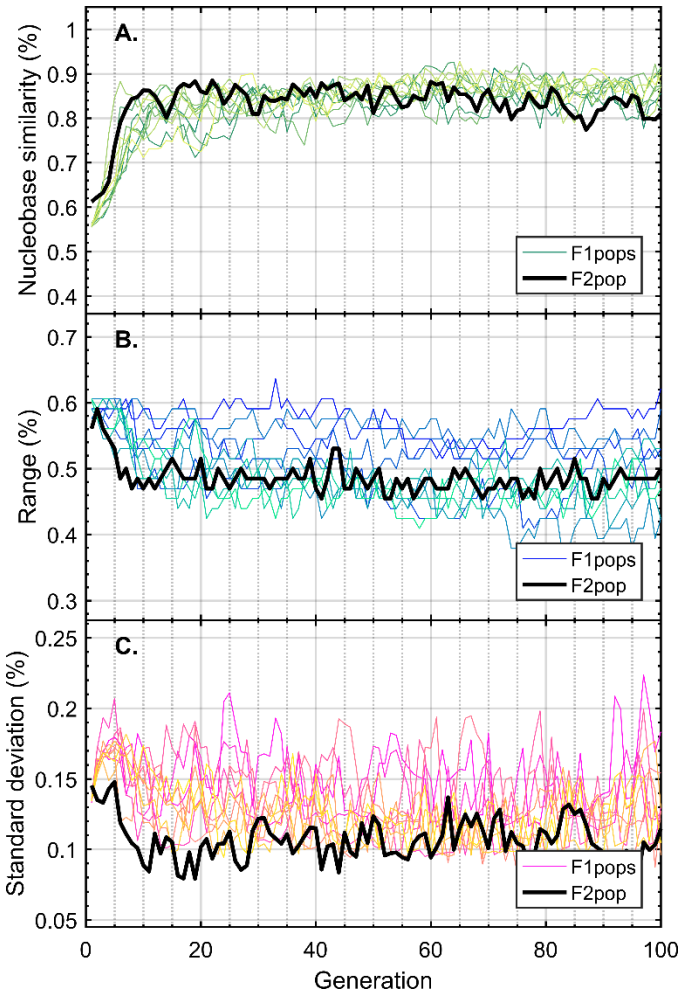


Figure 4.22: Similarity statistics for all pops over all generations based on (73). **A)** Average percentage of nucleotides that are identical within nodes across all individual solutions in the population. **B)** Range between maximum and minimum similarity for all individual solutions in the population at the given generation. **C)** Standard deviation of similarity scores between all individual solution in the population at the given generation.

Graphical analyses can be utilized to tailor future simulations by analyzing the effects of competition over iterative time scales to select the appropriate runtime parameters.

4.3.6 Runtime parameter analysis

When it comes to user experience with the genetic algorithm described here, there are two main metrics that must be considered: fitness of the final best solution, and time of simulation to generate the given solution. There are ten runtime parameters in Table 4.7 that have the ability to affect solution quality, while the simulation time depends heavily on the first three parameters. The user may adjust these numbers as they see fit in order to reach a high-quality solution in a reasonable amount of time, but it is not always immediately apparent what effect the changes will have.

a. Simulation time

Simulation time is highly variable. An estimate of the total simulation time can be taken from the ground up from several ballpark estimates. On average, the runtime per structure of an average structure in CC pairing environments will be 40-50 ms (DAO structure with CC bonds) on a standard PC. This number will vary widely based on the base parity, the number of node connections, and whether the model saves longitudinal fitness and population data (*savePopData, savePopFitness*). For a 64 bp (DAO) nanostructure of ~17,000 comparisons (41), this averages to ~3 µs *per comparison*. In a WC paring environment, this number will drop to ~2 µs, owing primarily to the symmetry of complement strand comparison (Operation B) in Figure 4.10, and the shortened vector subtraction in ruleMatrix (69). With 200 solutions over 500 generations, this simulation will take ~90 min in CC conditions and ~60 min in WC conditions. The speed of the model (t_{GA}) for comparison time t_{comp}, number of comparisons N (41), and runtime parameters *generations, popNum, popSize* and *numberOfElites* will approximately behave in the following way:

$$t_{GA} = (t_{comp})(N)(generations)(popNum)(popSize + numberOfElites) \qquad (72)$$

With this estimate, the speed and performance of the genetic algorithm can therefore be tailored to the needs of the user.

b. Final solution fitness

The fitness of the final solution will be influenced heavily by the type of competition and population dynamics present in the simulation. We investigate several parameters and their effects on final solution fitness through statistical analysis of repetitive tests using similar setup states.

b.i. Mutation rate

A population which is subject to low mutation rate may receive beneficial traits at a rate too slow to be useful to the user, while a population subject to elevated mutation rates may not be able to retain these traits after developing them in the first place. We perform statistical testing of the genetic algorithm for a DAO variant nanostructure, recording fitness and similarity statistics for ten simulations with each runtime setup (Figure 4.23). Three mutation rates were utilized, 2%, 5% and 10%. Solutions produced by both 2% and 10% mutation rates were both less diverse and less fit than solutions produced in the 5% condition. The fitness of 5% mutation falls at the lower extreme of the error bars for the other conditions, but the solutions are consistently clustered in this higher-fitness region. Results from the 2% and 10% conditions were much more variable in final result, but more internally similar. This suggests the mutation rates that are both too low and too high can lead to fitness niche creation, preventing more optimal solution from being found. It is evident that the computational model should therefore be run with a mutation rate of around 5%.

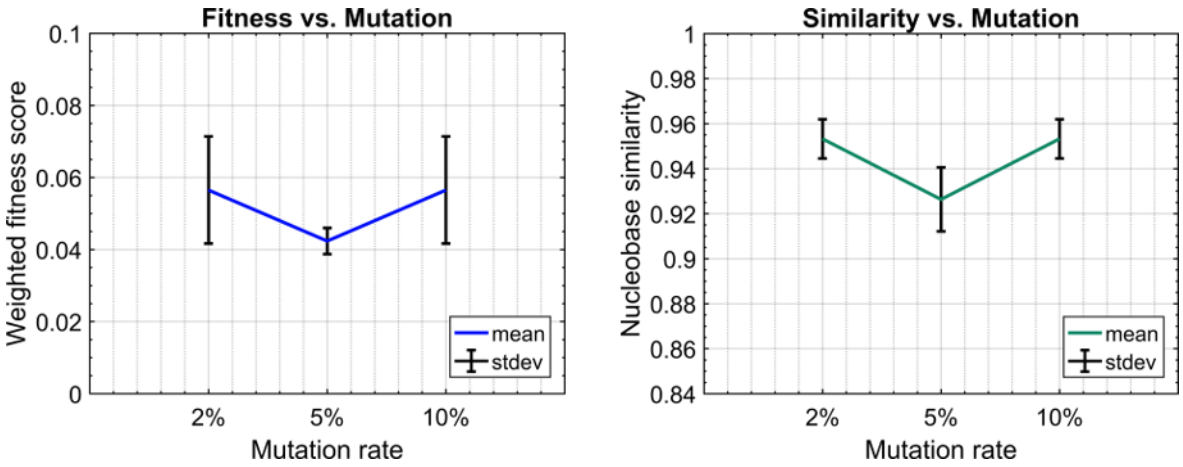


Figure 4.23: Mutation rate over iterative testing for DAO nanostructure variant at three separate mutation rates: 2%, 5% and 10% (low scores represent better solutions). All other runtime parameters were set to default values in Table 4.7. **A)** *F2pop* best fitness scores for and accompanying error bars for ten separate instances of running the genetic algorithm. Better solutions for 5% mutation, based on fitness function (71). **B)** Similarity calculations for the final generation of *F2pops* based on (73) and accompanying error bars show that a medial mutation rate actually increases population diversity compared with high and low rates.

b.ii. Population number and elitism

A similar analysis was performed to extrapolate the relationship between population size and population number (Figure 4.24). Simulations were performed for 200 solutions, segregated into varying numbers of populations (2,3,5, and 10). By definition, the number of elite solutions in each generation will have an effect independent of actual population size, and in order to decouple the effect of elite fraction in differing population scales, the experiments were repeated with both exactly 5 elites and 10% elites (rounded to integer values). Nanostructures with 10% elitism experienced much less change in fitness for differing population size. By contrast, these populations became much less internally similar as population sizes decreased (increasing population number). Simulations using 5 elite nanostructures experienced greater diversity as population size dropped, but this value became

asymptotic as the number of elites occupied 25% of the total population. A sharp increase in overall fitness (more negative fitness score) was exhibited at population numbers greater than three. For both conditions, 5 populations outperformed data collected from 2, 3 or 10 populations, with some question of statistical significance. As a general starting point, the model can be run with 5 populations with 10% elitism for best end fitness and overall solution diversity. Increased simulation times (generation number) may demonstrate better results in these conditions due to diversity retention and relatively smaller fitness niches in these conditions.

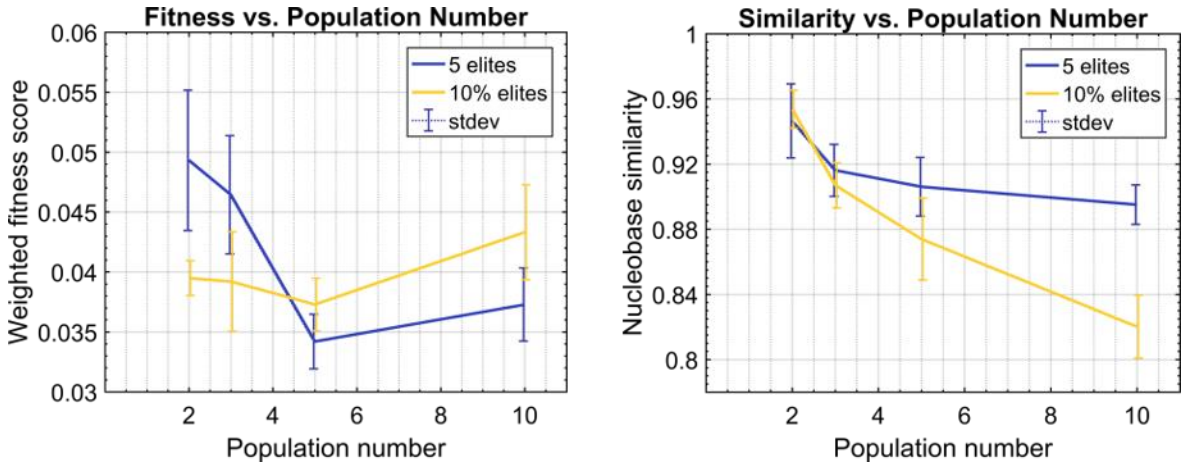


Figure 4.24: Population size and number with varying rates of elitism for a DAO variant nanostructure. Each data point represents the average of ten independent simulations. Each simulation had 5% mutation and 200 total solutions divided into the number of populations shown on the x-axis. All other runtime parameters were set to default values in Table 4.7. To decouple the effects of elitism from population size, the data were collected in two separate cases, one where there were 5 elites in every generation, and again where 10% of the next generation was an unaltered elite. Standard deviation was calculated for each data point (n=10). **A)** Average fitness scores (71) in the best solution for each of the ten independent *F2pops*. **B)** Average similarity scores for terminal generation *F2pops* (73).

Chapter 5

Silver(I) nanotechnology:

DNA nanostructures with Ag^+-mediated, conductive base pairing for self-assembling electronic arrays

"We select a possibility and we walk until we reach it. So, in a sense, we create it.

Let's leave it at that for now."

Roger Zelazny, The Guns of Avalon

5.1. Design of DNA nanostructures

DNA nanotechnology, and in turn DNA nanostructures, began with the idea of using small DNA units to assemble large, crystallographically-symmetrical lattices to serve as scaffolds for trapping biomolecules.[38,164] From the beginning, it was realized that the regularity and self-assembly of these lattices could be used to order matter on extremely small scales, with base-pair resolution of angstroms (axially) and nanometers (radially), and Seeman conceived of arranging electronic components on DNA scaffolds. The first experiments were conducted with tiles as small as possible, owing to the synthesis costs of DNA oligomers in the late 1980s and early 1990s. These early studies were approached with mathematical rigor that took into account the precise rotation, periodicity, and angular stress experienced by linear and branched DNA. Indeed, it was the organic-chemical-mindset that led to early successes in the field—certain rules were established and motifs discovered as researchers began to assemble a library of small tiles that could form an ever-increasing variety of nanoscale designs. The focus on geometry, topology, and organized assembly rules led this line of research to be called '*Structural DNA Nanotechnology*,' differing from diverging fields, such as: DNA origami and its descendants,[165] focused on taking large strands and imposing order through kinetic traps applied through short oligomeric DNA staples; DNA computing, which typically involves mechanical and biochemical computation using kinetic state changes over second- or millisecond-timescales;[166] RNA nanotechnology,[116] which typically focuses on the catalytic properties of RNA oligomer secondary structures; and more abstract, hybrid technological fields dedicated to molecular tweezer development,[167] gene silencers,[168] and even plasmonic device fabrication through integration of DNA into nanoscale waveguides.[169]

Whereas structural DNA nanotechnology has seen great successes, developing lattices and nanomaterials with linear,[170] parallel,[171] square,[172] hexagonal,[173] tetrahedral,[174] and even more abstract polyhedral geometries,[175] there has been significant delay in attaining any meaningful electrically-active

nanomaterials, owing primarily to DNA's *structural*, but not electrically-conductive nature. Indeed, nucleic acids demonstrate a wide variety of behaviors, forming enzyme-like species,[176] encoding information, forming natural crossovers, limited self-replication and transposon-based self-motility, but only one study has convincingly demonstrated the electrical usefulness of a naturally-occurring DNA variant, namely, guanine tetraplexes.[30] Many studies have worked to integrate nanomaterials into DNA lattices, either through: post-annealing nanoparticle or nanotube attachment;[22,177] post-assembly metalation[178] or use as a lithographic mask;[179] trapping of semiconductors in polyhedral cages;[45] redox activation of backbone free electrons;[180,181] or even fluorophore-based plasmon transfer;[169] but a robust system of electrical integration or nanowire assembly has yet to be established.

The promising G4 candidate may become a viable route to semiconductor self-assembly, but synthesis constraints of long guanine tetraplexes inhibit large-scale lattice development at this date. Metalation or ashing of DNA lattices takes advantage of the structural properties of DNA, but does not use its site-specific self-assembly to attain diverse nanomaterials with sub-10 nm resolution, which is the true promise of structural DNA nanotechnology. The introduction of a conductive metal base pair[62,94] offers the promise of atomic rearrangement within the double helix to forge conductive pathways into polynucleotides, but changes in structural chemistry and resulting thermodynamics make it difficult to predict the two measures of early DNA nanotechnology success: exact helical structure and kinetics of assembly. To allow for more flexible work with the rotational dynamics of DNA oligos and more precise control of kinetic traps (heterostructures), it was decided in this work that the assembly of 2D materials with embedded ions would begin with three minimal-nucleotide tile types: the single crossover (SX) Holliday junction (HJ) tile,[122] the T-junction (TJ) tile,[182] and the double crossover (DX) tile[148] (Figure 5.1). Each tile was chosen for its various structural properties: HJ and DX tiles are quite rigid, while TJ tiles are flexible and involve kissing loops. TJ and DX units involve short strands, while the HJ unit uses longer oligonucleotides. HJ and DX tiles form rigid, rectangular lattices, while TJ tiles can form a variety of

quadrilateral and circular structures. Each tile was tested with its literature-derived (canonical) analog, while small changes were made to force the necessity chelation of silver ions into critical lattice points. Experimental yield is discussed in terms of *design fidelity*, or adherence of the experimentally-derived nanostructure to its designed nanoscale morphology. We analyze the arrangement of individual tiles into their resulting lattices by size and shape of the resulting windows/bundles. We subsequently analyze the resulting arrangements of these tiles beyond the single-unit level. For structures of fixed size, an analysis can be quantitatively carried out by comparing micrographs of the fully polymerized structure to its designed shape; whereas lattices with unbounded growth are investigated for long-range order—grain size—before tearing or defect introduction. Over the course of the study, it became apparent that each tile has its advantages and drawbacks, and a number of modifications may be made on the promising results obtained from these basic units. Suggestions for future study and application are made with each type of nanostructure.

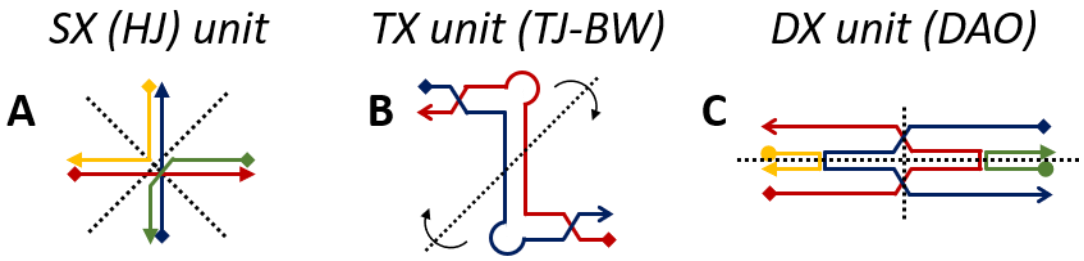


Figure 5.1: Designs of DNA nanostructure tiles used in this study. **A)** Single crossover (SX) tiles based on Holliday junctions involves branching between two vertically offset duplexes. At least four oligonucleotide segments participate in this junction. Tiles based on this junction generally contain four iterations of the SX unit. **B)** T-junction tiles based on sticky end/kissing loop junctions in planar units with two internal 90^{o} hairpin turns. Lateral strand length and SE/KL complementarity determines the type of structure that can be assembled. **C)** Double crossover (DX) tiles involve two adjacent (parallel or

antiparallel) duplexes that contain two consecutive, collapsed SX units. Parallel tiles can bind to themselves, while antiparallel tiles require SE-modified analogs to create an alternating 'AB' lattice.

5.2. Single crossover (SX) units: Holliday junctions and variants

5.2.1. General design

The most basic branching unit in DNA nanotechnology is the single crossover, whose structure mimics that of the Holliday junction in a meiotic dsDNA crossover event. The tile itself consists of a cross shape, where two perpendicular oligonucleotides are bound together (offset in the Z axis) by two bent oligos (Figure 5.2). The tile has two isomers, which can interchange freely in a monomeric state. The tile envisioned by Ned Seeman that began the DNA nanotechnology revolution involves the strand linking of four SX tiles into a single monomeric unit (Figure 5.2C).[122] While the SX unit is floppy and has isomeric structures, the full Holliday junction (HJ) DNA nanostructure is built from longer, weaving strands that force a single, rigid confirmation of crossovers at distances large enough to be visualized via AFM (14.2 nm windows), but small enough to be well below the persistence length of DNA (~35 nm**)**.[183] As a result, the HJ tile has been a very successful test case in the field of DNA nanotechnology. Here, we present several variations to this tile in order to incorporate dC:Ag^{+}:dC and other nanowire types into the lattices. As a template, the HJ tile is very amenable to alterations and serves as an excellent basis for stepwise alteration through chemical and topological modifications.

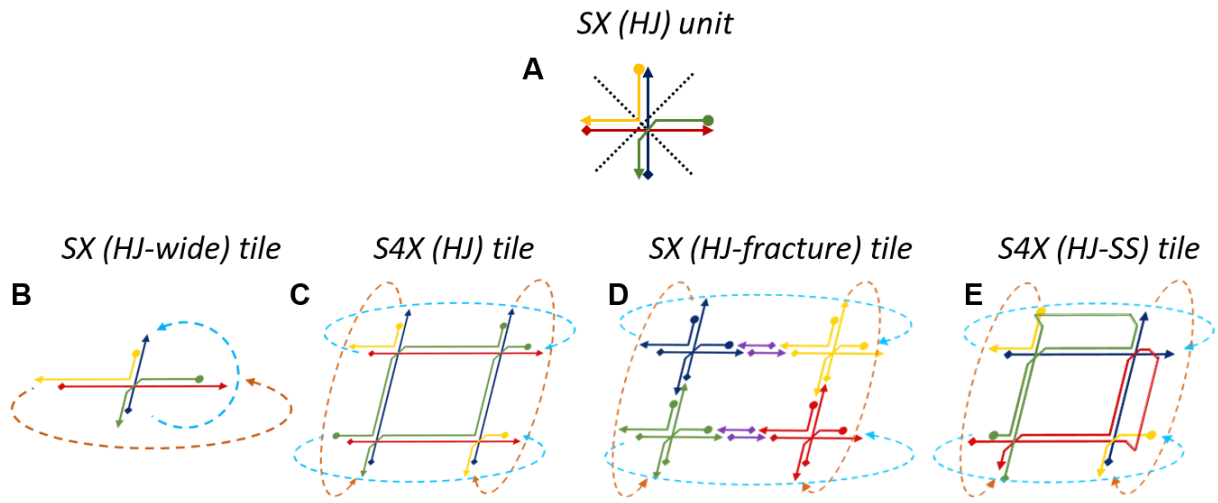


Figure 5.2: Types of single-crossover (SX), or Holliday junction (HJ), tiles. Colorful dotted lines show sticky end connectivity. **A)** The Holliday junction, whose geometry has been used to build a variety of tiles in the literature and in this manuscript. **B)** The Holliday junction tile can be functionalized with sticky ends to polymerize with itself, but the design is not rigid for long-range order. The standard minimal tile uses symmetrical arms of 1-2 helical turns (10-11 or 21 bp), but can be envisioned in an asymmetric design. **C)** The standard DNA nanostructure, built by Ned Seeman and colleagues, links four single crossover units in a rigid assembly using four straight oligo beams, two short L-shaped connectors, and two long, W-shaped winding strands.[122] D) The standard lattice can be cut into its components and rejoined using a variety of linker sequences with orthogonal chemistries. **E)** The standard lattice can also be modified to include fusion strands, connecting strands of opposite orientation to enable functionalization of a stacked layer through secondary annealing.

5.2.2. Original Seeman HJ Lattice

As envisioned by Ned Seeman, the HJ lattice is a rigid bundle of four single-crossover units. The structure utilizes eight oligonucleotides, four perpendicular 6-turn oligos that make up the frame, two short sticky-end-functionalized, L-shaped, 2-turn oligos, and two long, W-shaped 10-turn oligos that serve to weave the structure together (Figure 5.3A**,** Table 5.1).

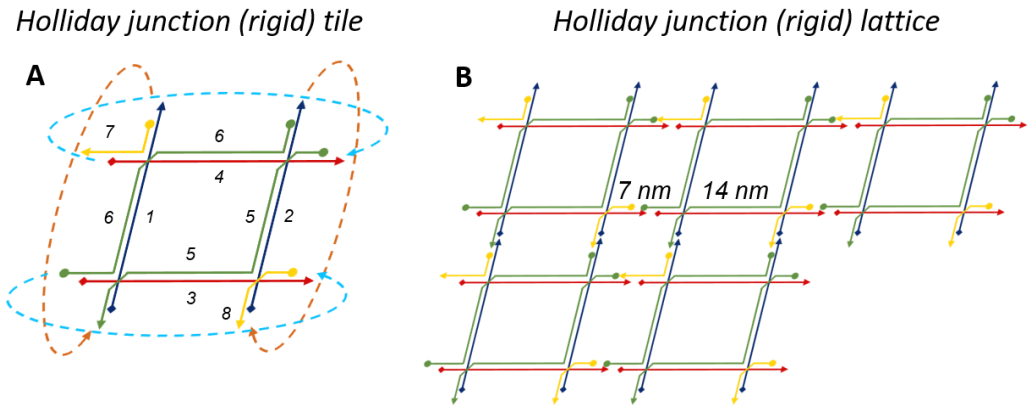


Figure 5.3: HJ original lattice design. **A)** The individual tile is shown with its sticky end connectivity (dotted lines) and sequence numbers (inset). **B)** When polymerized, the lattice possesses roughly rhomboid windows of 4x4 turns (14.2 nm) and 2x2 turns (7.1 nm) that are visible by AFM.

Table 5.1: HJ original sequences (sticky ends bold)[122]

Type	Sequence #	Nucleotide sequence
Beams	1	5'-**GTATG**-CTGATAGGACAATGAGTAGCTATTGGTGATCAACGTTAAGATACCAGTGGACGAATCG-3'
	2	5'-CAGTATGGACGTAGATACTGTGCTAACGATATTCGAACTAGCGTCATCGGACGATCAG-**AGACG**-3'
	3	5'-**CATTG**-GTAGTGCCTGTAATAATGTTGACTGCGGTTACCGTACTAATTGCTGTACCTGAGTGAG-3'
	4	5'-TGACAGCCTGTCGAGTAGATCGTATGAATAGATGGCATCGCTGTAAATCCTGTGTCAC-**CTCAC**-3'
W-shape	5	5'-GTGACACACCGATGACGCTAGTTCGAATATCGTTAGCACAGTATCTACGTGGTACAGCAATTAGTACGGTAACCGCAGTCAACATTATTACACCTATCAG-3'
	6	5'-CTGATCGTGGATTTACAGCGATGCCATCTATTCATACGATCTACTCGACACCACTGGTATCTTAACGTTGATCACCAATAGCTACTCATTGTGGCACTAC-3'

L-shape	7	5'-**CAATG**-CTCACTCACCATACTG-**CGTCT**-3'
	8	5'-**CATAC**-CGATTCGTGGCTGTCA-**GTGAG**-3'

We synthesized and imaged HJ lattices using these archival sequences, annealing 0.2 nmol each oligo together in 100 μL 1x MOPS (100 mM $NaNO_3$, 12.5 mM $MgSO_4$ and 10 mM MOPS), cooling the reaction in a water bath from 95 ^oC to RT over 48-72 hr, following the original reported annealing protocol.[122] The lattices were built using no sticky ends; sticky ends along one axis (1D lattices); and all sets of sticky ends (2D lattices) (Figure 5.4). Over a variety of experiments, polymerization was visualized as predicted; and it could be controlled to grow lattices in one or two dimensions over many hundreds of nanometers (1D tiles) or several microns (2D tiles). The yield, in terms of *design fidelity*, was reasonable: many lattices were formed with the appropriate periodicity. In addition to lattices, there were also aggregates of DNA which originated as lattices and tore on deposition (see Figure 5.4B,E, top of images). In terms of scale, lattices of the appropriate shape were observed to grow to lengths of many microns and widths of several hundreds of nanometers before the introduction of defects, or grain boundaries. The yield, in these respects, was somewhat lower than observed in the literature, as grain boundaries in the original manuscript extended into the micron scale in two dimensions, not one.[122] The differences in the experimental results here (and in subsequent tiles) can be attributed to differences in experimental setup and AFM.

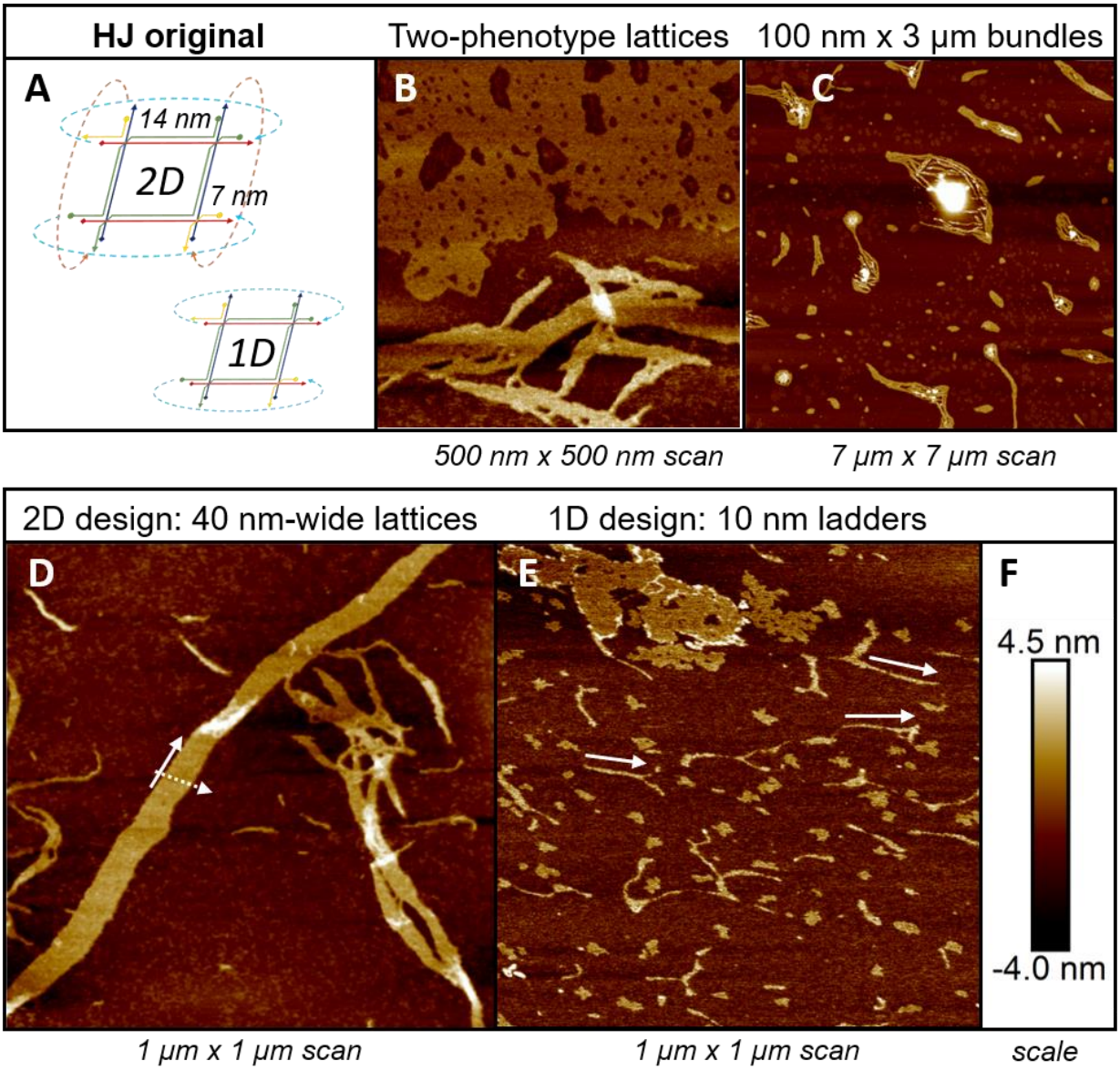


Figure 5.4: HJ original lattices. **A)** Design of the lattice using 1D and 2D designs, where sticky ends are selectively included to orient and constrain growth along one or two axes. **B)** 2D tiles show lattice assembly in agreement with the predicted shape, as well as a bundled mass of ssDNA that repeats across other HJ lattices. **C)** Over long range, the lattices can be seen to coagulate and form long spindles. **D)** Each spindle can grow to be ~40 nm wide and over 1 μm long, with lattice windows beginning to be visible. **E)** When designed for 1D growth, short ladder-like structures are attained, matching the

expected 10 nm width and growing between 80-200 nm in length. Overall, the lattice behaves as predicted and as reported by others. **F)** Scale bar for height sensor in AFM images shown.

5.2.3. HJ-fracture tiles with linkers

The Holliday junction structure, as first described by Mao, Sun and Seeman in 1999[122] consists of a tile with 4x4 helical turns in a center square, filled out by one helical turn extrusions from each vertex, with 5 bp sticky ends (see Figure 5.3). To introduce base pairing diversity into this structure, we cut the larger tile into its four component crosses. On the vertical lines, these cuts remained as 5 bp sticky ends, allowing hybridization. On the horizontal axes, the cuts were peeled back to create non-hybridizing sticky ends which could be filled with linkers. Each type of linker consisted of at least one double-stranded, internally complementary sequence, a non-canonical pairing feature, and the appropriate sticky ends for hybridization with the left and right halves of the tile. The total number of base pairs in the horizontal beams were corrected to multiples of 10.5 in order to preserve helical periodicity. With the introduction of additional base pairs on the horizontal axes, central windows are no longer rhomboid, but parallelogram in shape, creating a visual asymmetry that can be seen via AFM to determine the location, orientation, and fidelity of linker incorporation (Figure 5.5). To achieve this, 15 HJ lattice oligos (Table 5.2) and 20 linker oligos (Table 5.3) were utilized.

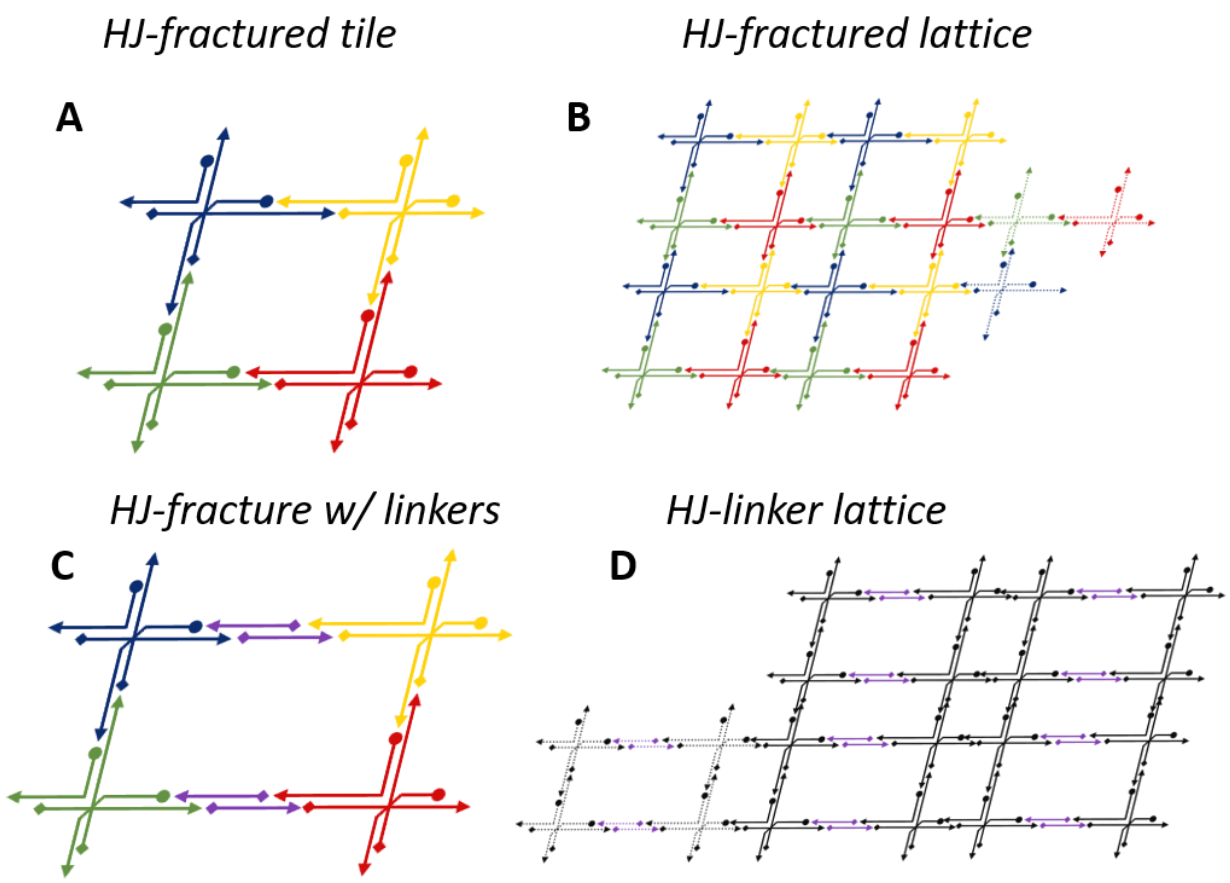


Figure 5.5: Design of the fractured Holliday junction lattice, based on original lattice by Mao, Winfree and Seeman (1999).[122] **A)** Sticky ends are created mid-edge by introducing 5bp cuts in the existing sequences used by Mao et al. **B)** A lattice can be assembled in piecewise fashion using 1-step or 2-step anneals, either adding all sequences together, or first annealing individual crosses and then mixing. Note that there exist three different windows, a large rhombus of dimension 14 x 14 nm, a small rhombus of size 7.1 x 7.1 nm, and parallelograms with dimension 7.1 x 14 nm. **C)** Linker chemistry is added to the horizontal arm by peeling back 12 base pairs in the center beams, creating two sticky ends of length 5 bp, and discarding terminal AT pairs. The linkers are held together with at least some region of WC, CC dsDNA, and may include stretches of ssDNA or G4 chemistry. Where appropriate, rotation is corrected to bundles of single helical turns (10.5 bp). **D)** Lattices created from these units are assembled in 3-step anneals, and are no longer topologically symmetrical. There exists one small rhombus of size 7.1 x 7.1

nm, parallelograms of size 14 x 7.1 nm, and a central parallelogram of larger size, depending on the linker length and stiffness. Some linkers will have differing size, owing to their mechanical properties.

Table 5.2: Holliday junction fractured lattice

Sequence Name	Attribute	Nucleotide sequence
HJ-1A	Split sequence (2x)	5'-GTATGCTGATAGGACAATGAGTAGCTATTGGTG-3'
HJ-1B	Split sequence (2x)	5'-TTAAGATACCAGTGGACGAATCG-3'
HJ-2A	Split sequence (2x)	5'-CAGTATGGACGTAGATACTGTGCTA-3'
HJ-2B	Split sequence (2x)	5'-TTCGAACTAGCGTCATCGGACGATCAGAGACG-3'
HJ-3A	Split sequence (2x)	5'-CATTGGTAGTGCCTGTAATAATGTTGACTGCGGTTA-3'
HJ-3B	Split sequence (2x)	5'-CCGTACTAATTGCTGTACCTGAGTGAG-3'
HJ-4A	Split sequence (2x)	5'-TGACAGCCTGTCGAGTAGATCGTATGAATAG-3'
HJ-4B	Split sequence (2x)	5'-ATGGCATCGCTGTAAATCCTGTGTCACCTCAC-3'
HJ-5A	Split sequence (3x)	5'-GTGACACACCGATGACGCTAGT-3'
HJ-5B	Split sequence (3x)	5'-ATCGTTAGCACAGTATCTACGTGGTACAGCAATTAGTACG GTAACC-3'
HJ-5C	Split sequence (3x)	5'-GCAGTCAACATTATTACACCTATCAG-3'
HJ-6A	Split sequence (3x)	5'-CTGATCGTGGATTTACAGCGATGCCATCTATT-3'
HJ-6B	Split sequence (3x)	5'-CATACGATCTACTCGACACCACTGGTATCTTAACGTTGA-3'
HJ-6C	Split sequence (3x)	5'-ATAGCTACTCATTGTGGCACTAC-3'
HJ-7	Original sequence	5'-CAATGCTCACTCACCATACTGCGTCT-3'
HJ-8	Original sequence	5'-CATACCGATTCGTGGCTGTCAGTGAG-3'

Table 5.3: Holliday junction fractured lattice linkers

Sequence Name	Attribute	Nucleotide sequence
L-A1	C11 linker, L-A2 complement	5'-ACGATCCCCCCCCCCC-3'
L-A2	C11 linker, L-A1 complement	5'-TCGAACCCCCCCCCCC-3'

L-A3	C11 linker, L-A4 complement	5'-TCAACCCCCCCCCCC-3'
L-A4	C11 linker, L-A3 complement	5'-ACCACCCCCCCCCCC-3'
L-C1	C32 linker, L-C2 complement	5'-ACGATCCCCCCCCCCCCCCCCCCCCCCCCCCCCCC-3'
L-C2	C32 linker, L-C1 complement	5'-TCGAACCCCCCCCCCCCCCCCCCCCCCCCCCCCCC-3'
L-C3	C32 linker, L-C4 complement	5'-TCAACCCCCCCCCCCCCCCCCCCCCCCCCCCCCC-3'
L-C4	C32 linker, L-C3 complement	5'-ACCACCCCCCCCCCCCCCCCCCCCCCCCCCCCCC-3'
L-D1	[G4-C_{11}-G4] linker, L-D2 complement	5'-ACGATTGGGTTAGGGTTAGGGTTAGGGTCCCCCCCCCCC-3'
L-D2	[G4-C_{11}-G4] linker, L-D1 complement	5'-TCGAATGGGTTAGGGTTAGGGTTAGGGTCCCCCCCCCCC-3'
L-D3	[G4-C_{11}-G4] linker, L-D4 complement	5'-TCAACTGGGTTAGGGTTAGGGTTAGGGTCCCCCCCCCCC-3'
L-D4	[G4-C_{11}-G4] linker, L-D3 complement	5'-ACCACTGGGTTAGGGTTAGGGTTAGGGTCCCCCCCCCCC-3'
L-E1	G-rich linker, L-E2 complement	5'-ACGGGGTTAGGGTTAGGGTGTAGCGATTGG-3'
L-E2	G-rich linker, L-E1 complement	5'-TCGAATGGGTTAGGGTTAGGGTACCAATCGCTAC-3'
L-E3	G-rich linker, L-E4 complement	5'-TCAACTGGGTTAGGGTTAGGGTGTAGCGATTGGT-3'
L-E4	G-rich linker, L-E3 complement	5'-ACCACTGGGTTAGGGTTAGGGTACCAATCGCTAC-3'
L-F1	T-rich linker, L-F2 complement	5'-ACGATTTTTTTTTTTTTTTTTTTTTTTGTAGCGATTGGT-3'
L-F2	T-rich linker, L-F1 complement	5'-TGGTATTTTTTTTTTTTTTTTTTTTTTTTTACTATTTT -3'

L-F3	T-rich linker, L-F4 complement	5'-ACTTCATTTTTTTTTTTTTTTTTTTTTGTAATTGCGGGT-3'
L-F4	T-rich linker, L-F3 complement	5'-ACCACTTTTTTTTTTTTTTTTTTTTTTACCAATCGCTAC-3'

To fully realize a fractured-tile lattice, each of the crosses in the lattice were labeled 1-4, going counterclockwise from the top left. Sub-tiles 1-4 and the linker components (A1,2-F3,4) were annealed over 48 hr from 95 ^{o}C to RTP in a water bath, with 1:1 Ag^{+} molarity where required, with 0.2 nmol each DNA strand in 10 mM MOPS, 100 mM $NaNO_3$ and 12.5 mM $MgSO_4$. Secondary anneals joined tiles [1]+[2]+[X1.2]+[X3,4] and [3]+[4] for linker X (A-F), cooling from 40 ^{o}C to RTP over 24 hr. Tertiary anneals combined the left and right halves of the lattice tiles [1,2,X1-2,X3-4] + [3,4], cooling from 40 ^{o}C over 24 hr.

a. G4-CC hybrid linkers

The most successful linking chemistry came from a G4-C_{11}-G4 spring linker, which utilized $[AGGGTT]_4$ repeats in order to create guanine tetraplex features of 3 nt in length (12 total dG bases). The complementary middle of the linker was bound with $[dC:Ag^{+}:dC]_{12}$, creating a lattice linking chemistry between WC topology, guanine tetraplexes (a known semiconductor)[30] and C_{12}, which we had previously shown to display enhanced conductance relative to its WC analog,[62] which may only polymerize when the linkers form ion-coordinating metal base pairs. Because of the addition of the linkers, the windows are no longer rhomboid, but rather form 4x7 helical turn parallelograms. Because the G4 units may be uncoiled under mechanical stress, the short side of the holes is estimated to be a precise 14.3 nm, while the long edge may be between 21.1 nm (tightly formed G4) and 33.3 nm (uncoiled ssDNA). The unraveling of linked G4 bundles would likely lead to the decay of any type of conductance across such a hybrid lattice—this may act as a confounding effect in a proposed active nanostructure, or it may present an opportunity for mechanical switching of electronic behavior to add or remove current

blockades. We can see in Figure 5.5 that this assembly was extremely successful, producing lattices with well-defined features and long-range order. Large, interwoven structures formed with widths between 80-200 nm (Figure 5.5 B,C) and lengths exceeding the AFM scan size. The images show the long and short axes of the lattice (Figure 5.5 E,F), and it can be clearly seen that where the lattice is stretched mechanically during deposition, the G4 sections of the linkers will expand to their uncoiled length, from 21 to 33 nm (Figure 5.5F). In the larger scans, holes can be seen at the very center of vertices between three or more different lattice axes, where tension exceeds the ability of the G$ springs to stretch (Figure 5.5C). In the largest image, the lattice holes can still be clearly seen in regular, window like repeats, again with larger windows near the vertices. An extreme close-up micrograph shows that the lattice windows are in good agreement (±1 nm) with the predicted lengths (Figure 5.5I).

This experiment is similar in yield to controls (Figure 5.4), with grain boundaries of microns in one axis and tens to hundreds of nanometers in the other. Tears in the lattice from deposition during imaging can account for the limitations in grain size and the web-like appearance of nanostructures in Figure 5.5C. In terms of design fidelity, the lattice periodicity and features behave exactly as designed, with a minority of material forming non-lattice or proto-lattice aggregates in the holes between grains. This type of yield can be considered quite high, and a softer deposition process might garner images of accurately-formed lattices with grain sizes extending many microns. This G4-C:Ag^{+}:C hybrid Holliday junction is a successful DNA nanostructure; and it is the first of its kind to incorporate multiple forms of orthogonal base chemistry into a single periodic lattice. As such, this design possesses great potential in both nanomechanical and nanoelectrical studies.

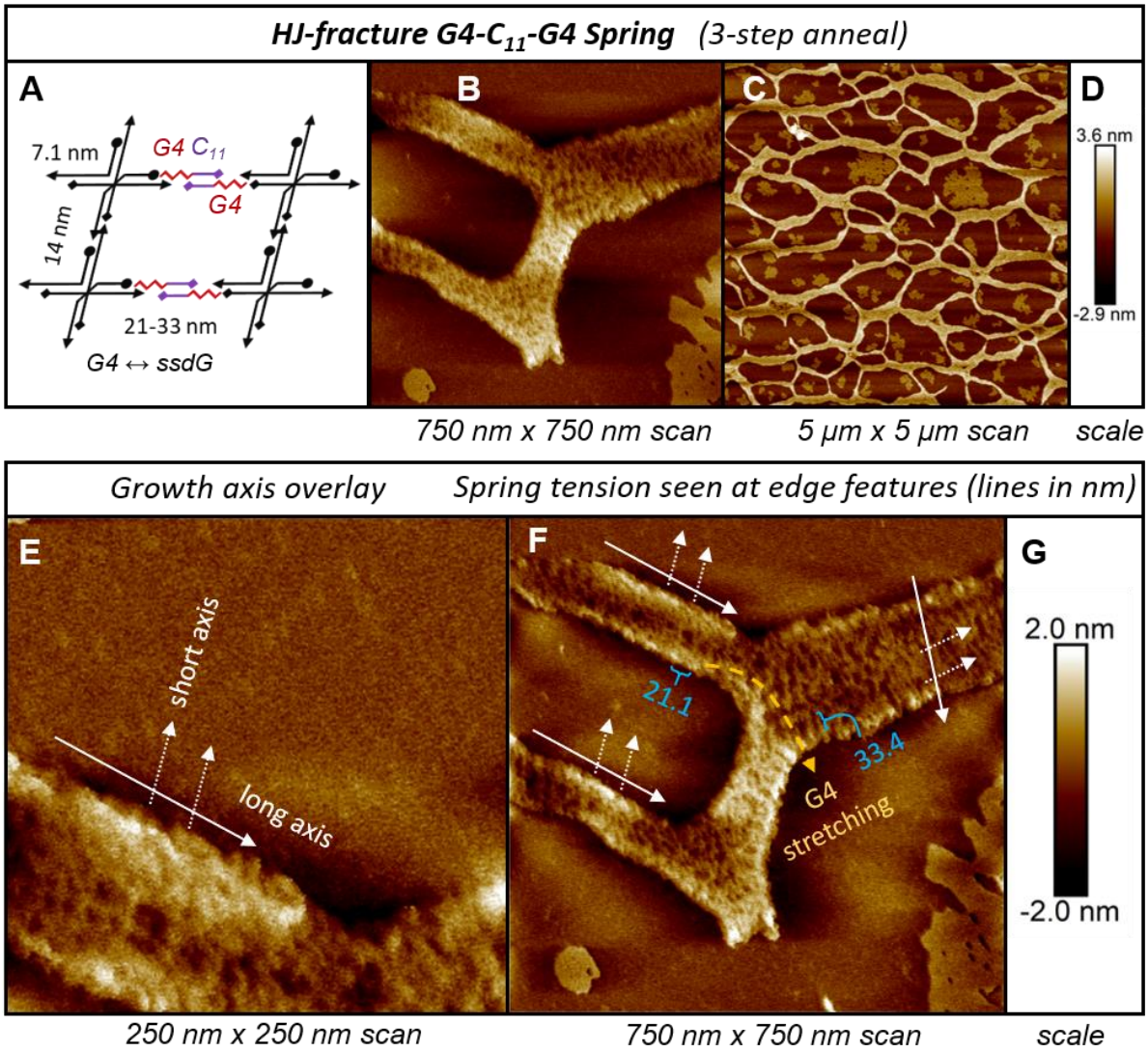

HJ-fracture G4-C11-G4 Spring (3-step anneal)
A
7.1 nm
G4 C11
G4
14 nm
21-33 nm
G4 ↔ ssdG
B
C
D
3.6 nm
-2.9 nm
750 nm x 750 nm scan
5 µm x 5 µm scan
scale
Growth axis overlay
Spring tension seen at edge features (lines in nm)
E
short axis
long axis
F
21.1
33.4
G4
stretching
G
2.0 nm
-2.0 nm
250 nm x 250 nm scan
750 nm x 750 nm scan
scale
Cropped zoom
Overlay lines (nm)
H
I
J
2.0 nm
-1.5 nm
18.7
14.3
6.6
22.7
12.6
22.1
8.1
7.2
8.3
8.3
7.3
21.7
15.4
14.8
20.7
23.8
7.0
7.8
9.2
21.5
75 nm x 75 nm crop
75 nm x 75 nm crop
scale

Figure 5.5: AFM analysis of HJ lattice with G4-C_{11}-G4 linkers. **A)** Predicted tile shape from a 3-step anneal, showing 21 x 14 nm large windows, 7.1 x 7.1 nm vertex windows, and their adjoining rectangles. **B)** Successful lattice with component windows visible for 750 x 750 nm micrograph. **C)** Long range order of the lattice with similar topology to B, noting that in vertices between interlinking lattices, G4 springs stretch to their 33 nm maximum. In some cases there are holes where mechanical forces during drying exceeded the strength of the sticky ends. **D)** scale bar for B,C. **E)** Close-up 250 nm x 250 nm micrograph shows lattice periodicity and growth axes, with the long axis depicted with a solid white arrow and the short axis with a dotted one (inset). **F)** Image from B shown again with a 6^{th} order flatten function to accentuate details. Growth axes are again marked with solid and dotted white arrows. Stretching of the long growth axis across a lattice vertex (orange line) is shown, and the spring-like unwinding of G4 increases window size from 21 nm to the predicted 33 nm with high fidelity (blue inset). **G)** scale bar for E,F. **H)** Zoomed-in reproduction of E shows parallelogram shape of large windows, square shape of the small windows, and the adjoining rectangles. **I)** Measurements of these edge lengths is overlaid (white lines, lengths in nm) and are in close agreement with the prediction in A for tightly coiled G4 in the linkers. **J)** scale bar for H,I. All images taken on a Bruker Dimension Icon AFM in a PeakForce QNM in Air mode using a 1 nm silicon nitride high-resolution tip. Scan parameters utilized low Z-piezo range (1-2 µm), PeakForce Engage Setpoint (50-80 mV) and were, unless otherwise noted, subjected to a 1^{st} order flatten function in Nanoscope Analysis v1.9.

b. Other linkers

Other linking chemistries were tested, with varying success. Many of the structures seemed promising, and additional trials or high-resolution AFM time could help promote these nascent designs. C_{11} linkers showed some lattice-like structure (Figure 5.6B-C), without the rigid lattice axes seen in the G4 hybrid. C_{32} linkers showed some promise in large lattice growth, with tearing in >5 µm lattices suggesting flat sheets in aqueous conditions (Figure 5.6F-G). The reaction in this assembly was tightly controlled to

minimize Ag^+ availability during linker assembly. We have shown that unitary Ag^+ promotes frame alignment in > 11 nt polycytosines, and it appears that this prevented overly pronounced stretching of the lattice in what would be an irreversible (or temporally unrealistic) relaxation of the C_{32} sliding. This structure shows promise for future functionalization and sequence optimization. Watson Crick linkers were tested as controls, and though the sequences appeared to have some incongruencies, the lattices were still able to assemble in a meaningful way, with clear growth axes and window periodicity (Figure 5.6J-K).

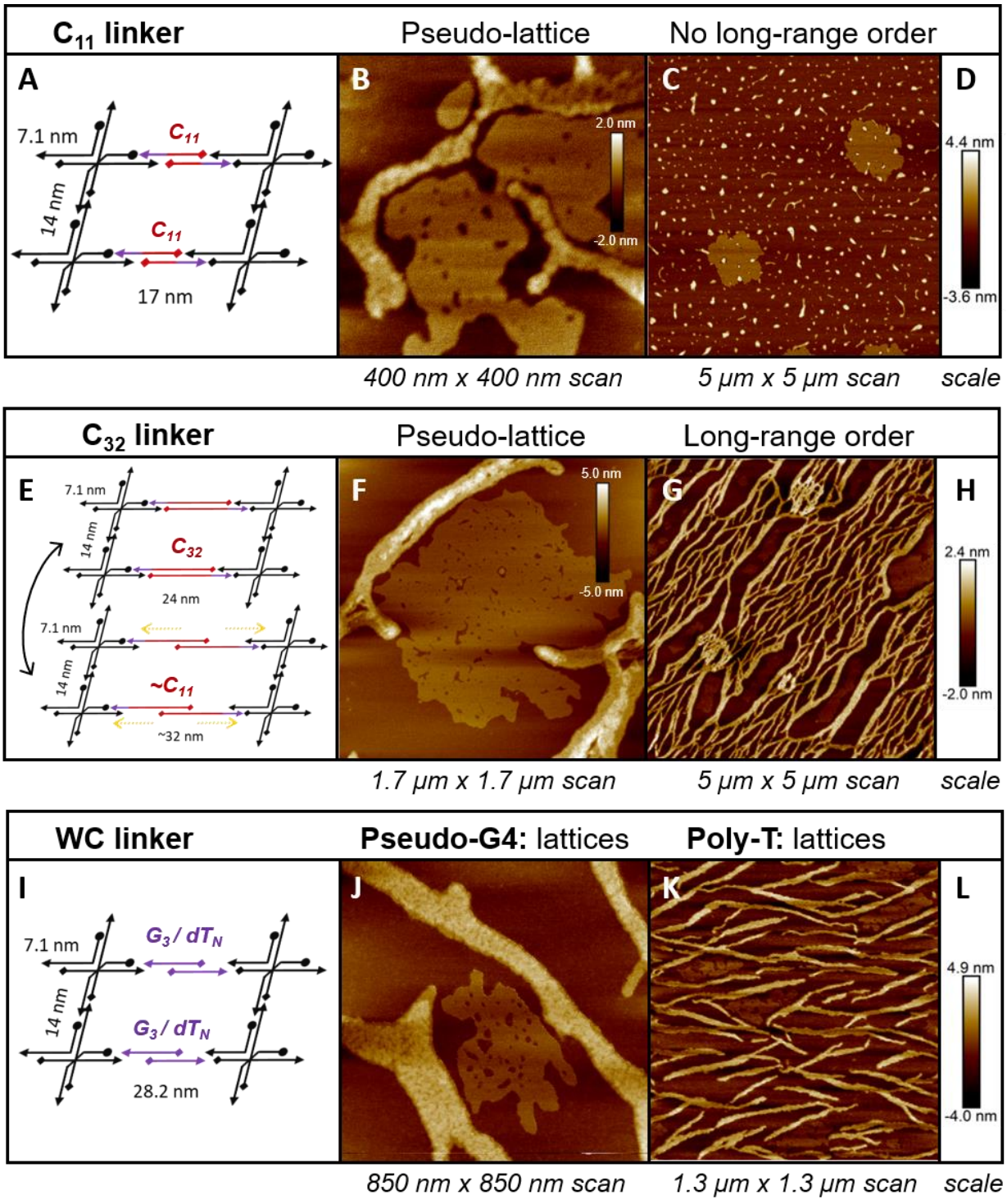


Figure 5.6: HJ fracture with linkers. **A)** Design and predicted lattice dimensions with an 11 bp polycytosine linker. **B)** Some pseudo-lattices are observed, but the images are predominated by disordered ssDNA (scale bar inset). **C)** This result is repeated over long range scans, with small, >500 nm

lattice-like shapes littered between supercoiled oligos. **D)** Scale bar for C. **E)** Design of a C_{32} linker, in which the known promiscuity of the C_{32} oligo makes asymmetrical, non-reversible stretching of the lattices possible. **F)** When annealed, ordered lattices are observed with a rounded profile, suggesting stacking of additional layers or inappropriately bound C_{30} molecules (scale inset). **G)** Over long range, lattices are oriented and linked across distances >5 µm, suggesting good lattice assembly. Frequent breaks in the short side of the lattices open holes that may be due to C30 stretching but are more likely the result of breakage during deposition, suggesting that many μm^2 sheets may be present in aqueous conditions. **H)** Height scale for G. **I)** Design of WC-pairing linkers and expected dimensions. J) Pseudo-G4 sequence *L-E1-4* forms lattices with good grain orientation, with an overall lattice width of approximately four tiles (~120 nm). **K)** PolyT linker *L-F1-4* shows long-range lattice assembly with significant background ssDNA detritus. **L)** Height scale for J,K.

The overall yield of these designs was somewhat lower than the controls in Figure 5.4 and the hybrid linkers in Figure 5.5. With the exception of C_{11}, the other linkers exhibited medium- (polyT, pseudo-G4) and long-range (C_{32}) grain development. It remains unclear if the nanoscale lattice features were exact replicates of the designs, but the uniformity of large lattice grains suggests a similar uniformity on the nanoscale, and indeed AFM images in Figure 5.6 hint at lattice windows of the appropriate scale (Figure 5.6). On the whole, the HJ fracture appeared to be quite successful, and this design may be amenable for functionalization by a wide variety of nanomaterials, including carbon nanotube-thiol compounds,[137] metal base pairs,[82] XNA,[184] nanoparticles,[22] and guanine tetraplexes.[30]

5.2.4. HJ superstrand tiles (HJ-SS)

a. Design considerations

The HJ "superstrand" tile is a hybrid design. It uses DNA nanotechnology as a substrate in the way that many studies suggest (such as an etch mask or waveguide);[169] however, this substrate is a direct

template for site-directed assembly of a secondary layer composed of orthogonal nanowires. To achieve this, a fusion of intermediate length arms in the HJ Seeman design (sequences 3 and 5: 3/5 fusion, as well as sequences 4 and 6: 4/6 fusion) are carried out using an ssDNA linker (Figure 5.7**,** Table 5.4). The sequence composition of this intra-strand linker allows for subsequent annealing with polycytosine oligos.

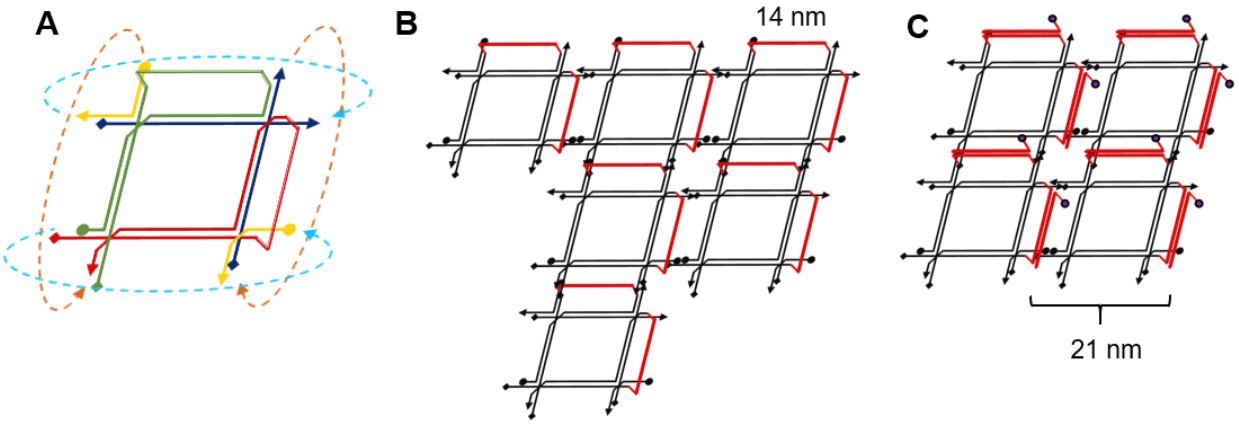


Figure 5.7: Lattice design using the HJ-SS tile. **A)** Monomeric tile showing sticky end connectivity; **B)** lattice shown with fusion strands extruded in Z (red); and **C)** lattices shown with the C30-FAM linker attached after secondary annealing. Lattices have 42 bp, 14.2 nm large windows and 21 bp, 7.1 nm small windows.

Table 5.4: HJ-SS fusion strands (sticky end bold)

Type	Design components	Nucleotide sequence
3/5 fusion	(HJ-3)(A_6)(C_{30})(A_6)(HJ-5)	5'- GTAGTGCCTGTAATAATGTTGACTGCGGTTACCGTACTAATTGCTGTACCTGAGTGAGAAAAAACCCCCCCCCCCCCCCCCCCCCCCCCCCCCCAAAAAAGTGACACACCGATGACGCTAGTTCGAATATCGTTAGCACAGTATCTACGTGGTACAGCAATTAGTACGGTAACCGCAGTCAACATTATTACACCTATCAG-3'
6/4 fusion	(HJ-6)(A_6)(C_{30})(A_6)(HJ-4)	5'-CTGATCGTGGATTTACAGCGATGCCATCTATTCATACGATCTACTCGACACCACTGGTATCTTAACGTTGATCACCAATAGCTACTCATTGTGGCACT

		ACAAAAAACCCCCCCCCCCCCCCCCCCCCCCCCCCCAAAAAATGACAG CCTGTCGAGTAGATCGTATGAATAGATGGCATCGCTGTAAATCCTGTGT CAC-3′
3/5 <u>fracture</u> w/ SE (A)	(**SE**)(HJ-3)(T_3)(A_6)(C_{30})(A_6)(T_3)(HJ-5A)	5′-**CATTG**GTAGTGCCTGTAATAATGTTGACTGCGGTTACCGTACTAATT GCTGTACCTGAGTGAGTTCAAAAAACCCCCCCCCCCCCCCCCCCCCCC CCCCCAAAAAATTTGTGACACA CCGATGACGCTAGTTCGAATA-3′
3/5 <u>fracture</u> w/ SE (B)	(HJ-5B)	5′-TCGTTAGCACAGTATCTACGTGGTACAGCAATTAGTACGGTAACCGC AGTCAACATTATTACACCTATCAG-3′

The fusion strands were approached from two perspectives: first using a single, 200 nt fusion strand ordered as an Ultramer® oligo from IDT DNA; and second as a series of sticky-end coordinated pieces to achieve the two-layer design. The first approach benefits from continuity—the superstrand component is a single unit, simplifying the kinetics of assembly. With fewer degrees of freedom and fewer distinct oligonucleotides, the design benefits from tighter reaction control and increased likelihood of single tile assembly. The fundamental drawback to this design comes from synthesis constraints: IDT DNA, Elim Biopharmaceuticals, and other commonly used synthesis companies have an oligomeric length cap of 200 nt. To achieve a true fusion strand, the sticky ends at the end of each sequence are clipped (sticky end inversion would produce an improper 5′-5′ junction). As a result, each fusion generates a loss of two sticky ends. The original HJ lattice has eight edge connections; the 3/5 fusion design has six edge connections; while the 3/5-4/6 fusion design benefits from only four connections. As a result, long range order may suffer.

The second approach instead retains the sticky end, while introducing a nick in the HJ-5 strand similar to what was carried out in the HJ fracture tile in *Section 5.2.3*. The polymerization of this tile benefits from eight functional sticky ends at the edges of the design, but introduces a ninth toehold directly beneath the superstrand component. The differences between these designs can be seen in Figure 5.8.

Experimental performance of these two approaches are compared both with and without their C30 linkers; and assembly of single tiles, 1D chains, and 2D lattices is assessed below.

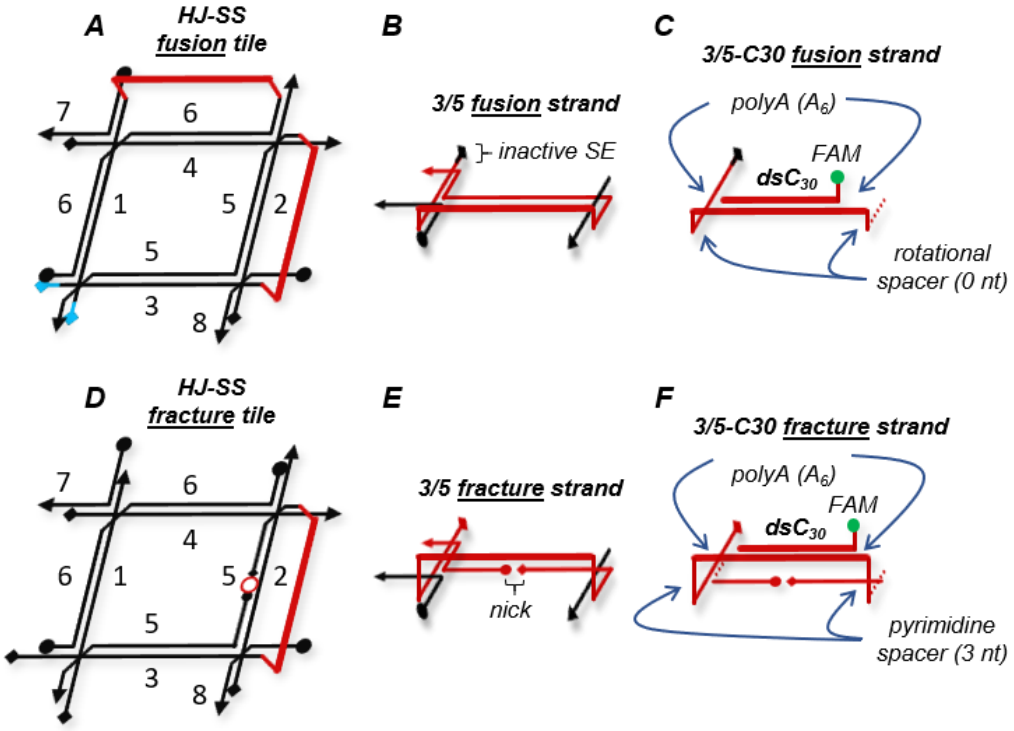


Figure 5.8: Two types of HJ-SS tiles are built, an HJ-SS fusion and an HJ-SS fracture tile. **A)** The HJ-SS fusion tile seen from above has two visible fusion strands (3/5 and 4/6, both red). Each fusion strand forces the inactivation of a terminal sticky end (blue) due to sequence length constriction. The strand numbers in the original Holliday junction design are shown for reference. **B)** When seen laterally (the Z axis is up), the fusion strand (red) can be seen winding around the lattice, in some parts annealing to itself. **C)** The annealing of C30 onto the fusion strand occurs in a high-tension area, as a result of design and synthesis constraints. There is no spacing between the two layers by design, but the A_6 caps to the superstrand region relieve some of the tension and allow for rotational freedom in the C30 duplex. The C30 sequence was synthesized with 3'-FAM (fluorescein) fluorophore to determine if association with the lattice was specific and discrete. The whole superstrand region is 42 nt long, or 14.3 nm, while the

C30 linker is 30 nt, or 10.2 nm, occupying just under three of the four possible helical rotations in the region. **D)** The HJ-SS fracture tile retains 8 functional sticky ends, but introduces a nick in strand 5 (red circle) to reduce the maximum oligo length to below 200 nt. The fracture strand is thus the coordination of two strands, HJ-3/5A and HJ-3/5B (see Table 5.4) bound 21 nt into the central beam of HJ-2. **E)** The nick occurs directly beneath the superstrand region. **F)** The relaxation of length constraints allows for two 3 nt pyrimidine (dT) spacers in the superstrand region to allow a vertical offset of the C30 duplex when bound. This is expected to reduce tension along the nanowire region and allow for greater rotational freedom.

b. Control experiments: lattices without nanowires

The HJ-SS substrate was tested before addition of any nanowire components: annealing was performed without the addition of the C30-FAM linker. In this experiment, the products were analyzed by AFM (Figure 5.9). Single, 14 nm windows were attained when fracture tiles (Figure 5.8D) were annealed without sticky ends (Figure 5.9D), though the windows were frequently opened (i.e. interrupted frame), irregularly-shaped, or trailing their longest oligomers. At this scale, it is clear that tightly-controlled windows are attainable, but they are not the predominant product of annealing. Lattices of a scale between 100-800 nm were attained when fusion tiles (Figure 5.8A) were annealed without the attachment linker (Figure 5.9E,F; Figure 5.7B). In this case, the experiment was carried out with only one fusion strand (3/5) to investigate assembly with six sticky ends. Unlike in the 0D case, the sticky-end-modified sequences were able to produce windows of the appropriate size with good regularity, and differential layers from mica deposition begin to become apparent. It is likely that the planar extrusions provided by the superstrands facilitate stacking, and perhaps growth, of lattices in the Z direction. Unlike AFM performed on other structures, most of the background noise in HJ-SS images consists of 14 nm squares, suggesting that the layers may be several iterations deep (see Figure 5.9E). All of the fusion

strands are composed of ss-polyC, and it is unclear whether it is the interactions between these sequences from acidification during drying that produces this stacking behavior, or whether the lattices are simply growing in an irregular fashion due to the complexity of the tile.

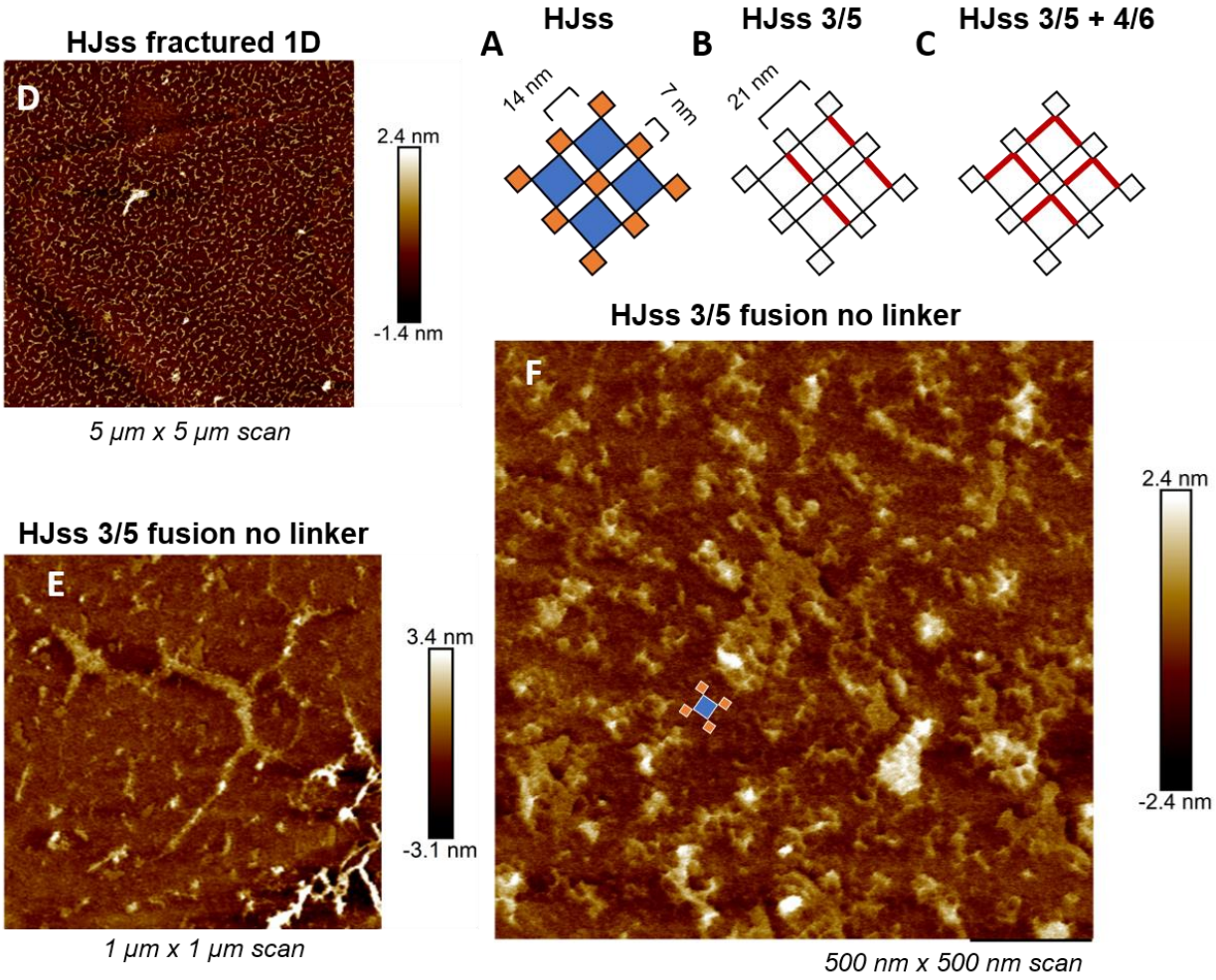


Figure 5.9: Control experiments for the Holliday junction superstrand tile. **A)** The underlying lattice pattern of the HJ-SS design mimics the original HJ tiles. Here four tiles are shown in tight association. There are windows of 14.2 nm in size (42 bp, blue) and 7.1 nm (21 bp, orange) **B)** The superstrand arms of the tiles are extruded above the lattice by steric rotation with no spacers (fusion tile) or by three nucleotides (fractured design) (Figure 5.8). This superstrand (red) sits above the tile and contains a 30 nt (10.2 nm) polycytosine as a secondary annealing site for functionalization. Here, the fusion of oligos HJ-3

and HJ-5 is shown, producing a regular nanowire adhesion site every 63 bp, or 21.3 nm. **C)** Design with two fusion strands, 3/5 and 4/6 is shown. This produces corners that may allow 'turning' of nanowires above the HJ substrate with similar regularity. **D)** HJ SS 3/5 fracture tile assembled without sticky ends. Windows are roughly rectangular in shape, and are frequently broken during deposition or partially unwound. This suggests that the tension along the superstrands makes the tile prone to disassembly in the absence of sticky end anchoring. **E)** When the 3/5 fusion tile is annealed with six sticky ends, regular lattices with windows are formed in the HJ fusion tile that resemble the standard HJ shape. **F)** Closer imaging of the lattices shows good agreement with the predicted feature sizes, and further shows multiple layers of lattices growing or stacking atop one another with identical orientation.

*c. HJ-SS **fusion** with duplex nanowires*

HJ-SS fusion tiles were investigated for duplex nanowire incorporation. A secondary anneal was performed by adding equimolar C30 and unitary Ag^+ to the lattices, followed by cooling from 40 ^oC over 48 hr. The stoichiometry of the reaction is difficult to control, as C30 will also form a regular nanowire with itself over long range in the presence of ionic excess (see Figure 2.2D).[91] It is expected that there will be insufficient C30 to achieve nanowire connectivity across all the lattices, and this may be corrected in the future by changing the sequence design to incorporate adenines (two-ring, wide nucleobases) as steric inhibitors at key sites. In general, the reaction was designed to starve C30 of ions in order to discourage self-annealing. Occupancy of the double-stranded fusion strand assembly appears to be < 40% from nanomechanical maps of the lattices (Figure 5.10).

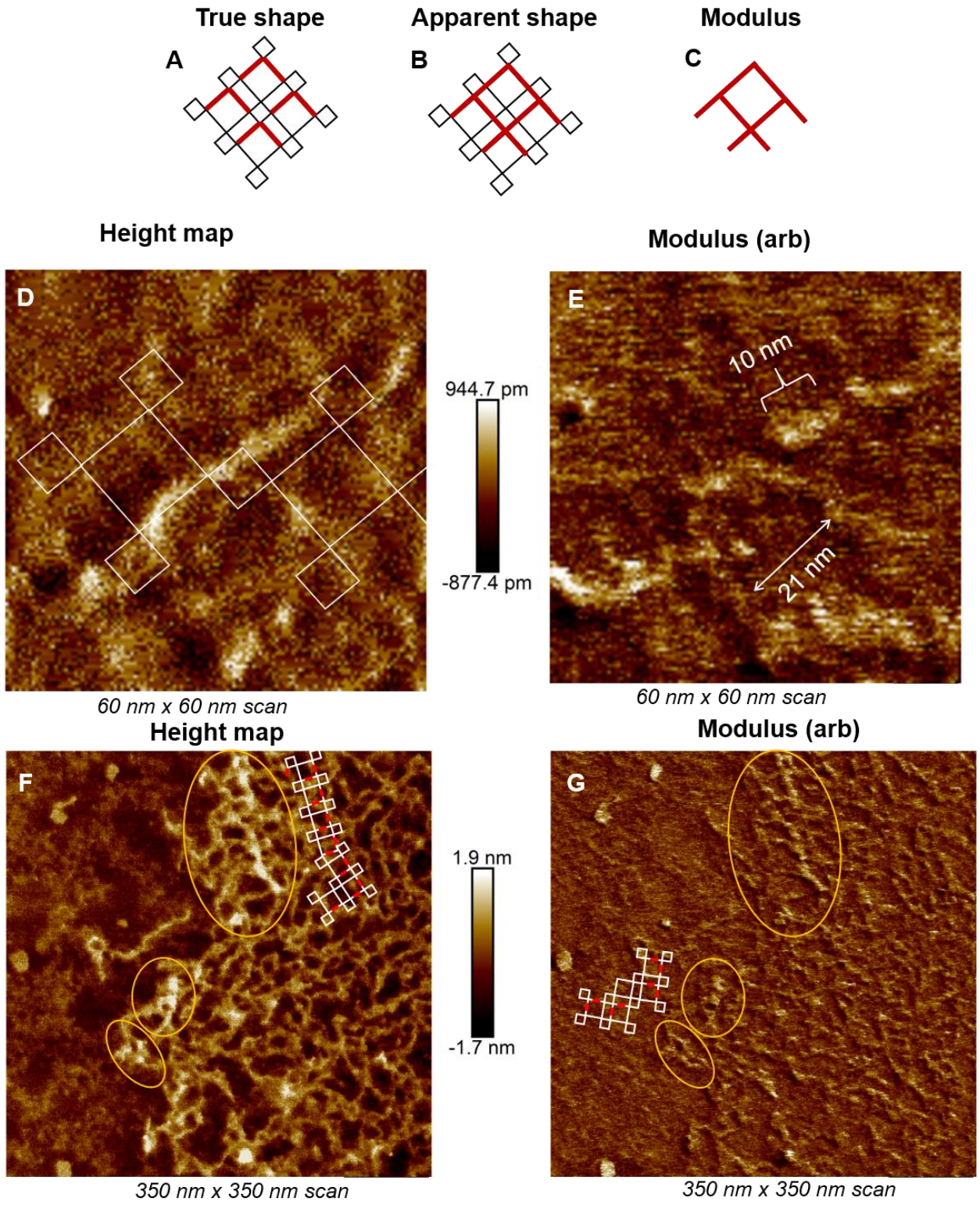


Figure 5.10: Experimental conditions for HJ superstrand 3/5, 4/6 fusion nanowire lattices after secondary annealing with equimolar C30. **A)** Predicted shape of the HJ lattice, with fusion strand C30

shown in red. The C30 duplex has an estimated length of 10.2 nm, while the spacing between parallel C30 nanowires is predicted to be 21.4 nm. **B)** The same lattice is shown, connecting the superstrands to mimic what is may be seen on an AFM, when the holes between sequences on the secondary layer may not be visible. **C)** The stiff regions formed from the superstrands are shown without the underlying lattice, showing the DMTmodulus profile predicted from nanomechanical AFM. **D)** Lattices retain the layered phenotype seen in Figure 5.9E,F, and mimic to some degree the shape in B (white overlay, scaled appropriately). A tall, linear feature is seen running across the image at the edge of a set of 14 nm windows . **E)** The same feature is investigated using nanomechanical AFM, and clearly shows bar-like features of 10 nm length that mimic the expected shape and regularity of the superstrand feature. The window between the superstrands here is 21 nm, in agreement with the mechanical prediction in C. **F)** Over longer range, the lattices appear misshapen with respect to the prediction. Closer analysis shows that, for many of the tiles, this is due to the extrusion and occasional unwinding of the fusion strands, likely as a result of closure of four of the eight sticky ends. Three regions of dsDNA nanowires are highlighted (orange), and in one case, the lattice (white) and superstrand position (red) is shown, mimicking the image with good agreement. **G)** When compared to the modulus of the same data, several superstrand regions can be seen with ~10 nm regular lines. The lattice orientation can be more clearly seen in this image. Overlays of the critical regions (orange) and lattices (red, white) are shown to complement F.

As with the controls, the full C30-HJ-SS structure is subjected to multiple levels of annealing. In order to distinguish between lattice stacks and superstrand duplexes, AFM was performed using PF-QNM (PeakForce Quantum Nanomechanical in Air, Nanoscope 9.1 and 9.4, Bruker)(Figure 5.10). The comparison of mechanical maps, including dissipation (force capacitance), adhesion and modulus with the height sensor teased out the differences between the lattice and the nanowire layers. Because the fusion strand has no spacing between the layers, there is a great deal of tension in the superstrand

sequence. When correctly annealed, the dsC30 regions will be stretched tautly, with rotational freedom, anchored by two polyadenine ssA_6 sequences to fill out 42 nt (Figure 5.10). With this in mind, the stiffness of the superstrand region is expected to be greater than that of the surrounding lattice. When the stiffness maps of C30-HJ-SS lattices are compared with height data, 10 nm lines appear with regular spacing across some (but not all) of the lattice area. In general, comparing the lattices to their predicted designs, it is clear that there are some regions that anneal with high fidelity, but where the fusion strands begin to unwind, there is a disruption in the lattice that affects tiles tens of nanometers away.

Fundamentally, the loss of four sticky ends causes improper anchoring of the tiles at each edge. Much of the disorder seen in Figure 5.10 can be traced to faulty connectivity—windows of the appropriate size and duplex nanowire functionality are regularly obtained across the lattice, but there is a roundness and rotation to the tiles that does not mimic the intended shape. This is a primary drawback of the HJ-SS fusion design, suggesting that a fracture tile may be more appropriate.

When the same analysis of HJ-SS 3/5 fusion tiles is carried out in a one-step anneal, a different lattice phenotype is attained. In the previous experiment, the second step of a two-step anneal is carried out at lower temperature (40 ^{o}C), and thus allows the tiles to fill gaps in the structure but also unwind with the excess thermal energy. By contrast, a one-step anneal spends only a relatively short period of time in the critical temperature zones (40-60 ^{o}C).[165] This introduces more point defects, but experimentally the periodicity of the overall lattice has a much higher fidelity (Figure 5.11). Whereas a two-step anneal produced a misaligned, whole lattice, performing a one step anneal generated a point defect replicated with long-range order: groups of eight tiles assembled in squares around the missing sticky end, leaving a hole at the center of a 70 nm pseudotile. Each arm of the pseudotile, on cross-section, has peaks corresponding to the ~14 nm lattice windows designed in the HJ standard tile. This structure demonstrates a repetitive, defined shape; and it may represent a means of altering or modifying lattice period through modification of sticky end availability. The regularity implies that future synthesis of a

>200 nt oligonucleotide fusion strand that still includes sticky end components may be a means to closed, uniform HJ-SS lattices.

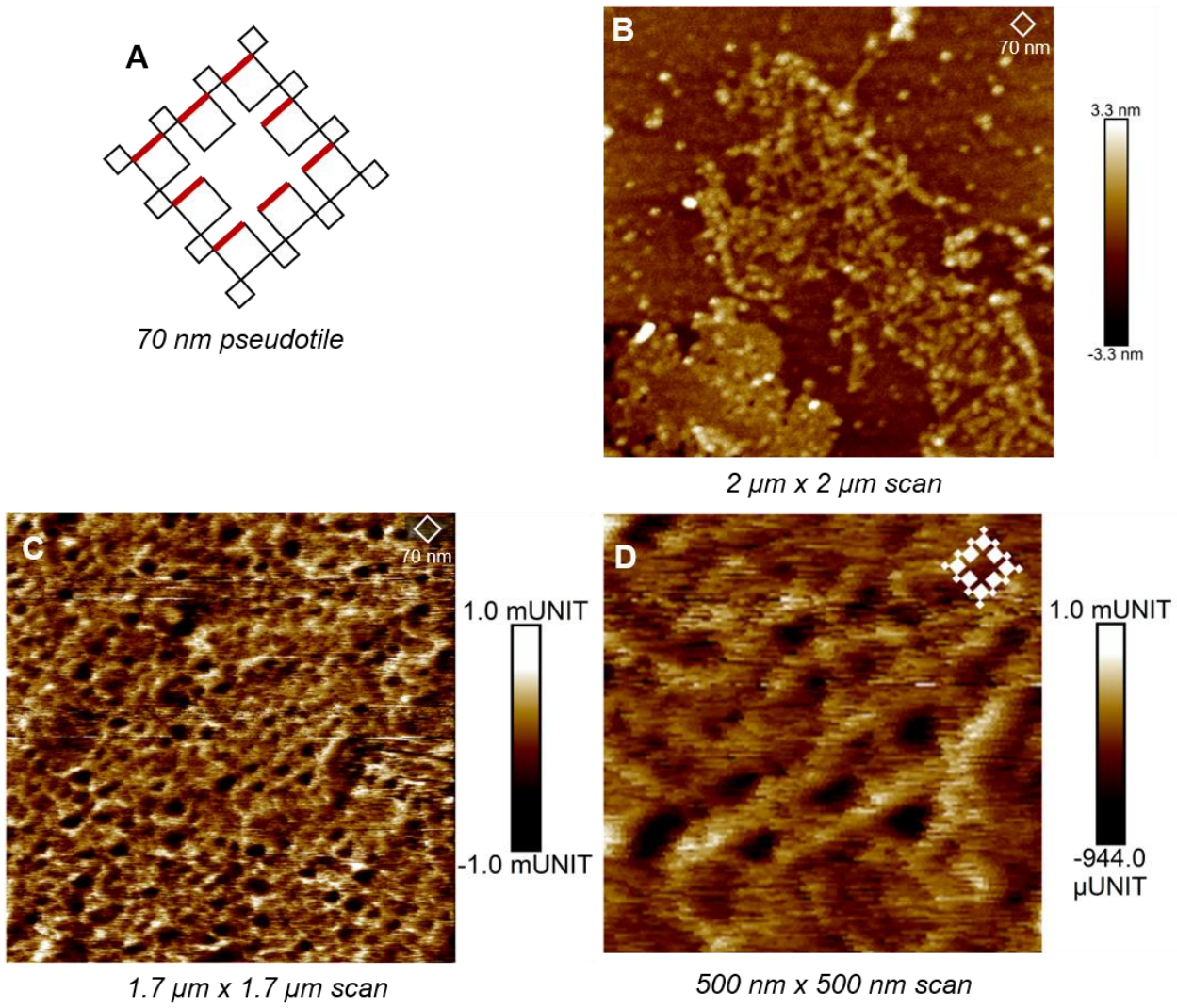


Figure 5.11: HJ-SS fusion pseudotile with C30 linkers in a one-stage anneal. **A)** As a result of the missing sticky end in HJ-3/5 fusion, lattices can form with a missing unit. The defect creates a pseudolattice with holes of ~70 nm in diameter. The regularity of the holes cannot be controlled, as some tiles may fill gaps even without the presence of a sticky end. Superstrand regions are marked in red. **B)** Lattices form with regular 70 nm windows (scale inset). **C)** Adhesion map of a similar structure shows highly regular pseudotile assembly with a periodicity of 63-75 nm (scale inset). **D)** Smaller windows within the

pseudotiles can be seen in an adhesion map that correspond to the designed lattice features (tile composite inset to scale). Windows are highly regular, and each edge contains two peaks 11-13 nm apart, corresponding to the HJ lattice period expected by the original design.

d. HJ-SS fracture with duplex nanowires

The HJ-SS fracture tile was tested in a one-step anneal with its accompanying C30-FAM linker. This tile performed much better than the fusion tile, retaining strong, periodic features expected in Holliday junction lattices. These lattices can be seen to grow in bundles larger than 20 μm, perturbed by deposition during imaging (Figure 5.12B). These large-scale features demonstrate the appropriate 14 and 7 nm features when imaged at higher resolution (Figure 5.12D-G). Due to the small features in a tightly-wound lattice, adhesion maps show more detail of the lattice features then force height maps (Figure 5.12 G/F, Figure 5.13 B/A, Figure 5.13 C). Adhesion force is likely less affected by changes in the voltage setpoint as the adhesion force is extracted from hysteresis of the tip trace, which occurs within a single engage and retract, and remains largely independent of the lift height of the cantilever. In adhesion maps of the HJ-SS fracture nanostructure, lattice features are clearly visible. When processed with Fourier filtering (fast Fourier transform, FFT and inverse FFT, in NanoScope Analysis 1.90, Bruker) to remove background noise, periodic features become quite pronounced (Figure 5.13D-F). *Overall, the fracture tile appears to generate a Holliday junction lattice of the specified dimensions.*

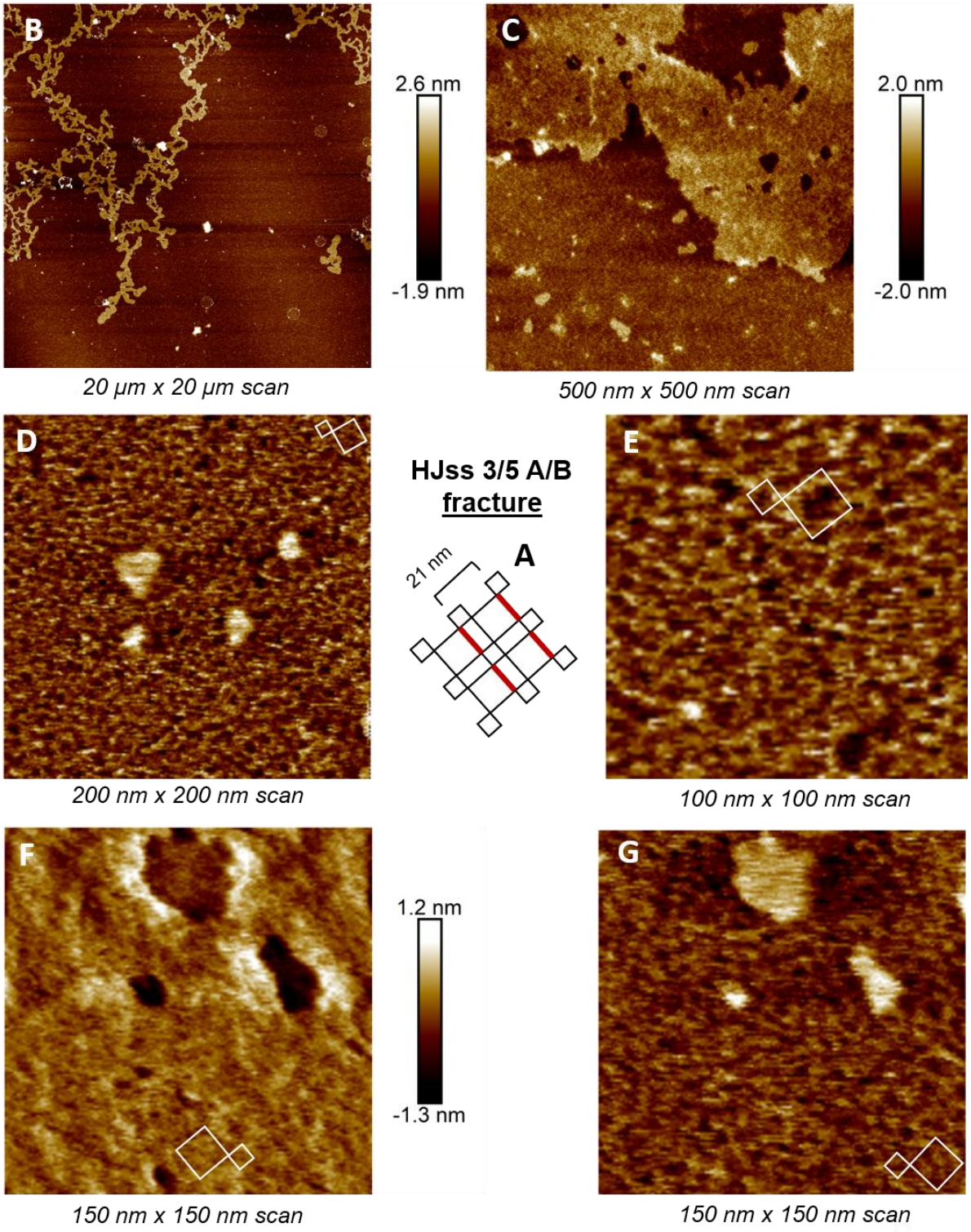


Figure 5.12: Successively zoomed AFM micrographs of HJ-SS 3/5 fracture tile with one-step C30-FAM anneal. **A)** HJ-SS fracture lattices contain 14 and 7 nm square windows with a vertically-offset

superstrand every 21 nm (red). **B)** Lattices are visualized over many microns, with large-scale tears likely a result of deposition during imaging. **C)** Closer investigation of large features in A show a basic lattice phenotype. **D)** Adhesion map of 200 nm image shows periodic lattice features with 14 and 7 nm windows as designed in the tile (inset to scale). **E)** An adhesion map at a 100 nm scale shows similar lattice features (inset to scale). **F)** and **G)** show a single scan with atomic force height and adhesion force, respectively (insets to scale). Regular lattice windows are observed across all structures.

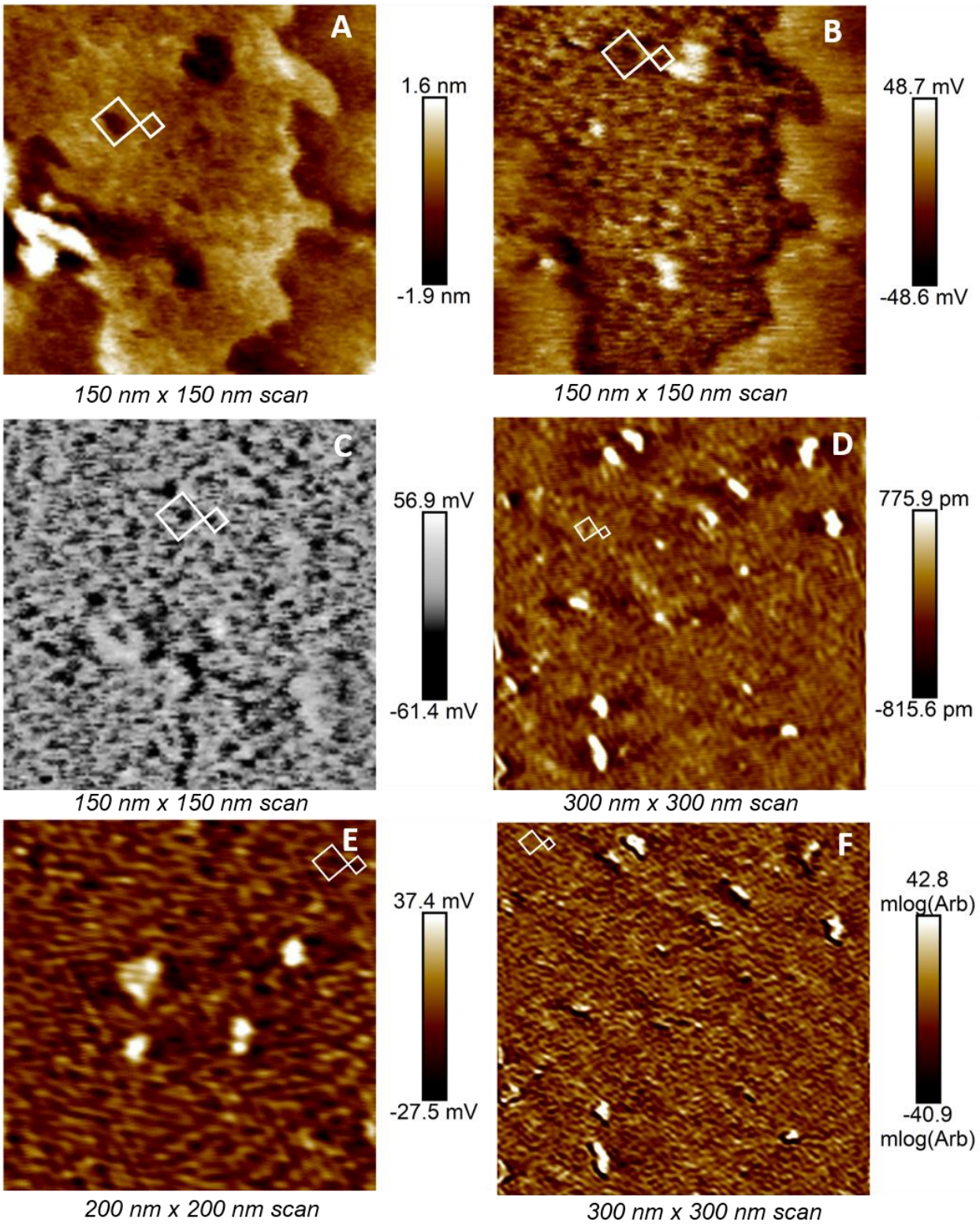

A
1.6 nm
-1.9 nm
150 nm x 150 nm scan
B
48.7 mV
-48.6 mV
150 nm x 150 nm scan
C
56.9 mV
-61.4 mV
150 nm x 150 nm scan
D
775.9 pm
-815.6 pm
300 nm x 300 nm scan
E
37.4 mV
-27.5 mV
200 nm x 200 nm scan
F
42.8
mlog(Arb)
-40.9
mlog(Arb)
300 nm x 300 nm scan

Figure 5.13: HJ-SS fracture lattices with varying degrees of image processing. **A)** and **B)** show the same scan in force height and adhesion force, respectively. Lattices demonstrate periodic windows of the predicted size (inset to scale). **C)** Adhesion map of a lattice shows high-resolution Holliday junction lattice of predicted shape (inset to scale). **D)** Force height image subjected to FFT filtering of background noise gives highly periodic lattice with 7 and 14 nm features (inset to scale). As a result of FFT, the 7 nm period predominates. **E)** FFT filtering of adhesion map shows both 14 and 7 nm features in good agreement with predicted shape (scale inset). **F)** FFT filtering of LogDMT Modulus mechanical map shows high-resolution 7 nm period with 14 nm larger windows (scale inset). Overall, lattices are well-formed and periodic, and hold up to Fourier analysis.

Geometric analysis of the HJ-SS fracture tile suggested that the nanostructure assembles as designed. It remained unclear from AFM whether the C30-FAM linkers were making their way into the structure at the appropriate sites. In order to understand whether the C30-FAM sequence was in fact bound specifically to its complementary C30 binding site, the products of annealing were imaged using PAGE with various reaction components. The results of C30-FAM were compared with nonspecific control pyrimidine T30-FAM. Gels were run on 10% polyacrylamide pre-cast gels (Thermo Fisher) at 135 V at room temperature, and shielded from light by aluminum foil to prevent photobleaching. Gels were imaged for fluorescein without staining, and again measured on a non-interfering channel after 30 min post-stain in gel red (Thermo Fisher) (Figure 5.14). In this experiment, distinct, high-MW features appeared in the C30-HJ-SS lattices, but not in HJ standard lattices (Figure 5.14A), and only at sizes greater than one tile (~270 bp). These distinct bands suggest specific and regular incorporation of C30 into lattices of small size, but are not seen in large, many kilobase lattices. This behavior is not seen in control sequence T30 (Figure 5.14C), confirming the need for polyC and HJ-SS sequences for this phenotype. Furthermore, depletion of band streaking associated with C30:C30 annealing is observed in the presence of HJ-SS lattices, suggesting that the lattices are able to compete kinetically with C30 self-

complementarity. This effect might be further mitigated by designing non-homogenous superstrand linkers in future studies.

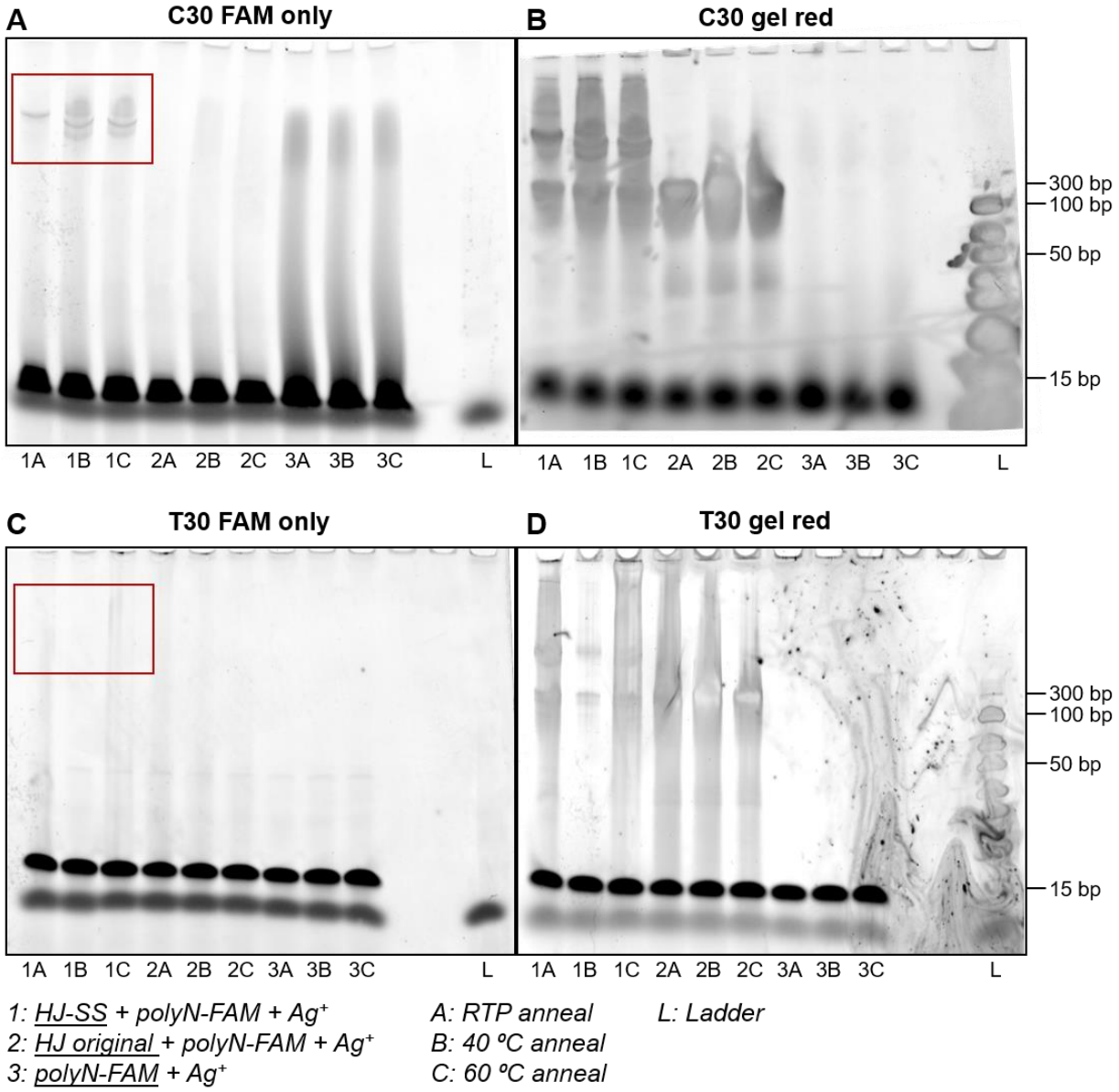


Figure 5.14: PAGE analysis of HJ-SS fracture design on specific binding with FAM-labeled polypyrimidine oligonucleotides. Across all gels, lanes i-iii (labeled 1) contain HJ-SS lattices with C30 binding sites; lanes iv-vi (labeled 2) contain HJ lattices without fusion strands or C30 binding sites; lanes vii-ix (labeled 3) contain just the polypyrimidine; and lane x (labeled L) contains a 10 bp DNA ladder. Lanes labeled 'A'

had secondary anneals with the polypyrimidine and 10x Ag^+ at room temperature; lanes 'B' were annealed at 40 ^oC; while lanes 'C' were annealed at 60 ^oC. Each well has a loading dye visible as the bottom feature. **A)** No-stain image taken of C30-FAM gel. Lanes 1A-1C show discrete bands at high molecular weight regions are formed, demonstrating specific and regular incorporation of the C30 sequence. A single band is formed in RT anneals (1A), while multiple bands separate for higher-temperature anneals. Lanes 2A-2C show no features in this range, confirming that the binding is specific to the fusion strand regions. Lanes 3A-3C show the expected streaking from polyC anneals. Notably, this streaking is greatly reduced in wells 1A-1C, which suggests that more C30 molecules are incorporating into the lattices than forming nanowire chains. In all wells, unfolded ssC30 can be brightly seen near the bottom of the gel. **B)** After staining, the typical HJ lattice profile can be seen between 200-300 bp (expected tile size is 270 bp, after accounting for ssDNA regions and sticky ends). This profile is seen in both HJ original and HJ-SS lanes, but not in C30-only wells. Bands corresponding to the discrete, high-MW bands in A are repeated here, showing the regular lattice structure. **C)** Unlike C30, T30 shown no polymerization under any conditions and is predominated by ssT30 in all wells. **D)** Wells 1A-1C show fewer bands and less streaking than in B, indicating a lack of T30 incorporation. Typical HJ lattice behavior is otherwise observed.

e. Conclusions from the HJ-SS lattice

In general, the incorporation of C30 into HJ-SS lattices was successful, but the overall yield was dependent on several factors. In nanostructures with unbounded growth, the best measure of yield is design fidelity—namely, the regularity and adherence of the lattice features to the prescribed shape. The size of the lattices (above several tens of subunits) represents a secondary measure of yield. The HJ-SS fusion tile produced two lattice phenotypes: badly connected arrangements of tiles without holes, attained through two-step annealing; and regular distributions of pseudotiles missing every ninth tile, produced in one-step annealing. These lattices grew to intermediate size, between 300 nm – 5 µm. The

yield of this experiment, in terms of design fidelity, was quite low. It was, however, determined that some adjustments could be made to the reaction to aid future experiments: first and foremost, the design requires all eight original sticky ends to function, and a fusion-style tile requires an oligonucleotide synthesis of more than 200 nt from a specialized company. Secondly, annealing with the C30-FAM linkers may be carried out in one-step together with lattice oligos in order to facilitate good periodicity.

The HJ-SS fracture tile showed very high yield. The lattice features appeared uniformly regular with 7 and 14 nm windows per the original design. The growth of these periodic lattices extended into the tens of microns. Finally, the lattices demonstrated the ability to specifically incorporate the C30-FAM linkers into their structure with distinct gel bands. In terms of both design fidelity and long-range order, the HJ-SS fracture lattice demonstrates high yield. It demonstrates that cytosine-enriched, Ag^{+}-mediated DNA nanowires can be incorporated into WC-pairing DNA tiles without great disruptions in overall lattice phenotype. Additional superstrand components to generate a nanowire boxframe may be possible. There also exists the possibility of G4 or other orthogonal nucleotide incorporation into the HJ-SS nanostructure for the construction of long-range lattices. The promise of the design comes in layering electronically-functional and electronically-resistive components on the Z-axis to create vertical circuit architectures. With current yields and better designs, this may be a realistic goal. As such, further work may be advised.

5.3. Kissing loop units: T-junction (TJ) arrays

5.3.1. General design

The T-junction tile was first imagined by Hamada and colleagues, and was reported in 2009.[182] The basic design revolves around mid-duplex branching, more commonly found in RNA, in which two complementary hairpins will bind to form a ‘kissing loop’ structure.[116] In this case, 5 nt loops are

incorporated into dsDNA helical bundles, allowing two full turns and a 90^{o} junction, or T-junction. The authors estimate that (6 bp) minor groove loops will cause only 4^{o} deviation in phase difference between the helical planes before and after the junction. In this estimate, they connect a 20 bp central beam (1 bp short of two helical turns), a non-pairing A:T complement (accounting for their extra nucleotide), and a 5 bp kissing loop to form their T-junction structure (Figure 5.15). The kissing loop binds to sticky ends terminal to the tile on either end, and the complementarity scheme of the SE-KL bonds determines the resulting structure. In general, as the T-junction part of the tile is facilitated by a single-stranded hairpin, a larger degree of flexibility can be expected in resulting nanostructures, which may both allow greater structural diversity and reduced yield.

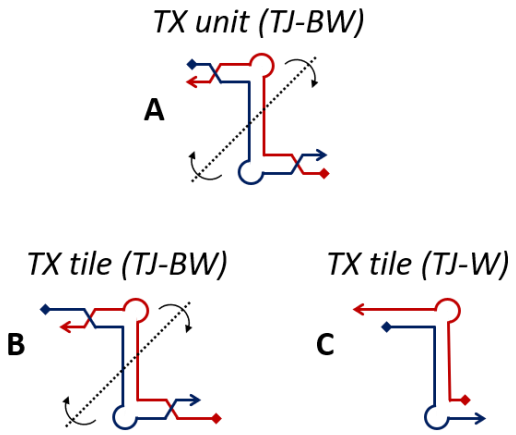


Figure 5.15: T-Junction tile family. **A)** General scheme of the T-junction tile shown without sticky ends. Connectivity is tailored by the orientation of sticky end/ kissing loop interactions. **B)** The brick wall tile (TJ-BW) is constructed by connecting the blue sticky end to the blue kissing loop, thereby causing the tile to flip each time a new monomer is added. In doing so, rectangular windows are assembled in layers exactly 180^{o} out of phase with one another, creating a brick-and-mortar-style lattice. **C)** Using asymmetrical arms and connecting like-colored SE/KL pairs, a wheel can be constructed that is 36-37 nm in size, consisting of 12 subunits.

This tile was chosen as a promising template for metal pair integration as it represents a minimal number of base pairs, on the order of 50 bp total, making the solution space much smaller for algorithmic sequence optimization. A design scheme was chosen whereby CC-bonds were selectively incorporated from the sticky ends inward, first demonstrating efficacy of assembly with a C:Ag^+:C-requisite SE-KL assembly to open the door for more integrated, algorithmic designs.

5.3.2. TJ brick wall tiles (TJ-BW)

The brick wall tile (TJ-BW) is capable of forming lattices over large areas. In ideal conditions, the original authors report multi-micron scale assemblies (when using substrate-assisted annealing)(Figure 5.16). The original sequences are first tested for consistency, and then the structure is modified to incorporate sticky ends which require the uptake of Ag^+ into metal base pairs.

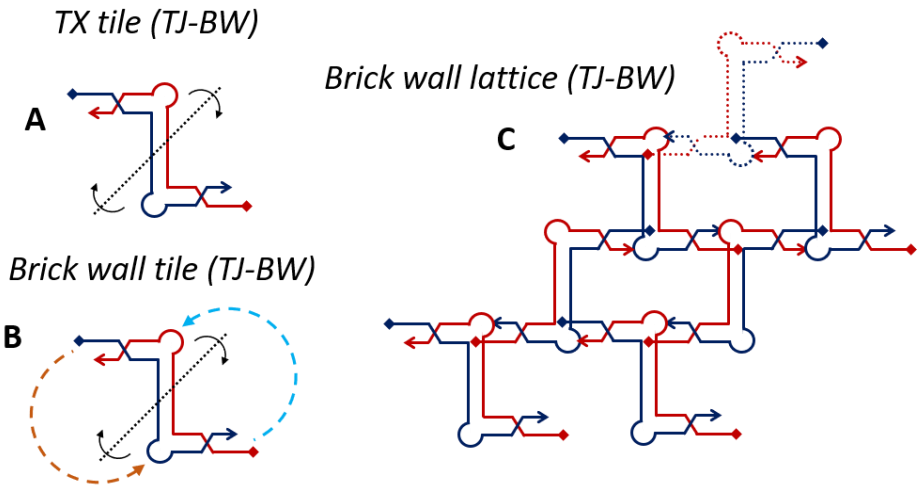


Figure 5.16: Brick wall tiles. **A)** General design of the TJ-BW design, with two oligos of 50 nt in length. **B)** Connecting like-colored SE-KL pairs, the design must be reflected across the Y-axis to bind with itself. **C)** When assembled into a lattice, the windows formed by the T-junction tiles resemble a brick wall. The original authors calculate that these rectangles are 8.3 x 9.6 nm in size.

a. Standard tile

Standard sequences from Hamada et al[182] are tested against their results (Table 5.5). The annealing protocol recommended by the authors, to correct for defects introduced by tile flexibility, is to perform an initial full-temperature anneal followed by 10-100 cycles of slow, low-temperature correction annealing. To do so, a lab-standard protocol was followed, in which 0.2 nmol of each ssDNA oligo was mixed together in 1x MOPS buffer, pH 7.5, with both Na^+ and Mg^{2+} (100 mM, 12.5 mM, respectively). Primary annealing was performed in a thermo cycler, and then the 10-60 correction cycles were performed. The resulting structures were deposited on freshly-cleaved mica and imaged in air, using both standard ScanAsyst and high-resolution HPP-SAA tips on a Bruker Dimension Icon AFM (Bruker). In all cases, brick wall lattices of long (but not wide) scale were observed (see Figure 5.17).

Table 5.5: TJ-BW Standard (Hamada et al)[182]

Sequence #	Nucleotide sequence (SE/KL underline)
TJ-BW (+)	5'-AGCCCTTGTGGTAGTTGGCACCAGAAGACCACGGTGGGCTTAACACCATC-3'
TJ-BW (-)	5'-CGACGGATGGTGTTAACCGTGGTCTTCTGGTGCCACGTCGACTACCACAA-3'

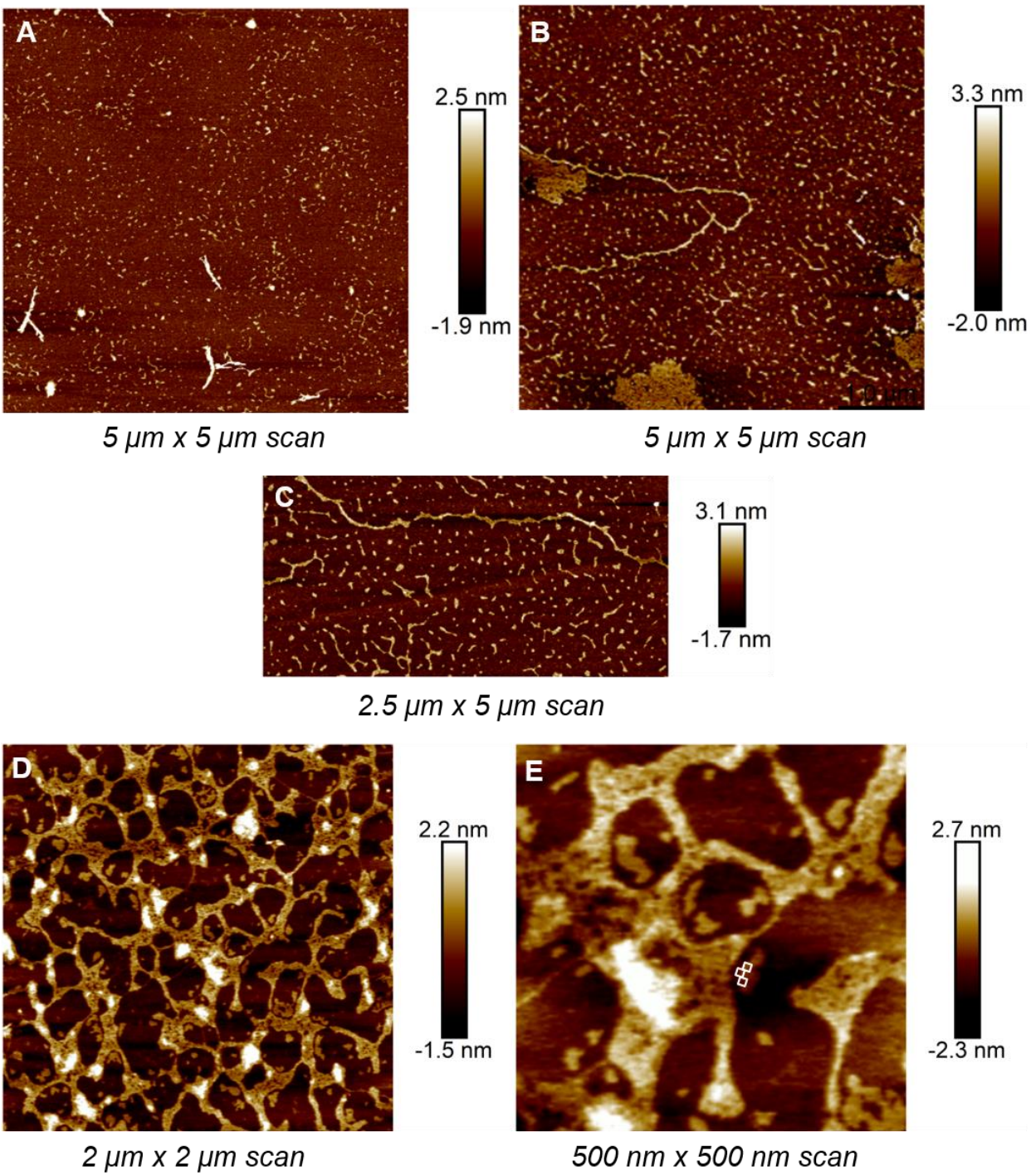


Figure 5.17: TJ-BW lattices using standard chemistry in MOPS buffer. Many-micron long, 100 nm-wide lattices are observed, and seem to be broken from deposition onto mica. Fluid AFM was chosen as a follow-up to these measurements, and applied to experimental structures where deposition yield was

low. **A)** Some large lattice pieces are observed, and many small components can be seen on the surface. **B)** Long spindles are observed, indicating that the maximum lattice growth size is quite large. Many lattice components and some unfolded, amorphous ssDNA bundles are also observed. **C)** A further lattice is observed in which growth direction is shown to be generally linear, though winding of the spindles is observed as a result of the tile flexibility. Successful CC-lattices should show similar features. **D)** Closer image and subsequent zoom **E)** of lattices shows lattice with windows of the appropriate size (scale inset, E). The lattices here seem to have many holes, likely forming during deposition as a result of the extremely flexible TJ tile.

Overall, this assembly is reasonably successful, and suggests that MOPS-based annealing does not interfere greatly with overall yield. No planar structures of rectangular dimensions are observed, suggesting differences in experimental method from the original authors. Nanoscale, single-unit features are in agreement with the design; however, observed grain sizes are quite small in these structures.

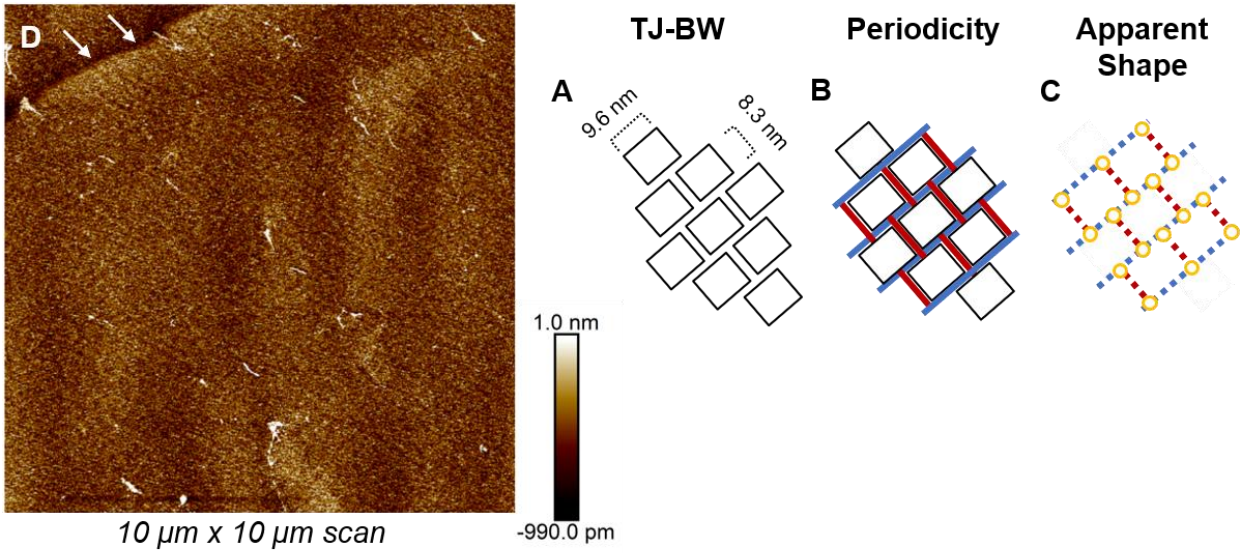
b. CC-tile

The T-junction brick wall tile was modified to include a polycytosine sticky end / kissing loop (SE/KL) interaction that would necessarily prevent lattice assembly in the absence of C:Ag^+:C bond formation. The original TJ-BW (-) sequence was used with a modified TJ-BW-5CC (-) complement (Table 5.6). Several annealing recipes were tested. Structures were either annealed with or without Ag^+ over 4 hr from 95 ^oC to 20 ^oC in a Thermo Cycler, followed by 10 correction cycles from 20 ^oC to 45 ^oC to 20 ^oC, with a temperature change of 6 min/^oC. After annealing, samples were cooled to 4 ^oC and stored. Samples without initial Ag^+ were subsequently infused with 5x $AgNO_3$ and subjected to 50 correction cycles. Without Ag^+, extremely fragile 1D T-junction polymers are able to form, and during corrective annealing, these tiles are able to form full lattices for a slower, gentler anneal.

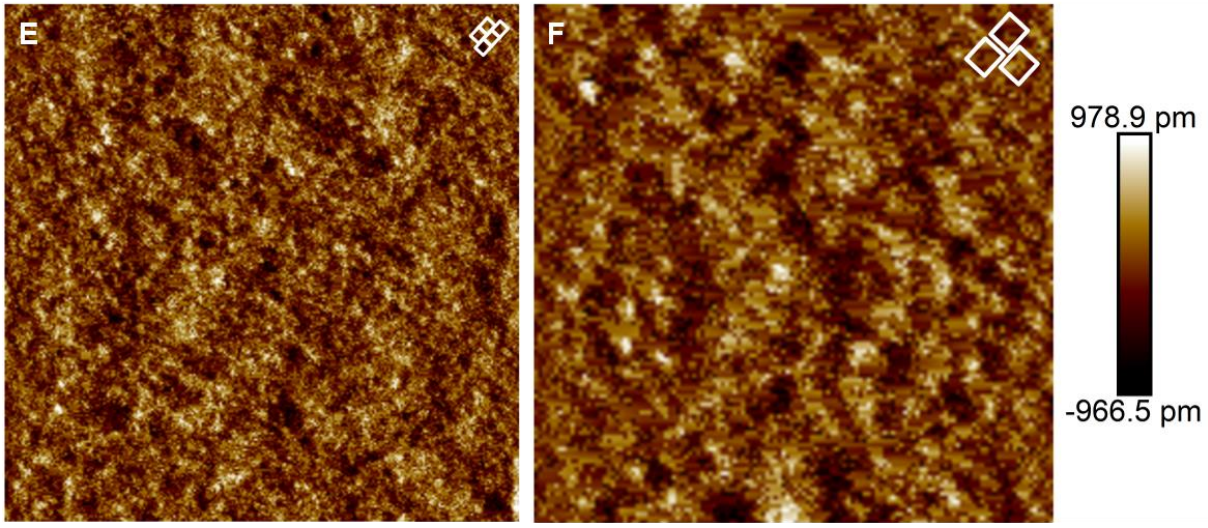
Table 5.6: TJ-BW-CC (5CC)

Sequence #	Nucleotide sequence
TJ-BW-5CC (+)	5'-**CCCCC**TTGTGGTAGTTGGCACCAGAAGACCACGGT**CCCCC**TAACACCATC-3'

Several resulting nanostructure phenotypes were obtained. Most importantly, structures subjected to correction cycles were able to produce flat sheets of brick wall lattices extending over many microns (Figure 5.18). In some cases the lattices appear to stack in layers, with underlying layers apparent (Figure 5.18D**,** Figure 5.19A). As image resolution is increased in lattices of this phenotype, the internal brick wall structure is clearly apparent with good growth axis orientation (Figure 5.18E-H), fitting the predicted dimensions (Figure 5.18A-C) with good agreement. In fact, the bricks in these lattices can also be seen to have the predicted six sticky end / kissing loop bundles at T-junction vertices, confirming the successful assembly of the designed shape. Overlay of the shape prediction clearly demonstrates the lattice periodicity in these images. In other samples, smaller lattices can be seen, and the component fibers of these ribbon-like structures pull away from the sheets over several brick wall iterations. The main difference in reaction between these two phenotypes seems to be time—as the lattices age they begin to fall apart, perhaps due to exonuclease activity or aqueous degradation. Lattices imaged more quickly after annealing (see Figure 5.18) were more likely to retain a flat-sheet phenotype, while lattices imaged more than 7 days after the last correction cycle seemed to start unwinding into spindle and ribbon phenotypes (Figure 5.19).

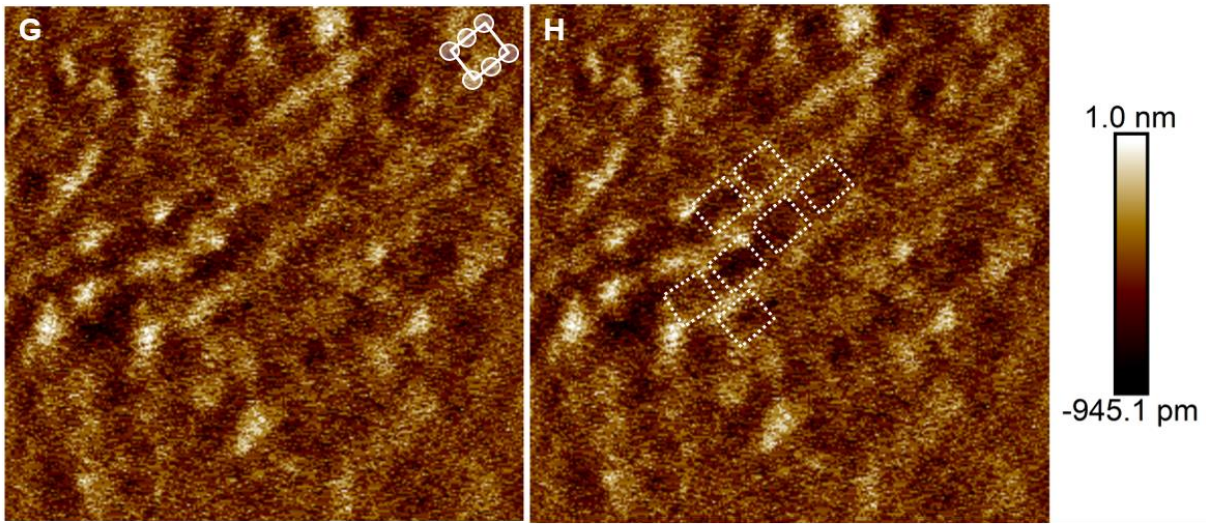


10 µm x 10 µm scan



350 nm x 350 nm scan

150 nm x 150 nm scan



100 nm x 100 nm scan

100 nm x 100 nm scan

Figure 5.18: AFM analysis of TJ-BW-CC nanostructure lattice parameters. **A)** Predicted arrangement of 9.6 nm x 8.3 nm brick wall windows with each new layer stacking 180^{o} out of phase with the underlying layer. **B)** Growth axes of the brick wall nanostructure can be easily identified: interlayer structure lines are continuous (blue) and occur every 8.3 nm, while intralayer lines corresponding to the 20 bp central beam of each T-junction tile (red) are interrupted by the windows. Within each brick layer, these lines have a periodicity of 9.6 nm. **C)** When visualized via AFM, the sticky end/kissing loop bulges will occur in bundles of six around each window, corresponding to the six SE/KL T-junctions that make up the given shape. On (blue) interlayer lines, these bulges (yellow) occur with 4.8 nm periodicity, with slight fluctuations expected for tile flexibility. An unsuccessful SE/KL junction will likely be less rigid and demonstrate reduced apparent height, accompanied by softer nanomechanical properties. **D)** Scan of TJ-BW-CC lattice annealed with a 10x correction cycle in the presence of Ag^{+}. Two large, >10 µm sheets appear to be stacked, with a 2 nm height gap between the layers (white arrows, inset). **E)** Closer scan with a 1 nm HiRes AFM tip on samples annealed without Ag^{+}, subjected to 10 correction cycles, then infused with 5x $AgNO_3$ and annealed for a subsequent 50 correction cycles. Lattice windows of the exact size and expected orientation (white bricks, inset) are apparent throughout the whole image. **F)** Sample with similar annealing recipe imaged at higher resolution, showing regular windows of the correct size across the image (white bricks, inset). Scale bar identical for E-F. **G)** Similar results are obtained with lattices annealed with Ag^{+} in the original anneal, though the long range order is reduced. In this scan, bricks with six bulges indicating SE/KL bundles (white brick, inset) are seen across the image. **H)** The same image is shown with overlay of the brick windows in their lattice position to demonstrate the predicted and experimental structures together. Images were subjected to 2-6^{th} order flatten functions to remove surface bow, while D received a lower order flatten to avoid artificial removal of layer stacks. All analysis performed in Nanoscope Analysis 1.9 (Bruker).

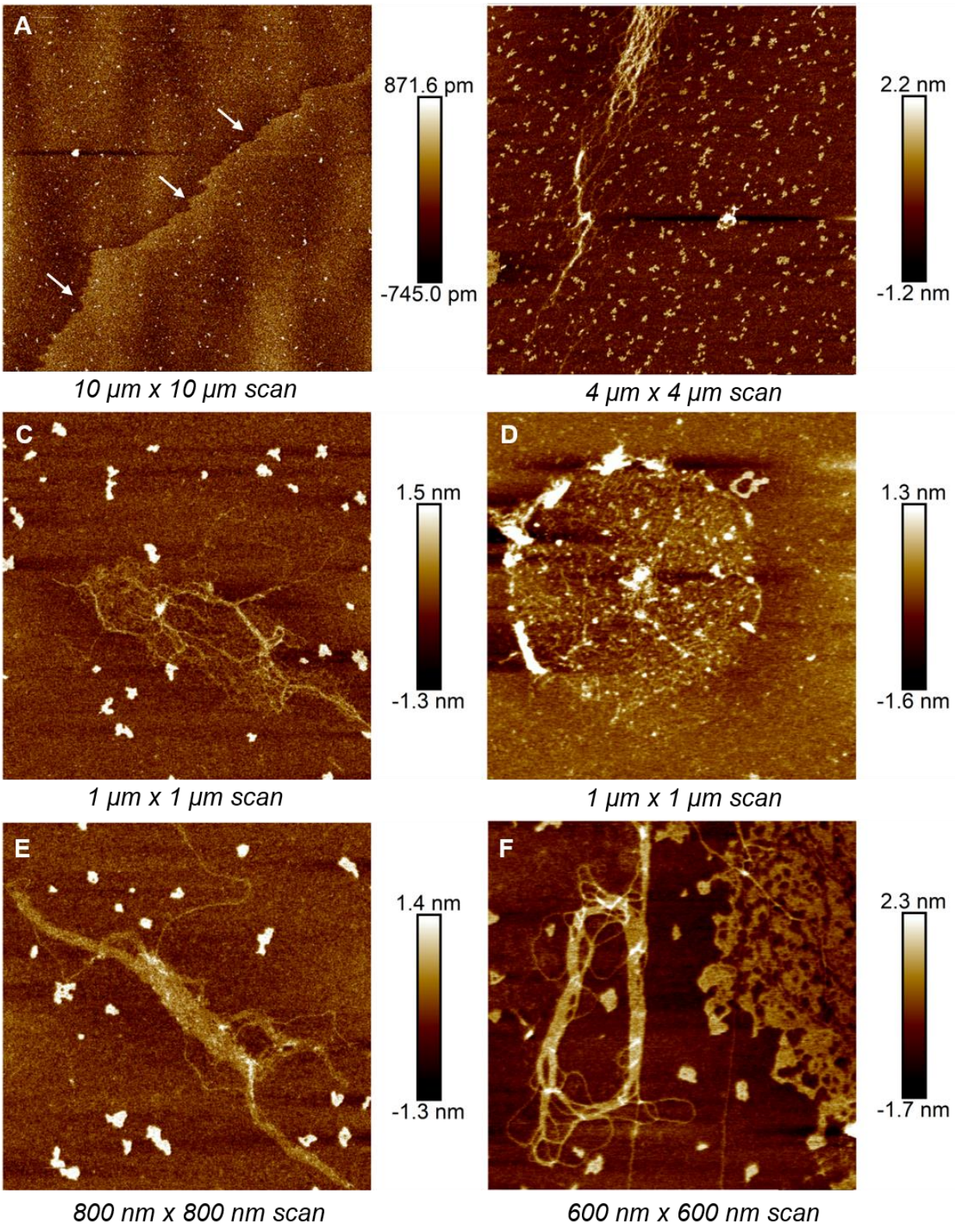

A
871.6 pm
-745.0 pm
10 μm x 10 μm scan
2.2 nm
-1.2 nm
4 μm x 4 μm scan
C
1.5 nm
-1.3 nm
1 μm x 1 μm scan
D
1.3 nm
-1.6 nm
1 μm x 1 μm scan
E
1.4 nm
-1.3 nm
800 nm x 800 nm scan
F
2.3 nm
-1.7 nm
600 nm x 600 nm scan

Figure 5.19: AFM micrographs of the TJ-BW-CC nanostructure. **A)** Anneal with Ag^+ and 10 correction cycles produces a flat stack of layers with a 2 nm height difference. Apparent lattice size is > 10 µm. Jagged edge of the top layer (white arrows) typical of T-junction tile assembly in which polymerization may occur on multiple concurrent axes. Growth axes visible, though the windows that can be seen are not small enough for brick wall features, but rather are aggregate lattice density lines of the kind also observed by Hamada.[182] **B)** The same reaction, when carried out without correction cycles, produces interwoven 20-30 nm-wide spindles of many microns in length. **C)** Micrograph of 60x-corrected, Ag^+-first structures diluted 100-fold produces a similarly bundled phenotype. In this image, the brick wall windows webbed between spindles are clearly visible. **D)** Circular lattices with salt edges can be seen in similarly-annealed structures, with windows apparent. Over long range, these salt droplet lattices can be seen repeated in many images (data not shown). **E)** 100x dilution of samples broke apart salt bundles into component lattices, and in this image, a ribbon-like lattice is observed with 7 brick wall layers (60 nm) apparent at its widest. The ribbon is at least 1 µm long. **F)** A similar lattice ribbon is observed, with component T-junction 1D chains extruded from the central ribbon and extending outwards over 2 µm (data not shown). Here an adjacent amorphous DNA aggregate sits on the surface, and the 1D T-junction spindles appear to cut through the DNA pile, suggesting that the amorphous DNA phenotype is a result of deposition, not proper annealing. All images subjected to 2-3rd order flatten functions and plane fits in Nanoscope Analysis 1.9 (Bruker).

To more closely analyze the periodicity of the lattices, images shown in Figure 5.18 were subjected to fast Fourier transform (FFT), filtered for high-density lattice points, and subjected to inverse FFT (Nanoscope Analysis 1.9, Bruker). These data show with much greater clarity the intended brick wall shape (Figure 5.20), with highly regular stacks of 8.3 nm x 9.6 nm rectangles. Hamada predicted a 90^o internal angle to each rectangle, while an analysis of the lattices in their paper shows approximately 50^o and 130^o corners, forming a rhomboid window. Analysis of Figure 5.20C-D shows a highly regular,

successful lattice with corner angles of 70°-75° and ~105°-110°. This important result suggests that the rotational dynamics of C:Ag^+:C-mediated SE/KL bundles are different from WC DNA. Though the lattice is still successful, the kissing loop rotates 20°-25° less around the T-junction, producing windows that are more square than rhomboid. This result suggests that the introduction of Ag^+-mediated bonds increases the overall duplex stiffness, producing a lattice with different overall shape, greater rigidity, but long range order similar to that found in the original design.

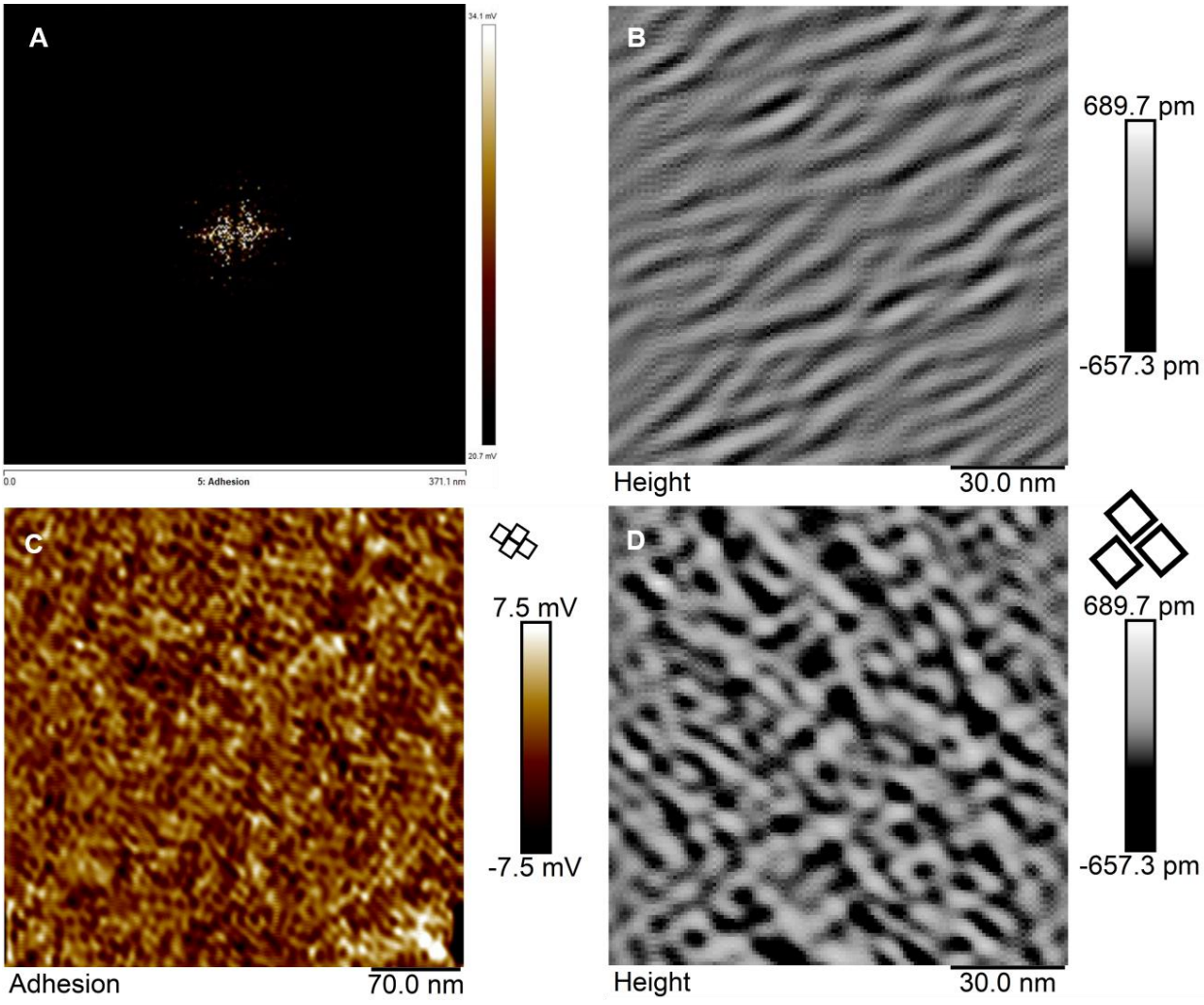


Figure 5.20: Fourier transform filtering of Figure 5.18 lattices. **A)** Fast Fourier transform of adhesion force map from Figure 5.18E using Nanoscope Analysis 1.9 (Bruker). Image is zoomed in to show central cluster and 8 outlying points. The rotation of these points is slightly different to that achieved by

Hamada and colleagues, suggesting altered lattice geometry. **B)** Surface roughness data from low density points that is filtered out of Figure 5.18F and subjected to Inverse FFT. This surface noise is shown—a conflation of these data and the data in D produce the image found in Figure 5.18F. **C)** Figure 5.18E adhesion map after noise is removed via manually selected FFT data filtering. Lattice period is clearly visible across the image, with ~8 nm x ~10 nm rectangles (see black boxes, inset) matching the predicted shape. Adhesion data are used as they are less subject to surface roughness, and instead are already integral transforms of the force curve that are not as sensitive to instrument feedback. **D)** Filtered data from Figure 5.18F, which involved the subtraction of data in B. Though the lattice is over somewhat lower quality, the period and brick layering are clearly visible across the image, matching predictions (inset black rectangles).

Overall, the T-junction brick wall tile seems highly amenable to metal base pair functionalization, as a variety of lattices were constructed in good agreement with predicted shape, period and phase. The long-range polymerization of the brick wall structure was predicated upon the ability of the tile to incorporate multiple C:Ag^+:C bonds into a nonlinear, sticky end / kissing loop bundle that did not behave in a standard, B-form dsDNA manner. The kissing loops that make up the T-junction tiles generate the turns that give the monomer its name, but also act as sticky end anchors. As a result, the bond between SE/KL complements is not expected to form a traditional duplex, but a pseudo-hairpin heterostructure. The single-tile features seem to agree with the design, while grain sizes appear to be quite large. With the incorporation of Ag^+-bonding into this type of structure, it is apparent that orthogonal base pairs can participate in, and further mediate, nonlinear polynucleotide assemblies with diverse geometries. Future studies may incorporate such kissing loops into larger lattices to allow 3D layer extrusion, incorporation of Ag^+ nanowires into existing structures using known polyC SE/KL interactions, and even the anchoring of other nanomaterials into unfilled lattice kissing loops. Furthermore, the successful assemble of a TJ-BW lattice with CC sticky ends suggests that algorithmic optimization of these

sequences should generate tiles that are amenable to longer nanowire elements, perhaps functionalizing whole growth axes or branching patterns within a larger lattice with electrical conductivity.

5.3.3. TJ wheel tiles (TJ-W)

The T-junction wheel tile is built asymmetrically to introduce assembly of SE/KL bundles along an irregular quadrilateral (Figure 5.21). When sticky ends are taken into account, the smaller arm of the tile comprises approximately one helical turn (11 bp), while the longer arm contains two full helical turns (21 bp). Over the rotation of the wheel, some strain along the inner axis will arise: each helical turn comprises a non-integer number of base pairs (10.45), while the short arm contains an integer value of 11 base pairs. Over twelve tiles, this will introduce ~45^{o} of extra rotation, or 12% of an extra turn. This strain may be alleviated by the non-helical structure of SE/KL bundles, but may also introduce asymmetries in ring formation including: breakage, rings with one fewer tile, and/or flaring of the wheels to locally reduce ring torsion. The T-junction tile is by default quite flexible, and previous studies show 17% overall yield, which was likely reduced due to AFM tip-induced damage to the rings.[182] This structure was chosen for modification due to its fixed size—unlike brick wall tiles or Holliday junctions, the resulting nanostructure has a defined shape, and data analysis can be carried out in a more quantitative manner.

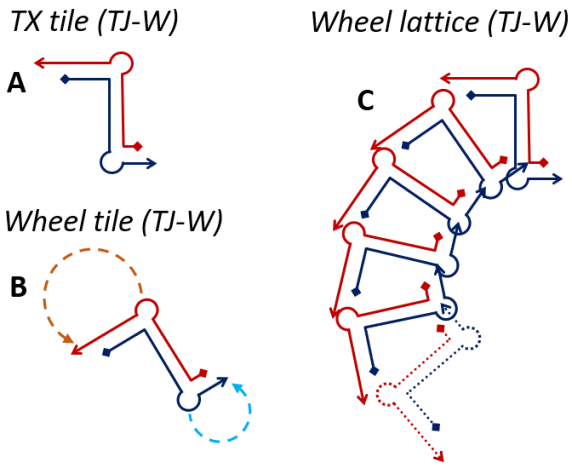


Figure 5.21: T-junction wheel tile and resulting nanostructure. **A)** TJ-W tiles are asymmetrical in shape, while still possessing the same two-SE, two-KL base design. **B)** The longer arm (red) is self-complementary, while the shorter arm (blue) adheres to itself, creating angular torsion on the structure. **C)** When annealed together, twelve tiles assemble to form a wheel structure of ~37 nm in diameter. A smaller hole can be observed that is ~20 nm in size, though this hole may not be apparent via AFM.

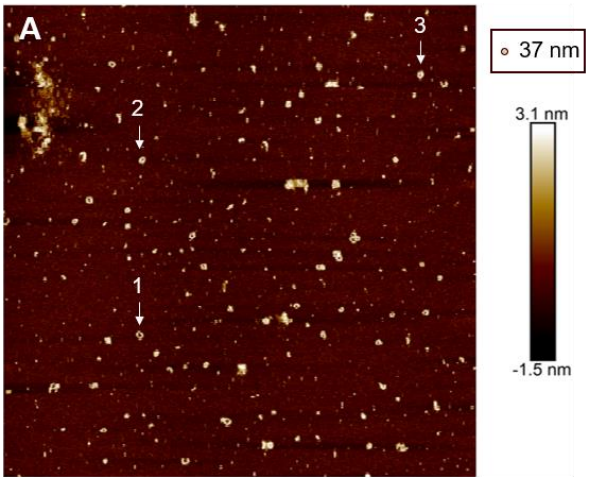
a. Standard tile

The standard T-junction wheel was assembled using literature sequences (Table 5.7). In order to test the robustness of the nanostructure in different annealing conditions, the wheels were built in both Tris and MOPS buffers, with and without Na^+, and with and without Ag^+. In all cases, annealing was carried out from 95 ^oC to 45 ^oC in a Thermocycler, with 50 correction cycles from 20 ^oC to 45 ^oC to 20 ^oC, with a temperature change of 6 min/^oC. Results showed many wheels (Figure 5.22) with different phenotypes, levels of aggregation, and varying surface distribution. A quantitative study of surface ring diameter (Figure 5.22A) showed an average size of 37.4 nm with a standard deviation of 5.5 nm. Hamada measured 36.0 ± 2.5 nm, while the predicted size was ~37 nm. While the achieved shape is in close agreement with calculations, the high degree of variation attained here, and to some degree in Hamada's study, demonstrates that the rings are not, in fact, static structures, but rather subject to

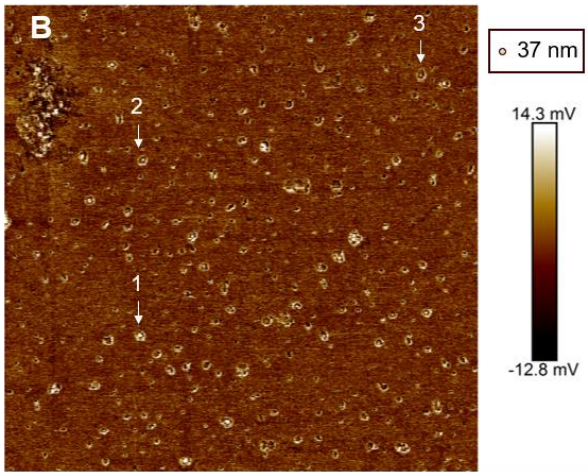
addition or subtraction of tiles. In general, most rings are near to 37 nm in Feret diameter, while others are dramatically different, at ~30 nm or ~45 nm, suggesting a tile addition or subtraction, rather than fluctuation of the overall shape. Furthermore, a number of rings were observed with elliptical shape, such as in Figure 5.22F, which may indicate a tile omission, or more likely a strain-induced rearrangement of the ring shape or loss of a single oligonucleotide (half-tile) to correct the 45^{o} torsional strain built into the design. Adhesion force micrographs of the wheels shows high contrast with the background surface, with the holes showing the surface characteristics of the underlying mica. These force maps give greater clarity to height scans that may otherwise conflate the holes with surrounding wheel tiles. In sum, the T-junction wheel structure can be assembled in various buffers with reasonable yield, though the resulting structures are highly subject to variation by defect or depositional strain.

Table 5.7: TJ-W Standard (Hamada et al)[182]

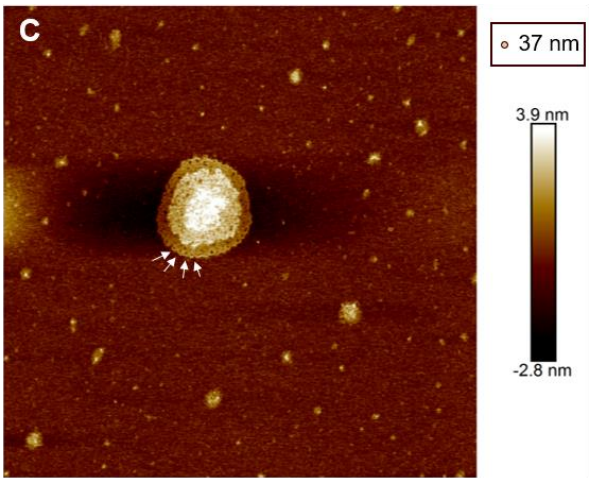
Sequence #	Nucleotide sequence
TJ-W-Std (-)	5'-CGCTTCGTTTGCGGAACTGGAGATGAGCCATTACCGAGTAGGTGGACAGACC-3'
TJ-W-Std (+)	5'-TCCACGGTCTGCTACTCGGTAATGGCTCATCAAGCGTCCAGTTCCGCAAACG-3'



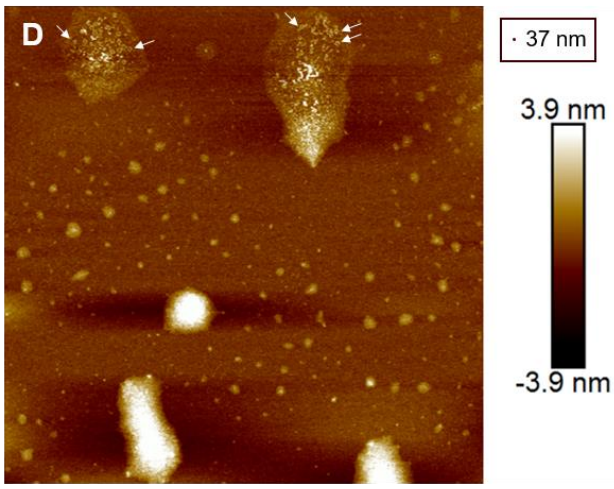
3 μm x 3 μm scan



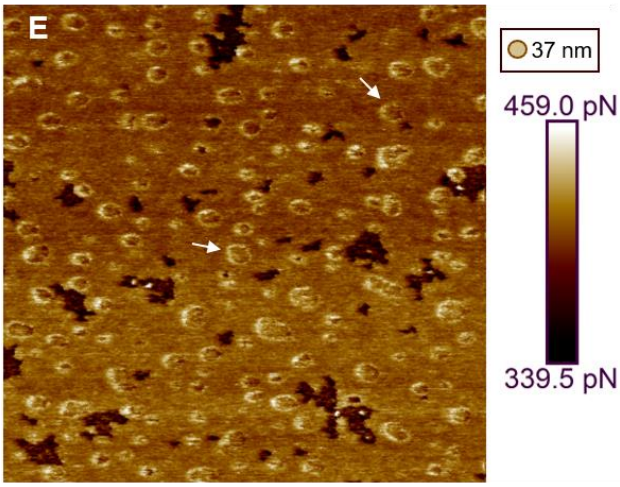
3 μm x 3 μm scan



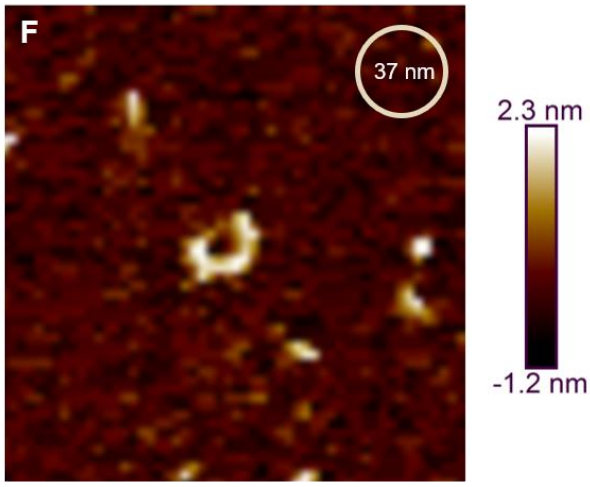
3 μm x 3 μm scan



5 μm x 5 μm scan



1 μm x 1 μm scan



200 nm x 200 nm scan

Figure 5.22: AFM micrographs of T-junction wheels with standard base chemistry. **A)** Anneal in Tris Na^+/Mg^{2+} with no surface dilution. Ring shapes are clearly visible on the surface, with a hole at the center of many (but not all) wheels. A random sample of 20 wheels was measured, in the same manner as by Hamada,[182] and an average diameter of 37.4 nm with a standard deviation of 5.5 nm was calculated. The predicted diameter is ~37 nm (inset to scale), and Hamada et al measured 36 ± 2.5 nm for the same nanostructure. Three wheel structures (white inset, 1-3) are highlighted to illustrate continuity. Wheels can be seen that include too many or too few T-junction units, and occasionally broken rings can also be observed. **B)** The same image is shown with its nanomechanical adhesion map, more clearly demonstrating the holes at the center of each ring. The same wheel structures and scale are inset. In this image, aggregated wheels can be clearly identified by their regions of reduced adhesion at the edges and centers of the shapes, while the height map conflates the objects. **C)** Annealed in Tris Na^+/Mg^{2+} without dilution or vigorous mixing, a mass aggregate of wheel structures can be seen. Each wheel is identifiable by the hole at its center (examples illustrated by arrows, inset). The rings proceed in layers with successive stacks up to three wheels in height at the center of the bundle. **D)** Similar conditions in MOPS buffer produce aggregates that are less tightly bound (arrows inset). **E)** When annealed without Na^+ and diluted on the surface, good separation between the rings can be seen (arrows inset) in an adhesion force micrograph. **F)** A high resolution scan of a single wheel in these conditions shows reasonable agreement with the expected shape (inset), with a lower density top left edge, due to a missing tile, movement of the ring during scanning, or the loss of a single oligonucleotide (half-tile) during deposition to alleviate torsional strain. This ring measures 37 nm at its widest (Feret) diameter and 27 nm at its shortest. This type of wheel object suggests a good amount of structural heterogeneity in this flexible tile assembly. All samples annealed with 50 correction cycles in their respective buffers.

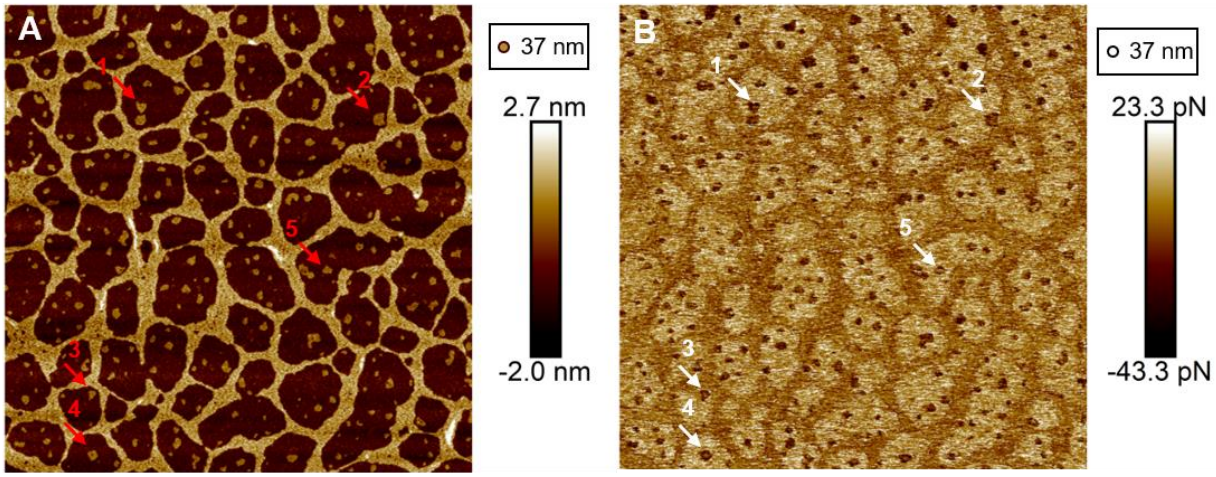
b. CC-tile

Two variations of the TJ-W tile were designed, the first with 6 CC mismatches endemic to both sticky ends, forcing dC:Ag^+:dC bond formation for rings to assemble, and the second with an additional 6 mismatches extended into the tile center (Table 5.8). Modifications to the 12CC wheel occurred on only one strand, allowing the TJ-W-12CC(+) strand to anneal with the TJ-W-6CC(-) complement. Annealing was carried out in the same conditions as the standard tile, with 2x Ag^+ added to the reaction. The panel of images in Figure 5.23 shows that wheel nanostructures are produced in these reactions with reasonable yield. As before, adhesion force maps of the structures served to enhance contrast between them, the substrate, and other debris (Figure 5.23A,B). The rings obtained in the 6CC design follow the same size heterogeneity, with some appearing to be larger than their intended size (scales inset in Figure 5.23), and others possessing an astigmated phenotype due to either ring torsion or loss of a half tile or full tile during annealing and deposition. Ring windows can be seen in Figure 5.23D-F, and faintly in C, showing the structural similarity between the modified wheels, the standard wheels, and the prediction.

Table 5.8: TJ-Wheel-CC (6CC, 12CC) (CC mismatches bold)

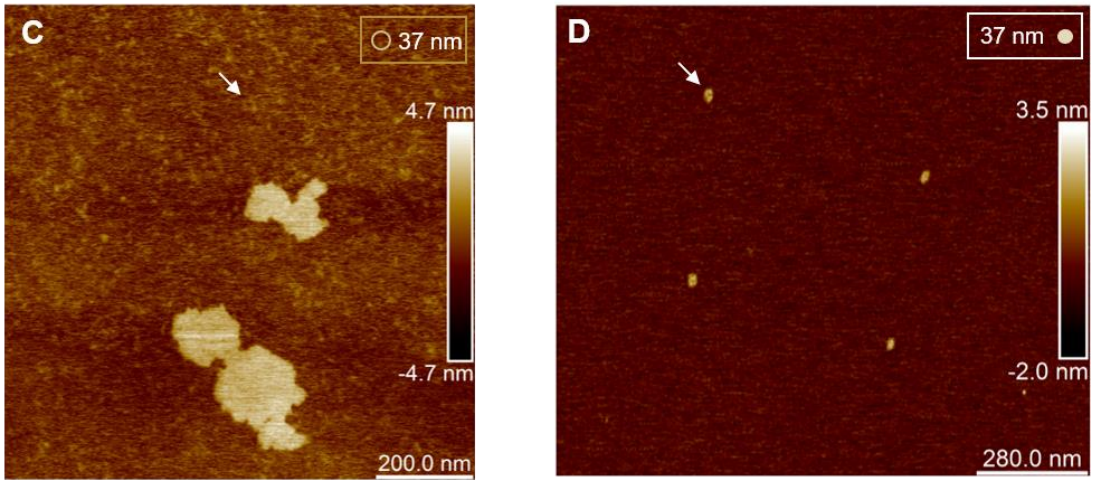
Sequence #	Nucleotide sequence
TJ-W-6CC (-)	5'-C**C**CTTCGTTTGCGGAACTGGAGATGAGCCATTACCGAGTAG**C**T**CC**ACAGACC-3'
TJ-W-6CC (+)	5'-TCCACGGTCTGCTACTCGGTAATGGCTCATCAA**C**C**C**TCCAGTTCCGCAAACG-3'
TJ-W-12CC (+)	5'-TCCAC**CC**TCT**C**CTACTC**CC**TAAT**CC**CTCATCAA**C**C**C**TCCA**C**TTCC**C**CAAAC**C**-3'

MOPS(-)(+)7

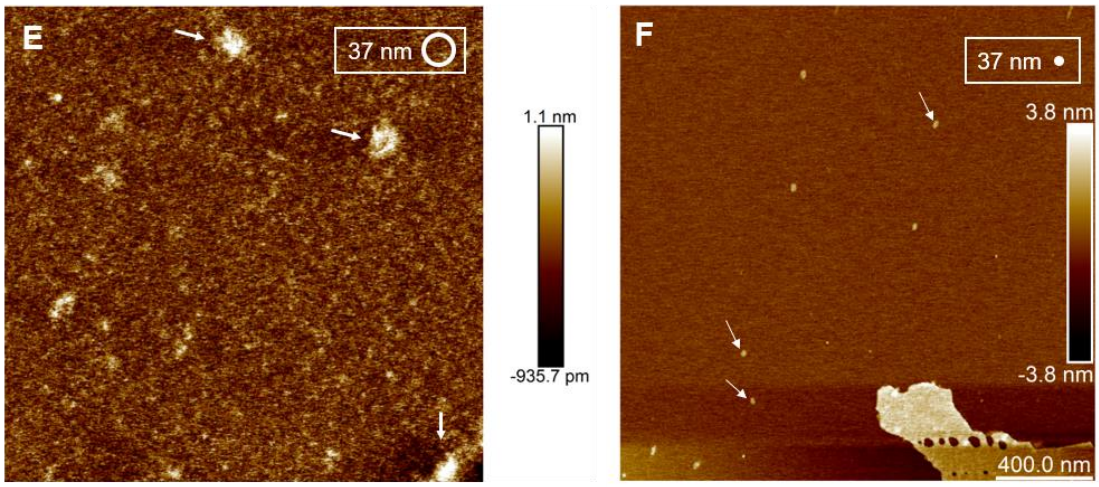


2 μm x 2 μm scan *2 μm x 2 μm scan*

Tris(-)(+)7



1 μm x 1 μm scan *1.5 μm x 1.5 μm scan*



650 nm x 650 nm scan *2 μm x 2 μm scan*

Figure 5.23: AFM micrographs of T-junction wheels with 6 dC:Ag^{+}:dC bonds. **A)** HiRS-SAA tip scan of wheels annealed in MOPS Mg^{2+} with no Na^{+}. Expected ring size (inset) matches wheels in the windows between aggregated DNA. Five structures (red arrows inset, 1-5) are illustrated. **B)** The same image is shown as an adhesion force map. Wheels are much clearer, with background DNA possessing different mechanical properties. The wheels shown in A (white arrows inset, 1-5) are more apparent, with emergent holes at the center of the rings. **C)** Annealed in Tris Mg^{2+} without Na^{+}, a number of rings can be seen faintly against the background (arrow inset), obscured by tall features at the center of the image. **D)** In the same reaction diluted on the imaging surface twofold, four wheels can be seen with faint windows at the center (arrows, inset). They possess an astigmatic shape similar to the ring in Figure 5.22F. **E)** A similar sample shows three rings with the expected holes (predicted size inset), and several smaller ring-like structures that failed to anneal, were broken on deposition, or broken during scanning. **F)** A wider angle scan of D shows more rings of similar size distributed across the imaging surface. All samples annealed with 2x Ag^{+} molarity and 50 correction cycles.

When analyzed quantitatively, the resulting rings appear to match predictions with a Feret diameter of 38.7 nm and a standard deviation of 3.6 nm (Figure 5.24). The size range is tightly clustered around the predicted shape, with a smaller error than in the standard rings. The images shown in Figure 5.24 also show some smaller debris from half-formed or broken rings, which were not included in the calculations.

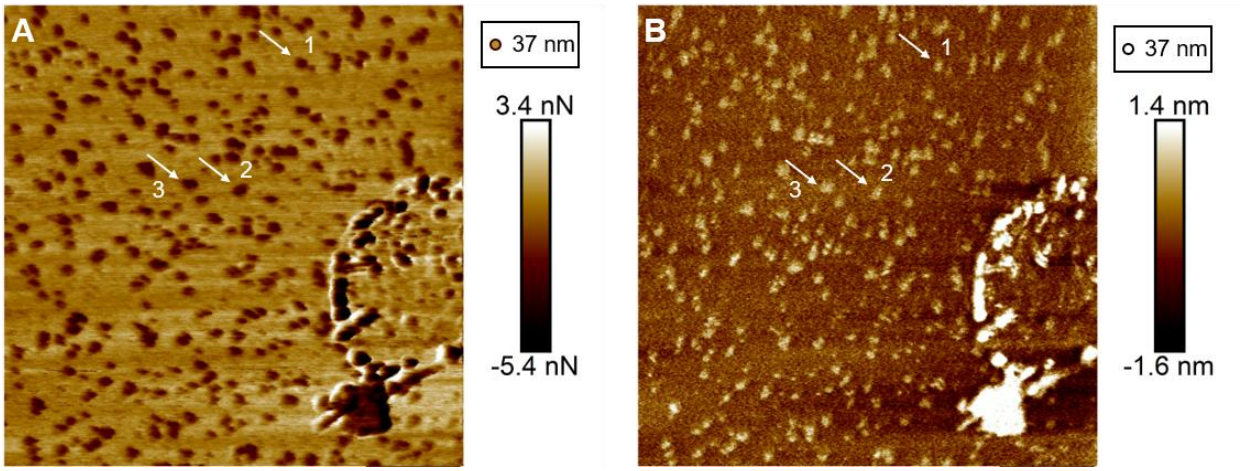


Figure 5.24: T-junction wheels with CC bonds annealed in Tris Mg^{2+} without Na^{+} over 50 correction cycles. **A)** Adhesion map of TJ-W nanostructures. Objects in micrograph appear circular in nature and fit the expected dimensions of wheel structures (inset). **B)** Height map of the same micrograph shows hole-like features at the center of the wheels. Three wheel structures are identified (white arrows) to demonstrate the shape change between height and mechanical maps. A random sample of 20 wheels, taken in the same manner as Hamada,[182] gives an average wheel diameter of 38.7 nm with a standard deviation of 3.6 nm. The predicted diameter is ~37 nm (inset), while Hamada measured diameters of 36 ± 2.5 nm in their study.

Overall, the yield of orthogonal-chemistry ring-like nanostructures is quite good. This assembly uses Ag^{+} bonding as a necessary component to successfully polymerization, and, as with the brick wall tiles, does so with a nonlinear SE/KL bundle that departs from a traditional b-form helix. While nanoscale, single-tile features were not observed, the high fidelity of the ring diameter to designed dimensions implies that tile assembly was carried out without large deviations. The T-junction tile provides a template for both fixed-shape (wheel) and open-ended (brick wall) nanostructure assemblies. Both design types have a place in the creation of novel nanoelectronic structures, with rings potentially acting as monodisperse nanomaterials that can be embedded on surfaces or into larger brick-wall- or Holliday-junction-style

lattices. The incorporation of silver ions into these T-junction nanostructures was successful, and may provide a powerful template for future algorithmically-driven designs.

5.4. Double crossover (DX) units: rigid DFX and DAO structures

Double crossover (DX) tiles comprise a family of highly-successful structures that have come from the Seeman lab. First described in 1998, these structures are built around parallel double helices which exchange strands at rotational periods in multiples of 10.5 bp.[148] The key difference between the DX tile and the HJ tile lies in the frequency of the crossover—DX crossovers occur in pairs of two, forcing the structure to remain axially rigid in the direction of helical alignment. The second crossover collapses Holliday junction X-like structures into parallel, bound helices.

5.4.1. General design

The DAO and the DFX tiles both use the same type of crossover symmetry, with (A)ntiparallel helices and (O)dd numbers of helical bundles (DAO) (Figure 5.25). Both types of tiles involve sequence exchange between adjacent, rigid DNA helices. DFX involves a large number of base pairs (189 nt per helix), while the DAO molecule utilizes much shorter strands (32 nt per helix). They are used in different ways to incorporate Ag^+ conductive pathways. Due to sequence design constraints, DAO becomes the better design for conductive polymer tiles.

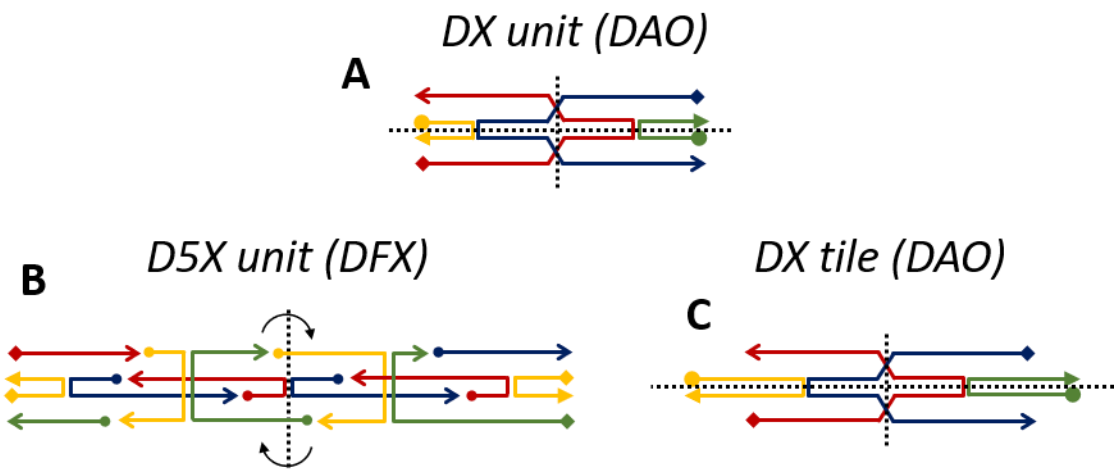


Figure 5.25: DX tiles designed and utilized in this work. **A)** General double crossover, antiparallel, odd bundled tile geometry (DAO). **B)** The DFX tile employs five crossovers with this symmetry, while the DAO tile in **C)** employs two crossovers. DAO subunit shown with sticky ends that allow lattice polymerization.

5.4.2. Five crossover tiles (DFX)

The DFX nanostructure was first designed by Dr. Risheng Wang.[138,185] It consists of a DX-style unit, with two parallel double helices exchanging strands over five crossover points. The structure is comprised of twelve oligonucleotides comprising 189 nt from end to end. The overall measured length of this structure is 61.4 nm ± 6.1 nm and is therefore visible via AFM.[185] This structure has been used as a platform for carbon nanotubes, avidin-bound plasmonic nanoparticles, and has been used as a bridge between gold nanodots. Traditional sequences from the literature were modified with Dr. Wang in order to introduce three highly-cytosinated regions in the structure: C_{10}, C_8W_4, and $[C_2W]_5$ (Table 5.9). Each of the regions contains 10 dC:Ag^+:dC bonds (Figure 5.26), and the modified structure is consequently named DFX-3C_{10}.

Table 5.9: DFX[185] to DFX-3C_{10} sequence modifications (alterations bold)

Region	Sequence #	Nucleotide sequence
$[C_2W]_5$	6*	5'-TCCTGCTTCGTGTCCAGGTAACTGCGA--**CCACCTCCACCTCCA**--3'
	5*	5'-**TCCACCTCCACCTCC**—TCGCAGCCTTCAGAGCAGCGTAGAAGAAGCTGGAAGGT GTTCCACACAGTCGCTCTCAGGCGT-3'
C_{10}	3*	5'-CGGTGAATGGCACCACCGACAC—**CCCCCCCCCC**—AACGACTATCGCACTCAA GAAACCACTGGGTACAAGCACAAC-3'
	4*	5'-TCGTT--**CCCCCCCCCC**—GTGTCGGATGAGCATGGGTCCCGTTGTGCTTGTACC CAGTGGTTGAACACCTTCCAGCT-3'
C_8W_4	10*	5'GGGACCCATGCTCATGTAT--**CCTTCCCTCTCC**—GTTGACCACGCACAGAGTCCA ACTGCCTGGTCGTCTCTGGAGC-3'
	9*	5'-GTCAAC--**CCACACCCAACC**—ATACTGGTGCCATTCACCGGCTCCAGAGACG ACCAGGCAGTCATCTCCTTCGGACA-3'

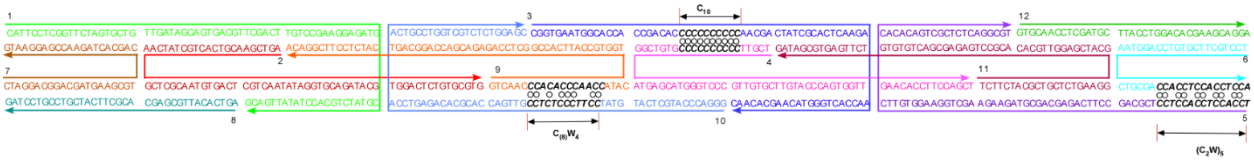


Figure 5.26: Design of DFX-3C_{10} structure with three high-cytosinated regions, each containing 10 dC:Ag^+dC bonds. Two of the regions (C_{10}, C_8W_4) are inside the structure, while the last region ($[C_2W]_5$) lies at the terminal end of the structure.

As a test of silver ion incorporation, the structure was annealed in a two-stage process. The twelve component oligonucleotides were ordered from IDTDNA and reconstituted to a concentration of 100 µM (10 µL / nmol) in MilliQ water. In a single tube, 2 µL (0.2 nmol) of each sequence was mixed together in a PCR tube with 100 µL MOPS buffer (10 mM MOPS, 100 mM Na^+, 12.5 mM Mg^{2+}), suspended in a 2 L

water bath, and cooled from 95 °C to 25 °C in a styrofoam chamber over 48 hr. A secondary anneal was carried out, with either 0x, 1x, or 10x Ag^+/CC pair added (1x = 6 nmol), cooling from 40 °C or 60 °C over 48 hr.

While the structure itself is quite rigid, 10-15 bp holes introduced when Ag^+ is not present may affect the overall cohesion of the structure, reducing yield. The purpose of the secondary anneal was to introduce equimolar silver to the ion incorporation sites and close non-rigid areas of the nanostructure. Resulting products were deposited on freshly-cleaved mica, dried, washed 3x in distilled water (wicking off water to remove salts), and subsequently imaged on a Bruker Dimension Icon AFM.

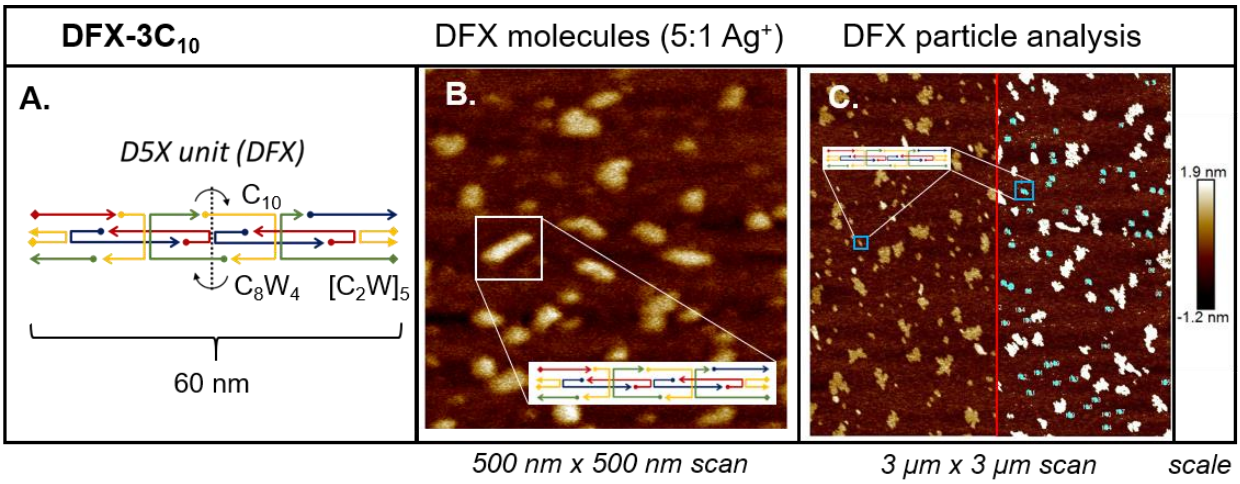


Figure 5.27: AFM analysis of DFX-$3C_{10}$ assembly. **A)** Design of the DFX variant, with the altered regions marked. **B)** Typical best micrograph of DFX-3C10 molecule after 2-step annealing in 5:1 Ag^+:CC, with approximately 10 DFX candidates present (inset). **C)** Image thresholding for particle size counting (red line shows image halves before and after particle count applied), with DFX candidates marked (inset). Scale bar shows approximately 3 nm in height difference. Image taken with 2 nm silicon nitride tip (ScanAsyst) on freshly-cleaved, dry mica.

AFM images were processed using ImageJ in order to extract particle size analysis of DFX-like objects (Figure 5.27C, Figure 5.28). Firstly, the images were exported as 500 DPI *.tif* files, cropped to 1668x1668

pixels, and scaled to the appropriate line resolution (1668 pixels/ 5000 nm, etc.). Images were then set to 8-bit greyscale and subjected to thresholding and background removal. Particle analysis was carried out in broad terms, setting the acceptable particle range to 30-10,000 nm^2 (Table 5.10). Results were exported to Microsoft Excel and further pared by their aspect ratio and Feret diameters, or minimum and max line length bounded by the object. Particles larger than 20 nm were marked as DNA events, while particles between 50-75 nm max Feret diameter and <50 nm min Feret diameter with aspect ratios of >1.4 were marked as DFX events. The expected diameter of the DFX object of ~60 nm x ~6 nm is in agreement with this bound, with nanostructure width a complex conflation of tip nominal width, surface pitch, and surface dissipation (force capacitance). In general, this analysis is bounded for 'roughly elliptical objects of near 60 nm length.' Close analysis of high-resolution images suggests that this analysis overestimates the number of successful events by approximately 50%. All AFM micrographs were subjected to similar deposition and analysis techniques.

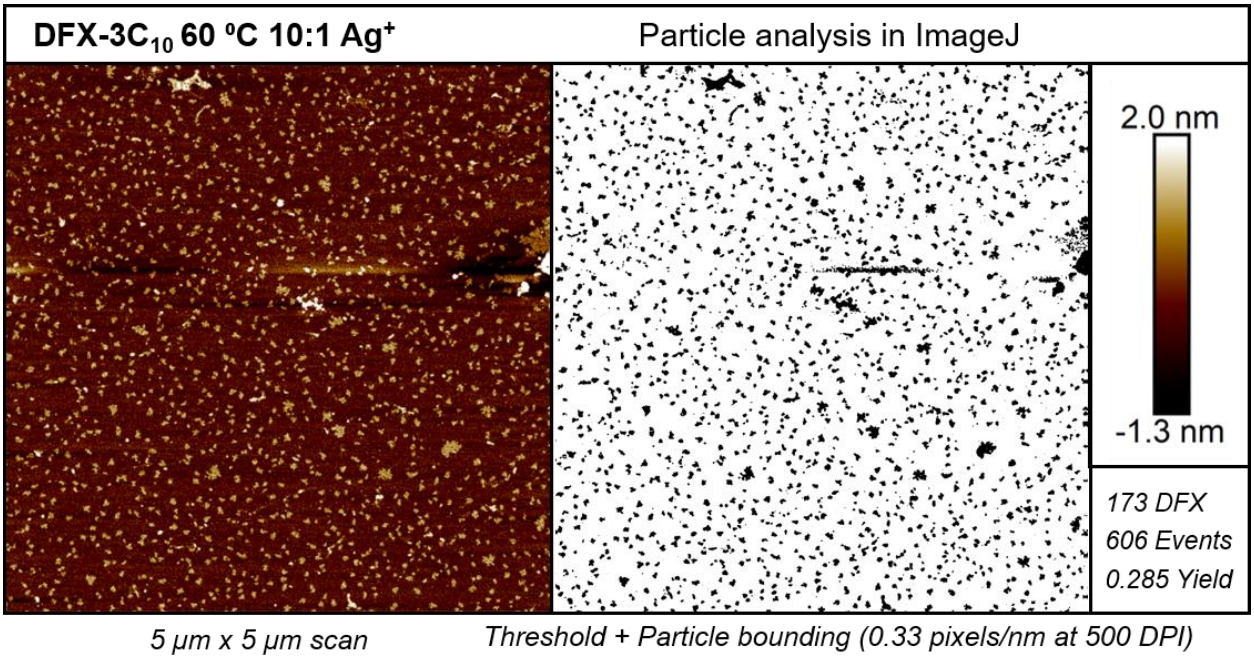


Figure 5.28: DFX-$3C_{10}$ structure with three $[dC:Ag^+:dC]_{10}$ regions with 10:1 Ag^+ added in secondary anneal with slow cool to RTP from 60 °C imaged via AFM. Raw, first-order flattened image (left) analyzed

in ImageJ for particles (right) produced 606 total events, with 173 probable DFX objects for an overall yield of 0.285 (inset).

Table 5.10: DFX Analysis Parameters

Particle Analysis	**Setting**
Image Resolution	500 DPI
Pixel Count per Edge	1668 pixels
Min Particle Area	30 nm²
Max Particle Area	10000 nm²
Successful DFX Parameters	
Min Aspect Ratio	1.4
Max Feret Min Diameter	50 nm
Min Feret Max Diameter	50 nm
Max Feret Max Diameter	75 nm
DFX-like Object Parameters	
Min Feret Max	20 nm

Table 5.11: DFX Yield via AFM

Reanneal Temp	**Ag^+ ratio**	**# Images**	**Total Area**	**Total Events**	**Num Pass**	**Pass Rate**
40 °C	0:1	3	9 um	1008	133	0.132
40 °C	1:1	4	11 um	1953	134	0.069
40 °C	10:1	7	21 um	2407	276	0.115
60 °C	0:1	3	12 um	2313	164	0.071
60 °C	1:1	4	14 um	1572	102	0.065
60 °C	10:1	5	13 um	1455	364	0.250

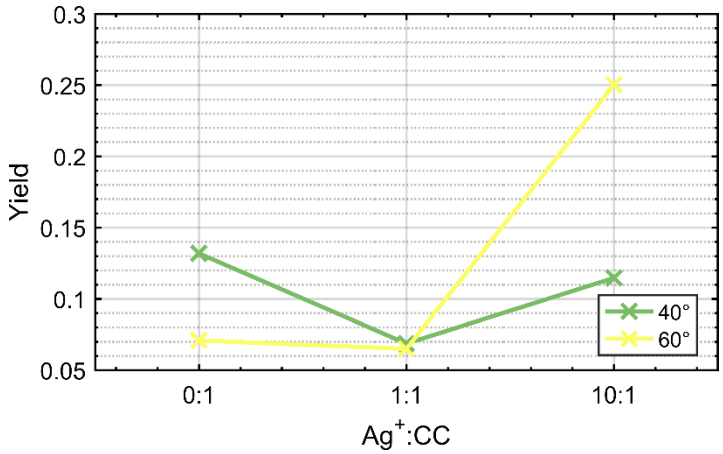


Figure 5.29: DFX yield in with variable Ag^+ to mismatch ratio (x-axis) from AFM image analysis (Table 5.10). Ag^+ is added in a secondary anneal at 40 °C or 60 °C over 48 hr. Best yield is observed in 10:1 Ag^+ 60 °C (see Table 5.11), suggesting successful ion incorporation.

In this analysis, the nanostructure yield was between 7-13% for re-annealed DFX in the absence of Ag^+. Intuitively this suggests that rigid DX tiles will be less successful in assembling when the component helices are disrupted. These data are also supported by DFX annealed in a 1-step reaction, where Ag^+ is included during primary annealing. In these conditions, some structure forms, but the size is highly irregular, indicating that base parity in non-optimized sequences for large structures is disrupted through the addition of the dC:Ag^+:dC bond. In both temperature regimes, the addition of equimolar Ag^+ reduced the yield averaged across the images, though some 2 µm micrographs did experience 20% overall yield (~10 DFX events). In all cases, the addition of 10x Ag^+ caused higher yield than 1x Ag^+, and in 60 °C annealing conditions, the total yield was 25%, with a total of 364 DFX-like objects. This is more than 3x greater than the yield of 0x Ag^+ at that temperature.

Aqueous Ag^+ may either form a pseudo-covalent bond with a cytosine mispair or an electrostatic interaction with negatively charged phosphate groups on the DNA backbone. When the molarity of Ag^+ is low, many ions may be trapped in backbone interactions and fail to reach the mismatch sites. By

contrast, molar excess of Ag^+ may allow the ions to penetrate the mostly-formed structures and promote bond formation. A temperature increase during this secondary anneal had a dramatic effect on yield in excess Ag^+ conditions, causing a rise from 11% to 25%. This suggests that helix penetration may indeed be limited by energetics, and that greater strand flexibility and lower backbone affinities experienced at higher temperatures may, in fact, promote structural alignment and dC:Ag^+:dC bond formation. This suggests that the DFX molecule is amenable to alteration through silver ion pairing, and that the effects are not an artifact of counterion stabilization.

Low overall yield is likely due to disruptions from holes left by unpaired CC bonds that are larger than the helical turn length of 10.5 bp, as well as a generally-reduced WC reaction favorability in the presence of Ag^+. The DFX molecule may be much more successful if subjected to sequence optimization for dimer reduction in silver-pairing conditions. Due to the number of component oligonucleotides and the overall lack of structural features visible to AFM, this optimization was not carried out. The DFX molecule serves as a good test case for ion incorporation, but other crossover tiles with greater shape diversity and shorter component strand lengths may be better initial test cases for electronic functionalization. However, if electronic behavior and yield is improved in other structures, the closed topological shape of the DFX molecule may provide a strong candidate for fixed-size device fabrication and site-specific nanomaterial integration with gold nanoparticles, carbon nanotubes and other DNA phenotypes.

5.4.3. Double crossover tiles (DAO)

To assist with the design of successful DNA nanostructures with the dC:Ag^+:dC base pair, a collaboration was carried out with Dr. Ruojie Shah and Prof. Nadrian Seeman at NYU. A set of successively more functionalized DX tiles were designed with non-standard helical elements to demonstrate incorporation of novel pairing regions. At this stage in the research process a full computational model was available for the first time. Consequently, all of the oligonucleotide sequence design in this section was carried

out algorithmically using rational design principles and connectivity prescribed through the collaboration.

a. General design

The (D)ouble crossover (A)ntiparallel tile with (O)dd helical periodicity was first described in 1999 by Winfree et al.[148] This DNA tile consists of two adjacent duplexes of 64 bp in length that exchange strands at two crossover points (Figure 5.30A). When functionalized with 5 nt sticky ends, a tile (A) may bond with a similar tile with differing sticky ends (B) (Figure 5.30B). These monomers chain together to form a rigid lattice without visible windows (Figure 5.30C).

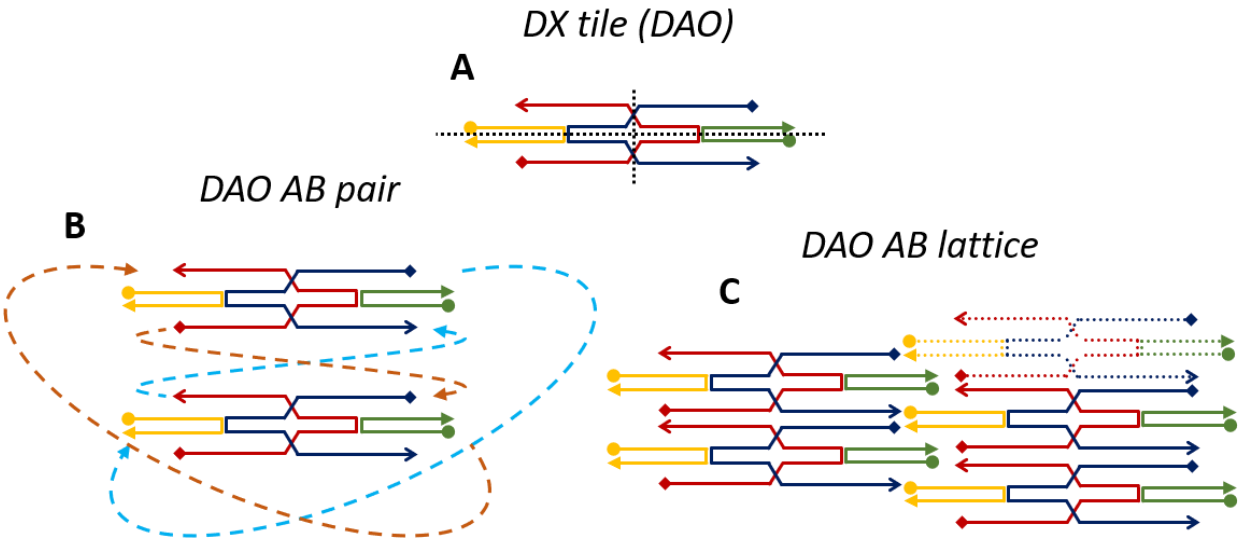


Figure 5.30: Design of DAO tile and AB lattice. **A)** The basic DAO tile, as designed by Winfree et al,[148] consists of two antiparallel duplexes in similar orientation that exchange strands at two crossover points. **B)** The tile does not possess the ability to polymerize with itself in an antiparallel fashion, but can be paired with a similar tile containing different sticky ends to form a two-unit (AB) pair. **C)** These pairs polymerize into sheets without visible holes or windows.

As a result of discussions with Dr. Shah, a set of conserved regions and pairing rules was established as essential components of the DAO tile, while other base pairs are open to sequence modification (Figure

5.31). A basic frame was established and modified in a stepwise fashion to produce more exotic lattices. The introduction of a CC mismatch was first carried out between the duplexes as a non-helical spacer. The vision for this base pair was to serve as a conductive bridge between helices with potentially different electronic properties. The embedding of an RNA-like cross pair would serve as a novel type of nanostructured component, and may present a relatively small perturbation in the overall tile shape. Several iterations of designs were carried out with differing numbers of CC mismatches, structural hairpins, and, most importantly, a titration of nucleotides into the central beams (N5,N14 in Figure 5.31) to account for possible rotational differences in dC:Ag^+:dC helices suggested by TJ-BW results in Figure 5.20.

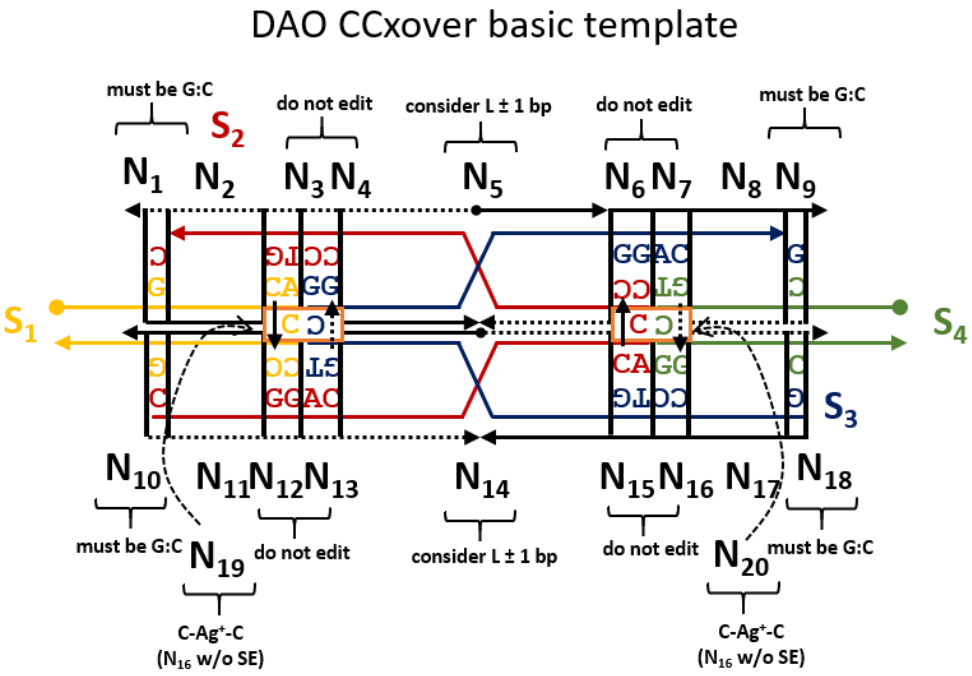


Figure 5.31: Template for modification of the DAO tile with CC bonds. The sequences (SX) are marked as well as nodes for computational design (NX). Design considerations at each point are marked. Design is based on discussions with Dr. Ruojie Shah.

b. DAO-CCxover tile

The first application of the design frame in Figure 5.31 was to fill in the sequences algorithmically without structural modification. Oligonucleotides were computationally derived using a standard genetic algorithm setup (default values in Table 4.7) to modify only the regions marked editable in Figure 5.31 (Table 5.12). Analysis of the fitness criteria is shown in Table 5.13, where there are only five dimers of size 5 nt. Structures were annealed in 1x MOPS (+)(+) 7 with varying ion availability. Assembly is assayed using AFM and PAGE (Figure 5.32). Yield across all Ag^+ environments is high, with uniform DAO tiles of the appropriate size and shape. Some AFM micrographs of the 10x Ag^+ condition shows apparent tile size to exceed the designed 12 nm, but corroborating gel evidence shows that this is likely an artifact of the imaging process or possibly Ag^+ salt accumulation on the DAO surface.

Table 5.12: DAO-CCxover tile sequences (CC bold)

Sequence #	Nucleotide sequence
DAO-CCx-S1	5'-GTAACACAC**C**CAGATAG-3'
DAO-CCx-S2	5'-CTATCTGGACTAAGTAGACAATCA**C**CCAAACTATGACATCCTGTGTTAC-3'
DAO-CCx-S3	5'-GTGAATCCTGATTGTCTACTTAGT**C**GGATGTCATAGTTTGGACTTGTTG-3'
DAO-CCx-S4	5'-CAACAAGT**C**GGATTCAC-3'

Table 5.13: DAO-CCxover modeling results

Histogram bins for n-mers	1-mers	2-mers	3-mers	4-mers	5-mers
Dimers	1630	518	168	44	5
Purine repeats	18	8	10		
Pyrimidine repeats	16	10	8	2	
AT repeats	18	8	14		

GC repeats	36	4	4
Guanine repeats	18	4	

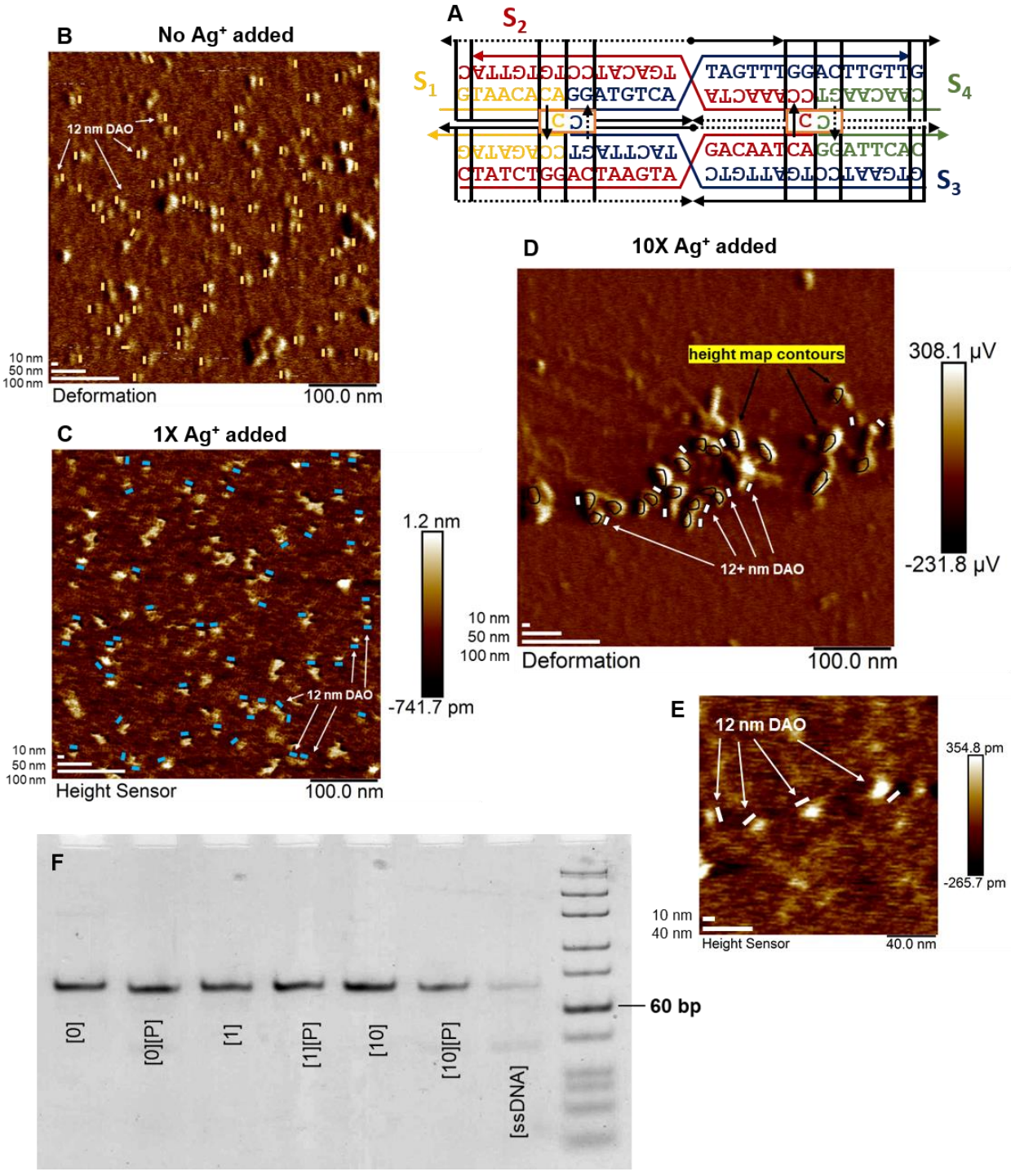


Figure 5.32: Experimental results of the DAO-CCxover tile. **A)** Design of the DAO-CCxover tile is carried out algorithmically (see template in Figure 5.31). A CC mismatch is introduced at the center of each crossover, adding a non-helical spacer with a dC:Ag^+:dC base pair to join the parallel duplexes. Sequences are color coded, and algorithmic design nodes are cordoned by black lines. **B)** The tile is able to successfully assemble in the absence of aqueous Ag^+ (AFM deformation map shown), indicating that the crossover perturbation is not sufficient to disrupt normal DAO assembly. This also demonstrates the viability of a tile with sequences derived entirely from computational modeling. Scale of DAO tile inset (yellow). **C)** In the presence of unitary Ag^+, DAO tiles assemble at the same size scale and apparent yield as the no Ag^+ control. Scale inset adjacent to successful structures (blue). **D)** In the presence of 10x excess Ag^+, tiles assemble with slightly larger apparent size, likely due to Ag^+ aggregation near the structures during drying. Predicted size is inset (white), and contours from the atomic force height (black circles) are overlaid on the deformation force map shown. **E)** Atomic force height is shown for 10x Ag^+ conditions, with four successful DAO tiles shown (scale inset, white). **F)** PAGE results of the different experimental conditions show uniformly assembled DAO tiles. Ion availability is indicated: [0], [1] and [10] for no Ag^+, 1x Ag^+, and 10x Ag^+ conditions, respectively. Tiles are also subjected to silver chloride precipitation [P] without perturbation of the structures. The control lane with only unpaired oligos [ssDNA] shows some unfolded product as well as poorly formed proto-structures assembled in the gel lanes. Stain carried out using StainAll (Thermo Fisher) in collaboration with Dr. Roujie Shah in the laboratory of Prof. Ned Seeman.

As a basic template for modification, the DAO-CCxover tile seems to behave as designed. Importantly, the introduction of the non-helical CC crossover does not interfere with tile formation. It remains unclear, however, whether the ions are successfully accumulating within the complex. To assay this nanostructure further for Ag^+ incorporation, more CC bonds are introduced in critical regions of the nanostructure.

c. DAO-CCclamp tile

Design is identical to the DAO-CCxover tile (Figure 5.32A), with the addition of two features. First, sticky ends are added to allow polymerization between tile pairs. More fundamentally, the crossover region is mutated to contain three CC mismatches in each duplex. The conserved sequence [CAGG:GTCC] is changed into [CACC:CTCC]. With the inclusion of the nonhelical C:Ag^+:C bond at the core of the crossover, there is a total of seven C:Ag^+:C bonds in each crossover region. It is expected that this level of disruption will impede formation of a 2D lattice without successful incorporation of Ag^+.

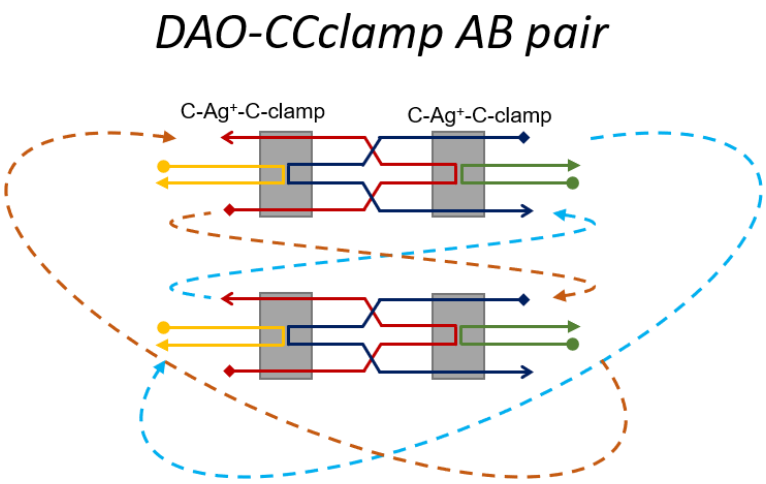


Figure 5.33: Design of the DAO-CCclamp AB tile pair. The 'CC clamps' are marked in silver, and represent a mutation of the [CAGG:GTCC] conserved region into [CACC:CTCC]. The non-helical C:Ag^+C bond at the core of each crossover is retained from the DAO-CCxover design (Figure 5.32A).

All sequences in this structure were designed computationally with the genetic algorithm parameters in Table 4.7. To reduce synthesis costs associated with oligonucleotide ordering and purification, the computational model was modified to allow a single node to include two versions of itself, or two sequences. This allowed for the swapping of sticky ends and the assay of fitness across both sticky ends without changing the internal design of the tile.

Analysis of T-junction brick wall lattices demonstrated that the rotational period of Ag^+ DNA duplexes may differ slightly from B-form DNA. It was hypothesized that the increased bond strength from the metal pair,[77] in tandem with metallophilic interactions between adjacent ions,[78] would render the duplex stiffer and increase its rotational period. Structural studies of stacked dC:Ag^+:dC bonds were carried out via NMR, not X-ray crystallography, and as a result there is no information on the rotational period of this orthogonal pair. Successful nanostructures in this work used polycytosine linkers with rotational spacers (HJ fracture, HJ-SS), or tiles with flexible secondary structures that may absorb differences in stiffness without great deviation (TJ-BW, TJ-W). The DAO tile, on the other hand, is a far more rigid tile that depends primarily on helical rotation for successful assembly. As a result, Dr. Shah suggested the titration of additional base pairs into the DAO tile to account for any deviations from 10.5 bp/ helical turn that might arise. To perform this titration, optimal sequences for tiles of standard dimension were computationally derived (Table 5.14); all nodes other than N5 and N14 were marked uneditable (see Figure 5.31); nodes N5 and N14 were expended from 12 bp to 13 bp; and the structures were subjected to a second round of computational optimization. In this way, four distinct tiles were produced with two sets of sticky ends and two distinct rotational periods from only eight oligonucleotides (Table 5.14). In table 5.14, the sequences are named by their color in Figure 5.33 (red, blue, green, yellow), their tile type (A2,B2), and the length of the central beam (12 nt, 13 nt) where applicable. The combination of oligos and tiles to produce the desired lattices is shown in Table 5.15. Computational scores for the standard size tiles (AB2-12nt) can be found in Table 5.16, while results for the longer tile pair (AB2-13nt) can be found in Table 5.17.

Table 5.14: DAO-CCclamp tile sequences (SE underline, CC bold)

Sequence #	Nucleotide sequence
DAO-A2yellow	5'-<u>CTCTA</u>GTGATT**C**A**CCC**TTGATG<u>GTATG</u>-3'

DAO-B2yellow	5'-AGAACGTGATT**C**A**CCC**TTGATGAGTGA-3'
DAO-A2green	5'-TCACTCTGTAA**C**T**CCC**TAGTTCTCTTG-3'
DAO-B2green	5'-CATACCTGTAA**C**T**CCC**TAGTTCGAGAT-3'
DAO-A2blue12	5'-GAACTACCTCTTAGTCATGTTT**C**T**CCC**ATAGTAGACATTCCACTTACAG-3'
DAO-A2blue13	5'-GAACTACCTCTACTATGTTGTTT**C**T**CCC**AAGTTTGATACATCCACTTACAG-3'
DAO-A2red12	5'-CATCAACCACAAACATGACTAA**C**A**CCC**AATGTCTACTATCCTCAATCAC-3'
DAO-A2red13	5'-CATCAACCACAAACAACATAGTA**C**A**CCC**ATGTATCAAACTTCCTCAATCAC-3'

Table 5.15: DAO-CCclamp tile annealing mixtures

Tile name	*Nucleotide sequence mix (first appearance bold) (95 ^{o}C to 25 ^{o}C, 72 h)*
DAO-A2-12nt	**DAO-A2yellow, DAO-A2green, DAO-A2blue12, DAO-A2red12**
DAO-B2-12nt	**DAO-B2yellow**, **DAO-B2green**, DAO-A2blue12, DAO-A2red12
DAO-A2-13nt	DAO-A2yellow, DAO-A2green, **DAO-A2blue13, DAO-A2red13**
DAO-B2-13nt	DAO-B2yellow, DAO-B2green, DAO-A2blue13, DAO-A2red13
Lattice name	*Tile mix (50 ^{o}C to 25 ^{o}C, 48 h)*
DAO-AB2-12nt	DAO-A2-12nt, DAO-B2-12nt
DAO-AB2-13nt	DAO-A2-13nt, DAO-B2-13nt

Table 5.16: DAO-CCclamp AB2-12nt tile pair modeling results

Histogram bins for n-mers	1-mers	2-mers	3-mers	4-mers	5-mers	6-mers	7-mers	8-mers
Dimers	2093	727	243	97	**18**			
Purine repeats	21	14	3					
Pyrimidine repeats	17	8	4	3	2	2	1	1

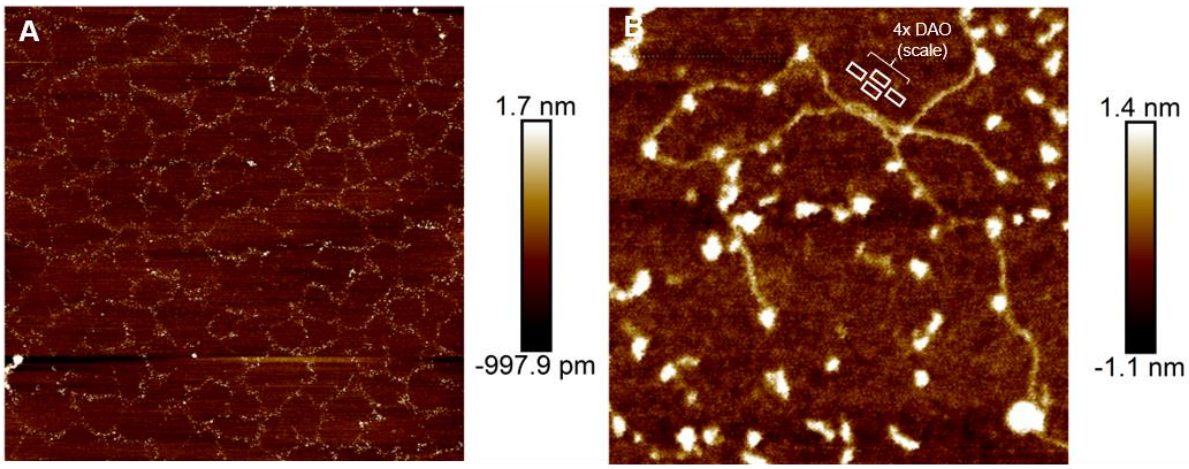
AT repeats	21	14	13
GC repeats	42	5	4
Guanine repeats	15	1	

Table 5.17: DAO-CCclamp AB2-13nt tile pair modeling results

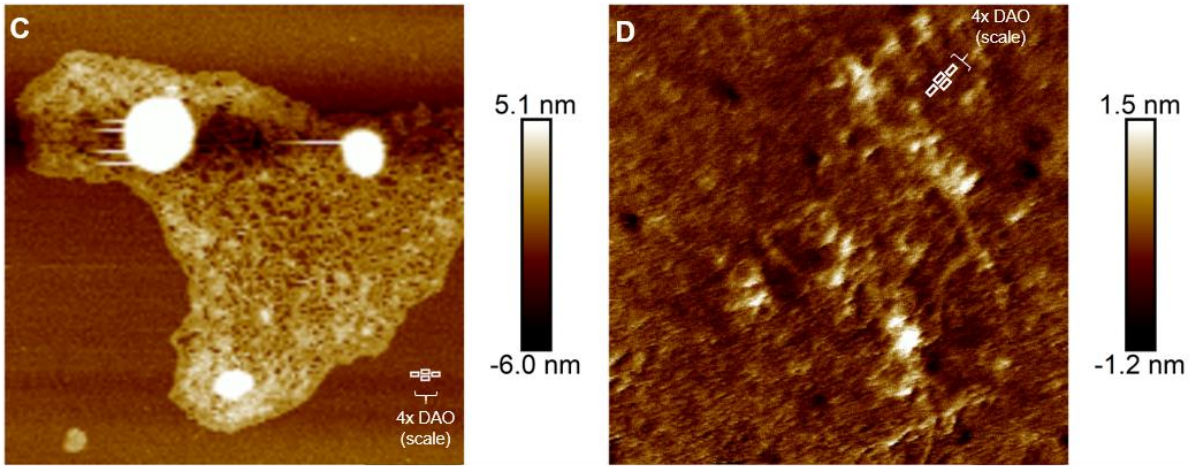
Histogram bins for n-mers	1-mers	2-mers	3-mers	4-mers	5-mers	6-mers	7-mers	8-mers
Dimers	2243	734	248	102	**23**			
Purine repeats	24	12	4					
Pyrimidine repeats	19	8	5	2	2	1	2	1
AT repeats	17	18	13					
GC repeats	42	5	4					
Guanine repeats	15	1						

The structures were annealed according to the recipes in Table 5.15 in MOPS (+)(+) 7 buffer (10 mM MOPS buffer, 12.5 mM Mg^{2+}, 100 mM Na^{+}, pH 7.0). Structures were initially cooled from 95 ^{o}C to 25 ^{o}C in a water bath over 72 h. AB tile pairs were reannealed at from 50 ^{o}C to 25 ^{o}C in an insulated bath for 48 h, *OR* subjected to 20 cycles of 20 ^{o}C to 45 ^{o}C to 25 ^{o}C with a rate of 6 min/^{o}C in a tabletop PCR machine. Tiles were supplied with 10x Ag^{+} per tile, with 14 CC mismatches in each monomer. Both 12 nt (standard) and 13 nt (+1 nt titration) tile pairs were built, and were subsequently assayed by high-resolution AFM. Results for 12 nt tiles can be seen in Figure 5.34.

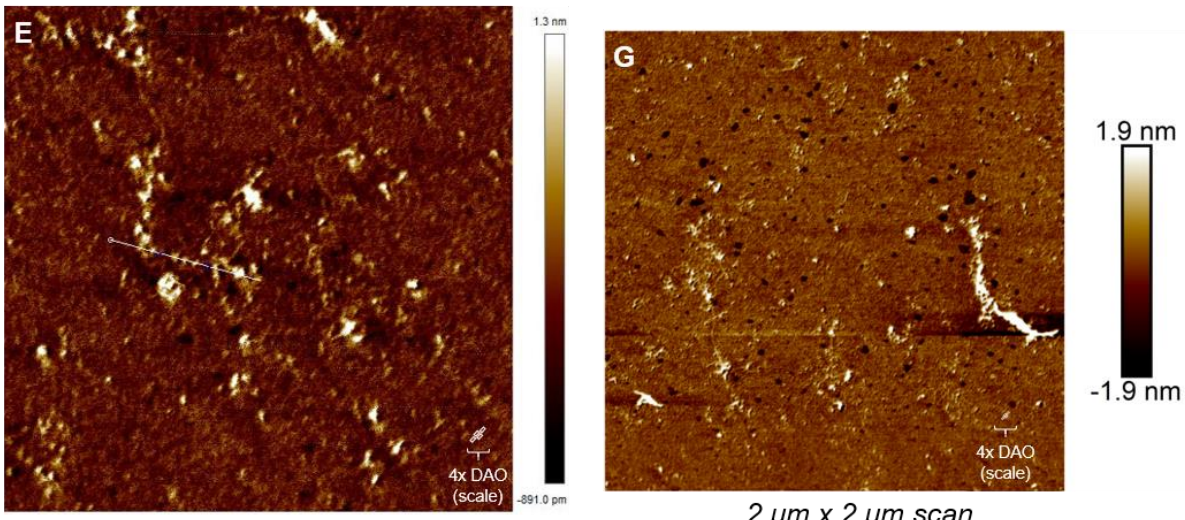
DAO-CCclamp AB2-12nt



10 µm x 10 µm scan *300 nm x 300 nm scan*



1.5 µm x 1.5 µm scan *500 nm x 500 nm scan*



1 µm x 1 µm scan *2 µm x 2 µm scan*

Figure 5.34: AFM micrographs of DAO-CCclamp AB 12nt tile lattices. Images A-C are obtained from water bath anneals, while imaged D-G are obtained from correction cycle PCR annealing. Scale of four bound DAO tiles (white, inset) is shown in images B-E and G. **A)** Evidence of widely-spaced, interrupted lattice is found in most images with fragile collections of AB pairs woven in a loose network and torn upon deposition. **B)** On closer inspection, individual tiles are connected axially to form linked bundles of DAO. No lattice orientation is visible, though some tile polymerization can be seen. These one- and two-tile wide assemblies hold together the larger structure, but it is clear that not all sticky ends are being used. **C)** At reduced image dilution, the bundles of DAO form a closely packed shape, again without clear lattice orientation. **D)** By contrast, networks annealed with correction cycles seem to be able to overcome kinetic barriers to lattice formation in order to produce small but two-dimensional assemblies of accurately placed DAO tiles. Network shown here extends between 200 nm and 300 nm in diameter. **E)** Though lacking an underlying long-range lattice foundation, bundles of well-formed lattices can be seen with appropriately-sized features. Line section shows height and spacing regularity within a small area in F. **F)** Height profile of line section in E. Set on a 45^{o} angle to the lattice orientation, feature width is not representative of the overall nanostructure, but height and periodic profiles of this section indicate good agreement with the design. YScale is from -1 nm to +1.5 nm in relative displacement. **G)** At a wider angle, these lattice bundles extend in a similar fashion across the surface suggesting uniformity of the lattice bundle phenotype.

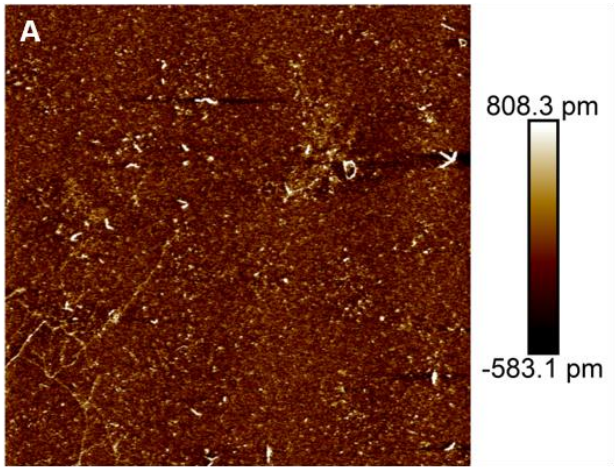
Overall, there are two distinct products in the 12 nt tile: water bath anneals produced a loosely-knit tile assembly that does not constitute a successful lattice; while correction cycle annealing produces a tightly bound lattice with round bundles of ~300-500 nm arranged across the surface. The ability of correction cycle annealing to produce a more regular lattice suggests that there is a high kinetic barrier to overcome relative to the original reports in Winfree and Seeman's manuscript.[148] This suggests that the 12 nt central beam is not an appropriate length for regular assembly: while the rotational

perturbation within a single tile is not sufficient to inhibit assembly of monomers, the accumulated rotational distortion as a result of dC:Ag^+:dC bonds in the CC clamps prevents more than one of the sticky ends on each side of the tile from being used at the same time. In this manner, water bath assemblies are only able to produce loose networks of non-lattice webs. By contrast, the thermally-cycled correction annealing produces far better lattices of proper shape, suggesting that the added coaxing allows bond displacement of Ag^+ or a WC pair in the right places along the structure to allow tight coordination. In this way, two distinct yields are attained: bath anneals show good design fidelity in the monomeric range, but prevent long-range order. Correction cycle anneals allow long-range order, but likely cause displacement of ions or bonds within monomers to allow this growth. As a result, the 12 nt tile is a reasonably successful first step, but requires length-based alterations before a lattice of high nanoscale and lattice-scale yield is attained.

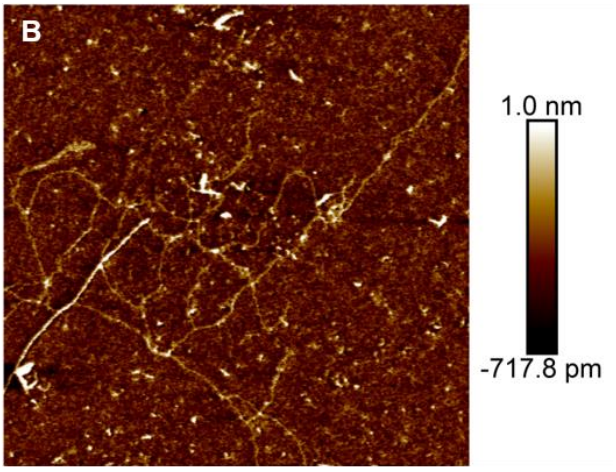
To account for differences in rotational period due to stiff dC:Ag^+:dC bond contribution, an extra base pair is titrated into both duplexes between the crossover (see scheme in Figure 5.31). AFM of these structures demonstrates more tightly-packed pseudo-lattices (Figure 5.35), as well a smaller, well-formed DAO AB lattices (Figure 5.36). In cases with pseudo-lattice formation, bundles of 2D lattices with some defects form the core of a cluster of DAO tiles radiating out with some (but insufficient) agreement to orientation (Figure 5.35D). These lattice nuclei (Figure 5.35E-F) represent a proto-lattice core with poorly-bound satellites, but contain well-formed DAO tiles with a clear growth axis inside a certain radius (< 500 nm). Within these pseudo-lattices, there continue to be spindle-like arrays of linear DAO, though at a higher density than in the 12 nt tiles (Figure 5.35 A,B vs. Figure 5.34B). By contrast, some imaging substrates contain higher yield lattices with good design fidelity: small tiles can be seen of the correct size and orientation that adjoin via sticky end to form micron-long sheets (Figure 5.36). Though these lattices appear to have a high yield, long-range order, and few defects, they remain thin

(2-4 tiles wide). The 13 nt DAO CCclamp tile is more successful than the 12 nt tile, but still may contain some rotational strain that prevents wide, flat sheets from forming.

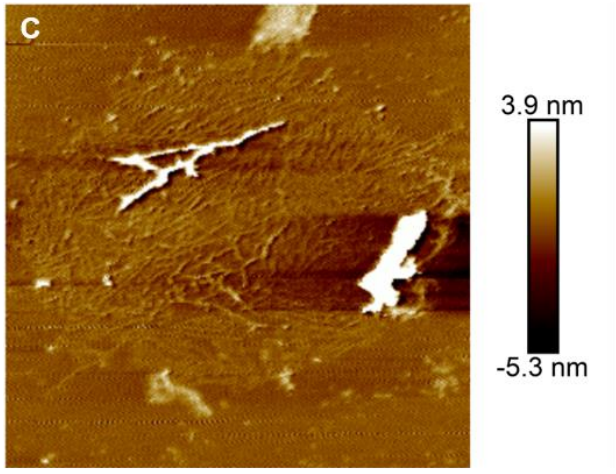
DAO-CCclamp AB2-13nt



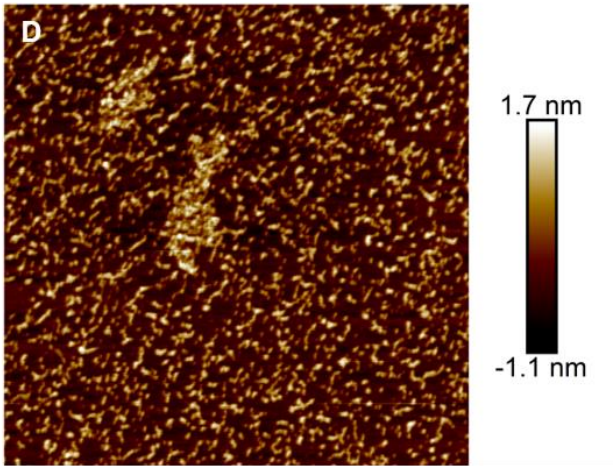
2 μm x 2 μm scan



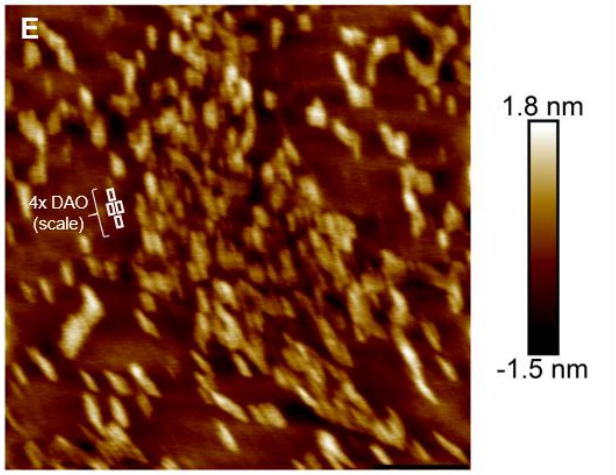
750 nm x 750 nm scan



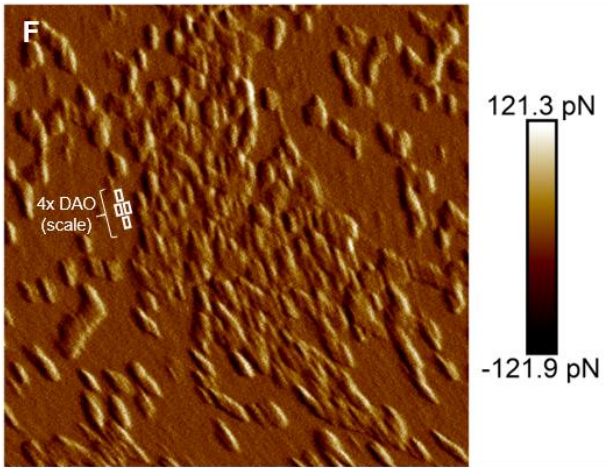
1 μm x 1 μm scan



2 μm x 2 μm scan



500 nm x 500 nm scan



500 nm x 500 nm scan

Figure 5.35: AFM micrographs of pseudo-lattice DAO-CCclamp-13nt tiles. **A)** When annealed with correction cycles, a mix of the two phenotypes attained in 5.34 is produced. Thin, one- and two-tile fibers connect groups of closely packed DAO tiles. **B)** A close-up of one spindle-like assembly shows lack of growth axis, but higher connectivity than in Figure 5.34A-C. **C)** When annealed in a water bath, a tightly-packed spindle assembly can be seen in which a growth axis may have been present in solution, but has torn on deposition. Lattice was at no point tightly packed. **D)** An extreme manifestation of the lattice-like assemblies in Figure 5.34 D, E and G, the tiles form a very-tightly knit, oriented structure with defects preventing fully woven lattices. The center of the image shows a lattice nucleus 500 nm tall, and the subsequently adjoined tiles are arranged tightly but not fully connected. **E)** A close-up image of a lattice nucleus is shown, with DAO tiles clearly visible (4x DAO tile scale inset, white). The image suffers from stretching due to cantilever feedback, but shows a well-knit lattice approximately 500 nm tall. **F)** PeakForce error (tip trace force with retrace subtracted) of the same lattice gives an extremely-clear view of the individual tiles (scale inset, white). Overall this lattice contains some defects, but is the beginning of a well-formed sheet.

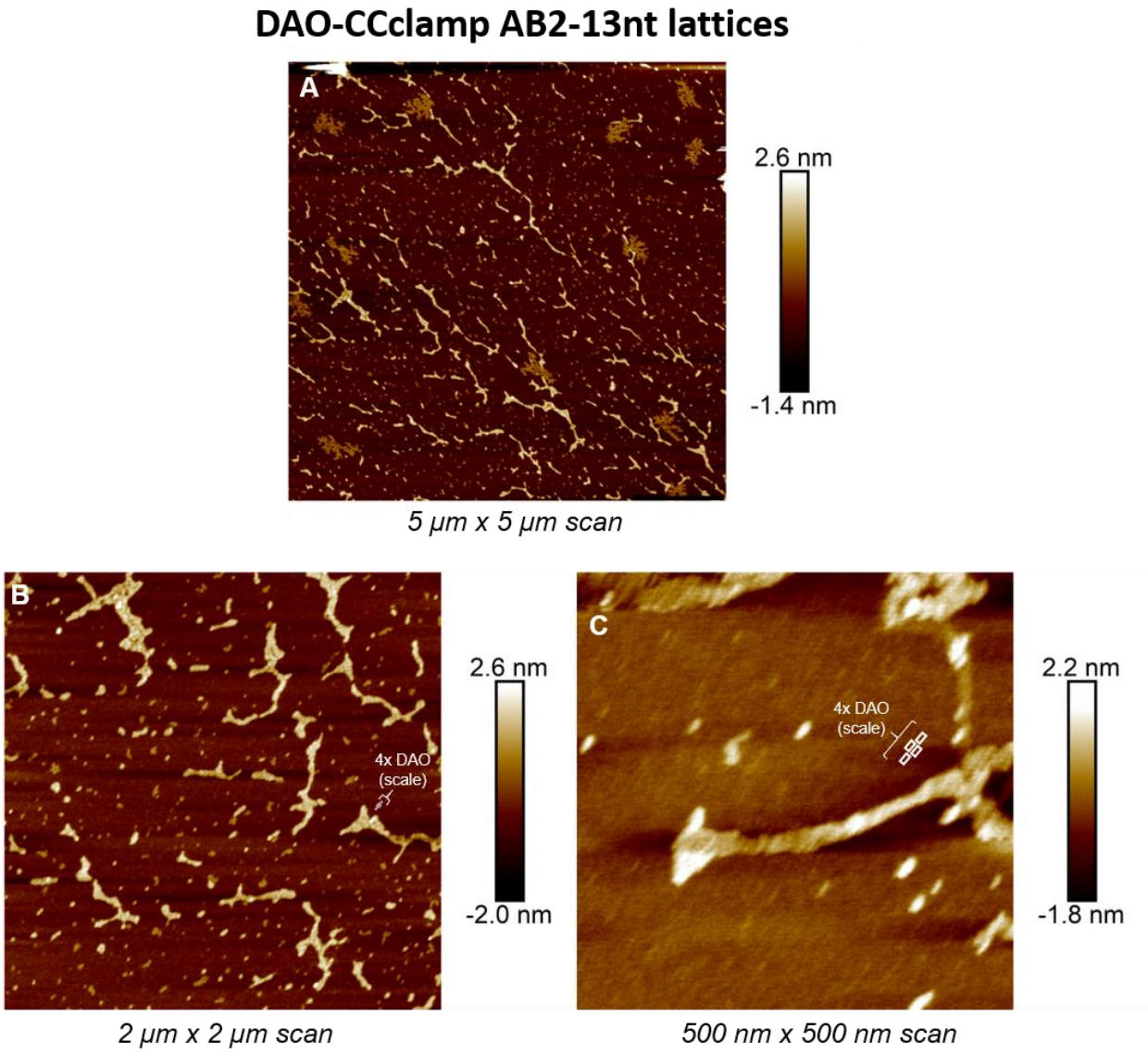


Figure 5.36: AFM micrographs of successful DAO-CCclamp-13nt lattices. **A)** Lattices are apparent over long range, having fragmented during drying. Presence of a well-defined growth axis in all the lattices (brighter features) suggests that they were part of a single structure in solution. **B)** Closer analysis of the structures shows defect-free assemblies: the lattices are closely packed with no windows. Edges of the tiles are apparent on close inspection (4x DAO tiles inset to scale, white). **C)** Close image shows 2-4 tile wide lattices extending at least 500 nm. Individual tile nodes are ~12 nm in size, compared to a predicted 10.2 nm.

d. DAO-CCclamp + hairpin tile

In keeping with original experiments by Winfree and Seeman[148] that utilized a hairpin secondary structure to produce differential height profiles on lattices in AFM, here we perform a similar experiment on the CCclamp tile base, replacing the A2 tile with an AH design (Figure 5.37). Similar to the experimental results from the CCclamp tile, a single base pair was added to each duplex to form a 13 bp central arm between crossover clamps. All sequences were designed algorithmically, and can be seen in Table 5.18, along with their composite fitness score (Table 5.19). The structure is annealed in tandem with the DAO-CCclamp-B2-13 tile and mixed in a secondary reaction in an insulated water bath, cooling from 50 ^{o}C over the course of 72 h.

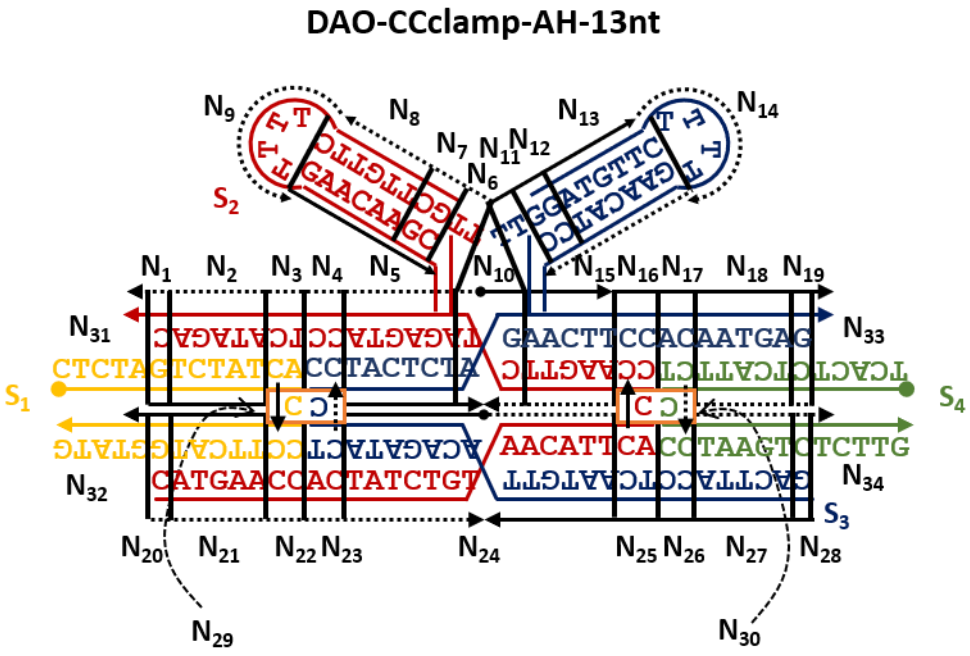


Figure 5.37: DAO-AH13 tile. Computational node numbers are indicated in black, directionality with arrows (solid for template strand, dotted for complement strand). Design differs from Figure 5.31 in the number of nucleobases in N24 and N5-N10-N15 (13 nt instead of 12 nt), the presence of a CC-clamp around the crossovers (N3/N4, N16/N17, N22/N23, N25/N26), and the secondary structure of nodes N6

through N14. General spacing and conserved regions derived from discussions with Seeman lab member Dr. Shah and the original report of DAO-H by the Seeman laboratory.[148]

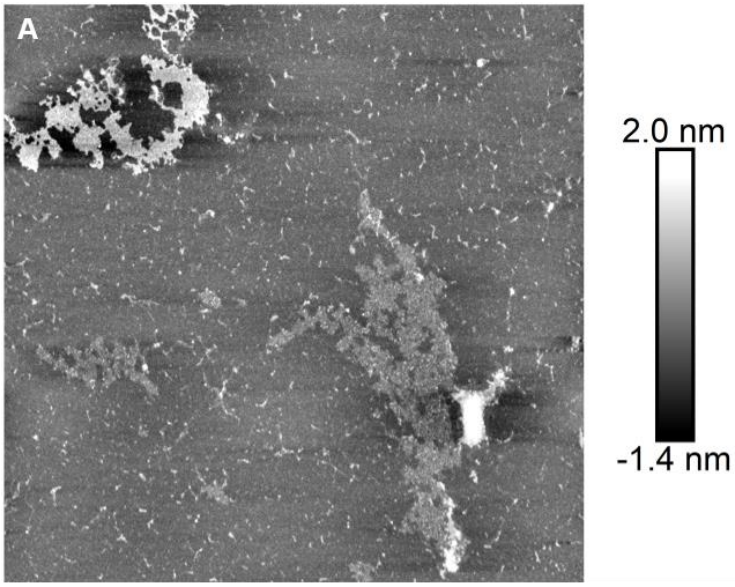
Table 5.18: DAO-CCclamp AH-13nt tile sequences (SE underline, CC bold)

Sequence #	Nucleotide sequence
DAO-AH yellow	5'-CTCTAGTCTAT**C**A**CCC**TTCATGGTATG-3'
DAO-AH green	5'-TCACTCTCATT**C**T**CCC**TAAGTCTCTTG-3'
DAO-AH yellow (blunt end)	5'-GTCTAT**C**A**CCC**TTCATG-3'
DAO-AH green (blunt end)	5'-CTCATT**C**T**CCC**TAAGTC-3'
DAO-AH red13	5'-CATGAA**CC**A**C**TATCTGTAACATT**C**A**CCC**AAGTTCTAGTTGCTTGTTCTTTTGAACAA GCAGTACCTCATAGAC-3'
DAO-AH blue13	5'-GACTTA**CC**T**C**AATGTTACAGATA**C**T**CCC**TACTCTAGATTGGATGTTCTTTTGAACAT CCACTTCCACAATGAG-3'

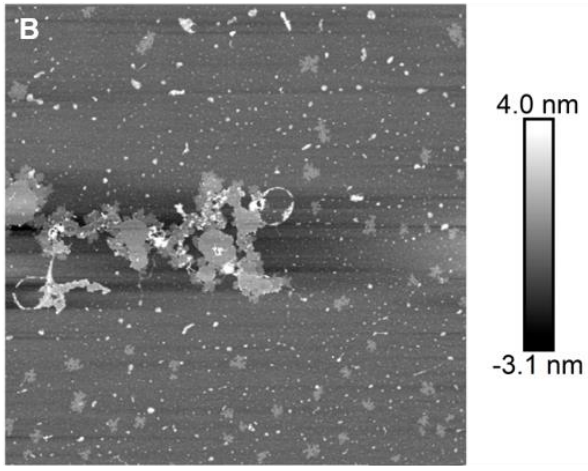
Table 5.19: DAO-CCclamp AH-13nt tile modeling results

Histogram bins for n-mers	1-mers	2-mers	3-mers	4-mers	5-mers	6-mers	7-mers	8-mers
Dimers	3719	1248	413	146	40	2		
Purine repeats	27	8	11					
Pyrimidine repeats	19	7	7	5	2	3	2	1
AT repeats	24	22	14	2				
GC repeats	52	9	14					
Guanine repeats	22	2						

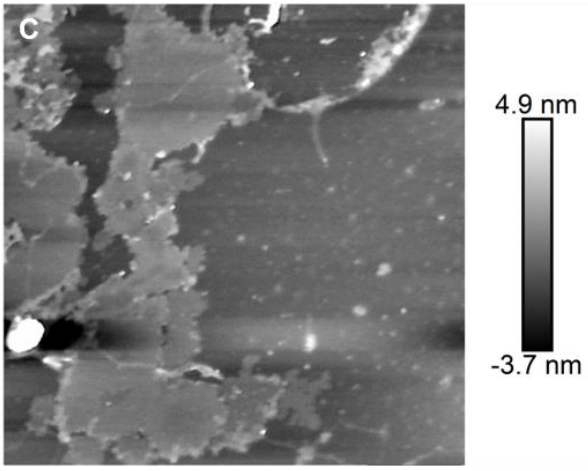
Experiments are carried out in two stages: DAO-B2-13nt tiles are annealed with 10x Ag^+ per standard reaction (water bath anneal cooling 95 ^oC over 72 h). DAP-AH-13nt tiles are annealed with and without both sticky ends and Ag^+, allowing for assessment of Ag^+ and AH tile dependence in lattice formation. AB pairs are mixed in subsequent reaction over 48 h. Results of the Ag^+/SE experimental condition can be found in Figure 5.38, while control experiments can be seen in Figure 5.39.



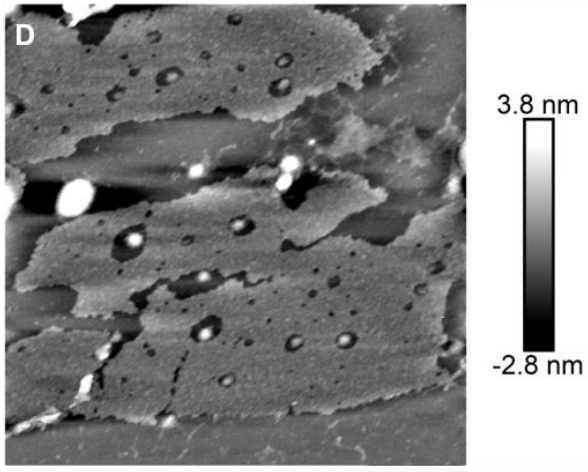
5 μm x 5 μm scan



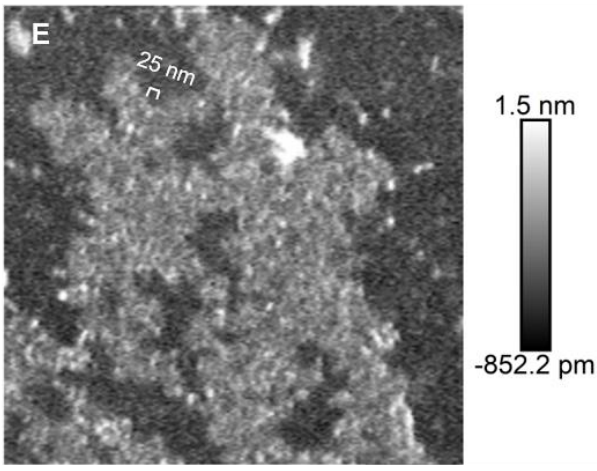
10 μm x 10 μm scan



2 μm x 2 μm scan



2 μm x 2 μm scan



1 μm x 1 μm scan

Figure 5.38: DAO-AH-13nt tile annealed with excess Ag^+. AH-13 tiles and B2-13 tiles are annealed separately and mixed in secondary reaction to form lattices. **A)** Cohesive lattice formation seen across many microns. Two distinct lattices can be seen, one upper left and the other center right. Though perturbed upon deposition, these lattices can be seen to grow up to 4 µm in overall size and up to 900 nm without grain disruption. **B)** Over a 10 µm scan, flat lattices can be seen clustered with few defects. Lattice sheets are approximately 1 µm in diameter. **C)** A close-up of B shows that these sheets are extremely flat and possess few defects. **D)** A smaller lattice sheet is imaged that has incurred more damage from deposition. Spectral period within the lattice is 11.9-12.6 nm, with a DAO tile size prediction of 12.5 nm. **E)** Single lattice sheet with defects shows an emergent hairpin grain of 25 nm (scale inset) across the nanostructure, with a predicted period of 25 nm.

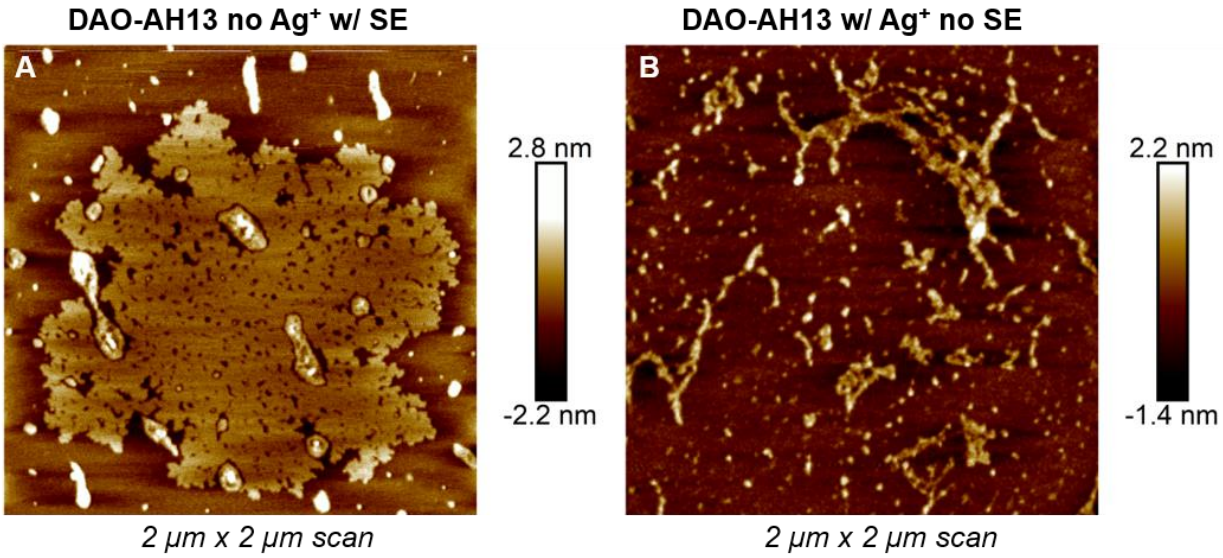


Figure 5.39: Control experiments for DAO-AH13 tile. B2-13 tile and various control AH-13 conditions are mixed in secondary reactions. **A)** Lattice annealed with B13-Ag^+ and AH13 without Ag^+ is able to coordinate sticky ends, but nanoscale features are scrambled with defects. Holes occur with 20-30 nm frequency, corresponding to a defect at the core of each AH13 tile unable to close the crossover region due to missing Ag^+. B tiles are annealed with Ag^+ and are able to coordinate some lattice development,

but grains are restricted to single tiles. **B)** By contrast, B13-Ag^+ tiles annealed with AH-blunt tiles are able to coordinate some stacking of disparate monomers, but are unable to grow meaningful sheets.

When annealed with both sticky ends and Ag^+, cohesive lattice growth is observed up to 4 μm, with depositional defects creating ~1 μm grains (Figure 5.38A-C). The lattice observed in Figure 5.38A is in good agreement with literature images of hairpin lattices (see Winfree, 1998, Figure 4A).[148] Close inspection of trace lines across these lattices shows a DAO tile period of 11.9 – 12.6 nm, a value in good agreement with the predicted 12.5 nm tile size (Figure 5.38D). Hairpin grains are not as dramatic as expected, but oriented hairpin features are observed in some lattices with a spacing of 25 nm, which corresponds to the predicted value (Figure 5.38E). Importantly, control experiments do not demonstrate the same features. When the AH-13 tile is deprived of Ag^+, sticky ends are able to coordinate a disordered lattice with defects occurring every 20-30 nm, or at the site of each AH-13 tile. These tiles are unable to clamp around the critical crossover regions due to the lack of Ag^+ during primary annealing, and are consequently left to open and close at random. The stabilizing effect of closed hairpins and leftover Ag^+ in the B2-13 tile solution allows for these tiles to function as linkers, but cohesive, defect-free lattices are not produced. By contrast, AH-13 tiles deprived of sticky ends do not form any lattices; rather these tiles coordinate with B2-13 tiles in short-range 1-2 tile stacks without a clear growth axis. This aggregation is expected with charged DNA tiles in the presence of excess stabilizing counterions (Ag^+, Na^+, and Mg^{2+}), and the lack of cohesive growth orientation underscores that these interactions do not mimic lattice behavior.

In sum, the DAO-AH13/B2-13 complex forms lattices at *high yield*: defects are minimal in sheets, with single grains extending to near-micron scale; while long-range order is observed across several micrometers. Future work could focus on attaining better resolution of the hairpin grain, but these lattices appear to be of much greater yield than other DAO-CC tiles described in this section. The increase in yield can likely be traced to the rotational flexibility afforded by AH-13 tiles containing two

hairpins with internal nicks. The introduction of these secondary features relaxes the overall strain on the lattice that may be imposed through C:Ag^+:C binding; and the increase in yield here suggests that further titration of 1-2 nt into the central arms of the DAO-CC tiles could further improve yield in CCxover and CCclamp AB pairs. Overall, this tile design represents a strong basis for future silver(I) DNA nanotechnology: its formation is predicated upon Ag^+ incorporation; rotational strain introduced by orthogonal binding chemistry is addressed through nucleotide titration and hairpin features; and it is the first nanostructure design with sequences fully developed by rationally-directed evolutionary search algorithms.

e. Future work

The DAO-CCxover tile was a first attempt to incorporate a non-helical C:Ag^+:C into a DNA nanostructure. It was shown that the presence of these Ag^+ binding sites did not disrupt the overall formation of the DAO species, both with and without Ag^+: the perturbation to tile structure was insufficient to prevent assembly in the absence of Ag^+ coordination (Figure 5.32F). The DAO-CCclamp tile, by contrast, required a titration of 1 nt into the central arm oligonucleotides to assemble lattice-like species, and even then the long range order was curtailed—likely due to rotational strain introduced by seven C:Ag^+:C bonds in close proximity at each crossover site (Figure 5.35E, Figure 5.36A). The DAO-CC-H tile type was met with much higher yield—lattices grew without defect across many microns, and were, in fact, only able to do so in the presence of Ag^+ and the correct sticky ends. As a direct conclusion, it may be surmised that the introduction of two hairpins into the upper duplex of the DAO-CCclamp-A13 tile allowed for strain relaxation and rotational correction across the nanostructure as a whole (Figure 5.38A). Indeed, structural analysis might conclude that the effective nicks produced by hairpin pinching were not completely closed, but rather remained open and flexible by an unknown number of base pairs (Figure

5.37). The start of each hairpin (Nodes N6, N11, Figure 5.37) includes, by design, two non-interacting nucleotides to promote this effect.

Future work would build on the success of these designs. The next tile, DAO-AB3, would incorporate the 'gap1' criterion from the bottom left of each tile (N22, N11, N12), through the nonhelical crossover (N19), and across the top middle and right of the tile (N4,N5,N6,N7,N8,N9,N23) (Figure 5.40). This design will harness the ability of the CC-crossover region to act as a non-helical continuation of ion stacking between duplexes. It is expected that if charge transport can be measured across this gap, this nanostructure could provide the basis for real nanoelectronic architectures. Work on the HJ fracture G4 tile (*Section 5.2.3.a*) suggests that exchange of central node base pairs for stacked G4 could introduce switching behavior into electronically conducting lattices for the construction of molecular transistors.

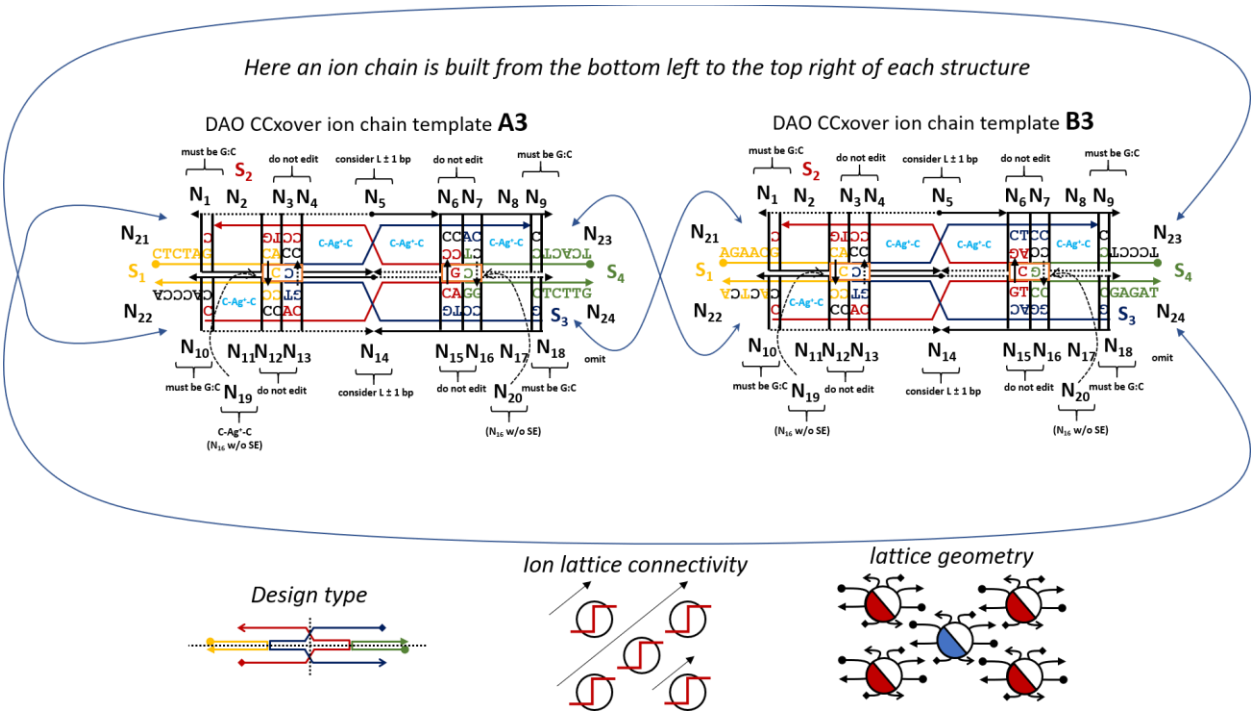


Figure 5.40: Design of a future DAO-AB3 tile pair with embedded Ag^+ extending from bottom left, through the crossover bond, and into the top right of the structure (blue ionic regions inset). Lattices would have an ion connectivity diagonal to the growth direction.

5.5. Unsuccessful nanostructure assemblies

Not all assemblies were successful. Here we illustrate the retired designs, important in their failure for informing what will and will not work in orthogonal chemistry structure design.

5.5.1. SX flare tiles (HJ-wide)

The flare or wide tile is a variation of the fractured Holliday Junction unit in which the horizonal arm consists of two helical turns to each side of the center, while the vertical beams consist of one helical twist (Figure 5.41). The purpose of this assembly was to use an early-generation genetic algorithm to devise non-interacting sequences to allow a 'gap1' C:Ag^+:C pattern in the vertical line (Table 5.20). Had this structure been successfully assembled, it would have consisted of nanowire-like sequences on the vertical axis and insulating AT-rich DNA on the horizontal axis.

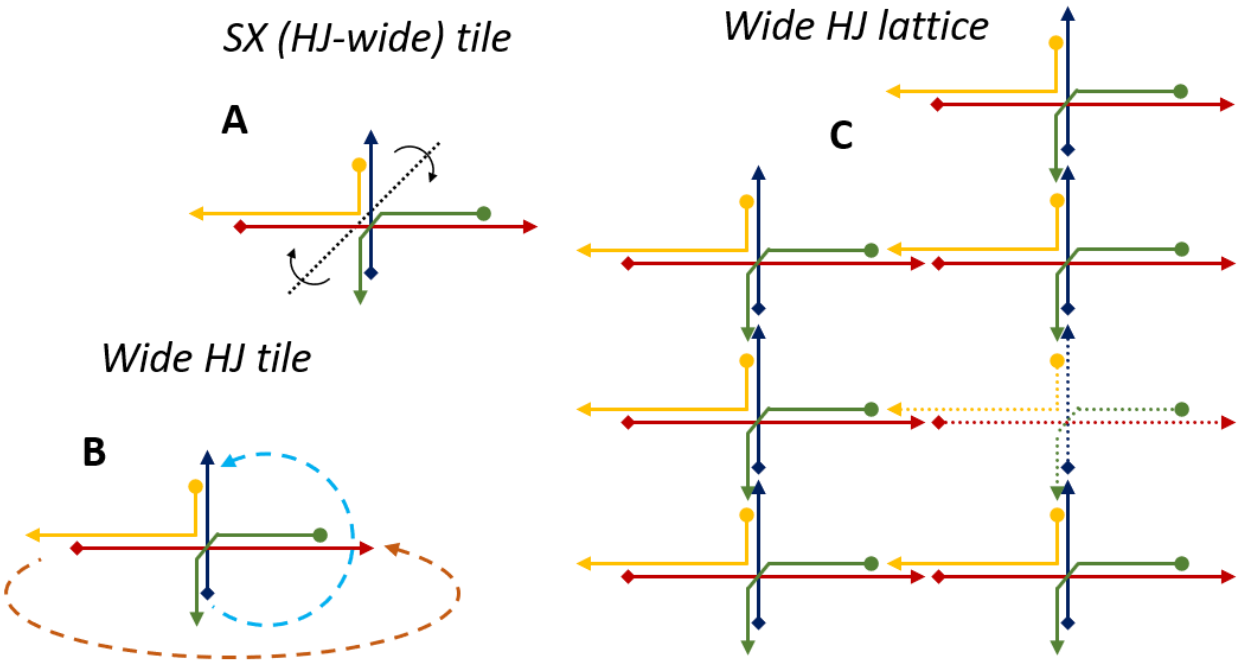


Figure 5.41: Design of the (unsuccessful) SX wide Holliday junction tile. **A)** Designed as a flared cross, the only internal symmetry is rotational. **B)** Sticky ends connect the top to the bottom, and the left to the right, without a mirror-flip of the structure. **C)** The lattice is designed to assemble in uniform rectangles,

with the vertical axis enriched with dC:Ag^{+}dC bonds. The lack of structural rigidity may have contributed to unsuccessful assembly.

Table 5.20: HJ Wide Cross (12CC) tile sequences (SE underline, CC bold)

Sequence #	Nucleotide sequence
HJW-1	5'-CAAAGAATCTTGACGAATTTACTTACAAATACACTACATAAC-3'
HJW-2	5'-T**C**T**CC**G**C**T**C**A**C**A**CC**A**C**G**CC**A**C**-3'
HJW-3	5'- A**C**A**C**T**CC**C**CT**C**C**AGTAAATTCGTCAAGATT-3'
HJW-4	5'-CTTTGGTTATGTAGTGTATTTGTAT**C**T**C**A**C**C**CC**-3'

The assembly of this structure was carried out in 100 µL MOPS (+)(+) 7.5 buffer (10 mM MOPS, 100 mM $NaNO_3$, 12.5 mM $MgSO_4$, pH 7.5) and silver ions were supplied via $AgNO_3$ at a 2:1 Ag^{+}:CC ratio for 12 sequence mismatched. Reactions were carried out in a water bath, cooling from 95 ^{o}C to RTP over 72 hr, with 0.2 nmol each DNA oligo. The structures attained had two phenotypes in the same way that the HJ fracture with linkers structure did, but the finger-like structures that indicated lattices in successful structures were rounder and possessed a multi-tiered shape with no discernable lattice morphology (Figure 5.42). Flat, round aggregates without a growth axis were present in higher concentration here than in successful species. The presence of the first, finger-like phenotype with long, directional features suggests that some ordered growth was able to occur, but no regular windows were attained. The structure could likely be corrected for rotation, rigidity and sequence promiscuity, but does not represent the highest priority nanostructure.

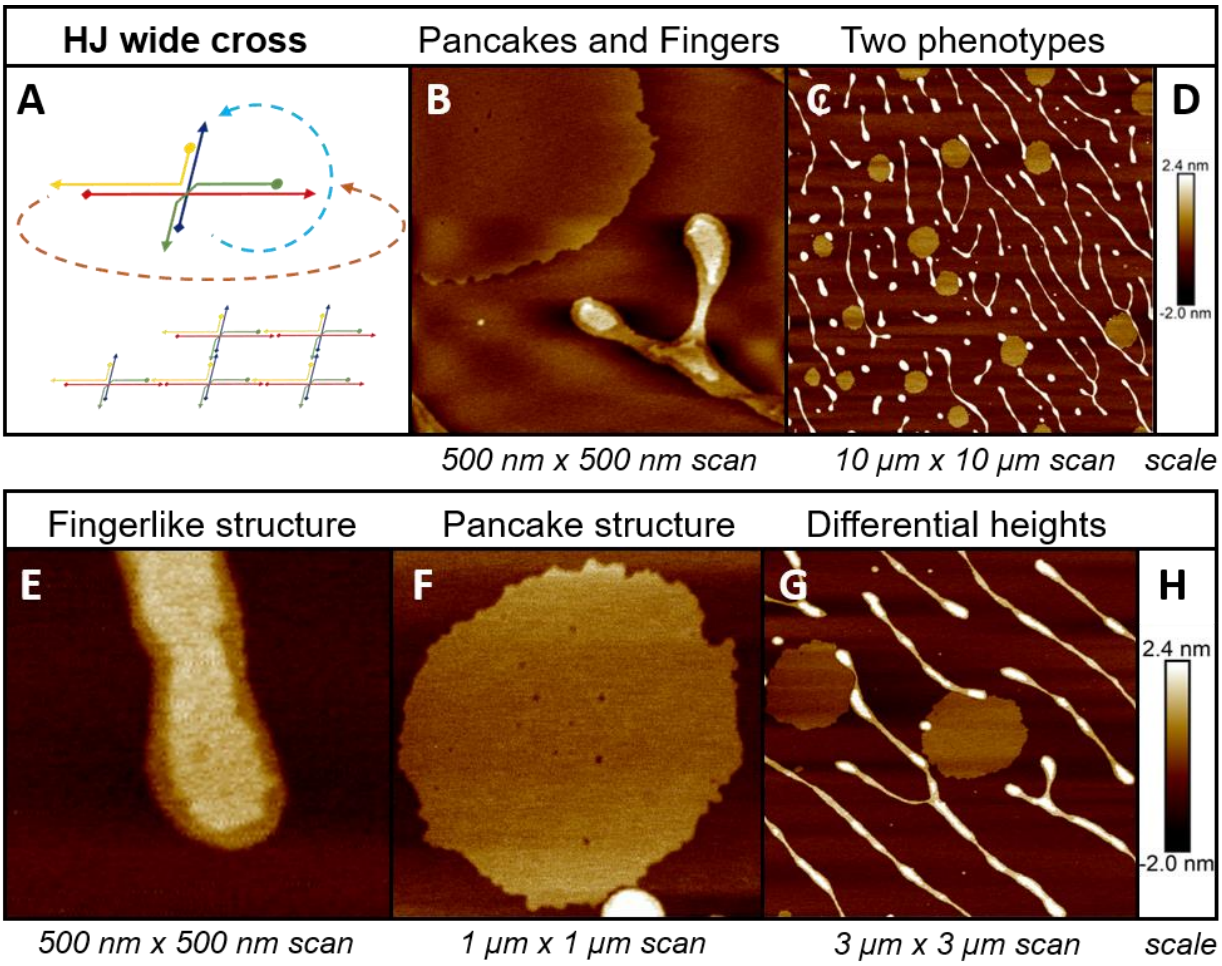


Figure 5.42: Results from the HJ wide cross tile. **A)** Expected lattice shape; **B)** two-phenotype assembly with flat, pancake-like aggregates and spindly, interconnected fingers. **C)** This type of assembly is repeated with a great deal of regularity over tens of microns. **D)** Scale bar for B-C. **E)** Fingerlike structures that indicated lattices in other nanostructure assemblies are rounded and lack defined windows, while **F)** Pancakes are shorter than 2 nm and do not have a discernible growth axis, but are roughly circular with a diameter of ~1 μm. **G)** The two phenotypes interact during deposition but are not grown together in solution. Their interaction seems steric and not driven by any apparent attractive force. **H)** Scale bar for E-G.

5.5.2. TJ-BW with CC Hairpin

In keeping with the hairpins used in early DX tile experiments,[148] a secondary structure was added to the T-junction brick wall lattice to allow for better resolution of nanoscale features. This experiment utilized the original TJ-BW tile sequences with the addition of a 16 bp, 7CC hairpin (Figure 5.43). This hairpin was designed using a search algorithm of all possible 16 bp duplexes with more than 40% C:Ag^+:C bonds. This was carried out before the development of the evolutionary search algorithm described in Chapter 4. Sequences can be found in Table 5.21. Experiments were carried out in MOPS(+)(+)7 (10 mM MOPS, 100 mM Na^+, 12.5 mM Mg^{2+}, pH 7.0) with and without 10x Ag^+. Results of the negative controls can be seen in Figure 5.44, while Ag^+ anneals can be seen in Figure 5.45.

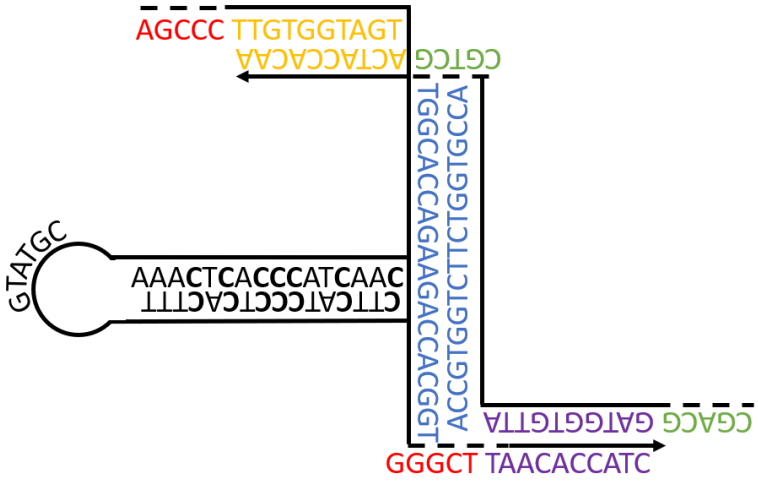


Figure 5.43: Design of the (unsuccessful) TJ-BW-H tile. Tile is based on sequences from the Hamada manuscript. Two intrastrand sticky end/ kissing loop pairs with WC chemistry (red, green) appear with literature-derived sequences. An orthogonal hairpin using optimal 16 bp sequences is designed by an early-generation, non-evolutionary, stacking-search algorithm.

Table 5.21: TJ-BW Hairpin (7CC) tile sequences (tetraloop italic, SE/KL underline, CC bold)

Sequence #	Nucleotide sequence
TJ-BW Hairpin (+)	5'-AGCCCTTGTGGTAGTTGGCACCAGAA**C**TT**C**AT**CCC**T**C**A**C**TTT(*GTATGC*)AAA**C**T**C**A**CCC**A T**C**AA **C**GACCACGGTGGGCTTAACACCATC-3'
TJ-BW Hairpin (-)	5'-CGACGGATGGTGTTAACCGTGGTCTTCTGGTGCCACGTCGACTACCACAA-3'

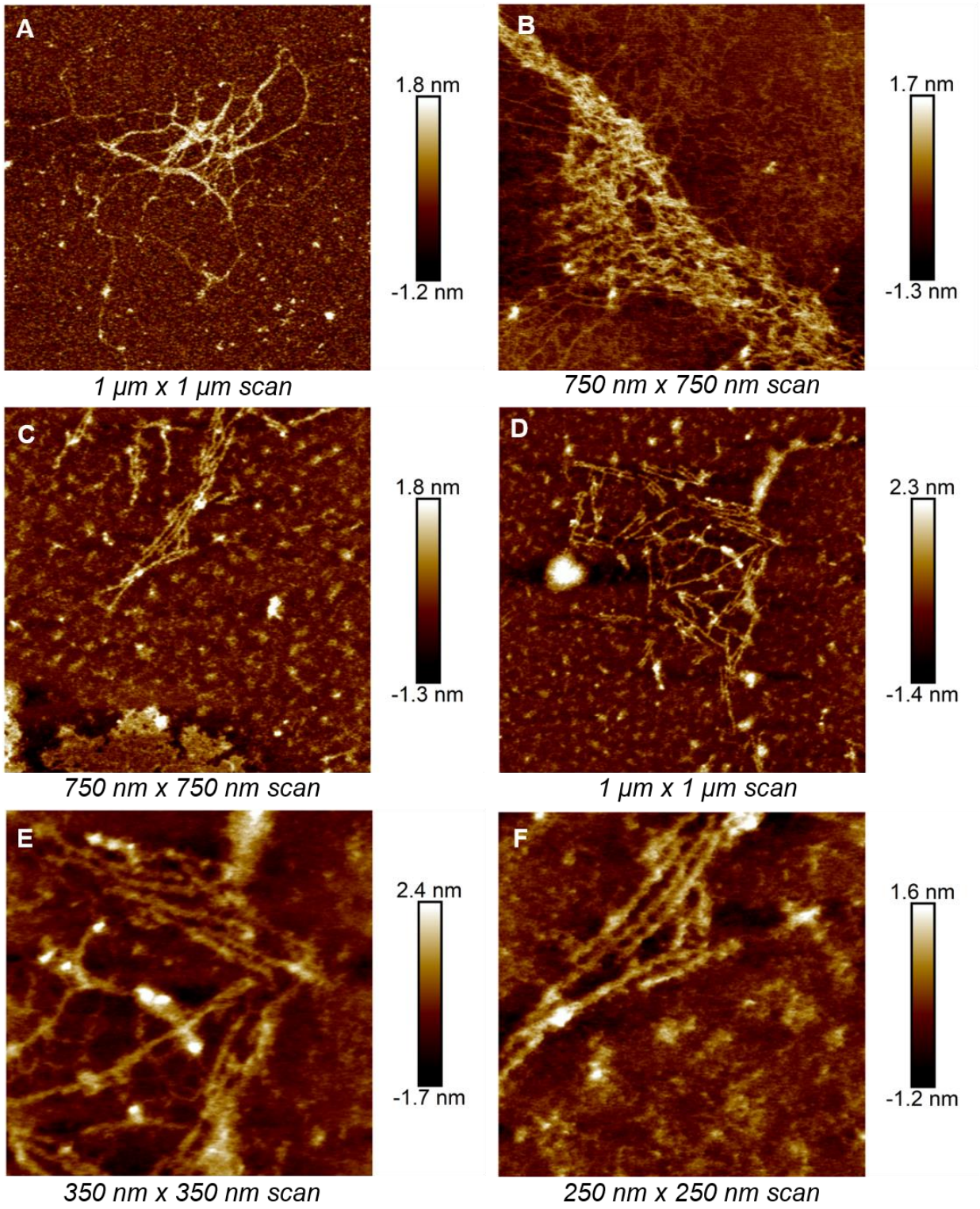


Figure 5.44: AFM micrographs of TJ-BW-H tile annealed without Ag^+. **A)** and **B)** show loose networks of pseudo-lattices no wider than 100 nm. **C)**, **D)**, **E)** and **F)** show in increasing detail a group of high

resolution (using HPP-SAA 1 nm tips, Bruker; Z-Range 1 µm, PeakForce Engage Setpoint 0.05 V) junctions of no more than three TJ-BW-H tiles wide at any point. Along the unpeturbed axis (no hairpin, Y axis in Figure 5.43) growth extends to ~500 nm. Images are resolved to visualize individual T-Junctions, and evidence of successful hairpin formation is absent, as necessitated by Ag^+-negative controls.

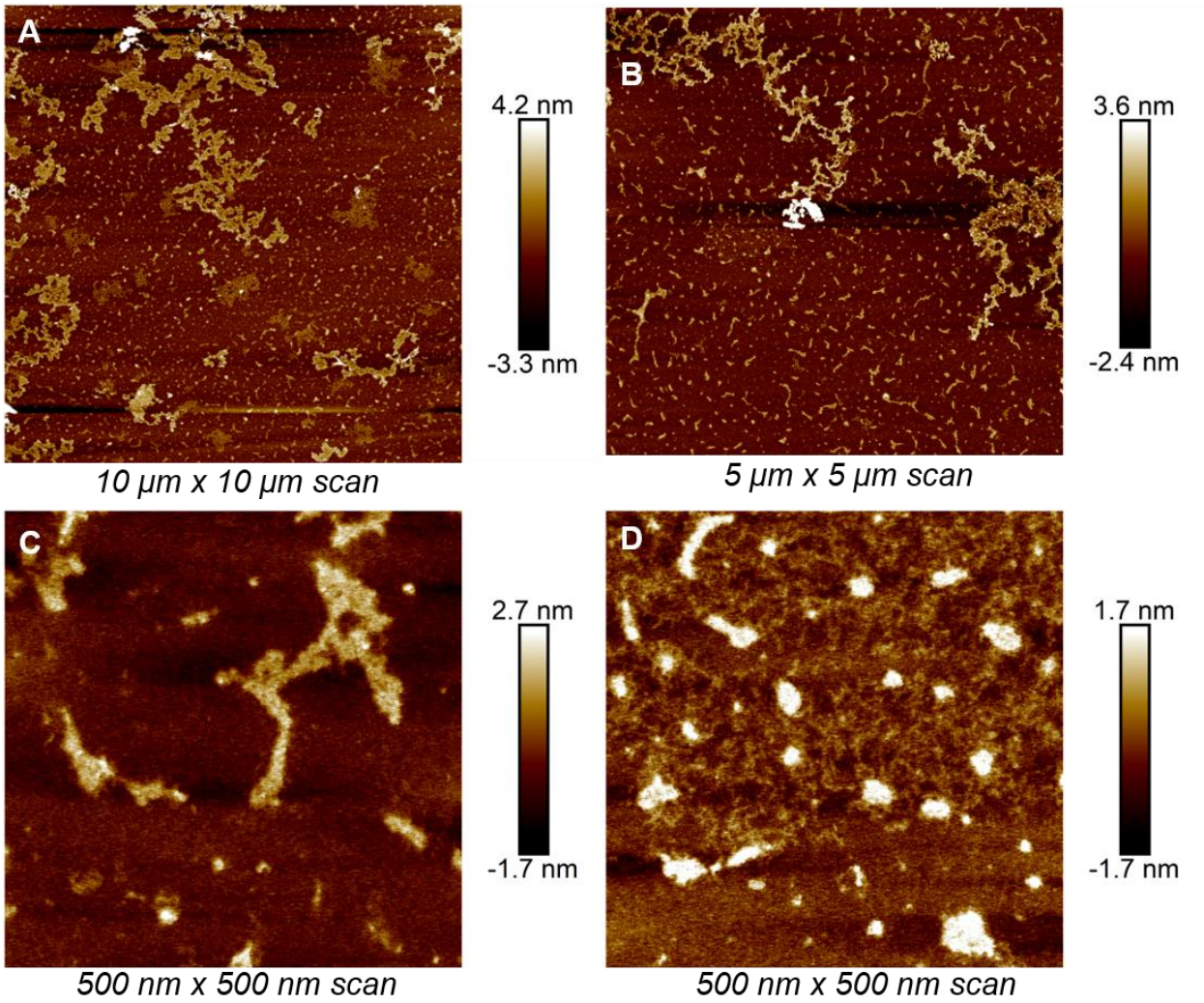


Figure 5.45: AFM micrographs of TJ-BW-H tiles with environmental Ag^+ after secondary anneal and correction cycling. **A)** Lattices can be seen to grow many microns in overall size, but no more than 1 µm on a specific growth axis. **B)** Closer imaging reveals lattice-like nanostructures along with lattice fragments. **C)** Similar to Figure 5.44E, individual T-junctions can be seen. Here the windows have been

plugged by the hairpins with some change in overall height due to the size of the secondary structure. Like both lattice-fragments and lattice bundles in B, these structures exhibit nanoscale design fidelity but overall disorder with respect to the growth axis. **D)** Higher-density regions display more tightly-packed bundles of successful TJ-BW lattices, but again lack the long-range order expected in the tile design.

The negative controls in Figure 5.44 demonstrate two key facts about the TJ-BW-H lattice: firstly, the individual tiles are able to join in the expected orientation; but secondly, the presence of a hairpin disrupts growth along at least one significant axis. Experimental conditions in Figure 5.45 show the same two characteristics: tiles are able to assemble correctly within grains of 10-20 nm; but the long-range order is disrupted by modifications made to the tile. Significantly, the presence of Ag^+ and the subsequent folding of the hairpin reduce the overall lattice strain. Instead of disjointed bundles (see Figure 5.44D), lattices are able to assemble across many microns. The perturbation caused by the hairpins does introduce regular curtailing of grain boundaries, but the lattice continues to polymerize within a single network. Overall, design fidelity on the nanoscale is high, and long-range yield can be described as low. Simple modifications to the tile could correct these deficiencies: shortening of the hairpin to less than the tile size; addition of the hairpin 5-6 nt into one of the central strands to extrude it orthogonally, rather than 11 nt for in-plane folding; and use of an AB tile such as is used in the DAO structures in *Section 5.4.3*. Single BW window resolution in Figures 5.18-5.20 obviated the importance of a hairpin feature for lattice resolution. This tile remains, however, promising and able to be fixed. It is consequently a candidate for future development.

5.5.3. TJ-Ladders

Hamada and colleagues produce a one-dimensional ladder that is relatively easy to identify via AFM.[182] This ladder is produced by using brick wall tiles with different sticky end geometry. An intermediary experiment in this work was to modify the ladder design to incorporate a string of fifteen CC bonds into

the upper arm of the structure. The sticky ends in the modified design were flipped to be cross-strand rather than intra-strand specific, with the intention of isolating the C:Ag^+:C string to one portion of the structure (Table 5.22). In doing so, the connectivity of the nanostructure itself was broken, allowing monomers but no ladders to form (Figure 5.46B). There was furthermore a C:Ag^+:T bond in a sticky end—an untested component suggested by thermodynamic studies but not subjected to the annealing analysis seen in Chapter 2. Overall, there were too many modifications to the original structure in one design step, and the underlying problem was not isolated. Success with the brick wall and ring species (*Section 5.3.2.b* and *Section 5.3.3.b*) precluded further study on the ladder design and subsequent corrections to sequence, topology, or both.

Table 5.22: TJ-Ladder (15CC) tile sequences (SE/KL underline, CC bold, CT*)

Sequence #	Nucleotide sequence
TJ-L (+)[182]	5’-GTGTCGAGTAACACGAAGCCAACCTCTAAATCTCCACGACACTCAGCATCCGATTTG-3’
TJ-L (-)[182]	5’-TGCAGTCAAATCGGATGCTGATGGAGATTTAGAGGTACTGCATGGCTTCGTGTTACT-3’
TJ-L-CC (+)	5’-T**CCCT*C**T**C**T**CCCCCCCCC**T**C**AAAAATACGCAATGAACGAAAGCTGTAAATACTCAAA-3’
TJ-L-CC (-)	5’-CTTTCGTTTCAGTATTTACAGTTCATTGCGTATTTT**CC*G****CC**AT**C**A**CCCCCCCCC**A**C**A-3’

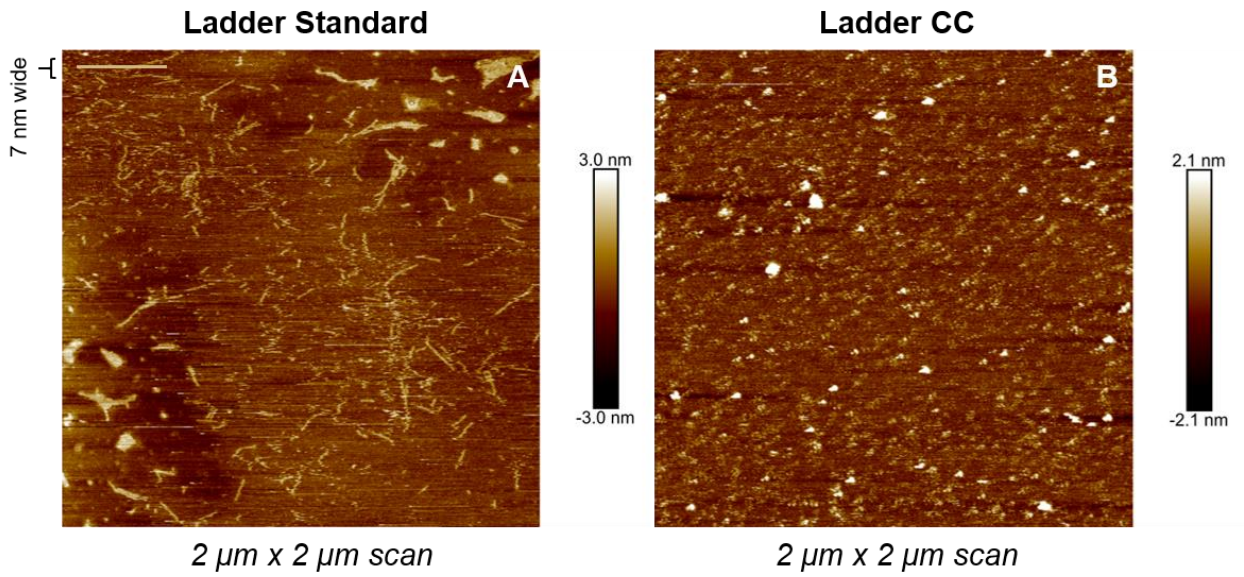


Figure 5.46: Results with the TJ-L tile. **A)** Ladders formed from the original sequences designed by Hamada et al are able to form one-dimensional ladders (scale inset). **B)** The orthogonal TJ-L tile produces only monomers.

References

1. Feynman, R. P. *APS Ann Meet (1960)*

2. Binnig, G. & Rohrer, H. *US Patent 4343993A* (1982).

3. Bennig, G. K. *US Patent 4724318A* (1988).

4. Harriott, L. R. Limits of Lithography. *Proc IEEE* **89**, 366-374, (2001).

5. Schmidt, V., Wittemann, J., Senz, S. & Gosele, U. Silicon Nanowires: A Review on Aspects of their Growth and their Electrical Properties. *Adv Mater* **21**, 2681–2702, (2009).

6. Rurali, R. Colloquium: Structural, electronic, and transport properties of silicon nanowires. *Rev Mod Phys* **82**, 427-449, (2010).

7. Peterson, I. Lithography Defects: Reducing and Managing Yield Killers Through Photo Cell Monitoring. *Yield Manag Sol* **2**, 17-24, (2000).

8. Montal, O., Dotan, K., Mebarki, B., Cai, M. P. & Ngai, C. DUV inspection and defect origin analysis for 22nm spacer self-aligned double-patterning. *Solid State Technol* **53**, 16-18, (2010).

9. Badger, K. D., Qi, Z. J., Gallagher, E., Seki, K. & McIntyre, G. Illuminating extreme ultraviolet lithography mask defect printability. *J Micro-Nanolith Mem* **12**, (2013).

10. Vieu, C. *et al.* Electron beam lithography: resolution limits and applications. *Appl Surf Sci* **164**, 111-117, (2000).

11. Forster, A. *et al.* Anionic surfactants for defect suppression in 193-nm lithography-Study of the adsorption process by ellipsometry and streaming potential measurements. *Colloid Surface A* **371**, 8-13, (2010).

12. National Academies of Sciences, E. & Medicine. *NASA Space Technology Roadmaps and Priorities Revisited*. (National Academies Press, 2016).

13. *International Technology Roadmap for Semiconductors*, <http://www.itrs.net/Links/2013ITRS/Home2013.htm> (2014).

14. McInerney, P., Adams, P. & Hadi, M. Z. Error rate comparison during polymerase chain reaction by DNA polymerase. *Mol Biol Int* **2014**, 1-8, (2014).

15. Hojfeldt, J. W., Blakskjaer, P. & Gothelf, K. V. A cleavable amino-thiol linker for reversible linking of amines to DNA. *J Org Chem* **71**, 9556-9559, (2006).

16. Sproat, B. S., Beijer, B., Rider, P. & Neuner, P. The Synthesis of Protected 5'-Mercapto-2',5'-Dideoxyribonucleoside-3'-O-Phosphoramidites - Uses of 5'-Mercapto-Oligodeoxyribonucleotides. *Nucleic Acids Res* **15**, 4837-4848, (1987).

17. Li, P., Medon, P. P., Skingle, D. C., Lanser, J. A. & Symons, R. H. Enzyme-Linked Synthetic Oligonucleotide Probes - Nonradioactive Detection of Enterotoxigenic Escherichia-Coli in Fecal Specimens. *Nucleic Acids Res* **15**, 5275-5287, (1987).

18. Cox, W. G. & Singer, V. L. Fluorescent DNA hybridization probe preparation using amine modification and reactive dye coupling. *Biotechniques* **36**, 114, (2004).

19. Wagenknecht, H. A. Fluorescent DNA base modifications and substitutes: Multiple fluorophore labeling and the DETEQ concept. *Ann N Y Acad Sci* **1130**, 122-130, (2008).

20. Rinker, S., Ke, Y., Liu, Y., Chhabra, R. & Yan, H. Self-assembled DNA nanostructures for distance-dependent multivalent ligand–protein binding. *Nat Nanotechnol* **3**, 418, (2008).

21. Park, S. H. *et al.* Programmable DNA self-assemblies for nanoscale organization of ligands and proteins. *Nano Lett* **5**, 729-733, (2005).

22. Wang, R., Nuckolls, C. & Wind, S. J. Assembly of heterogeneous functional nanomaterials on DNA origami scaffolds. *Angew Chem* **51**, 11325-11327, (2012).

23. Lyonnais, S. *et al.* DNA-carbon nanotube conjugates prepared by a versatile method using streptavidin-biotin recognition. *Small* **4**, 442-446, (2008).

24. Qin, W. J. & Yung, L. Y. L. Nanoparticle-DNA conjugates bearing a specific number of short DNA strands by enzymatic manipulation of nanoparticle-bound DNA. *Langmuir* **21**, 11330-11334, (2005).

25. Prody, G. A., Bakos, J. T., Buzayan, J. M., Schneider, I. R. & Bruening, G. Autolytic processing of dimeric plant virus satellite RNA. *Science* **231**, 1577-1580, (1986).

26. Isambert, H. The jerky and knotty dynamics of RNA. *Methods* **49**, 189-196, (2009).

27. Belkum, A. V., Abrahams, J. P., Pleij, C. W. & Bosch, L. Five pseudoknots are present at the 204 nucleotides long 3'noncoding region of tobacco mosak virus RNA. *Nucleic Acids Res* **13**, 7673-7686, (1985).

28. Wang, A. H. J. *et al.* Molecular structure of a left-handed double helical DNA fragment at atomic resolution. *Nature* **282**, 680, (1979).

29. Leroy, J.-L., Guéron, M., Mergny, J.-L. & Hélène, C. Intramolecular folding of a fragment of the cytosine-rich strand of telomeric DNA into an i-motif. *Nucleic Acids Res* **22**, 1600-1606, (1994).

30. Livshits, G. I. *et al.* Long-range charge transport in single G-quadruplex DNA molecules. *Nat Nanotechnol* **9**, 1040-1046, (2014).

31. Sen, D. & Gilbert, W. Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature* **334**, 364, (1988).

32. Cox, J. C. *et al.* Automated selection of aptamers against protein targets translated in vitro: from gene to aptamer. *Nucleic Acids Res* **30**, e108, (2002).

33. Bickle, M. B. T., Dusserre, E., Moncorge, O., Bottin, H. & Colas, P. Selection and characterization of large collections of peptide aptamers through optimized yeast two-hybrid procedures. *Nat Protoc* **1**, 1066-1091, (2006).

34. Min, K. *et al.* A simple and direct electrochemical detection of interferon-gamma using its RNA and DNA aptamers. *Biosens Bioelectron* **23**, 1819-1824, (2008).

35. Cox, J. C. & Ellington, A. D. Automated selection of anti-protein aptamers. *Bioorg Med Chem* **9**, 2525-2531, (2001).

36. Zhou, Q. *et al.* Aptamer-Containing Surfaces for Selective Capture of CD4 Expressing Cells. *Langmuir* **28**, 12544-12549, (2012).

37. Kwa, T. *et al.* Reconfigurable microfluidics with integrated aptasensors for monitoring intercellular communication. *Lab Chip* **14**, 1695-1704, (2014).

38. Seeman, N. C. Nucleic-Acid Junctions and Lattices. *J Theor Biol* **99**, 237-247, (1982).

39. Rothemund, P. W. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297-302, (2006).

40. Dietz, H., Douglas, S. M. & Shih, W. M. Folding DNA into twisted and curved nanoscale shapes. *Science* **325**, 725-730, (2009).

41. Perrault, S. D. & Shih, W. M. Virus-Inspired Membrane Encapsulation of DNA Nanostructures To Achieve In Vivo Stability. *ACS Nano* **8**, 5132-5140, (2014).

42. Ding, B. Q. *et al.* Gold Nanoparticle Self-Similar Chain Structure Organized by DNA Origami. *J Am Chem Soc* **132**, 3248, (2010).

43. Zhang, Q. *et al.* DNA origami as an in vivo drug delivery vehicle for cancer therapy. *ACS Nano* **8**, 6633-6643, (2014).

44. You, M. *et al.* DNA "nano-claw": logic-based autonomous cancer targeting and therapy. *J Am Chem Soc* **136**, 1256-1259, (2014).

45. Wang, X. *et al.* An Organic Semiconductor Organized into 3D DNA Arrays by “Bottom-up” Rational Design. *Angew Chem* **56**, 6445-6448, (2017).

46. Gorby, Y. A. *et al.* Electrically conductive bacterial nanowires produced by Shewanella oneidensis strain MR-1 and other microorganisms. *Proc Natl Acad Sci U S A* **103**, 11358-11363, (2006).

47. Ishii, S., Kosaka, T., Hori, K., Hotta, Y. & Watanabe, K. Coaggregation facilitates interspecies hydrogen transfer between Pelotomaculum thermopropionicum and Methanothermobacter thermautotrophicus. *Appl Environ Microbiol* **71**, 7838-7845, (2005).

48. Reguera, G. *et al.* Extracellular electron transfer via microbial nanowires. *Nature* **435**, 1098-1101, (2005).

49. Wang, Y., Tang, Z. Y., Tan, S. S. & Kotov, N. A. Biological assembly of nanocircuit prototypes from protein-modified CdTe nanowires. *Nano Lett* **5**, 243-248, (2005).

50. Nishinaka, T. *et al.* Conductive metal nanowires templated by the nucleoprotein filaments, complex of DNA and RecA protein. *J Am Chem Soc* **127**, 8120-8125, (2005).

51. Wildt, B., Mali, P. & Searson, P. C. Electrochemical template synthesis of multisegment nanowires: Fabrication and protein functionalization. *Langmuir* **22**, 10528-10534, (2006).

52. Neisser, M. & Wurm, S. ITRS lithography roadmap: 2015 challenges. *Adv Opt Technol* **4**, 235-240, (2015).

53. Braun, E., Eichen, Y., Sivan, U. & Ben-Yoseph, G. DNA-templated assembly and electrode attachment of a conducting silver wire. *Nature* **391**, 775-778, (1998).

54. Storm, A., Van Noort, J., De Vries, S. & Dekker, C. Insulating behavior for DNA molecules between nanoelectrodes at the 100 nm length scale. *Appl Phys Lett* **79**, 3881-3883, (2001).

55. Kasumov, A. Y. *et al.* Proximity-induced superconductivity in DNA. *Science* **291**, 280-282, (2001).

56. Endres, R. G., Cox, D. L. & Singh, R. R. P. Colloquium: The quest for high-conductance DNA. *Rev Mod Phys* **76**, 195-214, (2004).

57. Porath, D., Bezryadin, A., de Vries, S. & Dekker, C. Direct measurement of electrical transport through DNA molecules. *Nature* **403**, 635-638, (2000).

58. Shapir, E. *et al.* Electronic structure of single DNA molecules resolved by transverse scanning tunnelling spectroscopy. *Nat Mater* **7**, 68-74, (2008).

59. Kratochvilova, I. *et al.* Conductivity of natural and modified DNA measured by scanning tunneling microscopy. The effect of sequence, charge and stacking. *Biophys Chem* **138**, 3-10, (2008).

60. Hihath, J., Chen, F., Zhang, P. M. & Tao, N. J. Thermal and electrochemical gate effects on DNA conductance. *J Phys-Condens Mat* **19**, (2007).

61. Xu, B., Zhang, P., Li, X. & Tao, N. Direct conductance measurement of single DNA molecules in aqueous solution. *Nano Lett* **4**, 1105-1108, (2004).

62. Toomey, E. *et al.* Comparison of Canonical versus Silver(I)-Mediated Base-Pairing on Single Molecule Conductance in Polycytosine dsDNA. *J Phys Chem C* **120**, 7804-7809, (2016).

63. Rakitin, A. *et al.* Metallic conduction through engineered DNA: DNA nanoelectronic building blocks. *Phys Rev Lett* **86**, 3670-3673, (2001).

64. Fischler, M. *et al.* Chain-like assembly of gold nanoparticles on artificial DNA templates via 'click chemistry'. *Chem Commun (Camb)*, 169-171, (2008).

65. Timper, J. *et al.* Surface "click" reaction of DNA followed by directed metalization for the construction of contactable conducting nanostructures. *Angew Chem* **51**, 7586-7588, (2012).

66. Richter, J. *et al.* Nanoscale Palladium Metallization of DNA. *Adv Mater* **12**, 507-510, (2000).

67. Gu, Q., Chen, C. & Haynie, D. T. Cobalt metallization of DNA: toward magnetic nanowires. *Nanotechnol* **16**, 1358-1363, (2005).

68. Eidelshtein, G. *et al.* Synthesis and Properties of Novel Silver-Containing DNA Molecules. *Adv Mater* **28**, 4839-4844, (2016).

69. Zikich, D., Liu, K., Sagiv, L., Porath, D. & Kotlyar, A. I-Motif Nanospheres: Unusual Self-Assembly of Long Cytosine Strands. *Small* **7**, 1029-1034, (2011).

70. Russell, C. *et al.* Gold nanowire based electrical DNA detection using rolling circle amplification. *ACS Nano* **8**, 1147-1153, (2014).

71. Pearson, A. C. *et al.* DNA origami metallized site specifically to form electrically conductive nanowires. *J Phys Chem B* **116**, 10551-10560, (2012).

72. Uprety, B., Gates, E. P., Geng, Y., Woolley, A. T. & Harb, J. N. Site-specific metallization of multiple metals on a single DNA origami template. *Langmuir* **30**, 1134-1141, (2014).

73. Yamane, T. & Davidson, N. Complexing of Desoxyribonucleic Acid (DNA) by Mercuric Ion. *J Am Chem Soc* **83**, 2599-&, (1961).

74. Tanaka, K., Yamada, Y. & Shionoya, M. Formation of silver(I)-mediated DNA duplex and triplex through an alternative base pair of pyridine nucleobases. *J Am Chem Soc* **124**, 8802-8803, (2002).

75. Miyake, Y. *et al.* MercuryII-mediated formation of thymine-HgII-thymine base pairs in DNA duplexes. *J Am Chem Soc* **128**, 2172-2173, (2006).

76. Ono, A., Torigoe, H., Tanaka, Y. & Okamoto, I. Binding of metal ions by pyrimidine base pairs in DNA duplexes. *Chem Soc Rev* **40**, 5855-5866, (2011).

77. Torigoe, H. *et al.* Thermodynamic and structural properties of the specific binding between Ag(+) ion and C:C mismatched base pair in duplex DNA to form C-Ag-C metal-mediated base pair. *Biochimie* **94**, 2431-2440, (2012).

78. Dairaku, T. *et al.* Structure Determination of an AgI-Mediated Cytosine–Cytosine Base Pair within DNA Duplex in Solution with 1H/15N/109Ag NMR Spectroscopy. *Chem: Europ J* **22**, 13028-13031, (2016).

79. Yamaguchi, H. *et al.* The structure of metallo-DNA with consecutive thymine-Hg-II-thymine base pairs explains positive entropy for the metallo base pair formation. *Nucleic Acids Res* **42**, 4094-4099, (2014).

80. Liu, X. *et al.* Three novel silver complexes with ligand-unsupported argentophilic interactions and their luminescent properties. *Inorg Chem* **45**, 3679-3685, (2006).

81. Clever, G. H., Kaul, C. & Carell, T. DNA--metal base pairs. *Angew Chem* **46**, 6226-6236, (2007).

82. Clever, G. H. & Shionoya, M. Metal-base pairing in DNA. *Coord Chem Rev* **254**, 2391-2402, (2010).

83. Megger, D. A. & Muller, J. Silver(I)-mediated cytosine self-pairing is preferred over hoogsteen-type base pairs with the artificial nucleobase 1,3-dideaza-6-nitropurine. *Nucleos Nucleot Nucl* **29**, 27-38, (2010).

84. Zikich, D. *et al.* Ag+-Induced Arrangement of Poly (dC) into Compact Ring-Shaped Structures. *Int Rev Biophys Chem* **1**, 1-6, (2010).

85. Kaul, C., Muller, M., Wagner, M., Schneider, S. & Carell, T. Reversible bond formation enables the replication and amplification of a crosslinking salen complex as an orthogonal base pair. *Nat Chem* **3**, 794-800, (2011).

86. Urata, H., Yamaguchi, E., Nakamura, Y. & Wada, S. Pyrimidine-pyrimidine base pairs stabilized by silver(I) ions. *Chem Commun (Camb)* **47**, 941-943, (2011).

87. Megger, D. A. *et al.* Contiguous metal-mediated base pairs comprising two AgI ions. *Chem: Eur J* **17**, 6533-6544, (2011).

88. Pinheiro, V. B. & Holliger, P. Towards XNA nanotechnology: new materials from synthetic genetic polymers. *Trends Biotechnol* **32**, 321-328, (2014).

89. Vanmeert, M. *et al.* Rational design of an XNA ligase through docking of unbound nucleic acids to toroidal proteins. *Nucleic Acids Res* **47**, 7130-7142, (2019).

90. Hoshika, S. *et al.* Hachimoji DNA and RNA: A genetic system with eight building blocks. *Science* **363**, 884-887, (2019).

91. Vecchioni, S. *et al.* Construction and characterization of metal ion-containing DNA nanowires for synthetic biology and nanotechnology. *Sci Rep* **9**, 6942, (2019).

92. Torigoe, H., Miyakawa, Y., Ono, A. & Kozasa, T. Thermodynamic Properties of the Specific Binding Between Ag+ Ions and C:C Mismatched Base Pairs in Duplex DNA. *Nucleos Nucleot Nucl* **30**, 149-167, (2011).

93. Lopez-Gomollon, S. & Nicolas, F. E. Purification of DNA Oligos by denaturing polyacrylamide gel electrophoresis (PAGE). *Methods Enzymol* **529**, 65-83, (2013).

94. Ono, A. *et al.* Specific interactions between silver(I) ions and cytosine-cytosine pairs in DNA duplexes. *Chem Commun (Camb)*, 4825-4827, (2008).

95. Kwa, T. *et al.* Reconfigurable microfluidics with integrated aptasensors for monitoring intercellular communication. *Lab Chip* **14**, 1695-1704, (2014).

96. Chen, J. *et al.* High-throughput platform for real-time monitoring of biological processes by multicolor single-molecule fluorescence. *Proc Natl Acad Sci U S A* **111**, 664-669, (2014).

97. Chen, J. *et al.* Coupling of mRNA structure rearrangement to ribosome movement during bypassing of non-coding regions. *Cell* **163**, 1267-1280, (2015).

98. Aitken, C. E., Marshall, R. A. & Puglisi, J. D. An oxygen scavenging system for improvement of dye stability in single-molecule fluorescence experiments. *Biophys J* **94**, 1826-1835, (2008).

99. Tsukanov, R. *et al.* Detailed study of DNA hairpin dynamics using single-molecule fluorescence assisted by DNA origami. *J Phys Chem B* **117**, 11932-11942, (2013).

100. Kozasa, T., Miyakawa, Y., Ono, A. & Torigoe, H. in *Nucleic Acids Sympos Ser* 197-198 (Oxford University Press).

101. Peyret, N., Seneviratne, P. A., Allawi, H. T. & SantaLucia, J., Jr. Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A.A, C.C, G.G, and T.T mismatches. *Biochemistry* **38**, 3468-3477, (1999).

102. Owczarzy, R., Moreira, B. G., You, Y., Behlke, M. A. & Walder, J. A. Predicting stability of DNA duplexes in solutions containing magnesium and monovalent cations. *Biochemistry* **47**, 5336-5353, (2008).

103. Tan, Z.-J. & Chen, S.-J. Nucleic acid helix stability: effects of salt concentration, cation valence and size, and chain length. *Biophys J* **90**, 1175-1190, (2006).

104. Misra, V. K. & Draper, D. E. A thermodynamic framework for Mg2+ binding to RNA. *Proc Natl Acad Sci U S A* **98**, 12456-12461, (2001).

105. Ritchie, C. M. *et al.* Ag nanocluster formation using a cytosine oligonucleotide template. *J Phys Chem C* **111**, 175-181, (2007).

106. Gwinn, E. G., O'Neill, P., Guerrero, A. J., Bouwmeester, D. & Fygenson, D. K. Sequence-Dependent fluorescence of DNA-Hosted silver nanoclusters. *Adv Mater* **20**, 279-283, (2008).

107. Vecchioni, S., Capece, M. C., Toomey, E., Rothschild, L. & Wind, S. J. Methods of Synthesis and Characterization of Conductive DNA Nanowires Based on Metal Ion-Mediated Base Pairing for Single-Molecule Electronics. *JSAME* **6**, 61-90, (2018).

108. Allawi, H. T. & SantaLucia, J. Thermodynamics and NMR of internal G⊙T mismatches in DNA. *Biochemistry* **36**, 10581-10594, (1997).

109. Funai, T. *et al.* Regulated incorporation of two different metal ions into programmed sites in a duplex by DNA polymerase catalyzed primer extension. *Angew Chem* **53**, 6624-6627, (2014).

110. Meggers, E., Holland, P. L., Tolman, W. B., Romesberg, F. E. & Schulz, P. G. A novel copper-mediated DNA base pair. *J Am Chem Soc* **122**, 10714-10725, (2000).

111. Muller, K. M. & Arndt, K. M. Standardization in synthetic biology. *Methods Mol Biol* **813**, 23-43, (2012).

112. Ho-Shing, O., Lau, K. H., Vernon, W., Eckdahl, T. T. & Campbell, A. M. Assembly of standardized DNA parts using BioBrick ends in E. coli. *Methods Mol Biol* **852**, 61-76, (2012).

113. Vecchioni, S. *BBa_K1218022*, <http://parts.igem.org/Part:BBa_K1218022> (2013).

114. Vecchioni, S. *BBa_K1218026*, <http://parts.igem.org/Part:BBa_K1218026> (2013).

115. Swasey, S. M., Leal, L. E., Lopez-Acevedo, O., Pavlovich, J. & Gwinn, E. G. Silver (I) as DNA glue: Ag+-mediated guanine pairing revealed by removing Watson-Crick constraints. *Sci Rep* **5**, (2015).

116. Geary, C., Rothemund, P. W. & Andersen, E. S. RNA nanostructures. A single-stranded architecture for cotranscriptional folding of RNA nanostructures. *Science* **345**, 799-804, (2014).

117. Elbaz, J., Yin, P. & Voigt, C. A. Genetic encoding of DNA nanostructures and their self-assembly in living bacteria. *Nat Commun* **7**, 11179, (2016).

118. Siddiqui-Jain, A., Grand, C. L., Bearss, D. J. & Hurley, L. H. Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc Natl Acad Sci U S A* **99**, 11593-11598, (2002).

119. Simonsson, T. G-quadruplex DNA structures--variations on a theme. *Biol Chem* **382**, 621-628, (2001).

120. Johannsen, S., Megger, N., Böhme, D., Sigel, R. K. & Müller, J. Solution structure of a DNA double helix with consecutive metal-mediated base pairs. *Nat Chem* **2**, 229-234, (2010).

121. Mallajosyula, S. S. & Pati, S. K. Conformational tuning of magnetic interactions in metal–DNA complexes. *Angew Chem* **121**, 5077-5081, (2009).

122. Mao, C. D., Sun, W. Q. & Seeman, N. C. Designed two-dimensional DNA Holliday junction arrays visualized by atomic force microscopy. *J Am Chem Soc* **121**, 5437-5443, (1999).

123. Ding, B., Sha, R. & Seeman, N. C. Pseudohexagonal 2D DNA crystals from double crossover cohesion. *J Am Chem Soc* **126**, 10230-10231, (2004).

124. Gu, H. & Breaker, R. R. Production of single-stranded DNAs by self-cleavage of rolling-circle amplification products. *Biotechniques* **54**, 337-343, (2013).

125. Montagnier, L., Aissa, J., Ferris, S., Montagnier, J.-L. & Lavalléee, C. Electromagnetic signals are produced by aqueous nanostructures derived from bacterial DNA sequences. *Interdiscipl Sci Comp Life Sci* **1**, 81-90, (2009).

126. Xiang, L. *et al.* Intermediate tunnelling–hopping regime in DNA charge transport. *Nat Chem* **7**, 221-226, (2015).

127. Venkataraman, L., Klare, J. E., Nuckolls, C., Hybertsen, M. S. & Steigerwald, M. L. Dependence of single-molecule junction conductance on molecular conformation. *Nature* **442**, 904-907, (2006).

128. Bruot, C., Xiang, L., Palma, J. L. & Tao, N. Effect of mechanical stretching on DNA conductance. *ACS Nano* **9**, 88-94, (2014).

129. Yamaguchi, H. *et al.* The structure of metallo-DNA with consecutive thymine-HgII-thymine base pairs explains positive entropy for the metallo base pair formation. *Nucleic Acids Res* **42**, 4094-4099, (2014).

130. Roche, S. Sequence dependent DNA-mediated conduction. *Phys Rev Lett* **91**, 108101, (2003).

131. Mallajosyula, S. S. & Pati, S. K. Toward DNA conductivity: a theoretical perspective. *J Phys Chem Lett* **1**, 1881-1894, (2010).

132. Penzo, E., Wang, R., Palma, M. & Wind, S. J. Selective placement of DNA origami on substrates patterned by nanoimprint lithography. *J Vac Sci Technol B* **29**, 06F205, (2011).

133. Scarabelli, D. *et al.* Fabrication of artificial graphene in a GaAs quantum heterostructure. *J Vac Sci Technol B* **33**, 06FG03, (2015).

134. Penzo, E. *et al.* Directed assembly of single wall carbon nanotube field effect transistors. *ACS Nano* **10**, 2975-2981, (2016).

135. Bouilly, D. *et al.* Single-molecule reaction chemistry in patterned nanowells. *Nano Lett* **16**, 4679-4685, (2016).

136. Park, H. *et al.* High-density integration of carbon nanotubes via chemical self-assembly. *Nat Nanotechnol* **7**, 787, (2012).

137. Palma, M. *et al.* Controlled formation of carbon nanotube junctions via linker-induced assembly in aqueous solution. *J Am Chem Soc* **135**, 8440-8443, (2013).

138. Penzo, E. *Directed biomolecular assembly of functional nanodevices*. (Columbia University, 2014).

139. Tu, X., Manohar, S., Jagota, A. & Zheng, M. DNA sequence motifs for structure-specific recognition and separation of carbon nanotubes. *Nature* **460**, 250, (2009).

140. Penzo, E. *et al.* Directed assembly of end-functionalized single wall carbon nanotube segments. *Nano Lett* **15**, 6547-6552, (2015).

141. Toomey, E. *Electrical Characterization of Ag(I)-Intercalated Poly-Cytosine DNA*. (Brown University, 2015).

142. Swasey, S. M., Leal, L. E., Lopez-Acevedo, O., Pavlovich, J. & Gwinn, E. G. Silver (I) as DNA glue: Ag+-mediated guanine pairing revealed by removing Watson-Crick constraints. *Sci Rep* **5**, 10163, (2015).

143. Venkataraman, L. *et al.* Single-molecule circuits with well-defined molecular conductance. *Nano Lett* **6**, 458-462, (2006).

144. Nogues, C. & Wanunu, M. A rapid approach to reproducible, atomically flat gold films on mica. *Surface Sci* **573**, L383-L389, (2004).

145. Maver, U., Planinšek, O., Jamnik, J., Hassanien, A. I. & Gaberšček, M. Preparation of Atomically Flat Gold Substrates for AFM Measurements. *Acta Chimica Slov* **59**, (2012).

146. Wang, J. C. Helical repeat of DNA in solution. *Proc Natl Acad Sci U S A* **76**, 200-203, (1979).

147. Franklin, R. E. & Gosling, R. G. The structure of sodium thymonucleate fibres. I. The influence of water content. *Acta Crystallogr* **6**, 673-677, (1953).

148. Winfree, E., Liu, F. R., Wenzler, L. A. & Seeman, N. C. Design and self-assembly of two-dimensional DNA crystals. *Nature* **394**, 539-544, (1998).

149. Zhao, Z., Jacovetty, E. L., Liu, Y. & Yan, H. Encapsulation of gold nanoparticles in a DNA origami cage. *Angew Chem* **50**, 2041-2044, (2011).

150. Douglas, S. M. *et al.* Rapid prototyping of 3D DNA-origami shapes with caDNAno. *Nucleic Acids Res* **37**, 5001-5006, (2009).

151. Kearse, M. *et al.* Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. *Bioinformatics* **28**, 1647-1649, (2012).

152. Oberortner, E., Cheng, J.-F., Hillson, N. J. & Deutsch, S. Streamlining the design-to-build transition with build-optimization software tools. *ACS Synth Biol* **6**, 485-496, (2016).

153. Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res* **31**, 3406-3415, (2003).

154. PrimerQuest® program (IDT, Coralville, IA, USA, 2018).

155. Saiki, R. K. in *PCR technology* (Palgrave Macmillan, London, 1989).

156. Kaur, H., Arora, A., Wengel, J. & Maiti, S. Thermodynamic, Counterion, and Hydration Effects for the Incorporation of Locked Nucleic Acid Nucleotides into DNA Duplexes. *Biochemistry* **45**, 7347-7355, (2006).

157. Xia, T. *et al.* Thermodynamic Parameters for an Expanded Nearest-Neighbor Model for Formation of RNA Duplexes with Watson−Crick Base Pairs. *Biochemistry* **37**, 14719-14735, (1998).

158. Nakano, S.-i., Fujimoto, M., Hara, H. & Sugimoto, N. Nucleic acid duplex stability: influence of base composition on cation effects. *Nucleic Acids Res* **27**, 2957-2965, (1999).

159. Holland, J. H. *Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence*. (MIT press, 1992).

160. Konak, A., Coit, D. W. & Smith, A. E. Multi-objective optimization using genetic algorithms: A tutorial. *Reliab Eng Syst Safe* **91**, 992-1007, (2006).

161. Wu, J.-S., Lee, C., Wu, C.-C. & Shiue, Y.-L. Primer design using genetic algorithm. *Bioinformatics* **20**, 1710-1717, (2004).

162. Liu, D., Park, S. H., Reif, J. H. & LaBean, T. H. DNA nanotubes self-assembled from triple-crossover tiles as templates for conductive nanowires. *Proc Natl Acad Sci U S A* **101**, 717-722, (2004).

163. Hush, N. & Cheung, A. S. Ionization potentials and donor properties of nucleic acid bases and related compounds. *Chem Phys Lett* **34**, 11-13, (1975).

164. Seeman, N. C. Macromolecular design, nucleic acid junctions, and crystal formation. *J Biomol Struct Dyn* **3**, 11-34, (1985).

165. Castro, C. E. *et al.* A primer to scaffolded DNA origami. *Nat Methods* **8**, 221, (2011).

166. Reif, J. H. The design of autonomous DNA nano-mechanical devices: Walking and rolling DNA. *Natural Comp* **2**, 439, (2003).

167. Pfitzner, E. *et al.* Rigid DNA beams for high-resolution single-molecule mechanics. *Angew Chem* **52**, 7766-7771, (2013).

168. Roh, Y. H. *et al.* A multi-RNAi microsponge platform for simultaneous controlled delivery of multiple small interfering RNAs. *Angew Chem* **55**, 3347-3351, (2016).

169. Gopinath, A. & Rothemund, P. W. Optimized assembly and covalent coupling of single-molecule DNA origami nanoarrays. *ACS Nano* **8**, 12030-12040, (2014).

170. Fujishima, K., Venter, C., Wang, K., Ferreira, R. & Rothschild, L. J. An overhang-based DNA block shuffling method for creating a customized random library. *Sci Rep* **5**, 9740, (2015).

171. Wang, M., Huang, H., Zhang, Z. & Xiao, S.-J. 2D DNA lattices constructed from two-tile DAE-O systems possessing circular central strands. *Nanoscale* **8**, 18870-18875, (2016).

172. Han, D. *et al.* DNA gridiron nanostructures based on four-arm junctions. *Science* **339**, 1412-1415, (2013).

173. Aldaye, F. A. & Sleiman, H. F. Sequential self-assembly of a DNA hexagon as a template for the organization of gold nanoparticles. *Angew Chem* **45**, 2204-2209, (2006).

174. Goodman, R. P. *et al.* Rapid chiral assembly of rigid DNA building blocks for molecular nanofabrication. *Science* **310**, 1661-1665, (2005).

175. Pan, K., Bricker, W. P., Ratanalert, S. & Bathe, M. Structure and conformational dynamics of scaffolded DNA origami nanoparticles. *Nucleic Acids Res* **45**, 6284-6298, (2017).

176. Pley, H. W., Flaherty, K. M. & McKay, D. B. Three-dimensional structure of a hammerhead ribozyme. *Nature* **372**, 68, (1994).

177. Maune, H. T. *et al.* Self-assembly of carbon nanotubes into two-dimensional geometries using DNA origami templates. *Nat Nano* **5**, 61, (2010).

178. Schreiber, R. *et al.* DNA origami-templated growth of arbitrarily shaped metal nanoparticles. *Small* **7**, 1795-1799, (2011).

179. Shen, B., Linko, V., Tapio, K., Kostiainen, M. A. & Toppari, J. J. Custom-shaped metal nanostructures based on DNA origami silhouettes. *Nanoscale* **7**, 11267-11272, (2015).

180. Yavin, E. *et al.* Protein–DNA charge transport: redox activation of a DNA repair protein by guanine radical. *Proc Natl Acad Sci U S A* **102**, 3546-3551, (2005).

181. O'Neill, M. A., Becker, H. C., Wan, C., Barton, J. K. & Zewail, A. H. Ultrafast dynamics in DNA-mediated electron transfer: Base gating and the role of temperature. *Angew Chem* **42**, 5896-5900, (2003).

182. Hamada, S. & Murata, S. Substrate-assisted assembly of interconnected single-duplex DNA nanostructures. *Angew Chem* **48**, 6820-6823, (2009).

183. Brinkers, S., Dietrich, H. R., de Groote, F. H., Young, I. T. & Rieger, B. The persistence length of double stranded DNA determined using dark field tethered particle motion. *J Chem Phys* **130**, 06B607, (2009).

184. Pinheiro, V. B. *et al.* Synthetic genetic polymers capable of heredity and evolution. *Science* **336**, 341-344, (2012).

185. Wang, R., Palma, M., Penzo, E. & Wind, S. J. Lithographically directed assembly of one-dimensional DNA nanostructures via bivalent binding interactions. *Nano Res* **6**, 409-417, (2013).