

ABSTRACT

Title of Dissertation: THE ROLE OF SEQUENCE IN THE
STRUCTURE OF SELF-ASSEMBLING 3D
DNA CRYSTALS

Maithili Saoji, Doctor of Philosophy, 2015.

Dissertation Directed By: Dr. Paul Paukstelis, Assistant Professor,
Department of Chemistry and Biochemistry.

DNA is a widely used biopolymer for the construction of nanoscale objects due to its programmability and structural predictability. DNA oligonucleotides can, however, exhibit a great deal of local structural diversity. DNA conformation is strongly linked to both environmental conditions and the nucleobase identities inherent in the oligonucleotide sequence, but the exact relationship between sequence and local structure is not completely understood. We previously determined the X-ray crystal structure of a DNA 13-mer that forms a continuously hydrogen bonded three-dimensional lattice through Watson-Crick and non-canonical base pairs. In the current work I examined how the sequence of the Watson-Crick duplex region influenced crystallization of this 13-mer. I screened all possible self-complementary sequences in the hexameric duplex region and found 21 oligonucleotides that crystallized. Sequence analysis showed that one specific Watson-Crick base pair influenced the crystallization propensity and the speed of crystal self-assembly. I

determined X-ray crystal structures for 13 of these oligonucleotides and found sequence-specific structural changes suggesting that this base pair may serve as a structural anchor during crystal assembly. I explored the crystal self-assembly and nucleation process and demonstrated that crystals grown from mixtures of two different oligonucleotide sequences contained both the oligonucleotides. These results suggested that crystal self-assembly is nucleated by the formation of Watson-Crick duplexes. Finally, I also examined how a single nucleotide addition to the DNA 13-mer leads to a significantly different overall structure under identical crystallization conditions. The 14-mer crystal structures described here showed that all of the predicted Watson-Crick base pairs were present, but the major difference as compared to the parent 13-mer structure was a significant rearrangement of non-canonical base pairs. This included the formation of a sheared A-G base pair, a junction of strands formed from base triple interactions, and tertiary interactions that generated structural features similar to tandem sheared G-A base pairs. The adoption of this alternate non-canonical structure was dependent in part on the sequence of the Watson-Crick duplex region. These results provided important new insights into the sequence/structure relationship of short DNA oligonucleotides and demonstrated a unique interplay between Watson-Crick and non-canonical base pairs that are responsible for crystallization fate.

THE ROLE OF SEQUENCE IN THE STRUCTURE OF SELF-ASSEMBLING 3D
DNA CRYSTALS

by

Maithili Saoji

Dissertation submitted to the Faculty of the Graduate School of the
University of Maryland, College Park, in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
2015

Advisory Committee:
Professor Paul Paukstelis, Chair
Professor Jason Kahn
Professor David Fushman
Professor Nichole LaRonde-LeBlanc
Professor Ian White, Dean's representative

© Copyright by
Maithili Saoji
2015

Dedication

I dedicate this thesis to my loving husband Pushkar, for his relentless support and encouragement and my little angel Arya. I also dedicate it to my mom, dad, aajee, Shon and Purva for their constant faith in me and for helping me change my dream into reality, and to aai, baba and Neha for the never-ending love and support. Last but not the least, I dedicate this thesis to my loving friend Sudeshna for being with me through thick and thin.

Acknowledgements

I want to thank Dr. Paul Paukstelis for his constant mentoring and support throughout my journey in his lab. My experience would not have been the same without him. I want to thank Dr. Jason Kahn, Dr. David Fushman and Dr. Nicole LaRonde for their support and guidance through all the committee meetings, Dr. Ian White for his time as the Dean's representative and Dr. Amy Beaven, Dr. Silvia Muro and Dr. George Lorimer for opening their labs to me for my various experiments. I would like to extend a special thanks to my past and present lab mates, Mello, Stephanie, Diana, Ron, Alessandra, Ryan, Orion, Logan and Flo for their help from time to time and also for making the lab a fun place to work.

The work presented in this document is adopted from the following research articles:

- Probing the role of sequence in the assembly of three-dimensional DNA crystals
Maithili Saoji, Daoning Zhang & Paul J. Paukstelis
Biopolymers. 2015 May 26. doi: 10.1002/bip.22688.
- Sequence-dependent structural changes in a self-assembling DNA oligonucleotide.
Maithili Saoji, Paul J. Paukstelis
Acta Crystallographica Section D (Accepted).

Table of Contents

Dedication	ii
Acknowledgements	iii
Table of Contents	iv
List of Tables	vi
List of Figures	vii
Chapter 1: Introduction	1
Chapter 1.1 DNA: A Structural Building Block	1
Chapter 1.2 Structural DNA Nanotechnology	3
Chapter 1.3 Structural Diversity of DNA	7
Chapter 1.4 Motivation for Research	8
Chapter 1.5 Scope of Research	10
Chapter 2: Probing the Role of Sequence in the Self-Assembly Process of 13-mer DNA Crystals	12
Chapter 2.1 Introduction	12
Chapter 2.2 G-A Pairs in the Duplex Not Required for Crystallization	17
Chapter 2.3 Sequence Requirements for Crystallization	20
Chapter 2.4 Structural Implications of the Duplex Sequences	25
Chapter 2.5 B7 Does Not Form Base Pairs in Absence of Cations	31
Chapter 2.6 Watson-Crick or Non-canonical Base Pairs as Drivers for Crystallization	35
Chapter 2.7 Influence of Sequence on Crystallizability, Structure and Self- Assembly	42
Chapter 2.8 Materials and Methods	44
1. Crystallization Screen	44
2. DNA Purification and Crystallization	44
3. Data Collection and Refinement	45
4. Sequence and Structure Comparison and Analysis	45
5. Nuclear Magnetic Resonance	46
6. Heterogeneous Crystallization	47
Chapter 3: Other Methods Used to Study the Cation Driven Assembly Process	48
Chapter 3.1 Dynamic Light Scattering	50
Chapter 3.2 Ultra-Violet (UV) Spectroscopy	51
Chapter 3.3 Fluorescence Spectroscopy	53
Chapter 3.4 S1 Nuclease Digestion	56
Chapter 3.5 DNase I Digestion	58
Chapter 3.6 Confocal Microscopy	60
Chapter 3.7 Materials and Methods	63
1. Dynamic Light Scattering	63
2. UV Spectroscopy	63
3. Fluorescence Spectroscopy	63
4. S1 and DNase I Digestion	64
5. Confocal Microscopy	65

Chapter 4: Sequence-Dependent Structural Changes in 14-mer DNA Oligonucleotide	66
Chapter 4.1 Introduction	66
Chapter 4.2 Overview of the 14-mer Structures.	69
Chapter 4.3 B-Form Duplex Region Capped by Sheared A-G Pairs.....	74
Chapter 4.4 Tertiary Interactions Fulfil a Structural Role to Generate Tandem G-A Base Pairs.....	76
Chapter 4.5 Triplex Junction.....	81
Chapter 4.6 Sequence Requirements for Alternate Crystal Form.....	83
Chapter 4.7 Materials and Methods	87
1. DNA Synthesis and Purification.....	87
2. Crystallization.....	87
3. Data Collection and Structure Determination.....	88
Chapter 5: Conclusion and Future Prospects	89
Bibliography	91

This Table of Contents is automatically generated by MS Word, linked to the Heading formats used within the Chapter text.

List of Tables

Table 1. Data collection and refinement for B7.....	18
Table 2. Crystallized oligonucleotide sequences.....	21
Table 3. Data collection and refinement for 12 sequence variants.....	26
Table 4. Transverse relaxation experiments on B7 in presence of Mg ²⁺	34
Table 5. Heterogeneous crystals for pairwise combinations of 13-mer DNA oligonucleotides	41
Table 6. Data collection and refinement for 14-mer structures.....	72
Table 7. Sequence-dependent crystallization.....	86

List of Figures

Figure 1. B-form DNA helix ¹⁰ (PDB ID: 1BNA).....	2
Figure 2. Representative 2D and 3D DNA-based architectures.	4
Figure 3. The lattice formed by the tensegrity triangle ²³	6
Figure 4 The crystal structure of d(GGACAGATGGGAG) (PDB ID 1P1Y) ⁵⁹	14
Figure 5. Solvent channels in the 13-mer lattice.....	15
Figure 6. The 5' GGA- motif in the structural context of the DNA 13-mer.....	16
Figure 7. Structure of B7 variant compared to the original 13-mer.....	19
Figure 8. Representative crystal morphologies for DNA oligonucleotides identified in the screen.	22
Figure 9. Nucleobase frequency distribution of crystallizing DNA oligonucleotides.	23
Figure 10. Superimposition of the 13-mer structures with and without the A5-T8 base pair.	29
Figure 11. Average RMSD comparison for the duplex region of A5-T8 and the non-A5-T8 structures.	30
Figure 12. 1D- ¹ H NMR spectra of B7 variant.	33
Figure 13. NMR signal broadening upon Mg ²⁺ addition.	34
Figure 14. Models for crystal pre-nucleation.....	36
Figure 15. Heterogeneous crystal from DNA mixtures.....	38
Figure 16. Heterogeneous crystal from ratios of DNA mixtures.....	39
Figure 17. Monitoring the assembly of B7 DNA using UV Spectroscopy.	52
Figure 18. Fluorescence intensity measurements on B7 DNA using nucleic acid specific fluorescent dyes.	55

Figure 19. Denaturing PAGE gel for samples digested with S1 nuclease.....	57
Figure 20. Denaturing PAGE gel for samples digested with DNase I nuclease.....	59
Figure 21. Confocal images for unlabeled B7 core and fluorescent shells.....	61
Figure 22. Confocal images to observe the B7 crystal assembly.	62
Figure 23. 13-mer and 14-mer crystals.	68
Figure 24. Overview of 14-mer crystal structure.....	70
Figure 25. Crystal packing.....	71
Figure 26. Structural overlap for the 14-mer structures.....	73
Figure 27. Structural comparison of A3-13 and A3-14 monomers.	75
Figure 28. Comparison of GA/AG motifs.	78
Figure 29. Potential hydrogen bond between A3 and T9.	79
Figure 30. The inter- and intrastrand stacking at the sheared G-A base pair.....	80
Figure 31. Purine base triples.....	82

Chapter 1: Introduction

Chapter 1.1 DNA: A Structural Building Block

DNA is the genetic material for all living organisms. Through the complementary Watson-Crick base pairing (adenine-thymine and guanine-cytosine), it can successfully store and transmit hereditary information¹. Remarkably, the properties that make DNA a powerful genetic material, also make it an attractive synthetic building block for nano and micrometer scale constructions². Through the cooperative interplay of extremely precise Watson-Crick base pairing, base-stacking, electrostatic, and hydrophobic interactions, complementary DNA oligonucleotides self-assemble in solution to form an antiparallel B-form double helical structure³. The stability of the DNA double helix and its predictable structural and physical properties, like the 50 nm persistence length, ~2 nm diameter, ~3.4 nm helical pitch, makes it desirable for creating nanostructures with predictable and precise geometries (Figure 1). Cations play an important role in the structure and folding of DNA⁴. Divalent cations, most commonly Mg^{2+} , are used in the self-assembly of DNA based nanostructures to overcome electrostatic repulsion between DNA duplexes by screening the negative charge on the backbone phosphates⁵⁻⁸. Additionally, the success of using DNA for synthetic constructions is related to vast size of DNA sequence libraries that can be created through the permutation of the bases at each position of the oligonucleotides and the ease of chemical synthesis of any desired sequences with high yields and purity at relatively low costs using phosphoramidite chemistry⁹. By exploiting these unique properties of DNA the field of structural DNA nanotechnology has advanced

to create highly sophisticated nanoscale architectures and to use DNA as a template for precise positioning of materials and molecules.

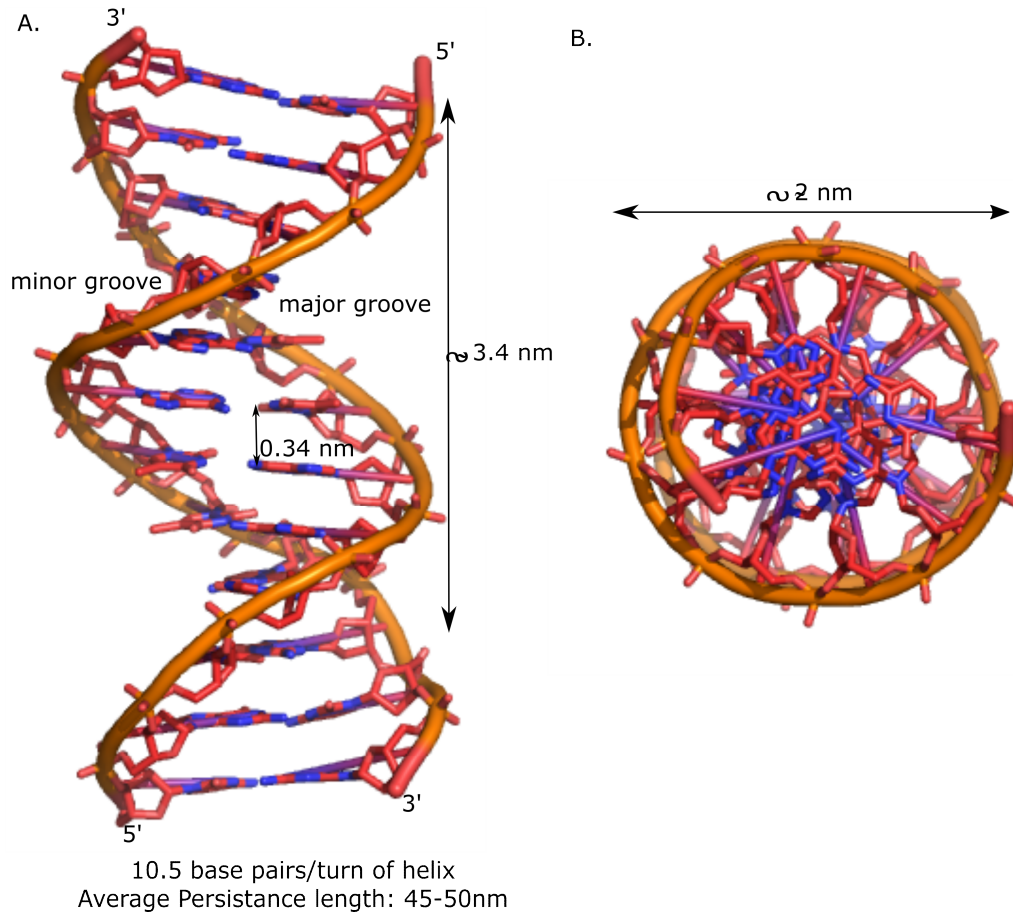


Figure 1. B-form DNA helix¹⁰ (PBD ID: 1BNA). **A.** The right handed B-form DNA double helix highlighting the physical parameters. **B.** Top view of the B-form double helix with the diameter.

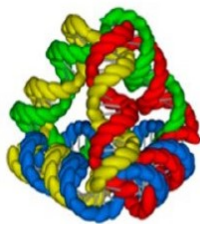
Chapter 1.2 Structural DNA Nanotechnology

DNA nanotechnology is a ‘bottom up’ construction process where dispersed DNA oligonucleotides self-assemble in solution to form complex architectures¹¹. The first artificially constructed DNA nanostructure was the immobile crossover Holliday junction¹² which was followed by a wide variety of more sophisticated architectures like the double-crossover (DX) DNA tiles¹³, triple-crossover (TX) tiles¹⁴, 4X4 tiles⁷, three-point-star tiles¹⁵ etc. Through an interplay of DNA helices and cross-over junction motifs, discrete higher order structures like the DNA cube¹⁶, polyhedra^{15,17-19}, nanotubes²⁰ and repetitive two-dimensional (2D) DNA lattices^{5,7,8,21,22} and three-dimensional (3D) tensegrity triangle based DNA crystals²³ were constructed (Figure 2 & 3).

Several non-periodic, spatially addressable and finite sized 2D and 3D DNA structures have been constructed using DNA origami technique^{6,24-27} where a long, circular, single-stranded genomic DNA, referred to as the scaffold-strand, is folded into various geometric shapes that are held together with short DNA oligonucleotides called the staple-strands. More recently the single-stranded DNA tile technique, in which single-stranded DNA tiles are interlocked with each other through local molecular connections, has been successfully used to by-pass the use of staple-strand in the construction of discrete DNA objects²⁸⁻³⁰ (Figure 2). DNA has also been successfully used as a building block for the construction of dynamic nanoscale devices and in molecular electronics and computing^{31,32}.



Chen et. al. Nature.



Goodman et. al. Science.



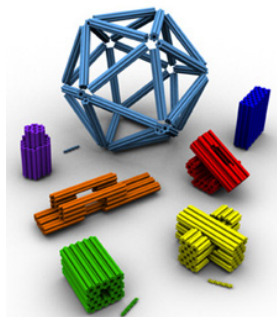
He et. al. Nature.



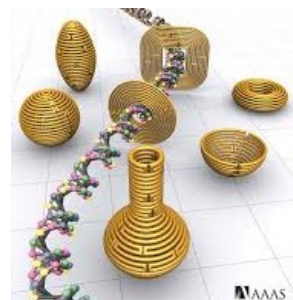
Aldave et. al. J. Am Chem.



Dietz et. al. Science.



Douglas et. al. Nature.



Han et. al. Science.



Wei et. al. Nature.

Figure 2. Representative 2D and 3D DNA-based architectures. The top panel shows the molecular model of a DNA cube¹⁶, DNA tetrahedron¹⁷, DNA dodecahedron¹⁵ and DNA biprism¹⁸ (left to right). The lower panel represents the sophisticated, discreet two and three dimensional architectures created using DNA origami²⁵⁻²⁷ (first three tiles) and single stranded tile technique²⁸ (last tile).

One of the long-standing goals of the DNA nanotechnology field has been the rational design and construction of periodic 3D DNA structures, or crystals, for use as molecular scaffolds. DNA based scaffolds as envisioned by Seeman could be used for ordering biological and non-biological guest molecules¹². This could provide a method for determining crystal structures of otherwise difficult to crystallize proteins. Despite the advances in the DNA nanotechnology field, this goal has not been realized. This is mainly because only one truly periodic 3D DNA array composed entirely of Watson-Crick base pairs has been described to date (Figure 3)²³. The Watson-Crick duplexes formed by complementary DNA oligonucleotides in solution are linear molecules, making the construction of periodic 3D array composed entirely of Watson-Crick duplexes challenging. To create truly periodic 3D lattices we need to branch away from linearity. The DNA crossover junctions provide branching in the construction of the lattices however, these motifs are flexible making it difficult to achieve precise molecular associations. In our lab we use ‘non-Watson-Crick’ base pairing motifs to circumvent the issue of flexibility of the crossover junctions. Non-canonical base pairs are rigid, could provide branching and controlled intermolecular interaction in multiple dimensions, for the construction of periodic 3D lattices. However the occurrence of these motifs is less predictable as compared to the Watson-Crick base pairs.

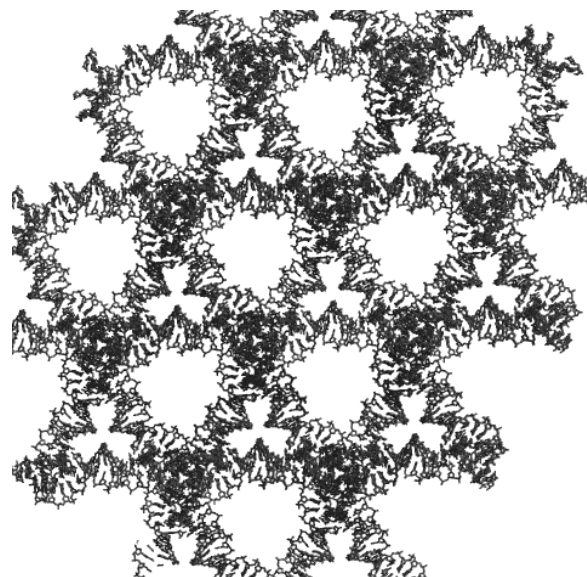


Figure 3. The lattice formed by the tensegrity triangle²³. The figure shows the X-ray crystal structure of the lattice formed by the tensegrity triangle based tile, determined by the Seeman group (PDB ID: 3GBI). It is a 3D periodic lattice composed entirely of Watson- Crick base pairing interactions.

Chapter 1.3 Structural Diversity of DNA

It has been known from some of the earliest structural studies that DNA can be both conformationally and structurally diverse³³. Depending on the environmental conditions, B-form DNA can undergo conformational transitions to the A- and the Z-forms³³⁻³⁶. Additionally, a variety of non-B-form DNA motifs have been characterized *in vivo* including DNA cruciform, hairpin structures, triplexes and quadruplexes³⁷⁻⁴³. There has been a growing appreciation for the use of these non-Watson-Crick structural motifs⁴⁴, including i-motifs⁴⁵, A-motifs^{46,47}, G-quadruplexes^{48,49} and parallel-stranded motifs^{50,51}, in the design of DNA nanoscale architectures and devices, as a way to provide structural diversity to nanoscale constructions. Watson-Crick duplexes are linear molecules and the use of non-Watson-Crick motifs provide rigid branch points that enable precise and controlled intramolecular interactions in 3D space. One of the major areas of DNA structural biology over the course of several decades was in understanding how non-Watson-Crick base pairs, or mismatches, could be accommodated in otherwise normal DNA helices, or are responsible for forming alternate DNA structures⁵²⁻⁵⁴. Non-canonical base pairs may be thermodynamically less stable in the context of a Watson-Crick helix⁵², but in certain sequences or other structural contexts they can be extremely stable⁵⁵ making them invaluable for DNA structural nanotechnology. However, the non-canonical motifs are not as predictable as the Watson-Crick base pairing.

Chapter 1.4 Motivation for Research

To date several techniques like UV and fluorescent spectroscopy⁵⁶, temperature controlled AFM assembly⁵⁷, single molecule FRET⁵⁸, have been employed to study the kinetics and thermodynamics of controlled complementary DNA self-assembly. These studies have enhanced the understanding of the DNA nanostructure assembly process, however, little is currently known about how periodic DNA arrays self-assemble into macroscopic crystals. A more thorough understanding of the assembly process would prove useful in the rational designing of DNA structures and help in optimizing the conditions for assembly, manipulation, and functionalization. This in turn will benefit both upstream design and downstream applications of DNA structural nanotechnology⁵⁶. The goal of this study was to understand the effect of nucleobase identities on DNA crystallization and structure while examining the assembly process of 3D DNA crystals.

Our lab previously determined a X-ray crystal structure of a continuously hydrogen bonded, self-assembling periodic 3D DNA crystal composed of both Watson-Crick and non-canonical base pairing interactions⁵⁹. The 13-mer DNA lattice has been successfully used as a scaffold for the incorporation of small guest molecules, as a model biomaterial solid for terahertz spectroscopy studies^{60,61} and also as a template for the design and development of 3D crystal lattices with expanded solvent channels^{62,63}. Here, I used this 13-mer DNA as a model to understand the relationship between sequence and structure. Specifically, I studied how the sequence in the duplex region of the 13-mer DNA influenced the crystallizability and the structure for these oligonucleotides and subsequently

performed the first experiments to examine the self-assembly process for these crystals.

Chapter 1.5 Scope of Research

I identified 21 crystallizing sequence variants of the 13-mer DNA. Sequence comparison of the crystallizing DNA oligonucleotides suggested Watson-Crick base pair preferences at specific positions in the hexameric sequence, some of which also correlated with crystallization speed. Structure determination of 13 of these variant crystals revealed that the sequences are structurally isomorphous, with relatively minor sequence-related differences restricted to certain nucleotide positions. Solution studies indicated that no base pairs are formed in the absence of additional divalent cations, and that assembly began immediately after the addition of Mg^{2+} . Finally, I demonstrated that I could generate heterogeneous crystals from mixtures of oligonucleotides that differed in the sequence of the duplex region, suggesting that the formation of Watson-Crick duplexes may be the initial step enabling crystal nucleation.

I also examined how a single nucleotide change in the DNA sequence was capable of promoting a significant change in the oligonucleotide interactions and the corresponding crystal structure, under identical crystallization conditions. Here, I described the crystal structures of four 14-mer oligonucleotides obtained by adding an adenosine at the 3' end of the DNA 13-mers. The 14-mer structures showed identical interactions in the Watson-Crick duplex region but a rearrangement of the non-canonical base pairs. Additionally, I also examined the role sequence played in the adoption of the alternate crystal form. The analysis suggested that the sequence in the 14-mer duplex region and the identity of the added nucleotide were necessary to promote this alternate structure. Remarkably, the added A14 residue from another

strand made tertiary contacts to the guanosine adjacent to a single sheared A-G pair, resulting in a conformation similar to tandem sheared G-A pairs. Together with a series of purine base triples, these interactions were responsible for the formation of the alternate crystal form. This study is a step forward in the understanding of the complex sequence/structure relationship of DNA oligonucleotides.

Chapter 2: Probing the Role of Sequence in the Self-Assembly Process of 13-mer DNA Crystals

Chapter 2.1 Introduction

Construction of a periodic 3D array composed entirely of Watson-Crick duplexes is challenging as the Watson-Crick duplexes are linear molecules. Repetitive periodic structures using immobile crossover junctions have been constructed however these junctions are flexible in nature¹¹. Sophisticated 3D constructions require more precise junctions. However, DNA crystal assemblies are not limited to Watson-Crick base pairing interactions. Non-canonical base pairs can provide structural diversity for the creation of branched DNA structures.

Hydrogen bonded non-Watson-Crick base pair combinations or Watson-Crick base pairs bonded through ‘non-Watson-Crick’ hydrogen bond donor-acceptor pairs are referred to as non-canonical base pairs. They occur naturally in RNA, the telomeric regions of DNA, and are associated with the structural and functional diversity of these molecules^{55,64}. Even though non-canonical base pairs are thermodynamically less stable in the context of a Watson-Crick helix⁵², they can be both extremely predictable and stable⁵⁵ in other structural contexts. There has been a growing appreciation for the use of non-Watson-Crick structural motifs in the design of DNA nanoscale architectures and devices.

One such parallel-stranded homopurine 5'-GGA motif was first identified and reported in a 13-mer crystal structure from our lab⁵⁹ (Figure 6A, B). In this crystal structure, the 13 nucleotide long DNA 5' GGACAGATGGGAG 3' is continuously hydrogen bonded through both Watson-Crick and non-canonical base pairs⁵⁹ and self-

assembles in the presence of Mg^{2+} . The crystal structure shows one DNA 13-mer base paired with three identical neighbors to form two distinct regions of base pairing (Figure 4A). The antiparallel Watson-Crick helical region is made up of a six base pair long self-complementary duplex composed of two sets of C4-G9 and A5-T8 Watson-Crick base pairs and two Type II⁶⁵ G6-A7 non-canonical base pairs about the central dyad axis (Figure 4B-D). Residues G1-A3 form parallel-stranded non-canonical homopurine base pairs with G10-A12 of another monomer (Figure 4E). Two sets of these non-canonical base pairs stack on a crystallographic symmetry axis to serve as a junction that links different helical layers into a continuously hydrogen bonded 3D lattice structure (Figure 4F). These interactions lead to the formation of solvent-occupied channels running throughout the length of the crystal. Two kinds of solvent channels with 300 \AA^2 and 360 \AA^2 cross-section area, run parallel and perpendicular to the six-fold symmetry axis respectively (Figure 5A, B). These solvent channels allow incorporation of guest molecules within the lattice.

The 13-mer crystals have also found application as a model biomaterial solid for terahertz spectroscopy studies^{60,61}. The GGA motif identified from the 13-mer structure has been successfully used in different sequence contexts to rationally design DNA crystals with expanded solvent channels to allow incorporation of protein guest molecules^{62,63}. My work was to understand how the nucleobase identities of the B-form duplex region influenced crystallization and in the process understand the assembly process for the model 13-mer DNA crystals, to help guide rational designing of 3D DNA crystals better suited for downstream applications as molecular scaffolds.

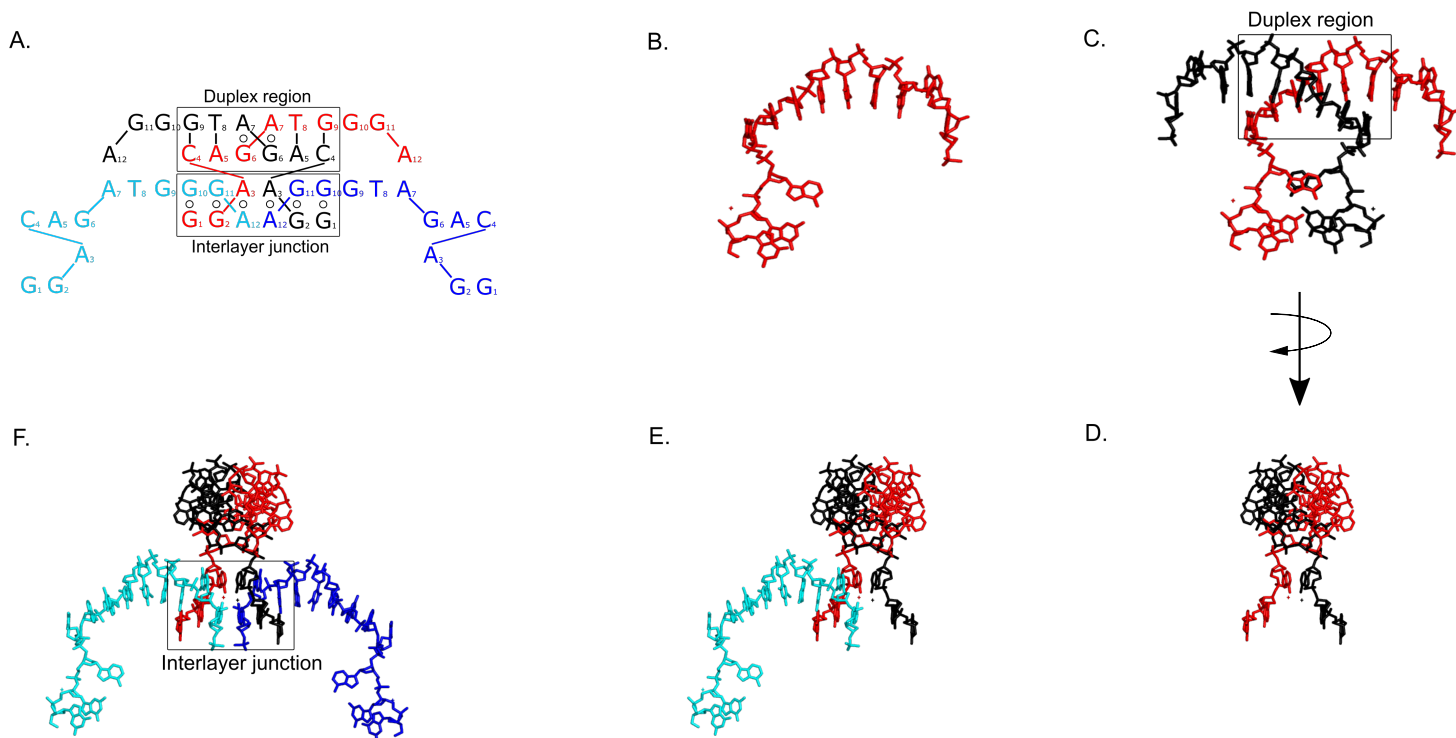


Figure 4 The crystal structure of d(GGACAGATGGGAG) (PDB ID 1P1Y)⁵⁹. **A.** Secondary structure showing the interaction between neighboring 13-mers (each strand colored differently). **B-F.** Show the generation of the continuously hydrogen bonded 3D lattice for the DNA 13-mer, through the formation Watson-Crick and non-canonical interactions.

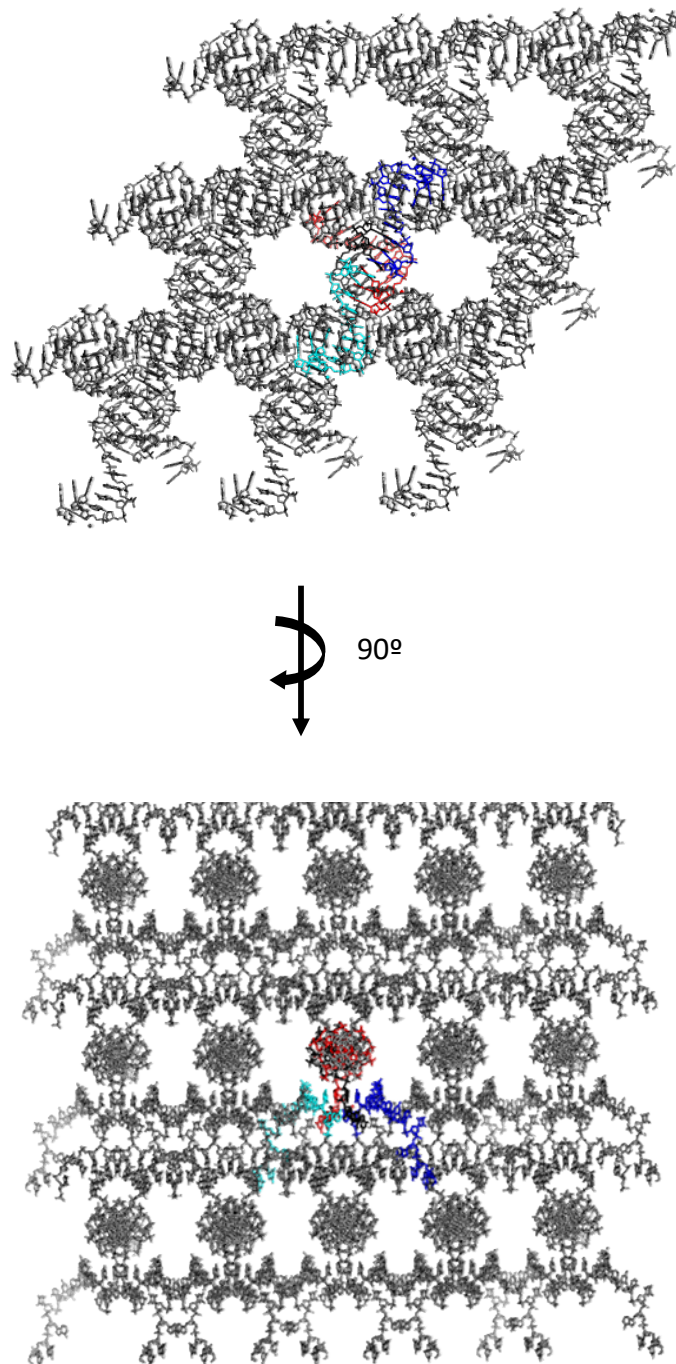


Figure 5. Solvent channels in the 13-mer lattice. The organization of the 13-mer in the 3D crystal lattice results in solvent channels of 300 \AA^2 and 360 \AA^2 running parallel (A) and perpendicular (B) to the six-fold symmetry axis respectively.

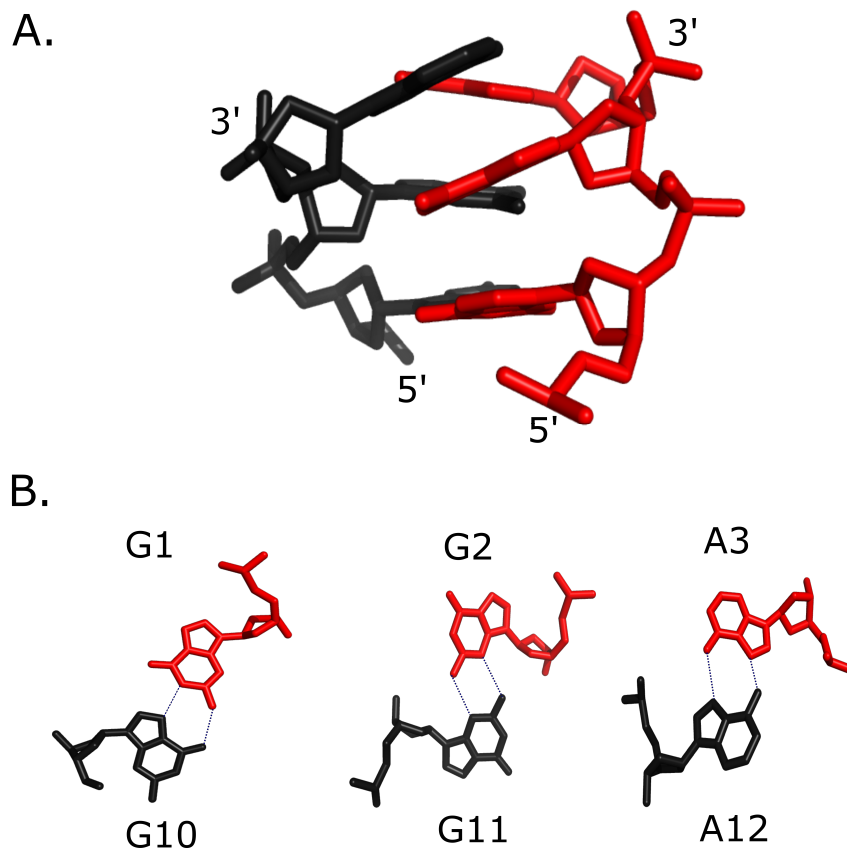


Figure 6. The 5' GGA- motif in the structural context of the DNA 13-mer. A. Two sets of GGA motifs (red and black) form parallel homopurine base pairs. **B.** Individual non-canonical base pairs of the GGA motif. In the first pair, the Watson-Crick edge of G1 (N1, N2) hydrogen bonds with the Hoogsteen edge of G10 (N7, N6). In the second pair both guanines, G2 (N2, N3) and G11 (N3, N2) are hydrogen bonded through the sugar edge. In the third pair, the Hoogsteen edges of both adenosines, A3 and A12, are base paired to each other through symmetrical N6-N7 hydrogen bonds. The hydrogen bonds are shown as dotted lines.

Chapter 2.2 G-A Pairs in the Duplex Not Required for Crystallization

In the original 13-mer structure, the six base pair duplex region contains central G6-A7 non-canonical base pairs flanked by two Watson-Crick base pairs (Figure 3A). The presence of these non-canonical base pairs did not significantly distort the overall B-form helical geometry. However, the large propeller angle of the G6-A7 non-canonical base pair results in an additional cross-strand hydrogen bond between N2 of G6 and O2 of T8. A similar hydrogen bond was previously reported in the d(CCAAGATTGG) duplex structure⁶⁶.

To determine if the duplex region could be composed entirely of Watson-Crick pairs, I tested an A7C substitution that converted this region to a self-complementary duplex. The A7C (also referred to as B7 in this document) oligonucleotide crystallized under the same conditions and required only the presence of Mg²⁺ cations to assemble. The B7 crystals were morphologically identical to the original 13-mer, crystallized in the same space group with one molecule in the asymmetric unit and had nearly identical unit cell parameters (Table 1). The crystal structure of B7 was highly similar to the original 13-mer (Figure 7) with the expected interactions leading to the formation of a Watson-Crick anti-parallel helical region and the parallel-stranded non-canonical junction. As in the original structure, the electron density of the G13 phosphate was observed, while the remainder of the nucleotide, which was located within the solvent channel, was disordered.

Table 1. Data collection and refinement for B7.

B7	
Data collection	
Space group	P6 ₄
Cell dimensions	
<i>a, b, c</i> (Å)	40.641 40.641 51.983
α, β, γ (°)	90, 90, 120
$I/\sigma I$	17.2 (2.8)
Number of reflections	2549 (386)
Completeness (%)	97.0 (98.1)
Redundancy	3.4 (2.9)
Refinement	
Resolution (Å)	2.16-35.20 (2.16-2.21)
Number of reflection	2287 (170)
R factor	0.213 (0.492)
R _{free}	0.248 (0.408)
Number of atoms	
DNA	254
Ion	1
Water	3
Bond lengths (Å)	0.01
Bond angles (°)	1.57
PDB ID	4ROK

*Values in parentheses are for the highest-resolution shell.

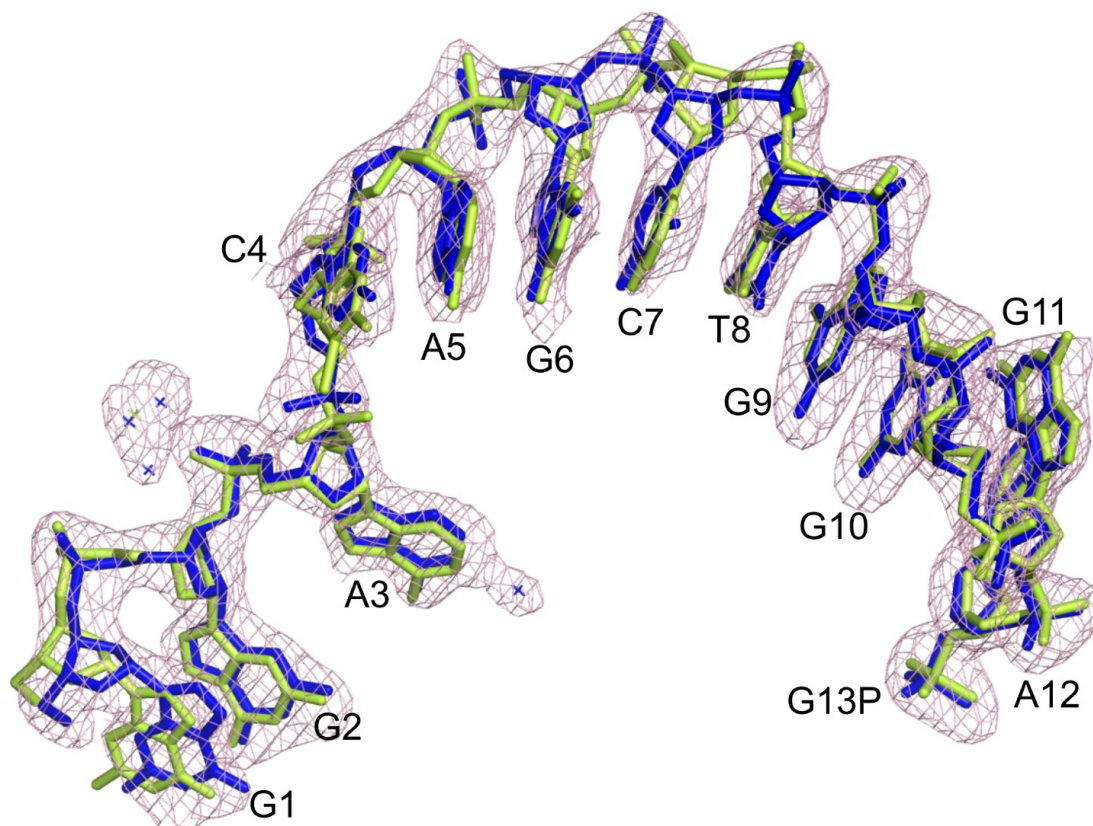


Figure 7. Structure of B7 variant compared to the original 13-mer. The original 1P1Y structure (green), superimposed on the B7 structure (blue). The sigma A-weighted electron density map ($2F_0-F_c$) of B7 variant is contoured at 1σ . Refined solvent molecules are shown as points with associated density. Only the phosphate of G13 (G13P) was present in the electron density for both structures.

Chapter 2.3 Sequence Requirements for Crystallization

To determine the sequence requirements within the duplex region necessary for crystal self-assembly, I screened all 64 possible self-complementary hexameric sequences. I maintained the GGA and GGAG sequences at the 5' and the 3' ends, respectively, to promote the formation of predictable non-canonical interactions required for the formation of the inter-layer junction that would hold the Watson-Crick duplexes in a crystal framework similar to the original 13-mer. This also helped me to correlate any observed structural changes to the sequence of the duplex region. Using just a single crystallization condition, I identified 20 new crystallizing DNA oligonucleotides along with the B7 variant as a positive control (Table 2). The size and morphology of the crystals ranged from large hexagonal pyramids that were identical to the B7 variant, to birefringent spherulites, clusters, and microcrystals (Figure 8). Crystals also appeared at different times and were classified into two groups. The fast group crystallized within 16 hours, and the slow group took more than 48 hours to crystallize. Out of the 21 crystallizing DNA oligonucleotides in the screen (including the B7), 10 belonged to the fast group and the remaining 11 belonged to the slower group (Table 2).

Table 2. Crystallized oligonucleotide sequences.

Designation	Sequence	Crystal morphology (Unpurified DNA)	Crystallization speed	Crystal morphology (purified DNA)
1p1y	GGA CAGATG GGAG	-	Fast	Hexagonal Pyramids
A1	GGA AAATTT GGAG	Microcrystals	Fast	Non-faceted hexagonal
A2	GGA AACGTT GGAG	Microcrystals	Fast	Hexagonal Pyramids
A3	GGA AAGCTT GGAG	Microcrystals	Fast	Hexagonal Pyramids
A4	GGA AATATT GGAG	Spherulites	Fast	No crystals
A7	GGA ACGCGT GGAG	Hexagonal Pyramids	Slow	No crystals
A12	GGA AGTACT GGAG	Spherulites	Slow	No crystals
B2	GGA ATCGAT GGAG	Hexagonal Pyramids	Slow	Hexagonal Pyramids
B6	GGA CACGTG GGAG	Microcrystals	Fast	Hexagonal Pyramids
B7	GGA CAGCTG GGAG	Hexagonal Pyramids	Fast	Hexagonal Pyramids
B8	GGA CATATG GGAG	Spherulites	Slow	Spherulites
B9	GGA CCATGG GGAG	Microcrystals	Slow	Hexagonal Pyramids
B11	GGA CCGCGG GGAG	Hexagonal Pyramids	Slow	Hexagonal Pyramids
C1	GGA CGATCG GGAG	Microcrystals	Slow	Non-faceted hexagonal
C2	GGA CGCGCG GGAG	Spherulites	Slow	Spherulites
C3	GGA CGGCCG GGAG	Intermediate morphology	Fast	Hexagonal Pyramids
D5	GGA GGATCC GGAG	Spherulites	Slow	Spherulites
E1	GGA TAATTA GGAG	Microcrystals	Fast	Hexagonal Pyramids
E2	GGA TACGTA GGAG	Hexagonal Pyramids	Fast	Hexagonal Clusters
E3	GGA TAGCTA GGAG	Hexagonal Pyramids	Fast	Hexagonal Pyramids
E4	GGA TATATA GGAG	Spherulites	Slow	No crystals
F1	GGA TTATAA GGAG	Cubic	Slow	No crystals

Bold: DNA oligomers with crystal structures determined

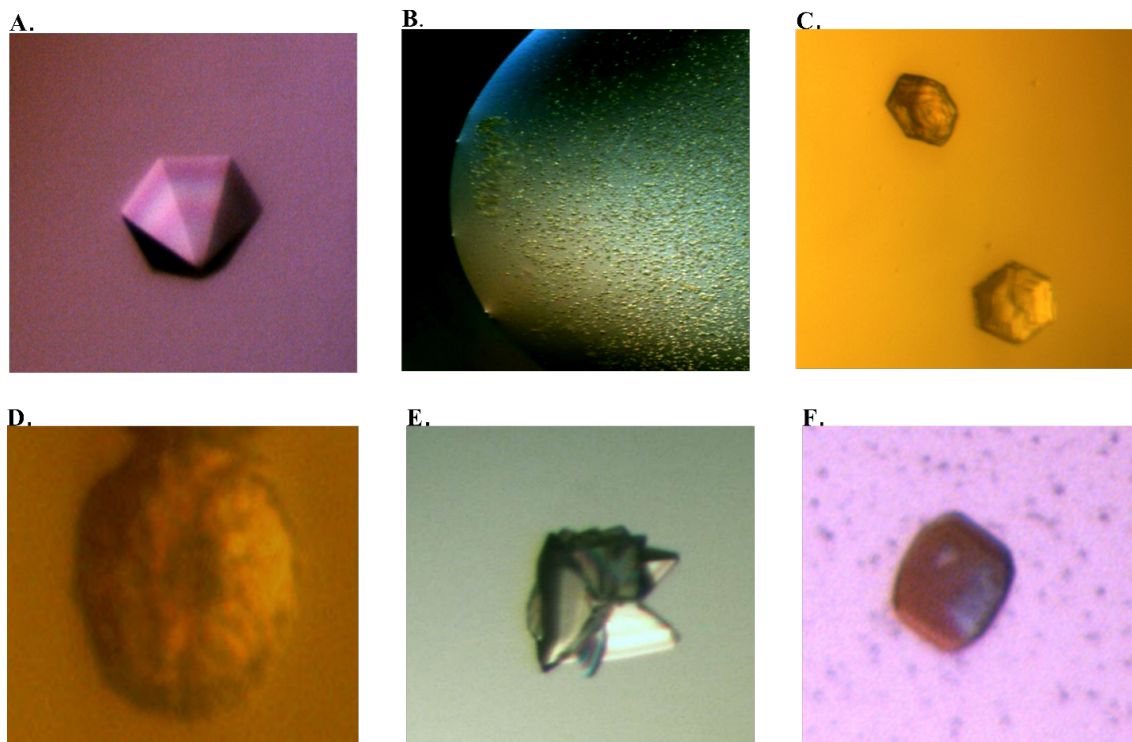


Figure 8. Representative crystal morphologies for DNA oligonucleotides identified in the screen (Table 2). A. Hexagonal pyramid. B. Microcrystals. C. Non-Faceted hexagon. D. Spherulite. E. Hexagonal clusters. F. Cubic

Sequence analysis of the crystallizing oligonucleotides revealed that all possible Watson-Crick base pairs were represented at least once at each of the six positions in the duplex region (Figure 9A). The strongest preference was for the A5-T8 base pair (11 of 21) with C4-G9 being the second most prevalent pair (8 of 21). Because of readily observable differences in the crystallization speed, I also compared sequences within the fast group (Figure 9B). Remarkably, 9 out of 10 oligonucleotides in the fast group had an A5-T8 base pair, while only 2 out of 11 from the slow group had the A5-T8 base pair. The most uncommon base pair was the G4-C9 base pair, with only one sequence containing this pairing. Notably, this belonged to the slow crystallization group and formed only birefringent spherulites, even after purification.

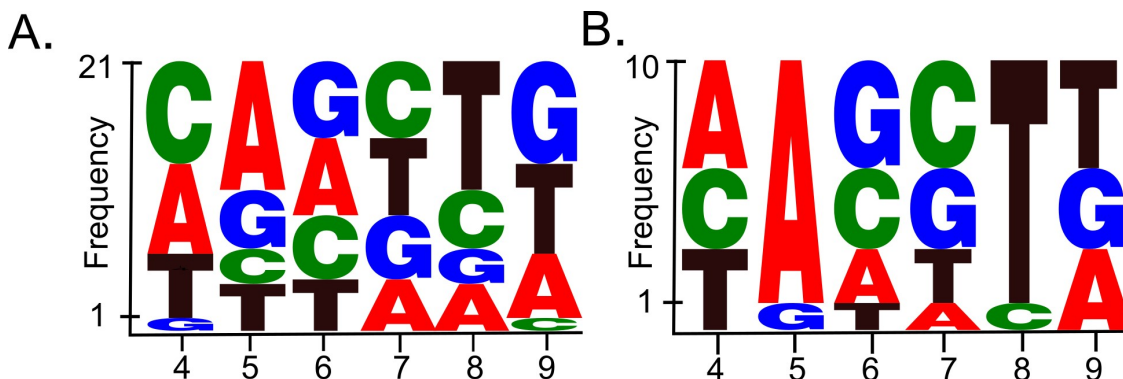


Figure 9. Nucleobase frequency distribution of crystallizing DNA oligonucleotides. Letter height reflects the frequency for each nucleotide position. **A.** Frequency distribution for the 21 DNA oligonucleotides identified in the initial crystal screen (including B7). **B.** Frequency distribution for the 10 DNA oligonucleotides belonging to the fast crystallizing sub-group (including B7).

The Ho laboratory took a similar screening approach to examine how sequence influences DNA structure⁶⁷. There, self-complementary duplexes containing d(CCnnnNNNGG) sequences showed sequence and environmental preferences for A-form, B-form, and 4-way junction structures. We anticipated that positive crystallization results from our study should correlate with sequences that preferentially take the B-form under crystallization conditions from their study. Interestingly, however, we found that only 3 of the 21 hexameric duplex sequences (A2, A12 and E1) were previously identified as preferential B-form sequences. Four of the sequences were identified as having A-form preferences (B9, B11, C1 and C3), while the remaining 14 had not been characterized. This indicates that the flanking regions, in our case the critical non-canonical base pairs, may play a significant role in how self-complementary sequences influence macromolecular conformation.

Chapter 2.4 Structural Implications of the Duplex Sequences

I solved the X-ray crystal structures of 12 different sequence variants to determine if the different duplex sequences had any observable structural implications. 14 out of the 20 new DNA oligonucleotides identified in the screen crystallized after gel purification (Table 2; Materials & Methods). From these, 12 gave morphologically similar hexagonal pyramidal crystals, while 2 yielded crystals that did not diffract x-rays. These 12 variants diffracted x-rays to high resolution limits of 2.03-2.39 Å, crystallized in the same space group, and had a narrow distribution of unit cell parameters (Table 3). Structurally, all variants were highly similar with the expected base pairing interactions. The average RMSD for the invariant 5' and 3' residues of all structures when compared to B7 was 0.31 Å, while the average RMSD for all backbone atoms compared to B7 was 0.49 Å. The G13 sugar and nucleobase was disordered in all the structures.

Table 3. Data collection and refinement for 12 sequence variants.

	A1	A2	A3	B2	B6	B9
Data collection						
Space group	P6 ₄	P6 ₄	P6 ₄	P6 ₄	P6 ₄	P6 ₄
Unit cell (a,b,c) (Å)	39.729 39.729 55.989	40.399 40.399 54.627	40.317 40.317 54.958	40.232 40.232 52.910	40.562 40.562 51.411	40.719 40.719 51.214
Resolution (Å)*	2.08-55.99 (2.08-2.19)	2.04-34.99 (2.04-2.15)	2.03-54.96 (2.03-2.14)	2.08-52.91 (2.08-2.19)	2.08-35.13 (2.08-2.20)	2.39-51.21 (2.39-2.52)
R _{merge}	0.051 (0.390)	0.034 (0.911)	0.030 (2.171)	0.021 (0.408)	0.025 (0.395)	0.030 (0.747)
I/σI	24.8 (3.5)	24.3 (1.3)	26.3 (0.9)	26.2 (1.7)	23.0 (1.4)	27.1 (1.8)
Number of reflections	3045 (426)	3261 (483)	3297 (482)	2873 (347)	2578 (321)	1953 (289)
Completeness (%)	99.2 (95.7)	99.6 (99.5)	99.3 (99.5)	96.0 (79.2)	88.5 (75.3)	99.6 (100)
Redundancy	5.2 (2.5)	3.8 (3.9)	5.0 (5.1)	3.1 (1.8)	2.8 (1.6)	5.0 (5.1)
Refinement						
Resolution (Å)	2.08-34.41 (2.07-2.13)	2.04-34.99 (2.04-2.09)	2.03-34.92 (2.03-2.08)	2.08-34.84 (2.07-2.13)	2.08-35.13 (2.08-2.13)	2.39-35.26 (2.39-2.45)
Number of reflections	2738 (182)	2952 (227)	2952 (191)	2574 (149)	2325 (137)	1743 (128)
R/R _{free}	0.200 0.239	0.226 0.259	0.221 0.251	0.225 0.265	0.216 0.219	0.210 0.252
Number of atoms						
DNA	254	254	254	254	254	254
Ion	2	1	1	1	1	1
Water	9	8	7	7	2	2
Bond lengths (Å)	0.01	0.01	0.01	0.01	0.01	0.01
Bond angles (°)	1.8	1.06	1.17	1.23	1.13	1.33
PDB ID	4RNK	4RO4	4RO7	4RO8	4ROG	4RON

*Values in parentheses are for the highest-resolution shell.

Table 3. Data collection and refinement (continued)

	B11	C1	C3	E1	E2	E3
Data collection						
Space group	P6 ₄	P6 ₄	P6 ₄	P6 ₄	P6 ₄	P6 ₄
Unit cell (a,b,c) (Å)	40.789 40.789 52.007	40.442 40.442 52.514	40.755 40.755 51.819	40.177 40.177 53.015	40.805 40.805 51.668	40.780 40.780 51.813
Resolution (Å)*	2.37-52.01 (2.37-2.50)	2.09-35.02 (2.09-2.20)	2.08-35.29 (2.08-2.19)	2.19-34.79 (2.19-2.31)	2.27-35.34 (2.27-2.40)	2.32-35.32 (2.32-2.45)
R _{merge}	0.025 (1.257)	0.029 (0.595)	0.030 (0.456)	0.024 (1.369)	0.037 (1.273)	0.030 (0.582)
I/σI	23.0 (0.9)	33.6 (1.8)	19.5 (1.2)	31.2 (1.1)	16.4 (1.0)	21.1 (1.6)
Number of reflections	1999 (291)	2944 (421)	2663 (322)	2540 (375)	2219 (321)	2079 (317)
Completeness (%)	97.9 (99.3)	99.7 (98.4)	89.1 (74.8)	99.7 (99.2)	97.7 (97.4)	96.3 (97.4)
Redundancy	3.8 (3.8)	6.1 (2.9)	2.8 (1.5)	5.5 (5.3)	2.8 (2.9)	2.9 (2.9)
Refinement						
Resolution (Å)	2.37-35.32 (2.37-2.43)	2.09-35.02 (2.08-2.13)	2.08-35.29 (2.07-2.13)	2.19-34.79 (2.19-2.24)	2.27-35.34 (2.27-2.33)	2.32-35.32 (2.32-2.38)
Number of reflections	1772 (128)	2655 (204)	2397 (143)	2285 (148)	2003 (140)	1878 (147)
R/R _{free}	0.237 0.301	0.221 0.250	0.220 0.244	0.222 0.231	0.202 0.243	0.202 0.231
Number of atoms						
DNA	254	254	254	254	254	254
Ion	1	1	1	1	1	1
Water	0	4	1	2	1	3
Bond lengths (Å)	0.01	0.01	0.01	0.01	0.01	0.01
Bond angles (°)	1.44	1.01	0.98	1.08	1.29	1.16
PDB ID	4ROO	4ROY	4ROZ	4RP0	4RP1	4RP2

*Values in parentheses are for the highest-resolution shell.

To examine if the sequence preference for the A5-T8 base pair observed in the initial screen translated into local structure differences, I aligned the structures containing the A5-T8 pair and those without and compared the sugar-phosphate backbone atoms at the six duplex positions (Figure 10 & 11). The average backbone RMSDs for those structures containing the A5-T8 base pair was 0.65 Å, while the average RMSDs for the group without the A5-T8 pair was 0.93 Å. Analysis of the backbone by nucleotide position suggested that the lower RMSD values for the A5-T8 group was not simply due to those structures having uniform nucleobase identities at these positions. The A5 and T8 backbone atoms had average RMSDs of 0.61 and 0.67 Å, respectively, while positions 4 and 6 with variable nucleobases had lower or similar average RMSD values (0.51 and 0.64 Å). In contrast, the non-A5-T8 structures had higher average backbone RMSDs for all but position 9, with positions 5 and 6 having the highest variability (1.25 and 1.10 Å, respectively). The lower average RMSDs around position 5 when these residues are adenosines indicated that the backbone conformation in these structures is much less variable suggesting that this nucleobase is important for establishing a local structure that fits the crystal framework. This is also supported by the observation that A5-T8 is present in 9 out of 10 fast crystallizing oligonucleotides.

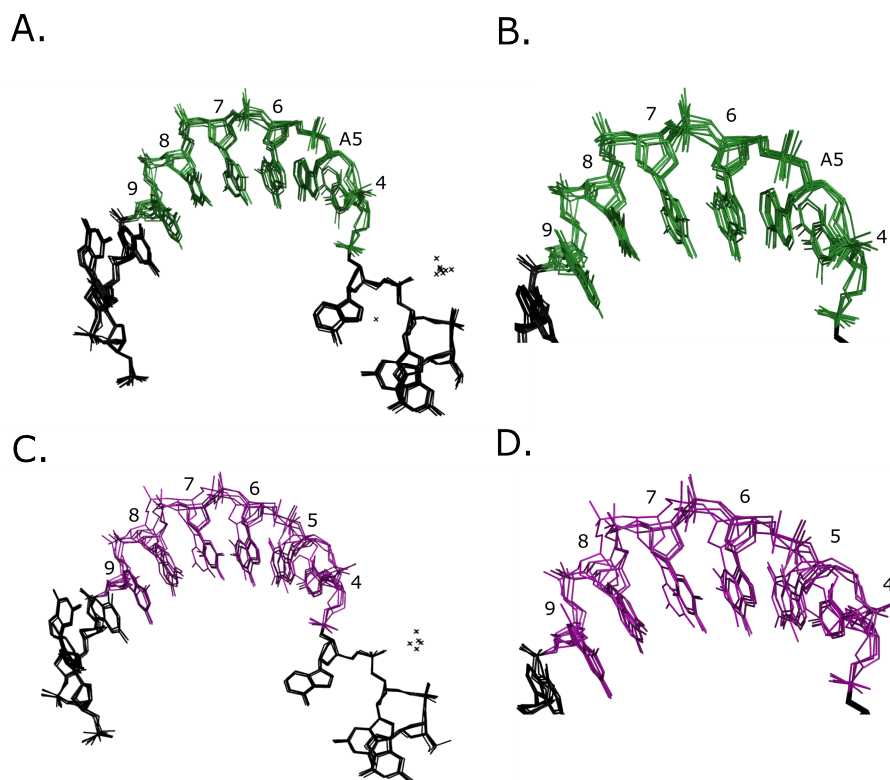


Figure 10. Superimposition of the 13-mer structures with and without the A5-T8 base pair. The constant G1-A3 and G10-G12 residues were used as a reference for alignment. **A.** The overlap between 8 structures (A1, A2, A3, B6, B7, E1, E2 and E3) having the A5-T8 base pair. **B.** The duplex region of the overlapped A5-T8 structures. **C.** 5 superimposed structures (B2, C1, C3, B9 and B11) lacking the A5-T8 base pair. **D.** The duplex region of the non-A5-T8 structures.

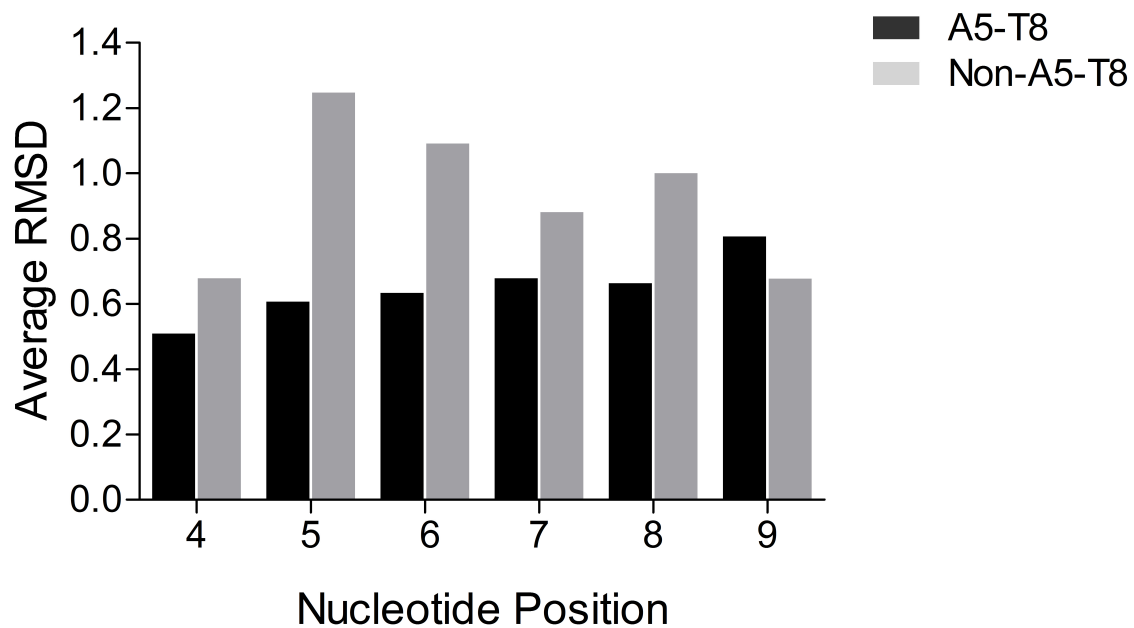


Figure 11. Average RMSD comparison for the duplex region of A5-T8 and the non-A5-T8 structures. The average root mean square deviation (RMSD) values, calculated by comparing all structures to each other in a pairwise manner, for backbone atoms for positions 4 through 9 of the duplex regions are grouped into A5-T8 (black) and non-A5-T8 (gray) groups.

Chapter 2.5 B7 Does Not Form Base Pairs in Absence of Cations

To begin to understand how these DNA 13-mers assemble into periodic crystal lattices, we examined the behavior of the oligonucleotides in solution. Using the B7 variant as a representative sequence, we examined the oligonucleotide by 1D- ^1H NMR with the assistance of Dr. Daoning Zhang. The spectrum showed sharp proton signals (Figure 12A) at conditions analogous to those used to store the oligonucleotide prior to crystallization (200 μM in water with 10% D_2O). Importantly, the lack of imino proton signals in the 9-14 ppm range suggested that the oligonucleotides did not form stable Watson-Crick or non-canonical base pairs in absence of cations. These results indicated that B7 is primarily a dispersed monomer under these conditions.

Even though the initial crystal growth was performed in a buffer containing 120 mM magnesium formate, 50 mM lithium chloride, 10% 2-methyl-2,4-pentanediol, B7 DNA crystallized in just a 120 mM solution of Mg^{2+} solution, demonstrating that divalent cations were the only minimal requirement for crystallization. To examine the effect of cation on the NMR spectra, we performed a Mg^{2+} titration on the B7 sample. We observed that addition of 5 mM Mg^{2+} led to significant signal broadening that became more severe with increasing Mg^{2+} concentration (Figure 13). This suggested the formation of high molecular weight assemblies. Some areas of the spectrum showed chemical shift changes upon addition of the cation. This was particularly true for the nucleobase proton region (Figure 12B) and the deoxyribose H1' region (Figure 12C). These chemical shift changes are likely the result of the altered environment upon base pairing

interactions, while the signal broadening is a result of the assembly into high molecular weight complexes. However, only two low intensity peaks in the imino proton region appeared after the addition of Mg^{2+} (Figure 12D). These peaks were most prominent at low Mg^{2+} and diminished at higher concentrations. The formation of discrete stable species through Watson-Crick or the non-canonical base pairs should likely have resulted in a larger number of imino signals with greater intensity. The limited number of imino peaks likely reflected the rapid base pairing and assembly of these oligonucleotides into high molecular weight structures.

To further investigate cation-induced assembly, we determined transverse relaxation times (T_2) for the oligonucleotide both above and below stoichiometric Mg^{2+} concentrations, again in collaboration with Dr. Daoning Zhang (Table 4). T_2 is inversely proportional to the tumbling time of the molecule in solution, resulting in lower T_2 values as molecular weight increases. Titration of sub-stoichiometric Mg^{2+} (1:2) resulted in a small but measurable (3 ms) decrease in T_2 . A similar decrease (3 ms) was observed as Mg^{2+} was increased to equimolarity. Interestingly, a significantly greater decrease in T_2 (19 ms) was observed as Mg^{2+} exceeded the DNA concentration (2.5:1), but did not change as Mg^{2+} was increased again (4:1). We could not accurately determine T_2 at higher Mg^{2+} ratios due to the significant line broadening that resulted in a loss of peaks at the longer decay times necessary for fitting. However, these results show that even low concentrations of divalent cations can induce formation of higher molecular weight structures.

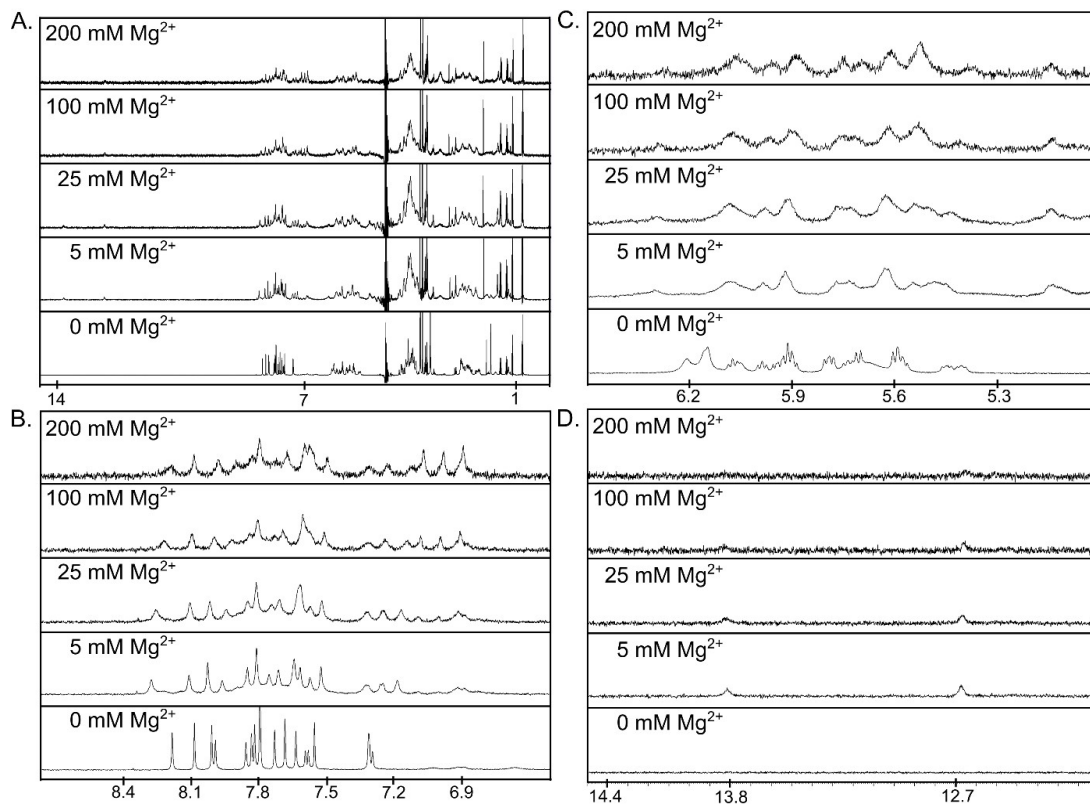


Figure 12. 1D-¹H NMR spectra of B7 variant. **A.** The complete spectrum for the B7 (A7C) variant in 0-200mM Mg²⁺ ions obtained in collaboration with Dr. Daoning Zhang. The base proton region (**B**), and the deoxyribose H1' region (**C**) shows broadening of peaks and chemical shift changes upon addition of increasing amounts of Mg²⁺ ions. **D.** The imino proton region lacks proton signals in the absence of Mg²⁺. Two low intensity signals appear at low Mg²⁺ concentrations, and disappear as Mg²⁺ concentration was increased.

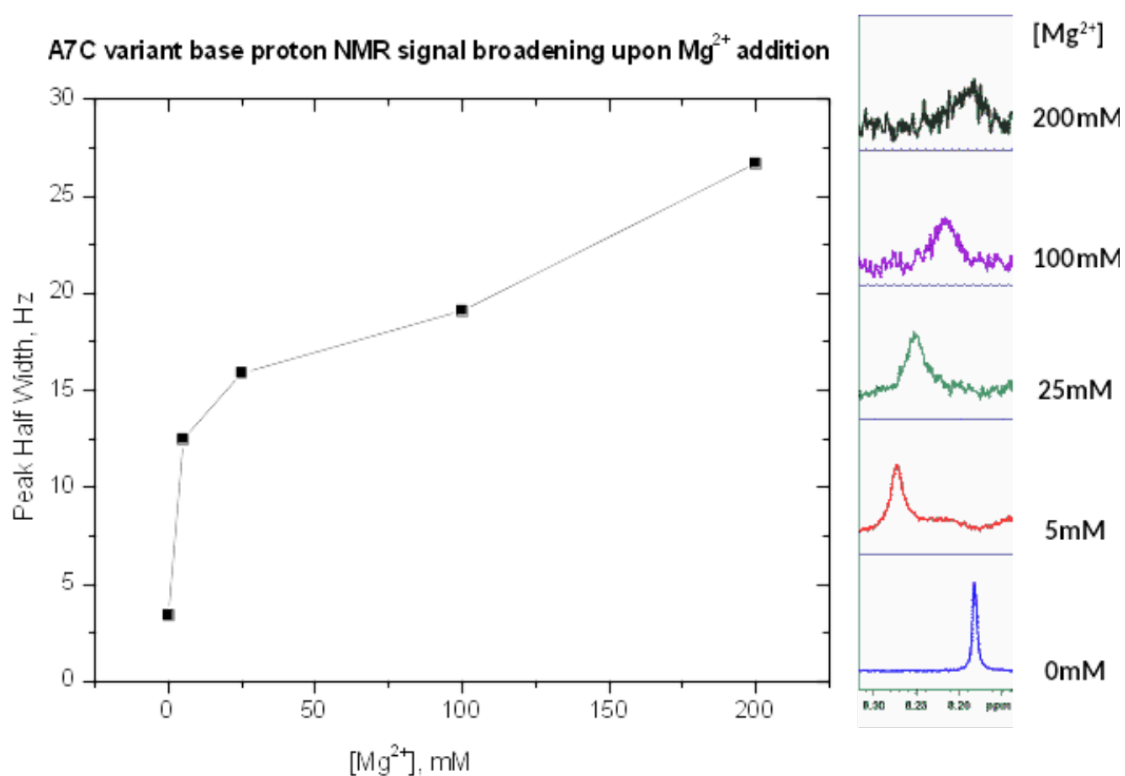


Figure 13. NMR signal broadening upon Mg²⁺ addition. Base proton resonance peaks were measured at different Mg²⁺ concentrations (right) in collaboration with Dr. Daoning Zhang. The width of the peak at half its total height is shown plotted against Mg²⁺ titration concentrations (left).

Table 4. Transverse relaxation experiments on B7 in presence of Mg²⁺.

DNA (mM)	MgCl ₂ (mM)	Mg ²⁺ /DNA	1/T ₂	T ₂ (ms)	Fit (R ²)
0.2	0	0	16.1	62	0.9986
0.2	0.1	0.5	16.87	59	0.9932
0.2	0.2	1	17.85	56	0.9997
0.2	0.5	2.5	26.85	37	0.9979
0.2	1	4	27.31	37	0.9776

Chapter 2.6 Watson-Crick or Non-canonical Base Pairs as Drivers for

Crystallization

Though there are many routes for crystal assembly and nucleation⁶⁸, one interesting possibility in this system is that the addition of divalent cations initially promotes intermolecular interactions that form pre-nucleation assemblies. In the absence of observable hydrogen bonding interactions in solution NMR experiments, we considered two simple scenarios for how base pairing could lead to pre-nucleation species and ultimately to crystal nucleation. The “duplex first” model would require the formation of base paired dimers through the specificity of the hexameric anti-parallel Watson-Crick duplex (Figure 14A). The dimers could then form non-canonical interactions through the free purine motifs at both ends of the duplexes, leading to crystal nucleation (Figure 14B). In the “non-canonical first” model the stability provided by the extensive base stacking interactions of the non-canonical region could initiate strand interactions, followed by the Watson-Crick interactions leading to the formation of a crystal nucleation site (Figure 14C). Analyzing these models from the perspective of mixed sequences, our analysis favored the duplex first model, as this process would always generate productive assemblies that could continue to propagate during crystal growth without base pair clashes. In contrast, depending on the associations formed during the assembly process, the non-canonical first model could lead to unproductive base pair clashes in the duplex regions that could potentially impact both crystal nucleation and growth (Figure 14D). Though we recognize that these models are not necessarily mutually exclusive and that the relatively long incubation times during crystal growth could allow for equilibrium

partitioning, the identification of multiple DNA sequences capable of forming isomorphous structures allowed us to test these models.

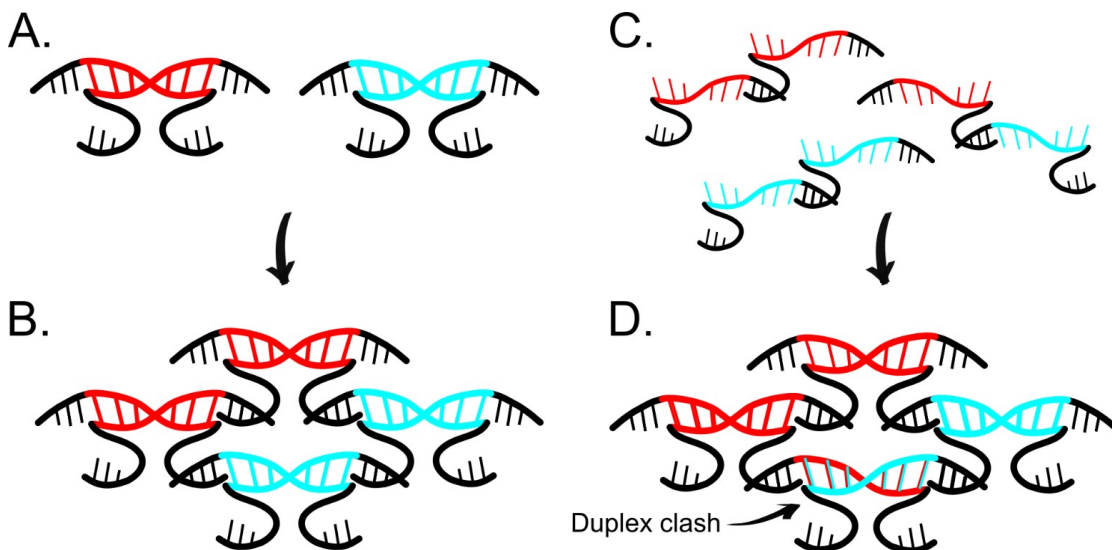


Figure 14. Models for crystal pre-nucleation. The duplex first and the non-canonical first models for crystal pre-nucleation explained by mixed sequence crystallization. DNA oligonucleotides differing in just the duplex sequence are shown in red and cyan. **A.** In the duplex first model, a mixture would partition first into respective complementary duplexes following the addition of Mg^{2+} . **B.** These duplexes could then assemble into heterogeneous higher-order structures through non-canonical pairings (black) leading to crystal nucleation. **C.** In the non-canonical first model, DNAs would assemble randomly through the non-canonical interactions, resulting in both homogeneous and heterogeneous dimeric structures. **D.** These dimeric structures could then assemble into higher-order structures, however, this could result in duplex clashes, as indicated.

To test the two models for crystal pre-nucleation, I created heterogeneous crystals from mixtures of different DNA oligonucleotide sequences identified from our screen. To detect both DNAs present in the heterogeneous crystals by gel electrophoresis I first tested mixtures in which one oligonucleotide contained a single nucleotide 3' extension. I confirmed that the addition of a 14th adenosine nucleotide did not prevent crystallization for 7 out of 10 extension variants I tested (A1, A2, A3, B6, B7, E1 and E3). However, for three of these (A1, A2 and A3), the addition of the 3' residue resulted in morphologically distinct crystals with different space groups and unit cell parameters (Figure 23B-E). These structures will be described in detail in Chapter 3. The remaining three 14-mer variants, all belonging to the slow group in the initial screen of 13-mers (B2, B11 and C1), did not crystallize under the tested condition.

Washed and dissolved crystals grown from two oligonucleotides with identical duplex sequences, but different lengths showed that the relative ratios of the two oligonucleotides in single crystals were consistent with the ratio of the oligonucleotides in the input mixture (Figure 15A and Figure 16). This indicated there was no preference for the longer or shorter oligonucleotide during crystal assembly, consistent with the 13 and 14 positions being oriented into the crystal solvent channels. Heterogeneous crystals grown with different duplex sequences of different length mixed together showed similar results (Figure 15B, C & Figure 16). Both oligonucleotides were present in all combinations of heterogeneous 13- and 14-mer mixtures that I tested, with only the E3-13/B7-14 combination showing some preference for incorporation of the 14-mer relative to the input mixture.

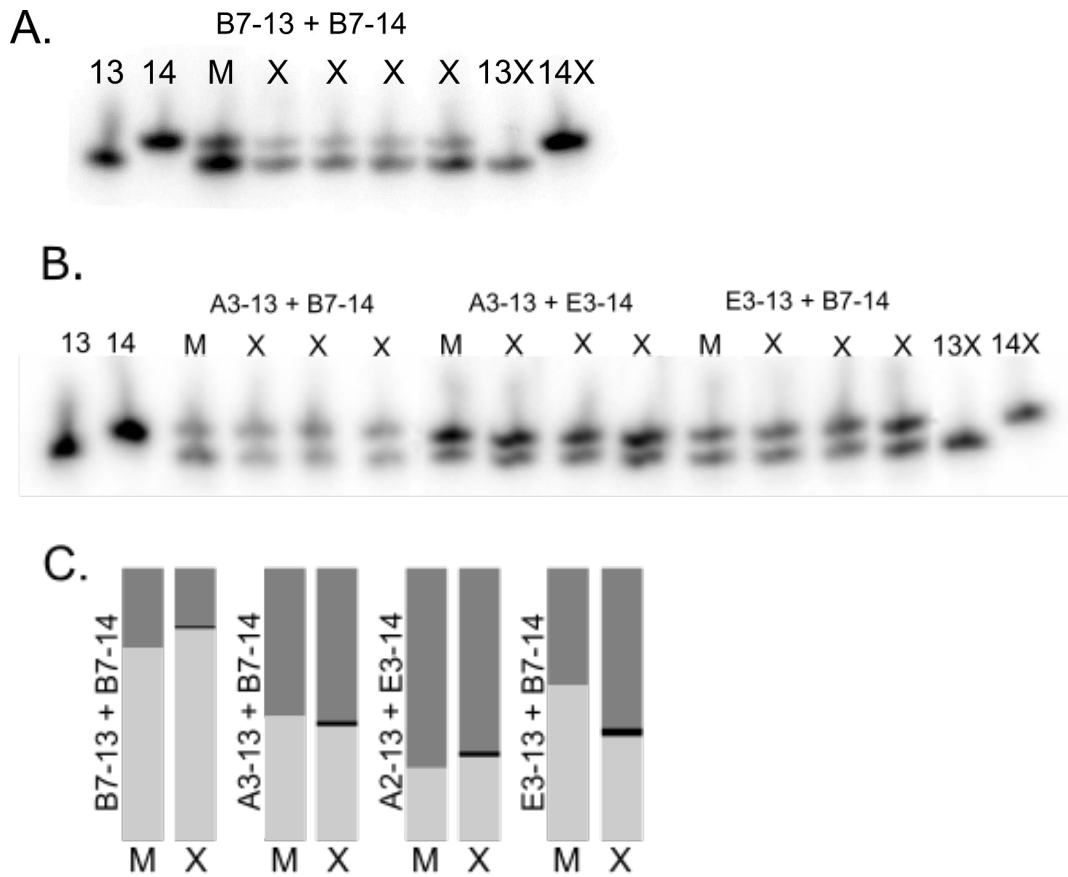


Figure 15. Heterogeneous crystal from DNA mixtures. **A.** Denaturing polyacrylamide gel for heterogeneous single crystals grown from a mixture of 13 and 14 nucleotide long DNAs with the same sequence in the duplex region (B7-13 and B7-14). **B.** A denaturing polyacrylamide gel for heterogeneous single crystals grown from a mixture of 13 and 14 nucleotide long DNAs with different sequence in the duplex region (indicated). For both gels, lanes designated 13 and 14 are the separate oligonucleotides; M, oligonucleotide mixture used for crystallization; X, single crystals grown from the mixture; 13X and 14X, single crystals grown from the individual oligonucleotides. **C.** Quantification of gels (A and B) showing the ratio of 13 nucleotide (light grey) and 14 nucleotide (dark grey) DNAs in the crystallization mixture (M) and averages of single crystals (X) obtained from the mixture. Black regions at the interface are standard deviations for four crystal (B7-13 + B7-14) or three crystals (remaining mixtures).

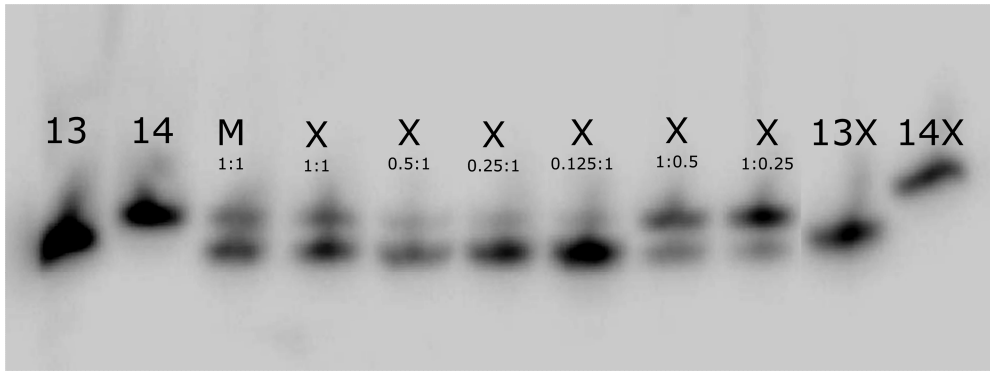


Figure 16. Heterogeneous crystal from ratios of DNA mixtures. Denaturing polyacrylamide gel for heterogeneous single crystals grown from a mixture of 13 and 14 nucleotide long DNAs with the same sequence in the duplex region (B7-13 and B7-14) and mixed in different ratios to each other. The ratio's are 14 mer DNA : 13 mer DNA based on the input mixture. Lanes designated 13 and 14 are the separate oligonucleotides; M, oligonucleotide mixture used for crystallization; X, single crystals grown from the mixture; 13X and 14X, single crystals grown from the individual oligonucleotides.

Finally, I screened 78 combinations through pair-wise mixing of the 13 DNA oligonucleotides for which the structures were determined. 74 out of 78 mixtures produced crystals of the typical size (Table 5) irrespective of the number of base pair differences in the mixed oligonucleotides. All the combinations gave me normal sized crystals. These results supported the model of crystal assembly by the formation of Watson-Crick pairs. The DNA mixtures did not inhibit crystallization in any appreciable way as expected from the non-canonical first model. While it is possible that some combinations could support crystallization through non-Watson-Crick pairings in the duplex region (similar to the original 13-mer structure), it is unlikely that all 74 of 78 pairwise mixtures could do so. Therefore, it is likely that both the initial crystal assembly steps and the propagation of crystal growth rely primarily on the formation of Watson-Crick base pairs. This also suggests that during crystal growth, the units incorporated into the crystals have already formed Watson-Crick duplexes. It is not clear if the incorporated unit is a duplex or some larger assembly. Further studies will be necessary to directly detect these base pairs either before or during the assembly process.

Table 5. Heterogeneous crystals for pairwise combinations of 13-mer DNA oligonucleotides

		A1	A2	A3	B2	B6	B7	B9	B11	C1	C3	E1	E2	E3
		AAA	AAC	AAG	ATC	CAC	CAG	CCA	CCG	CGA	CGG	TAA	TAC	TAG
E3	TAG													
E2	TAC													
E1	TAA													
C3	CGG													
C1	CGA													
B11	CCG													
B9	CCA													
B7	CAG													
B6	CAC													
B2	ATC													
A3	AAG													
A2	AAC													
A1	AAA													

78 total combinations
19 combinations with 3 base pair difference
35 combinations with 2 base pair difference
20 combinations with 1 base pair difference
4 combinations with no crystals
single nucleotide crystals

Chapter 2.7 Influence of Sequence on Crystallizability, Structure and Self-Assembly

This work has revealed many new DNA oligonucleotide sequences capable of adopting isomorphous crystal structures. Through analysis of sequence and structure I have found a strong preference for only one base pair position, A5-T8, which also correlates with crystallization speed. The structural differences between those structures with this base pair and those without, suggests that the A5-T8 pair provides a more uniform structural framework for crystallization. Based on these observations, we proposed that the A5-T8 base pair may function as a structural anchor that enables the hexameric duplex region to adopt an overall structure that is more amenable for crystallization. Interestingly, the overall magnitude of the differences between the A5-T8 and non-A5-T8 structures was relatively modest. This may be due in part to all the structures fitting into a common crystal framework, but it may also suggest that the significance of the A5-T8 pair may not be restricted to its contribution to the final structures. One interesting possibility is that this base pair identity is important during the crystal assembly process. Nucleobase identities at particular positions can have a significant influences on helix bending and breathing⁶⁹⁻⁷². Analysis using X3DNA⁷³ indicated there was little difference in the duplex helical bend angle in the final structures, however, it remains possible that the A-T base pairs might influence global helical parameters during the crystallization process that facilitates assembly. This may also explain the larger number of A5-T8 sequences that crystallized in the fast group of the initial screen.

Having identified sequences with isomorphous structures also provides new tools for understanding the crystal self-assembly process, particularly the nucleation

of these DNA oligonucleotides. Our mixed sequence crystallization results supported a model of nucleation through formation of Watson-Crick duplexes, followed by crystal growth through non-canonical pairs. Further, we established through NMR that the 13-mer oligonucleotide sequence does not form base pairs in solution without divalent cations, and that the addition of a divalent cation leads to the apparent rapid assembly into higher molecular weight complexes. Thus, these oligonucleotides may provide a unique system for further understanding how a dispersed biopolymer can self-assemble into macroscopic objects through a simple chemical trigger.

Chapter 2.8 Materials and Methods

1. Crystallization Screen

The 64 oligonucleotides representing all self-complementary sequences in the duplex region of the 13 nucleotide DNA oligomer were purchased from Integrated DNA Technologies (Coralville, IA), on the 100 nmol scale. Unpurified, desalted oligonucleotides were dissolved in water to 350 μM and screened for crystallization by sitting drop vapor diffusion. Crystallization drops were created using a Crystal Phoenix crystallization robot with 3-well Intelliplate crystallization trays (Art Robbins Instruments; Sunnyvale, CA). Each oligonucleotide was screened at three different ratios (1:1, 2:1, 3:1) of DNA to crystallization buffer (120 mM magnesium formate, 50 mM lithium chloride, 10% 2-methyl-2,4-pentanediol) to final drop volumes of 1 μl , 1.5 μl and 2 μl , respectively. The reservoir contained 100 μl of crystallization buffer. Trays were incubated at 22°C.

2. DNA Purification and Crystallization

The 20 molecules identified from the screen and the A7C variant were purchased from Integrated DNA Technologies (Coralville, IA) on the 1 μmol scale and were purified by 20% (19:1) polyacrylamide gel electrophoresis, electroeluted, and ethanol precipitated as previously described⁵⁹. The purified DNA samples were dialyzed against deionized water and the concentration was adjusted to 260 μM . The crystals were grown by sitting drop vapor diffusion in 24 well Cryschem M plate (Hampton Research, Aliso Viejo, CA). Prior to crystallization, DNA samples, in water, were heated at 95° for 2 minutes and cooled to room temperature. DNAs were

mixed (1:1) with crystallization buffer (120 mM magnesium formate, 50 mM lithium chloride, 10% 2-methyl-2,4-pentanediol) in a 4 μ l drop. The reservoir contained 400 μ l of crystallization buffer. The crystal plates were incubated at 22°C.

3. Data Collection and Refinement

Crystals were harvested by nylon loop, washed sequentially in crystallization buffer containing 30% and 40% 2-methyl-2,4-pentanediol, and flash-cooled in liquid nitrogen. Data were collected at Advanced Photon Source, Argonne National Labs, Sector 24-ID-C with an X-ray wavelength 0.979200 Å. Data were indexed and integrated using XDS⁷⁴, and scaled using Aimless⁷⁵. The isomorphous cell and space group for all datasets allowed direct refinement using the original 13-mer structure (PDBID: 1P1Y) as a starting model with individual nucleotide identities modified using Coot⁷⁶. Prior to refinement all atomic B-factors were reset and all sugar atoms were randomized to avoid biasing puckers. Refinement was performed using Phenix⁷⁷. Water molecules and ions were added manually during refinement. Following converged refinement in Phenix, all of the structures were run through the PDB-REDO⁷⁸ pipeline with single-group TLS (Translation/Liberation/Screw) and 10-fold cross-validation due to the small number of reflections in these datasets. Coordinates and structure factors were deposited in the Protein Data Bank⁷⁹ with accession codes provided in Table S2.

4. Sequence and Structure Comparison and Analysis

Depending on the time required for the appearance of crystals, the 21 crystallizing DNA oligonucleotides were classified into fast or slow crystallizing

groups, and their sequences were compared to calculate the frequency of nucleobases at each position of the duplex. For structural analysis, the variant structures were grouped depending on the presence or absence of the A5-T8 base pair. Structures were initially superimposed using the invariant residues (1-3, 10-12) to account for the slight differences in unit cell parameters. PyMol⁸⁰ was used to calculate RMSD values for backbone atoms (C1', C2', C3', O3', C4', O4', O5', P, O1P, and O2P).

5. Nuclear Magnetic Resonance

NMR spectra were recorded on a 600 MHz Bruker Advance III NMR spectrometer equipped with a CPTCI cryoprobe at a sample temperature of 298K. The purified B7 sample (200 μ M) was used in 500 μ L water containing 10% D₂O. 2 M MgCl₂ stock solution was added to NMR sample stepwise to reach target Mg²⁺ concentrations. At each titration point, a 1-D proton spectrum was acquired to observe signal chemical shift changes and/or signal intensity changes. The pulse program zgesgp was used to suppress water signal using gradient excitation sculpting⁸¹.

NMR transverse relaxation measurement were performed using the pulse program cpmgpr1d⁸². Each experiment consisted of a series of 1D measurement with at least 7 different decay times. Peak intensities plotted against decay time decreased exponentially and were fit to the equation: $M_{xy}(t) = M_{xy}(0)e^{-t/T_2}$ to determine values for T₂. A deoxyribose H2' signal at 1.7 ppm was chosen for peak intensity measurements because it was separated from other signals and had high intensity. MgCl₂ was added to the 0.2 mM B7 sample step wise, and at each addition point, a T₂ experiment was performed.

6. Heterogeneous Crystallization

For heterogeneous crystallization, two different DNAs were mixed and heated at 95° for 2', cooled to room temperature, and then mixed (1:1) with crystallization buffer in a 4 µl drop and incubated at 22°C. For polyacrylamide gel analysis, single heterogeneous crystals were washed in crystallization buffer three times and dissolved in water. Dissolved products were labeled with T4 polynucleotide kinase (New England Biolabs, Ipswich, MA) using 5 µCi/crystal ³²P γ-ATP (3000 Ci/mM; Perkin Elmer, Waltham, MA). The reaction was carried out at 37° for 30' and then terminated at 65°. 2 µl samples were run on a 20% acrylamide (19:1). The gels were exposed to the phosphor screen overnight and imaged using a Storm 860 phosphorimager (Molecular Dynamics, Sunnyvale, CA).

Chapter 3: Other Methods Used to Study the Cation Driven Assembly Process

I employed various direct and indirect methods to study the formation of higher order assemblies for the 13-mer DNA and to gather clues about self-assembly process. However, most of the used methods were only partially successful largely because of the high salt concentrations required to bring about the self-assembly process. We reasoned that the crystal pre-nucleation assemblies would be marked by the formation of stable intermediate species which could be distinguished from dispersed single-stranded DNA oligonucleotides based on their size, differential UV absorbance and specific dye binding capabilities. Dynamic light scattering is an effective method for fast and accurate size measurements of DNA nanostructures. As the presence of cations induces the 13-mer assembly, I wanted to monitor the formation of pre-nucleation complexes by measuring the hydrodynamic radii of the particles formed as a function of Mg^{2+} concentration. The challenge to this approach was to get an accurate size distribution of dispersed 13-mer DNA oligonucleotides in solution to compare with the particle sizes observed in presence of cations. Unfortunately, I was not able to get a consistent particle size distribution for the 13-mer in water probably owing to the low scattering from these short oligonucleotides or due to non-specific aggregation.

The differential absorbance of UV by single and double-stranded DNA has been traditionally used to monitor the assembly of DNA. The formation of double-stranded DNA duplex is associated with a hypochromic shift in the UV absorption due to the extensive base stacking interactions. I recorded the absorbance of the 13-

mer at 260 nm as a function of Mg^{2+} concentration. I observed a decrease in the absorbance with an increase in cation concentration but the results were unreliable due to an interplay of sample evaporation, aggregation and sample heating in the nanodrop during the course of observations. As a way to monitor the assembly process for the 13-mer crystals, I also explored the use of double-strand specific fluorescent dye, SYBR Green I. This method has been successfully used in quantitative PCR techniques to monitor the copy number of the amplified product. I established a baseline fluorescence using single-stranded 13-mer and then measured the fluorescence as a function of Mg^{2+} concentration. The biggest challenge to the use of SYBR Green I was the quenching of its fluorescence at the Mg^{2+} concentration required for promoting the DNA self-assembly.

Additionally, I probed the assembly process of the 13-mer DNA using indirect methods like S1 and DNase I endonuclease digestion. S1 nuclease is a single-strand specific endonuclease whereas DNase I does not show such specificity. I anticipated to see a protection against these endonucleases upon formation of higher order assemblies promoted by the addition of cations to the DNA sample. However, the narrow range of Mg^{2+} concentrations over which the endonucleases were active made it difficult for me to accurately and reproducibly draw clues about the self-assembly process. Finally I also tried to use 3' fluorescence labelled B7 sample in conjunction with an unlabeled sample to monitor the assembly process under a fluorescence microscope. This work has shown promising results and is being explored by other members in our lab.

Chapter 3.1 Dynamic Light Scattering

To study the particle sizes in the pre-nucleation assembly process, as a function of Mg^{2+} concentration, I recorded Dynamic Light Scattering (DLS) measurements on B7 in water and at 50 and 100 mM Mg^{2+} concentrations. The samples were preheated at 37°C for 2 minutes to get rid of non-specific hydrogen bond interactions. As a control, particle sizes were recorded in presence of 100 mM K^+ as it does not promote any crystal assembly. Similar readings were recorded on the original 13-mer (1P1Y), A7 sample which failed to crystallize in presence of Mg^{2+} (Table 2) and on EET11 sample (5' CAATCAGTCAG 3') which does not form any potential self-complementary dimers in presence of Mg^{2+} ions.

The B7 sample in water showed multiple peaks with average particle sizes of 2.3, 213 and 5416 nm. The observed sizes were larger than expected for a dispersed 13 nucleotide monomer. The readings for the sample upon addition of 50 mM Mg^{2+} and 100 mM Mg^{2+} were inconsistent in the number of observed peaks as well as the particle sizes. Similar inconsistencies were observed in the other three samples. As there were no clear differences in the sample and control particle sizes for the DLS experiment I concluded that it was not a useful technique to measure the self-assembly process of these 13-mer crystals. The inconsistent data could be from the low scattering from these short DNA molecules or due to the presence of non-specific aggregation.

Chapter 3.2 Ultra-Violet (UV) Spectroscopy

The most common real time measurement strategy to study the assembly process of complementary DNA strands include monitoring changes in UV absorbance based on the hypochromic effect of DNA. This phenomenon is attributed to the formation of extensive stacking interactions between the bases upon formation of the helical structure. To monitor the formation of higher order assemblies for the 13-mer DNA oligonucleotides, I measured the hypochromic shift in the absorption at 260 nm. I performed a time course experiment and recorded the absorption of B7 oligonucleotide as a function of Mg^{2+} concentration (Figure 17). B7 in storage conditions (absence of Mg^{2+}) was used as a negative control. The observation showed a significant drop in the absorption at 260 nm at the 3 hour time point for B7 samples in 100 mM and 200 mM Mg^{2+} with no further drop for the subsequent time points. This effect was more gradual for the B7 sample in 50 mM and 25 mM Mg^{2+} . The B7 samples in 5 mM and 10 mM Mg^{2+} showed a slight drop in the absorption immediately after mixing, but did not show any further change in the absorption in the course of the experiment. All the samples were measured again after 18 hours (overnight) with and without the application of heat and I observed an increase in the absorption for samples in Mg^{2+} concentration ranging from 25 mM – 200 mM upon application of heat. This result suggested that I could monitor association between the DNA molecules as a function of time and Mg^{2+} . This effect was most significant at higher Mg^{2+} concentrations (50 mM - 200 mM) and was reversed upon application of heat which would cause disruption of the DNA assemblies.

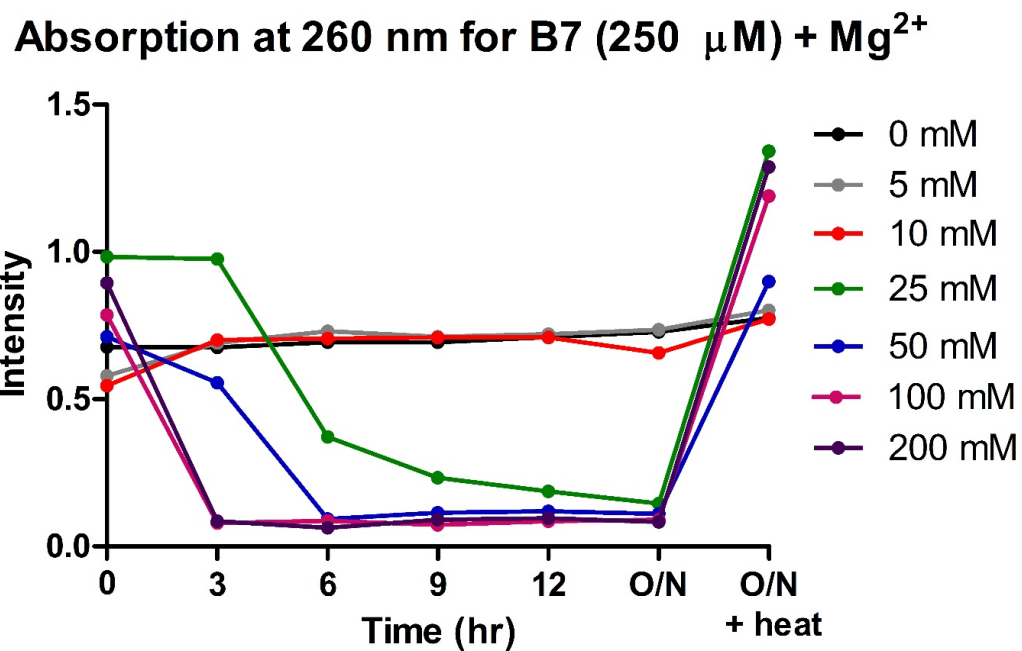


Figure 17. Monitoring the assembly of B7 DNA using UV Spectroscopy. 250 μ M B7 DNA sample was mixed with various concentrations of MgCl₂ (Colored lines) and the absorption intensity is measured at 260 nm as a function of time. After the overnight time point, the samples were heated at 95° for 2 minutes and then the absorption was measured again.

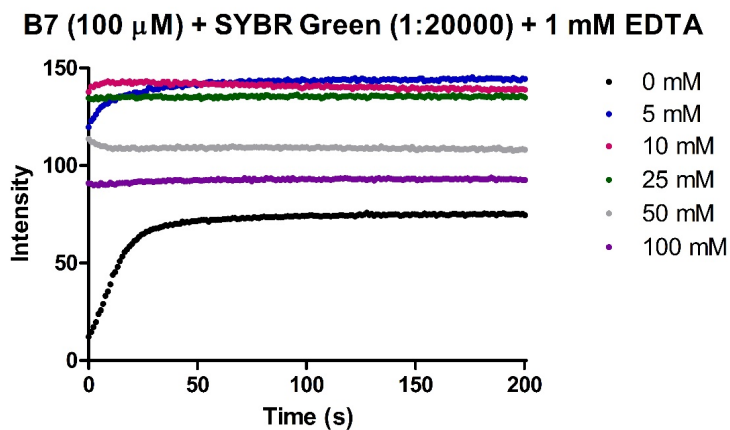
Chapter 3.3 Fluorescence Spectroscopy

One of the ways to monitor the assembly process of DNA oligonucleotides is to record the changes in emission signal from fluorescent dyes that are able to distinguish between dispersed single-stranded vs. double-stranded DNA molecules. I monitored the self-assembly process of B7 using SYBR Green I, a double-stranded DNA specific dye. It has an excitation maximum at 450 nm and an emission maximum at 520 nm. This dye is commonly used to monitor the copy number of the product in the quantitative polymerase chain reactions (qPCR). First, to determine the effect of SYBR Green I on the growth of B7 crystals, I performed post crystallization soaking along with co-crystallization and confirmed that SYBR Green I did not interfere in the self-assembly process for B7 and also did not have any visible effect on the crystal stability and geometry. I analyzed the assembly process by monitoring the fluorescence intensity as a function of Mg^{2+} concentration (Figure 18A). I used B7 oligonucleotide in absence of Mg^{2+} to establish a baseline and subsequently added increasing amounts of Mg^{2+} anticipating an increase in the fluorescence intensity with the formation of higher order assemblies. I used K^+ ions (results not shown) and SYBR Gold with B7 oligonucleotides as negative controls (Figure 18B).

I observed an increase in the fluorescence intensity upon increasing the Mg^{2+} concentration from 1 mM, to 2.5 mM and then to 5 mM. However, a drop in the fluorescence intensity was seen as the Mg^{2+} concentration was further increased to 10 mM, then to 100 mM and then again to 200 mM. Furthermore, through experiments conducted to measure the effect of Mg^{2+} on the SYBR green I fluorescence, it was found that the SYBR green I was active only over a small range of magnesium

concentration (optimum fluorescence at 1-5 mM Mg^{2+}), and higher Mg^{2+} concentrations were inversely related to the fluorescent intensity⁸³ (http://www.biofire.com/pdfs/LightCycler/LC_Exp_Design.pdf). To confirm this I used a self-complementary 13 nucleotide DNA as a control and performed fluorescence reading as a function of Mg^{2+} and observed the expected quenching of fluorescence intensity at 100-200 mM Mg^{2+} concentrations (data not shown here). So, I concluded that SYBR green I fluorescence spectroscopy was not a reliable method for monitoring the 13-mer DNA assembly process.

A.



B.

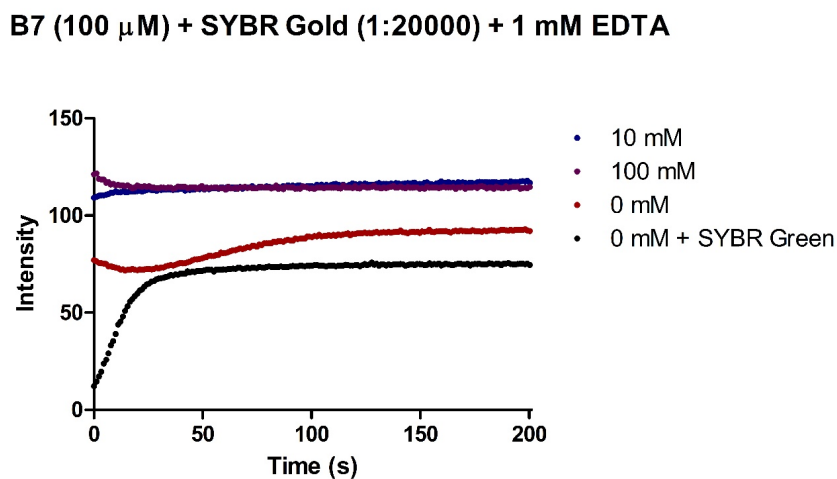


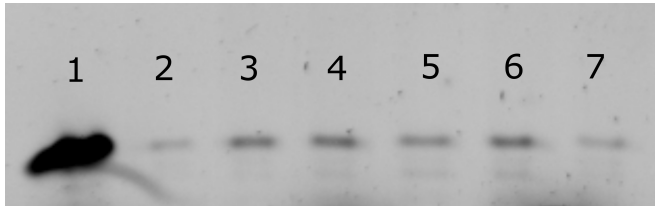
Figure 18. Fluorescence intensity measurements on B7 DNA using nucleic acid specific fluorescent dyes. Fluorescent intensity measurements were carried out as a function of time (reading every 3s up to 200s) on B7 DNA sample in presence of SYBR Green I, a double-strand specific DNA dye (**A**) or SYBR Gold (**B**) which binds to both single and double-stranded DNA. The time course experiment was carried at various concentrations of MgCl₂ (Colored lines).

Chapter 3.4 S1 Nuclease Digestion

S1 Nuclease is a single-strand specific endonuclease that hydrolyses the single-stranded regions in DNA duplexes or assemblies⁸⁴. As increasing the Mg^{2+} concentrations leads the free B7 oligonucleotides to assemble into duplexes and higher order assemblies, we believed that the process could be marked by an increased protection from S1 digestion which could then be monitored through gel electrophoresis. The heterogenous crystal experiments, described in Chapter 2.6, strongly suggest that the pre-nucleation event for the 13-mer crystals is driven by the formation of stable Watson-Crick duplexes, which would mean the S1 treatment would lead to the hydrolysis of the single-stranded non-canonical overhangs from one or both ends of the 13-mer. We anticipated that this would limit the assembly formation and propagation. For the S1 treatment, I incubated the oligonucleotide sample with Mg^{2+} for 4 hours and the samples were then digested with S1 nuclease. The samples were analyzed on a denaturing urea gel (Figure 19A). I performed similar experiment using EET (5' CAATCAGTCAG 3') as a control DNA oligonucleotide as it was unlikely to form any duplexes due to lack of any self-complementary sequence (Figure 19B).

The results showed that the B7 oligonucleotide was protected from digestion by S1 nuclease as the concentrations of Mg^{2+} increased. However, the band intensities were inconsistent at higher Mg^{2+} concentrations. The EET control showed complete S1 digestion at 0 mM and 100 mM, but small amount of protection was observed in 10 mM Mg^{2+} concentration. Due to these inconsistencies I did not use S1 nuclease for studying the assembly process of B7 oligonucleotides.

A.



1. B7 + no enzyme + No Mg^{2+}
2. B7 + No Mg^{2+}
3. B7 + 5 mM Mg^{2+}
4. B7 + 10 mM Mg^{2+}
5. B7 + 25 mM Mg^{2+}
6. B7 + 50 mM Mg^{2+}
7. B7 + 100 mM Mg^{2+}

B.



1. EET11 + no enzyme + No Mg^{2+}
2. EET11 + No Mg^{2+}
3. EET11 + 10 mM Mg^{2+}
4. EET11 + 100 mM Mg^{2+}

Figure 19. Denaturing PAGE gel for samples digested with S1 nuclease. The B7 sample at 300 μ M (A.) or EET sample at 413 μ M (B.) were incubated at various concentrations of $MgCl_2$ for four hours and digested with S1 at nuclease at room temperature for 10 minutes and visualized on a denaturing PAGE gel stained with SYBR Gold.

Chapter 3.5 DNase I Digestion

DNase I is an endonuclease that cleaves both single and double-stranded DNA oligonucleotides. I used DNase I to monitor the self-assembly process of B7 oligonucleotides, as a function of Mg^{2+} concentration. The formation of higher order assemblies for the DNA 13-mer in presence of Mg^{2+} , should theoretically make it inaccessible for digestion by DNase I enzyme which could then be resolved on a PAGE gel. We incubated the B7 oligonucleotide sample with water or various concentrations of Mg^{2+} for 4 hours and then the samples were digested with the nuclease. The samples were analyzed on a denaturing urea gel (Figure 20A). I performed similar experiment using T7 primer (5'-TAATACGACTCACTATA-3') as a control DNA oligonucleotide as it did not have any self-complementary sequence (Figure 20B, C).

The results showed that B7 oligonucleotide was protected from digestion by DNase I as the concentrations of Mg^{2+} increased, suggesting the formation of higher order assemblies. However, these results were not reproducible. Also, the control T7 DNA showed no digestion under identical reaction conditions and so we concluded that the DNase I experiments were not optimum for characterization of DNA assemblies. Reports show that DNase I is active over a relatively narrow range of Mg^{2+} concentrations and so the discrepancies in the results could be a result of partial or complete deactivation of the enzyme at high Mg^{2+} concentrations used in our reaction conditions⁸⁵.

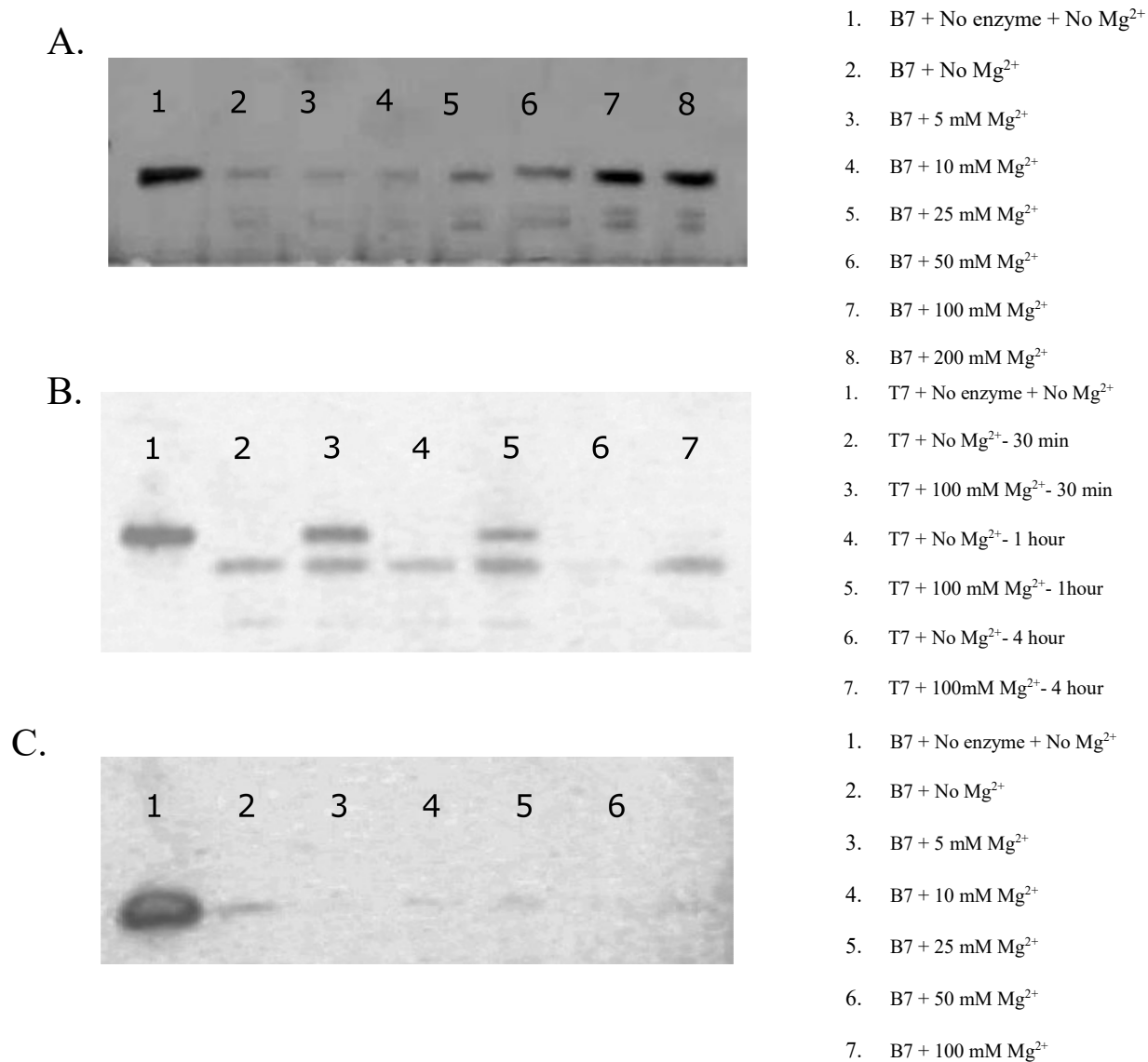


Figure 20. Denaturing PAGE gel for samples digested with DNase I nuclease. A. The B7 sample at 300 μ M was incubated at various concentrations of MgCl₂ for 4 hours and digested with DNase I nuclease for 10 minutes and visualized on a denaturing PAGE gel. **B.** The T7 sample was incubated with either 0 or 100 mM MgCl₂ for 1.5 hours and then digested at room temperature for up to 4 hours. Samples corresponding to 30 mins, 1 hour and 4 hour digestion time points are loaded onto the gel. **C.** The B7 sample at 300 μ M was incubated at various concentrations of MgCl₂ for 2 hours and digested with DNase I nuclease for 4 hours at room temperature and visualized on a denaturing PAGE gel. The gels were stained with SYBR Gold.

Chapter 3.6 Confocal Microscopy

I used a 3' fluorescein ($\lambda_{\text{exct}} = 494 \text{ nm}$ and $\lambda_{\text{emiss}} = 521 \text{ nm}$) labelled B7 to monitor the assembly process of the 13-mer crystals. It has been previously determined in our lab that a mixture of fluorescein labelled and an unlabeled B7 (1:9) can be used to grow a shell on a pre-grown unlabeled B7 crystal (core). We reasoned that we can use this mixture on the S1 nuclease treated or the DNase I treated crystals and gather clues about the assembly process. We also wanted to demonstrate the growth of fluorescently labelled shell around unlabeled microcrystals. We believed that this would give us a visual aid to observe the assembly process and help us understand the time scale at which the assembly took place.

I observed that both the S1 and the DNase I treated crystals were able to form a fluorescent shell within 2 hours when incubated in 120 mM Mg^{2+} (Figure 21A, B). No visual difference was observed in the rate of growth of the shell or its thickness between the untreated and the enzyme treated crystals (Figure 21C). In the seeding experiment, I observed appearance of the fluorescent shell around the non-fluorescent core within 2 hours (Figure 22A, C) and the appearance of fluorescent crystals in absence of the seed was seen after 5 hours of incubation (Figure 22B). However taking into account the previous S1/DNase I digestion experiments, these observations are not sufficient to unambiguously prove that they are a true representation of the assembly process and not the inhibition of the enzyme activity in high Mg^{2+} concentrations and so these experiments need to be explored further for studying the assembly process of 13-mer DNA crystals.

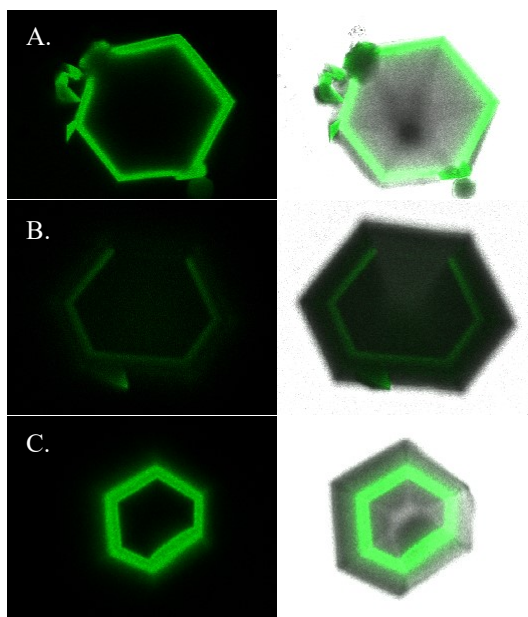


Figure 21. Confocal images for unlabeled B7 core and fluorescent shells. The unlabeled B7 crystals were grown to a standard size and then treated with either S1 nuclease (**A**) or DNase I (**B**) or no enzyme (**C**) for 1 hour at room temperature and then transferred to a mixture of unlabeled B7: fluorescein labelled B7 (9:1). After washing in buffer solution the crystals were imaged under a confocal fluorescence microscope. The images in panel **A**, **B** and **C** are the fluorescent image and the overlay of the fluorescent image with the bright field image.

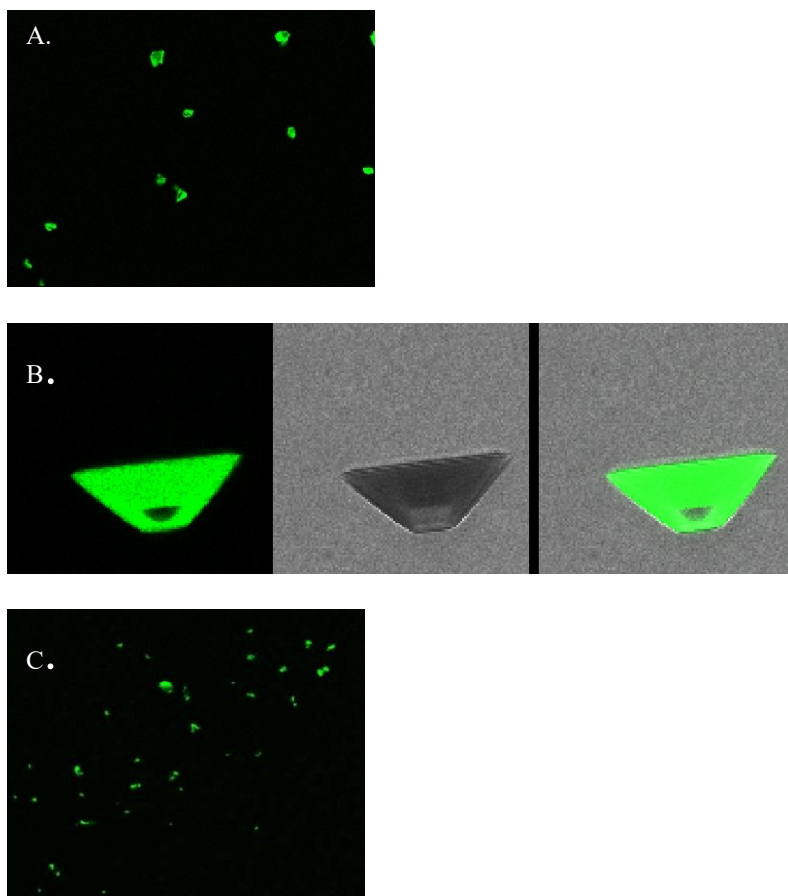


Figure 22. Confocal images to observe the B7 crystal assembly. A mixture of unlabeled B7: fluorescein labelled B7 (9:1) was used to observe the assembly process under confocal microscope. The assembly process was monitored in presence (**A**, **B**) or absence (**C**) of unlabeled B7 microcrystals as seeds for nucleation. Images were taken at various time points. In the presence of the microcrystal seeds the fluorescent shells started appearing as early as 2 hours (**A**) and were completely developed in the overnight sample (**B**) where as in the absence of the micro crystal seeds, the fluorescent assembly was visible at a 5 hour time point (**C**). The three panels in (**B**) indicate the fluorescence image, a bright-field image and an overlay of the two images going left to right.

Chapter 3.7 Materials and Methods

1. Dynamic Light Scattering

Unpurified B7, DET66, A7 and EET11 samples were dialyzed against deionized water over night and filtered through a 0.2 μm filter. 70 μl of samples were heated at 37° for 2 minutes and mixed with 7 μl of either deionized water, 500 mM or 1 M Magnesium formate or 1 M Potassium chloride to get a final concentration of 50 mM or 100 mM Mg^{2+} or K^{+} ions. The samples were then read in Zetasizer Ver. 6.20 (Malvern Instruments Ltd) in Dr. Silvia Muro's lab. The particle diameters were recorded three times for each sample maintained at 25°, with a material refractive index of 1.45 and 3s recording time.

2. UV Spectroscopy

Reactions were set up using 5 μl of 500 μM B7 DNA oligonucleotide with 5 μl of either water or MgCl_2 (10, 20, 50, 100, 200 and 400 mM). Samples were mixed and 1 μl of sample was used to measure the absorption at 260 nm using a NanoPhotometer (Denville Scientific Inc.) at 0, 3, 6, 9, 12 and 18 hour time points. Readings were also recorded by heating the sample at 95° for 1 minute after the 18 hour time point.

3. Fluorescence Spectroscopy

To monitor self-assembly process of 13-mer oligonucleotides using fluorescent spectroscopy, 90 μl of 100 μM B7 oligonucleotide sample in 1 mM EDTA was mixed with either 10 μl of water or 10X Magnesium formate solution (10, 25, 50, 100, 1000 mM) to get final concentrations of 0, 1, 2.5, 5, 10, 100 and 200 mM Mg^{2+}

ions respectively. The stock solution of SYBR Green I, a double-stranded DNA specific dye, was diluted 20X using DMSO and 0.1 μl of the diluted solution was added to the DNA- Mg^{2+} mixture to get a 20000X dilution. Upon mixing the fluorescence intensity ($\lambda_{\text{excit}} = 450 \text{ nm}$, $\lambda_{\text{emis}} = 520 \text{ nm}$) was recorded every 3s for a total of 200s. Similar readings were obtained for a non-specific DNA dye, SYBR Gold (20000X dilution), and using 10 μl of 100 or 1000 mM K^+ , instead of Mg^{2+} , with SYBR Green I (20000X dilution). To determine the effect of Mg^{2+} on the fluorescence of SYBR Green I, fluorescent intensity studies were performed on a 100 μM 13 nucleotide self-complementary duplex by mixing it with 10 μl of 100 or 1000 mM Mg^{2+} and SYBR Green (20000X dilution).

4. S1 and DNase I Digestion

For S1 and DNase I digestion experiments, 1.5 μl of 300 μM B7 oligonucleotide sample was incubated with 1.5 μl of either water or Mg^{2+} (10, 20, 50, 100, 200 mM) for 4 hours. The reaction mix was then digested with either S1 nuclease or DNase I (Thermo Scientific) in their respective reaction buffers (S1 1X reaction buffer: 40 mM Sodium acetate, pH 4.5, 1.5 M Sodium chloride and 10 mM Zinc sulphate) (DNase I 1X Reaction Buffer: 10 mM Tris-HCl, 2.5 mM Magnesium chloride, 0.5 mM Calcium chloride, pH 7.6) for 10 minutes at room temperature. The reaction was quenched using 5 μl of stop buffer (200 mM EDTA and 0.2% SDS) at 65° for 10 minutes for S1 and at room temperature for DNase I. 12 μl of the reaction mix was analyzed on a 20% polyacrylamide urea gel and stained using SYBR Gold (1:10000X dilution) in TBE buffer.

5. Confocal Microscopy

B7 crystals were grown in crystallization buffer (120 mM Magnesium formate, 50 mM Lithium chloride and 10% MPD) and were washed 2X in fresh buffer and transferred to a reaction mixture containing either 10 μ l buffer + no enzyme or 8 μ l buffer + 2 μ l S1 reaction buffer + 0.2 μ l S1 nuclease or 8 μ l buffer + 1 μ l DNase I reaction buffer + 1 μ l DNase I enzyme. The samples were incubated at room temperature for 1 hour and transferred to a fresh mixture of 2 μ l buffer + 2 μ l B7: fluorescein labelled B7 mixture (9:1) following a 2X washing step in buffer. The crystals were incubated for 2 hours and then imaged using the Leica SP5X confocal microscope in the imaging core facility of the cell biology and genetics department of University of Maryland.

Chapter 4: Sequence-Dependent Structural Changes in 14-mer DNA Oligonucleotide

Chapter 4.1 Introduction

DNA oligonucleotides are both conformationally and structurally diverse³³. Depending on the environmental conditions, B-form DNA can undergo conformational transitions to the A- and the Z-forms³³⁻³⁶. Additionally, a variety of non-B-form DNA motifs have been characterized in vivo including DNA cruciform, hairpin structures, triplexes and quadruplexes³⁷⁻⁴³. One of the major areas of DNA structural biology over the course of several decades was in understanding how non-Watson-Crick base pairs, or mismatches, could be accommodated in otherwise normal DNA helices, or are responsible for forming alternate DNA structures^{52,53,86}.

G-A base pairs have been one of the most well-characterized non-Watson-Crick base pairings that can be readily integrated into the B-form duplex^{66,87-94}. Structural studies revealed that these G-A base pairs can adopt up to four different base pairing combinations depending on the local sequence and environment⁹⁵. The two most prevalent types represented in the Nucleic Acid Databank⁹⁶ include the Type I pair involving the Watson-Crick edges of the bases, and the Type IV sheared G-A pair involving the guanosine sugar edge and the adenosine Hoogsteen edge^{65,94}. However, the type and stability of G-A base pair formed is highly dependent on the local sequence⁹⁷. The Type I pairing is favored for d(AGAT)₂ sequence due to an additional interstrand hydrogen bond between the N2 amino group of the paired G and O2 of the thymidine in the flanking A-T pair⁶⁶. The sheared G-A base pairs are favored in d(YGAR)₂ sequences. In nearly all cases the sheared G-A pairs are found

in tandem (GA/AG) and are thermodynamically quite stable within a canonical duplex due to the interstrand stacking between the sheared base pairs and the extensive intrastrand stacking between the sheared pairs and the flanking base pairs⁹⁵.

The 14-mer DNA oligonucleotide structures described in this section were obtained by adding an additional adenosine at the 3' of the DNA 13-mers described in Table 2. The 14th nucleotide was added for the easy resolution of the heterogenous crystals on a denaturing gel. Out of all the 14-mer DNA's screened for crystallization, the addition of the 14th nucleotide lead to a significantly different crystal habit under identical Mg²⁺ conditions for four DNA oligonucleotides (Figure 23B-E). Remarkably, the added A14 residue made tertiary contacts to the guanosine adjacent to a single sheared A-G pair from the neighbouring duplex, resulting in a conformation similar to tandem sheared G-A pairs. Together with a series of purine base triples, these interactions were responsible for the formation of the alternate crystal form. Here, I also examined the role sequence played in the adoption of the alternate crystal form. Our analysis suggested that the sequence in the 14-mer duplex region and the identity of the added nucleotide were necessary to promote this alternate structure.

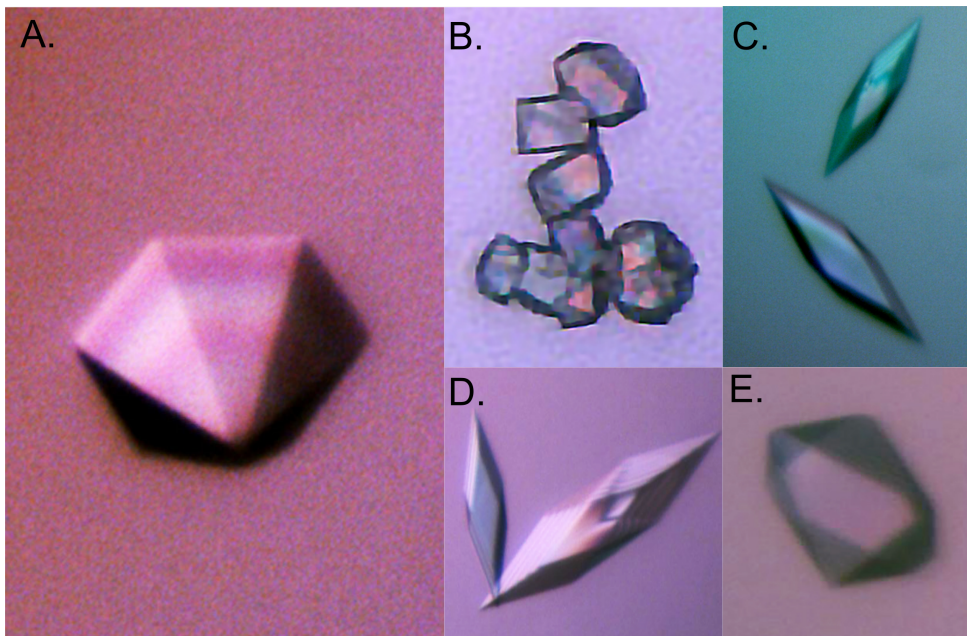


Figure 23. 13-mer and 14-mer crystals. A. The 13-mer DNAs crystallize with a hexagonal unipyramidal crystal habit. Under identical crystallization conditions the four 14-mer DNAs, A1-14, A2-14, A3-14 and A4-14 crystallize with habits shown in B-E, respectively. All of the 14-mers crystallized in the same space group with almost identical unit cell dimensions.

Chapter 4.2 Overview of the 14-mer Structures.

We determined the X-ray crystal structures of four 14-mer DNA oligonucleotides differing by one base pair in the self-complementary duplex region (Figure 24A). The structures were highly isomorphous to each other with an average RMSD of 0.40 Å for all identical aligned atoms, and 0.58 Å for the backbone atoms. All of DNAs crystallized with one molecule in the asymmetric unit, with crystal symmetry generating interstrand hydrogen bonding and base stacking interactions. Each strand in the crystal formed hydrogen bonds with 5 other strands to form two distinct regions of nucleobase interactions (Figure 24A, B). The B-form duplex region is formed from residues A3 through G10 of two strand, and the triplex junction is formed from G1 and G2 of one strand, G11-G13 of two different strands, and A14 of another strand. End-to-end stacking of the triplex junction regions lead to columns of pseudo-infinite coaxially stacked helices that interact only through the A14 residues (Figure 25). For convenience I will restrict the structural description to the A3-14 structure, noting that the only substantial differences in the other structures were the central base pairs of the duplex region (Table 7 & Figure 26).

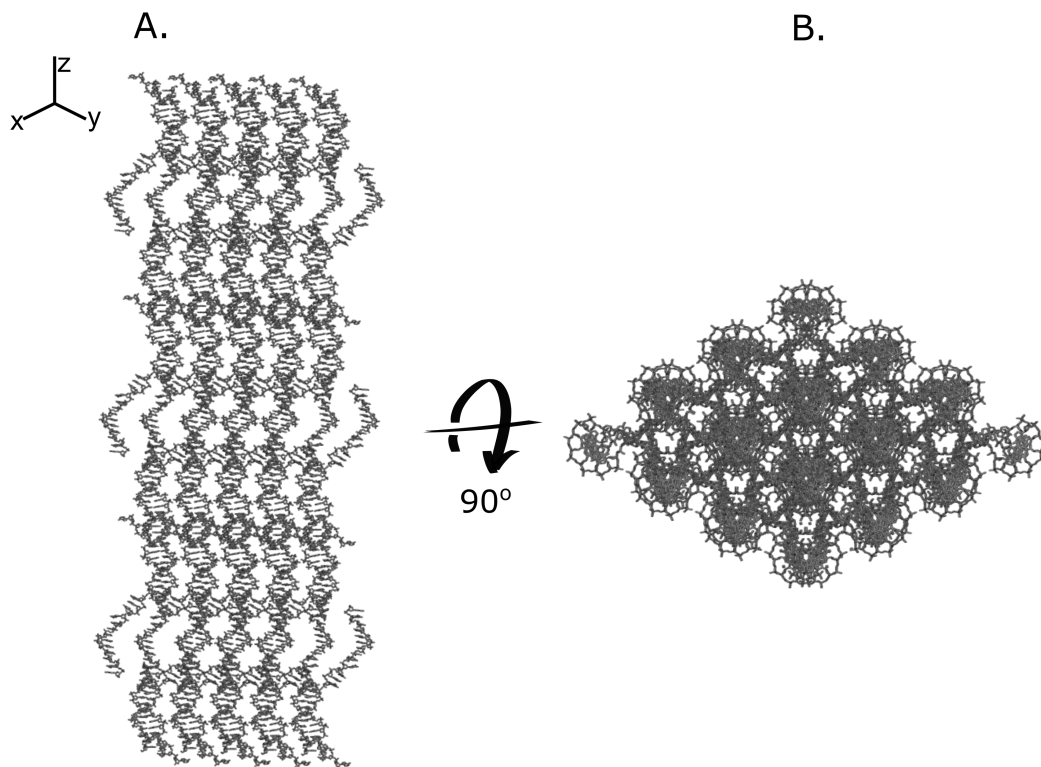


Figure 25. Crystal packing. **A.** The overall 3D arrangement of the 14-mer in the crystal lattice looking perpendicular to the three-fold symmetry axis, and **B.** looking down the three-fold symmetry axis, . Duplex regions are coaxially stacked, with adjacent duplexes interacting only through tertiary contacts via the added A14 nucleotide.

Table 6. Data collection and refinement for 14-mer structures.

	A1-14	A2-14	A3-14	A4-14	A3-14-Br
Data collection					
Wavelength (Å)	0.97919	0.97919	0.97919	0.97919	0.91940
Detector	ADSC Quantum 315	ADSC Quantum 315	ADSC Quantum 315	ADSC Quantum 315	Pilatus 6M
Space group	P3 ₁ 21	P3 ₁ 21	P3 ₁ 21	P3 ₁ 21	P3 ₁ 21
Number of crystals	4	4	3	2	1
Cell dimensions					
Avg. a, b, c (Å)	26.01 26.01 122.02	25.99 25.99 121.53	25.83 25.83 123.08	25.96 25.96 121.53	26.27 26.27 123.30
α, β, γ (°)	90, 90, 120	90, 90, 120	90, 90, 120	90, 90, 120	90, 90, 120
Resolution	2.03-22.52 (2.03- 2.09)	2.10-22.51 (2.10- 2.17)	2.15-41.02 (2.15- 2.23)	2.40-22.48 (2.40- 2.53)	1.99-122.34 (1.99- 2.09)
I/σI	13.9 (1.0)	16.3 (3.0)	14.7 (2.5)	8.9 (2.6)	14.8 (1.0)
CC _{1/2}	0.997(0.883)	0.987(0.978)	0.978(0.959)	0.993(0.954)	0.999(0.866)
R _{pim}	0.036 (0.45)	0.047 (0.12)	0.061 (0.15)	0.055 (0.13)	0.023 (0.38)
Number of reflections	3436 (268)	3035 (291)	2965 (281)	2131 (299)	3818 (507)
Completeness (%)	99.4 (97.3)	97.8 (97.4)	99.9 (99.9)	99.1 (98.7)	99.1(96.0)
Multiplicity	10.3 (6.9)	10.6 (4.6)	6.8 (4.5)	3.8 (3.5)	5.0 (3.8)
Anomalous Completeness (%)	N/A	N/A	N/A	N/A	95.2 (77.9)
Anomalous multiplicity	N/A	N/A	N/A	N/A	2.7 (1.8)
Refinement					
Resolution (Å)	2.03-22.52 (2.03- 2.07)	2.10-22.51 (2.10- 2.15)	2.15-41.02 (2.15- 2.20)	2.40-22.48 (2.40- 2.46)	
Number of reflections	3092 (206)	2716 (202)	2626 (178)	1898 (117)	
Average R _{free} ^a	0.2937	0.313	0.265	0.312	
R factor ^b	0.233 (0.502)	0.270 (0.454)	0.244 (0.421)	0.245 (0.523)	
R _{free} ^b	0.293 (0.538)	0.311 (0.416)	0.262 (0.606)	0.312 (0.641)	
Number of atoms					
DNA	293	293	293	293	
Ion	2	2	2	2	
Water	13	13	13	6	
Bond lengths (Å)	0.007	0.006	0.008	0.005	
Bond angles (°)	1.567	1.406	1.887	1.220	
PDB ID	5BZ7	5BZ9	5BXW	5BZY	

*Values in parentheses are for the highest-resolution shell.

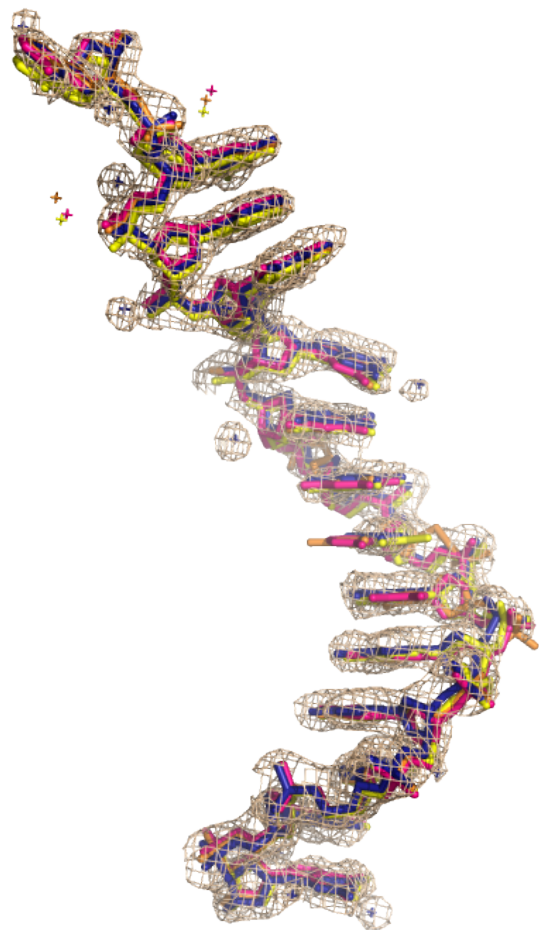


Figure 26. Structural overlap for the 14-mer structures. The three 14-mer structures, A1-14 (yellow), A2-14 (orange) and A4-14 (pink), are superposed with A3-14 structure (blue). The RMSD for the overlap is 0.40 Å. The weighted electron density map ($2F_o - F_c$) of A3-14 contoured at 1.0σ is shown in grey.

Chapter 4.3 B-Form Duplex Region Capped by Sheared A-G Pairs.

The B-form duplex region was formed through base pairing interactions between A3 and G10 of two DNA strands. The central six base pairs of the helix were composed of self-complementary base pairs A4-T9, A5-T8 and G6-C7. These six nucleotides were structurally isomorphous to the duplex region in the parent A3-13 structure with a RMSD of 0.69 Å (Figure 27). The sugar-phosphate backbone of residue A4 showed the greatest variability between the two structures (RMSD 3.40 Å for backbone atoms) and was the result of a significantly different conformations 5' of the A4 nucleotide. In all of the 13-mer structures we determined, residues G1-A3 were flipped out of the helical axis toward the major groove of the duplex where they were positioned to make non-canonical interactions with G10-A12 of another strand. In the 14-mer structures, A3 remained stacked with A4 and was base paired with G10 in a Type IV sheared base pair. This resulted in a duplex region containing six self-complementary base pairs flanked on either end by A-G pairs.

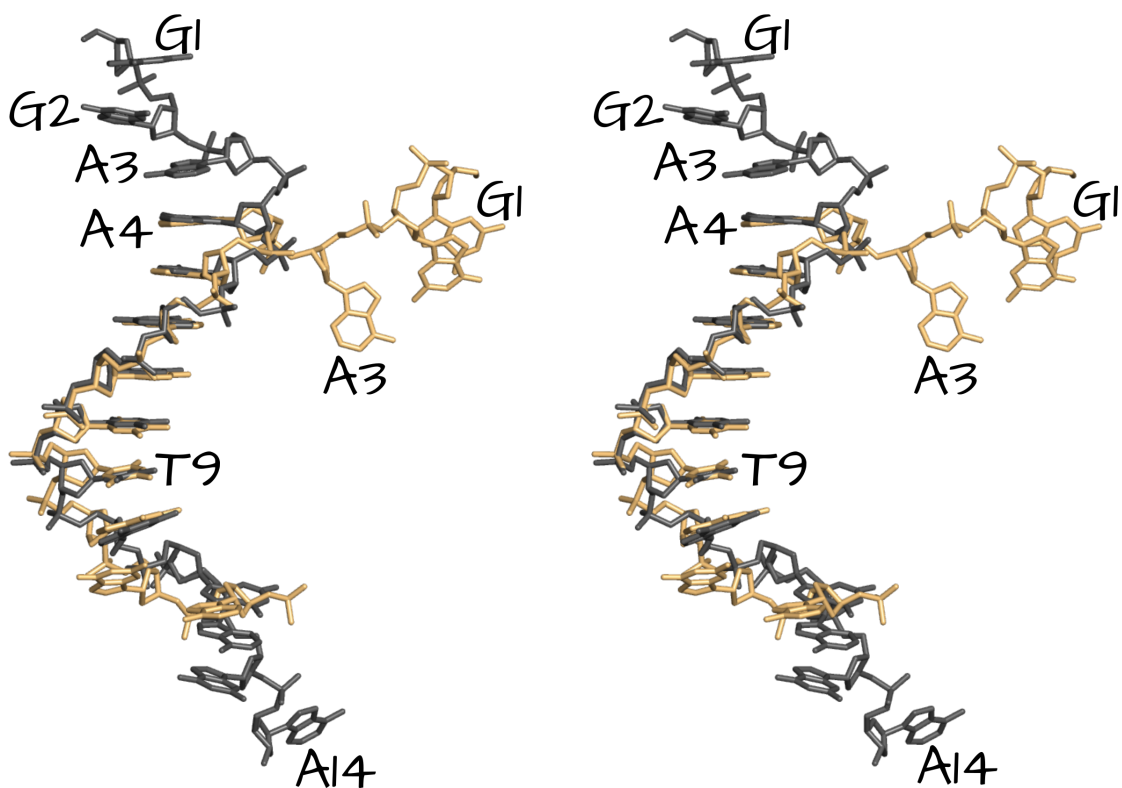


Figure 27. Structural comparison of A3-13 and A3-14 monomers. Stereo view of the X-ray crystal structures of A3-13 (orange) and A3-14 (grey) superposed using residues A4-T9 of the duplex region. The Watson and Crick region (A4-T9) was isomorphous in both structures whereas the 5' non-canonical region (G1-A3) and the 3' non-canonical region (G10-A14) shows a complete rearrangement in the 14-mer structure as compared to the parent 13-mer.

Chapter 4.4 Tertiary Interactions Fulfil a Structural Role to Generate Tandem G-A Base Pairs.

The secondary and tertiary structural environment surrounding the sheared A3-G10 base pair established a local structure that was highly similar to the tandem sheared GA/AG steps that have been previously observed in B-form helices⁹⁷⁻¹⁰¹, with the sheared A3-G10 base pair being structurally equivalent to the second base pair (Figure 28A, B). This base pair was formed through the Hoogsteen edge of A3 (N6, N7) and the sugar edge of G10 (N2, N3) and displayed the characteristic base pair buckling (Figure 28A). The non-planarity of the base pair lead to the formation of a potential interstrand hydrogen bond between N6 of A3 and O2 of T9, though the geometry is not ideal (Figure 29). Like previous solution structures, inter- and intrastrand stacking interactions played an important part in stabilizing the A3-G10 pairing. Despite the relatively large twist angle (60.3°) at the A3A4/T9G10 step, there was significant intrastrand stacking between A3 and A4 (4.85 Å² overlap based on polygon projections using X3DNA⁷³ (Figure 30B). Intrastrand stacking interactions were even more pronounced for the partner strand, with T9-G10 stacking having 8.10 Å² overlap. The overall 12.95 Å² overlap at this base pair step was the single most in the entire structure, suggesting that the capping A-G pairs provide significant stability to the duplex ends. Interstrand stacking interactions, one of the hallmarks of structures having tandem sheared G-A base pairs, were also present (Figure 30A). G2 of one strand stacks with G10 of the partner strand. Overall, the structural environment around the A3-G10 pair was remarkably similar to previous solution structures (Figure 28A), including the presence of several phosphate linkages in the

B_{II} conformer (G2, A3, G10) which is a hallmark of tandem sheared G-A structures⁹⁸. The major difference was the lack of the first G-A pair. Interestingly, tertiary contacts between A14 from a different column of coaxially stacked helices and G2 maintained base stacking interactions and lead to a similar overall structure.

The spacing between the G10 and G11 nucleobases allowed A14 to stack between the A3 and G11 nucleotides from the partner strands, while forming a base pair with G2 (Figure 28A, C). Unlike the tandem sheared G-A structures, this tertiary contact occurred between the Watson-Crick face of the A14 and the sugar edge of G2, with N6 of A14 in almost the identical position to the first G-A in the solution structures (Figure 28C). Along with these base pairing and base stacking interactions, A14 made additional 3' OH contacts with the A4 phosphate and was also involved in the base capping interactions with the A3 sugar. This is an example of a tertiary structure interaction providing a structural equivalence to a previously observed secondary structure motif.

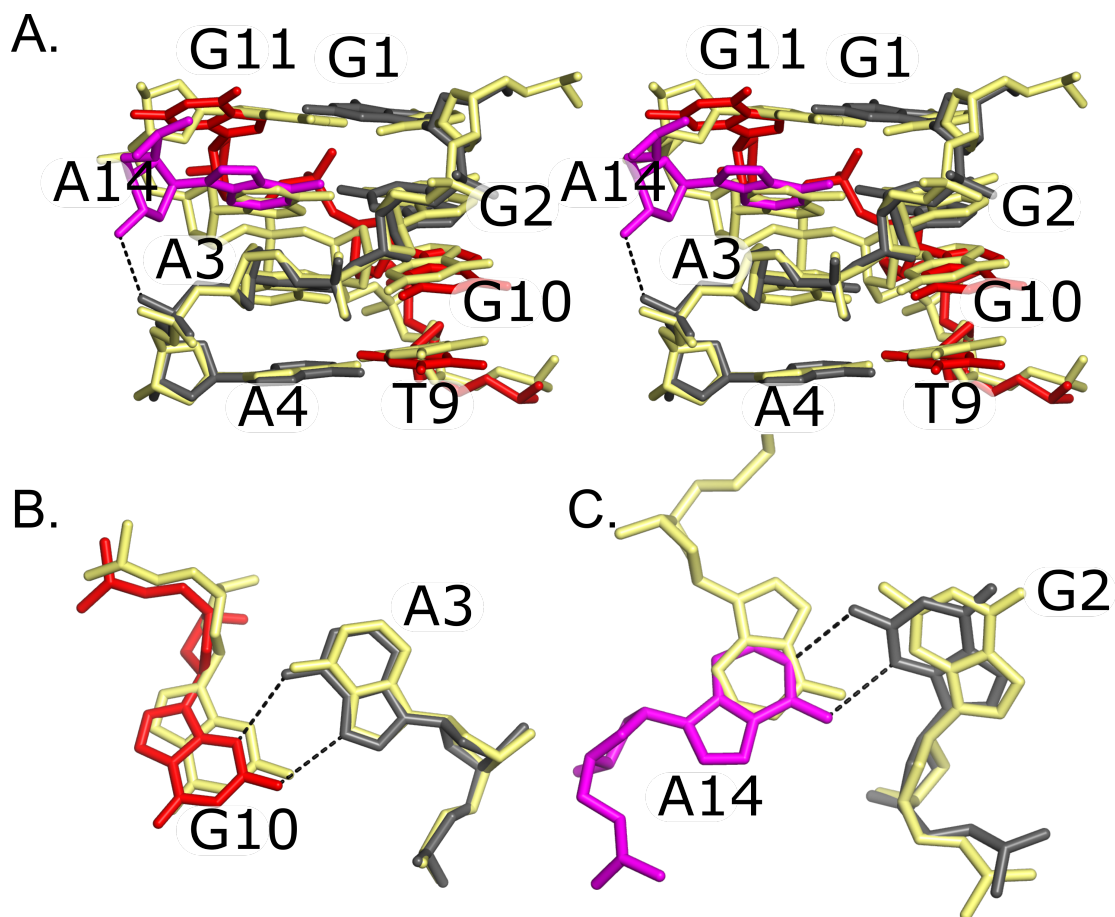


Figure 28. Comparison of GA/AG motifs. **A.** Stereo view of the tandem sheared GA/AG base pairs with the flanked Watson-Crick base pairs from a solution structure (yellow; PDB ID: 175D) superposed on G1-A4 from one strand (grey) and T9-G11 of the partner strand (red). A14 from a neighboring duplex is shown in magenta. A hydrogen bond between the A14 3'-OH and the A4 phosphate group is shown as a dashed line. **B.** The A3-G10 base pair from the 14-mer structure is highly similar to the second base pair in the tandem sheared GA solution structure. Hydrogen bonds are shown as dashed lines. **C.** The tertiary contact between A14-G2 mimics the first sheared GA base pair. The hydrogen bonds between the Watson-Crick face of A14 and the Hoogsteen face of G2 are shown as dotted lines.

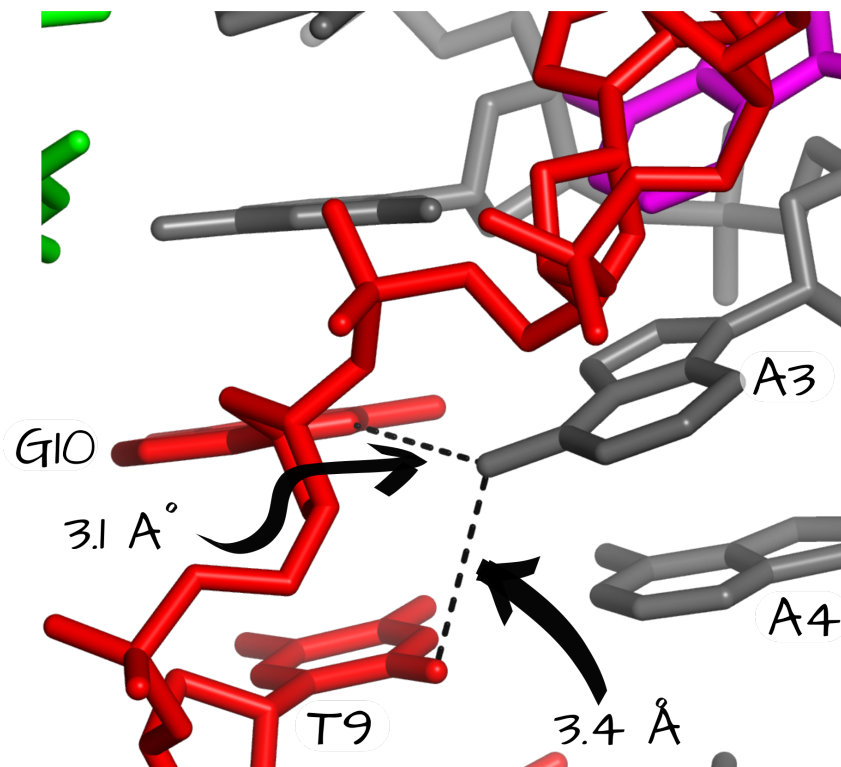


Figure 29. Potential hydrogen bond between A3 and T9. The sheared A3-G10 base pair leads to the formation of an interstrand hydrogen bond between N6 of A3 and O2 of T9, shown in dotted lines.

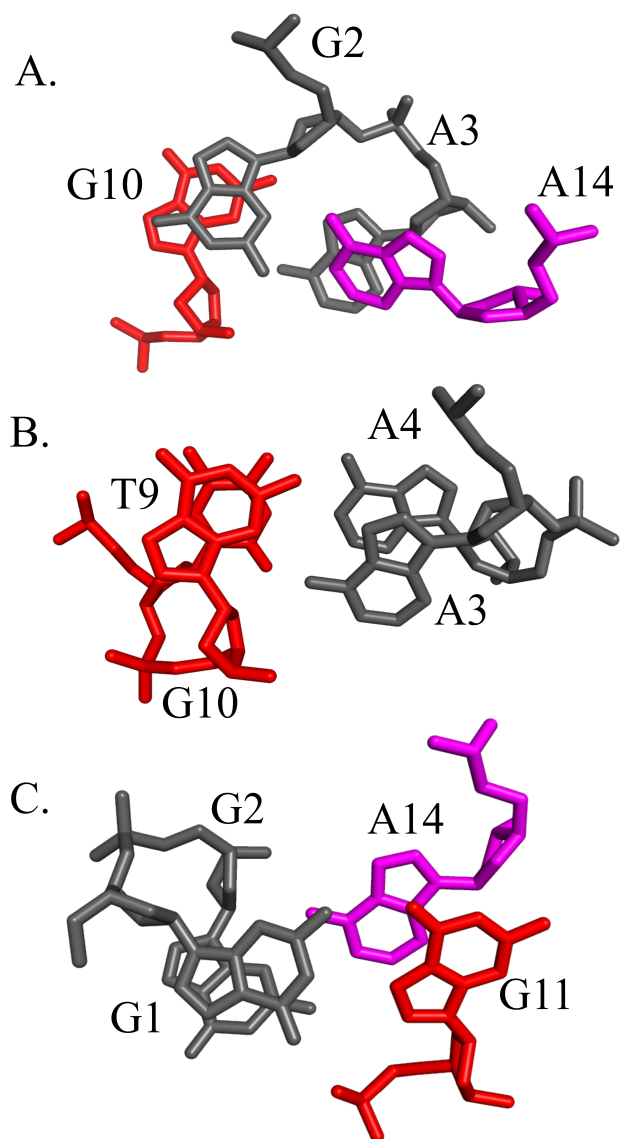


Figure 30. The inter- and intrastrand stacking at the sheared G-A base pair. A. The interstrand stacking between the G2 nucleotide of one strand (grey) and the G10 nucleotide of the partner strand (red) in a duplex and the A3 nucleotide of one duplex (grey) and the A14 nucleotide from a different duplex (magenta) involved in a tertiary mimic of tandem sheared G-A base pairs. **B, C.** Intrastrand stacking between the sheared A3-G10 base pair and the flanking T9-A4 base pair and G2-A14 base pair

and the flanking G1-G11 base pair respectively. The overlap values are a sum of both the base pairs, based on polygon projections obtained from X3DNA.

Chapter 4.5 Triplex Junction.

The triplex junction connected two duplex segments into pseudo-infinite coaxially stacked helices. Two distinct triple interactions were present within the junction. First, G2 was involved in a sugar-edge contact with A14 as described, but it also made a single hydrogen bond through O6 with G13 N2 from the next coaxially stacked duplex (Figure 31A). This interaction was mediated in part by the hydrogen bonding of a solvent ion that was within hydrogen bonding distance of G2 N1, G13 N1, G1 O6, and the G11 phosphate of the partner duplex. Next, G1 was base paired to G11 of the partner duplex through their Watson-Crick and Hoogsteen faces, respectively. A12 from a coaxially stacked duplex made a single hydrogen bond with G1 (Figure 31B). To our knowledge these kind of all purine triple interactions have not been previously observed in DNA. Additionally, they provide an atypical example of coaxial helical stacking. In this case, only the 5'-most residues (G1) are directly stacked, while the 3'-most residues (A14) are involved in the tertiary contacts that allow the parallel arrangement of adjacent duplex stacks. The purine base triples effectively “stitch” the four strands together at the major groove, without significant stacking interactions between the duplexes.

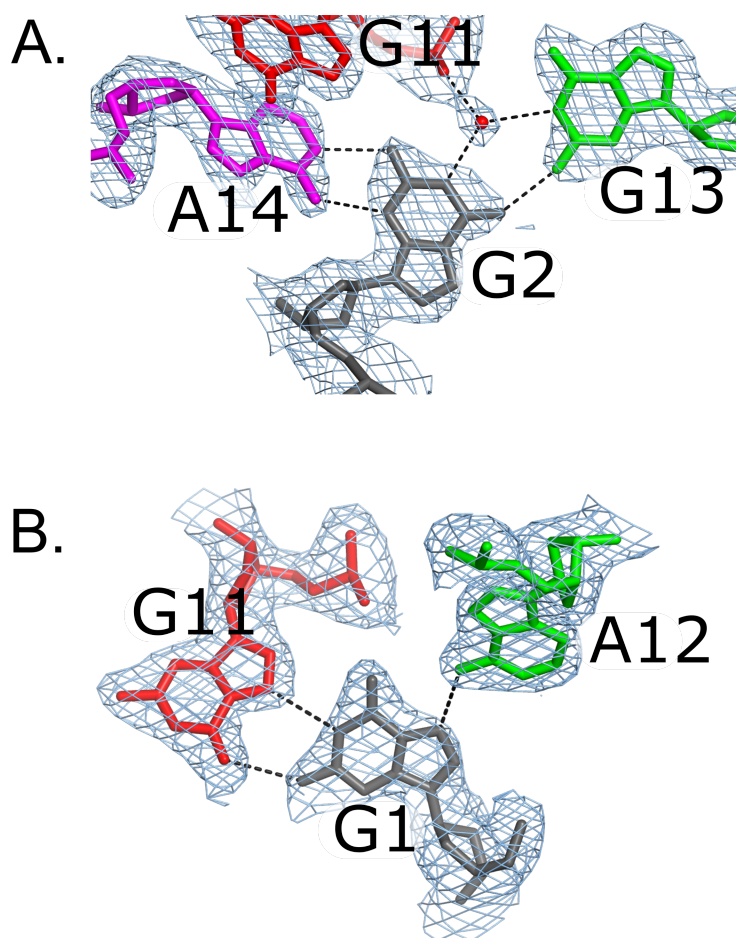


Figure 31. Purine base triples. The two purine base triple interactions **A**. The first base triple is mediated by the hydrogen bonding between the sugar edge of G2 and the Watson-Crick edge of A14 and with an additional hydrogen bond between O6 of G2 and N2 of G13 from the adjacent stacked duplex. Additional hydrogen bonding is mediated by the interaction of the water molecule with G2 N1, G13 N1, G1 O6, and the G11 phosphate of the partner duplex. **B**. The second base triple is formed between the Watson-Crick edge of G1 and the Hoogsteen edge of G11 from its duplex partner, along with the single hydrogen bond between G1 and A12 from a coaxially stacked duplex. In both **A**. and **B**. the hydrogen bonds are denoted by dashed lines and sigma A-weighted electron density map ($2F_o-F_c$) contoured at 1.0σ is shown in blue.

Chapter 4.6 Sequence Requirements for Alternate Crystal Form.

To understand why the addition of a single 3' adenosine could result in a significantly different structure under identical crystallization conditions, I set out to understand the sequence requirements for the alternate crystal form in the context of the determined crystal structures. I screened 30 variants of the 14-mer oligonucleotides by altering the nucleobase identities at positions mediating key interactions in the structures. I probed these interactions in three different groups. In the first group, I screened all four of the oligonucleotides described here, but with the different nucleobase identities at the added 14th residue. Out of the 12 DNA oligonucleotides screened, 7 crystallized with a hexagonal unipyramidal habit and the other 5 failed to show any crystals (Table 8). Notably, 3 of these sequences were the variants of A4-14, which did not crystallize as the 13-mer (Table 2). These results suggested adenosine at 14th position is a requirement for the alternate crystal form. This is consistent with our structural observations of the G2-A14 tertiary base pair. Simple modelling with different nucleotide identities at position 14 indicated that pyrimidines would be unable to pair with G2 without significant backbone clashes, while a guanosine at this position would present incompatible hydrogen bonding partners.

In the second group I examined six sequences with different self-complementary base pairs formed by positions 4 and 9, adjacent to the A3-G10 pairing (Table 8). I observed crystals in all cases, with sequences having the Y4-R9 base pair exhibiting the hexagonal crystal habit, while the other two sequences having a G4-C9 base pair forming microcrystals or irregular crystals. Though I could not

ascertain if the G4-C9 containing crystals belonged to one of the two crystal forms, these results indicated that the sequence rules observed for the formation of tandem sheared G-A base pairs⁹⁷ also applied to these crystals. Only sequences with a thymidine 5' of G10 adopted the alternate crystal form described here, though it is possible that a cytosine at this location could promote the alternate crystal form. The strong stacking interactions between T9 and G10, along with geometric constraints, were previously suggested as reasons for the presence of the pyrimidine 5' of guanines in the tandem sheared pairs^{97,98,101}. Our structural and crystal screen results support this analysis, but also indicate that a potential A3-T9 interstrand hydrogen bond (Figure S2) may help stabilize these structures. Notably, the hydrogen bond acceptor at O2 would be present in either pyrimidine at position 9.

Finally, I screened several sequence variants at the A5-T8 base pair. Based on the local sequence rules for the formation of the sheared G-A pair, I anticipated that this position should have little impact on the interactions necessary to form the alternate crystal form. Interestingly, 10 out of the 12 sequences screened in this group failed to crystallize, while the remaining two sequences formed only poor crystals (Table 8). This somewhat surprising result may be explained in several ways. First, this may indicate that the significant stacking interactions between T8 and T9 (7.56 \AA^2) are required to adopt the conformation necessary to form the alternate crystal form, though this does not appear to be the case for solution structures containing tandem sheared GA pairs. Second, in previous chapters we have established that the A5-T8 base pair was an important determinant for crystallization and crystallization speed in the context of 13-mers. It is possible that this base pair

may have a more fundamental role in the formation of the short self-complementary duplex that is a common feature of the 13-mer and 14-mer structures.

Altogether the sequence study and the structural observations strongly suggest that presence of A14-G2, A4-T8 and A5-T9 are all critical to the formation and stabilization of the alternate crystal form.

Table 7. Sequence-dependent crystallization.

Designation	Sequence*	Crystal Habit
A1-14-T	GGAAAATTTGGAG T	Hexagonal
A1-14-G	GGAAAATTTGGAG G	Hexagonal
A1-14-C	GGAAAATTTGGAG C	Hexagonal
A2-14-T	GGAAACGTTGGAG T	None
A2-14-G	GGAAACGTTGGAG G	Hexagonal
A2-14-C	GGAAACGTTGGAG C	Hexagonal
A3-14-T	GGAAAGCTTGGAG T	None
A3-14-G	GGAAAGCTTGGAG G	Hexagonal
A3-14-C	GGAAAGCTTGGAG C	Hexagonal
A4-14-T	GGAAATATTGGAG T	None
A4-14-G	GGAAATATTGGAG G	None
A4-14-C	GGAAATATTGGAG C	None
B6-14-A	GGACACGTGGGAGA	Hexagonal
B7-14-A	GGACAGCTGGGAGA	Hexagonal
E1-14-A	GGATAATTAGGAGA	Hexagonal
E3-14-A	GGATAGCTAGGAGA	Hexagonal
C9-14-A	GGAGAATTCGGAGA	Microcrystals
C11-14-A	GGAGAGCTCGGAGA	Clusters
B1-14-A	GGAA TATAT GGAGA	None
B2-14-A	GGAA TCGAT GGAGA	None
B3-14-A	GGAA TGCAT GGAGA	None
B4-14-A	GGAA TTAAT GGAGA	None
A5-14-A	GGAA CATGT GGAGA	None
A6-14-A	GGAA CCGGT GGAGA	None
A7-14-A	GGAA CGCGT GGAGA	None
A8-14-A	GGAA CTAGT GGAGA	Clusters
A9-14-A	GGAA GATCT GGAGA	Hexagonal
A10-14-A	GGAA GCGCT GGAGA	None
A11-14-A	GGAA GGCCT GGAGA	None
A12-14-A	GGAA GTA CTGGAGA	None

*Red indicate the position of sequence variability.

Chapter 4.7 Materials and Methods

1. DNA Synthesis and Purification

The four DNA 14-mers, designated A1-14: 5'-d(GGAAAATTTGGAGA); A2-14: 5'-d(GGAAACGTTGGAGA); A3-14: 5'-d(GGAAAGCTTGGAGA); A4-14: 5'- d(GGAAATATTGGAGA), were synthesized on the 1 μ mol scale (Integrated DNA Technologies, Coralville, IA) and were purified by 20% (19:1) polyacrylamide gel electrophoresis, electroeluted, and ethanol precipitated as previously described⁵⁹. The A3-14 (BrU9) oligonucleotide was synthesized using standard phosphoramidite chemistry on an Expedite 8909 DNA synthesizer (PerSeptive BioLabs) with reagents from Glen Research (Sterling, VA). The purified DNA samples were dialyzed against deionized water and the concentration was adjusted to 260 μ M. Oligonucleotides used to examine sequence effects on crystallization (Table 2) were synthesized on the 100 nmol scale, dissolved in deionized water, and used without purification.

2. Crystallization

The DNA oligonucleotides were crystallized by sitting drop vapour diffusion. Prior to crystallization, DNA samples (260 μ M) were heated at 95° for 2' and cooled to room temperature. Samples were mixed (1:1) with crystallization buffer (120 mM magnesium formate, 50 mM lithium chloride and 10% 2-methyl-2,4-pentanediol) in a 4 μ l drop. The reservoir contained 400 μ l of crystallization buffer. The crystal plates were incubated at 22°C. Crystals appeared in 16-20 hours and grew to an average size of 250x75x100 μ m.

3. Data Collection and Structure Determination

Crystals were harvested by nylon loop, washed sequentially in crystallization buffer containing 30% and 40% 2-methyl-2,4-pentanediol, and flash-cooled in liquid nitrogen. Native data sets were collected at Advanced Photon Source, Argonne National Labs, Sector 24-ID-E. Data were indexed and integrated using XDS⁷⁴, and scaled using Aimless⁷⁵. Several data sets had relatively low completeness due to crystal orientation. However, each crystal type was highly isomorphous with respect to the unit cell dimensions (RMSD of unit cell dimensions ≤ 0.1 Å) allowing the merging of observations from multiple crystals to improve completeness. Phases were initially determined using an A3-14 (BrU9) derivative with data collected at Advanced Photon Source beamline 24-ID-C. Phases were determined by single wavelength anomalous dispersion with the substructure sites identified by HySS in the Phenix crystallography package^{77,102}. Models were built in Coot⁷⁶. The other three 14-mer structures were solved by molecular replacement using the completed A3-14 structure as a search model. Refinement was performed with Phenix⁷⁷. Water molecules and ions were added manually during the refinement process. Following converged refinement in Phenix, all of the structures were run through the PDB-REDO pipeline⁷⁸ with 10-fold cross-validation applied due to the small number of reflections in these datasets. Average R_{free} for these 10 different test sets are reported in Table 1. Coordinates and structure factors were deposited in the Protein Data Bank⁷⁹.

Chapter 5: Conclusion and Future Prospects

Through my work, I have identified 12 new DNA oligonucleotide sequences, capable of adopting isomorphous crystal structures, starting from a single oligonucleotide sequence. This study opens up the sequence space and provides diversity in the design and construction of 3D DNA crystals. Importantly, the 13-mer crystals described in this study contains solvent channels that run throughout the length of the crystals, making them ideal candidates for use as molecular scaffolds. The identification of new sequences that form isomorphous structures will provide additional diversity for attaching and characterizing guest molecules for this and other applications.

Having identified sequences with isomorphous structures, I demonstrated that hybrid crystals could be grown from a mixture of two different DNA oligonucleotides. This provides a new tool for understanding the crystal self-assembly process, particularly the nucleation of these 13-mer DNA oligonucleotides. These oligonucleotides provides a unique system for further understanding how a dispersed biopolymer can self-assemble into macroscopic objects through a simple chemical trigger.

In this study I demonstrated how a single base pair (A5-T8) in the 13-mer structure is related to both crystallizability and the crystallization speed. Similarly, the 14-mer structure described here demonstrates how a single nucleotide change can lead to significantly different secondary and tertiary interactions and subsequently the overall structure. Additionally, through the sequence dependent crystallization study I

determined that adoption of the 14-mer structure is in part influenced by the neighboring sequence.

In our lab we are exploring the use of 13-mer oligonucleotides for developing “core-shell” crystals as drug delivery vehicles and for solid state catalysis. The added diversity provided by the newly identified isomorphous sequences, will help in attaching multiple guest molecules to the DNA to have a ‘multi-functional’ crystal. My work also proved useful in the crystal crosslinking project in our lab. The various sequences with isomorphous sequences proved to be a good system for studying the crosslinking efficiencies for various 13-mer sequences. The work on heterogenous crystals could be explored further to use these 13-mer crystals as ‘crystallization-aid’, where a small amount of crystallizing oligonucleotide promotes the crystallization of other difficult to crystallize oligonucleotides.

I believe that my work will have important implications for improving the characteristics of periodic 3D DNA crystals, which have been one of the most highly sought after DNA architectures. I also believe that my study is a step forward in understanding the assembly process for DNA oligonucleotides and provide insights in the sequence-structure relationship for 3D DNA crystals. This would prove useful in the rational designing of DNA structures and help in optimizing the conditions for assembly, manipulation, and functionalization. This in turn will benefit both upstream design and downstream applications of DNA structural nanotechnology.

Bibliography

1. Dahm, R. Discovering DNA: Friedrich Miescher and the early years of nucleic acid research. *Hum. Genet.* **122**, 565–581 (2008).
2. Jones, M. R., Seeman, N. C. & Mirkin, C. A. Programmable materials and the nature of the DNA bond. *Science* **347**, 1260901 (2015).
3. Aldaye, F. A., Palmer, A. L. & Sleiman, H. F. Assembling Materials with DNA as the Guide. *Science* **321**, 1795–1799 (2008).
4. Lipfert, J., Doniach, S., Das, R. & Herschlag, D. Understanding Nucleic Acid–Ion Interactions. *Annu. Rev. Biochem.* **83**, 813–841 (2014).
5. Winfree, E., Liu, F., Wenzler, L. A. & Seeman, N. C. Design and self-assembly of two-dimensional DNA crystals. *Nature* **394**, 539–44 (1998).
6. Rothmund, P. W. K. Folding DNA to create nanoscale shapes and patterns. *Nature* **440**, 297–302 (2006).
7. Yan, H., Park, S. H., Finkelstein, G., Reif, J. H. & LaBean, T. H. DNA-templated self-assembly of protein arrays and highly conductive nanowires. *Science* **301**, 1882–4 (2003).
8. He, Y., Chen, Y., Liu, H., Ribbe, A. E. & Mao, C. Self-assembly of hexagonal DNA two-dimensional (2D) arrays. *J. Am. Chem. Soc.* **127**, 12202–3 (2005).
9. Roy, S. & Caruthers, M. Synthesis of DNA/RNA and Their Analogs via Phosphoramidite and H-Phosphonate Chemistries. *Molecules* **18**, 14268–14284 (2013).
10. Drew, H. R. *et al.* Structure of a B-DNA dodecamer: conformation and dynamics. *Proc. Natl. Acad. Sci. U. S. A.* **78**, 2179–2183 (1981).

11. Seeman, N. C. DNA in a material world. *Nature* **421**, 427–431 (2003).
12. Seeman, N. C. Nucleic acid junctions and lattices. *J. Theor. Biol.* **99**, 237–47 (1982).
13. Li, X., Yang, X., Qi, J. & Seeman, N. C. Antiparallel DNA Double Crossover Molecules As Components for Nanoconstruction. *J. Am. Chem. Soc.* **118**, 6131–6140 (1996).
14. LaBean, T. H., Yan H., Kopatsch J., Liu F., Winfree E., Reif J. H., Seeman N. C. Construction, Analysis, Ligation, and Self-Assembly of DNA Triple Crossover Complexes. *J. Am. Chem. Soc.* **122**, 1848–1860 (2000).
15. He, Y., Ye T., Su M., Zhang C., Ribbe A. E., Jiang W., Mao C. Hierarchical self-assembly of DNA into symmetric supramolecular polyhedra. *Nature* **452**, 198–201 (2008).
16. Chen, J. H. & Seeman, N. C. Synthesis from DNA of a molecule with the connectivity of a cube. *Nature* **350**, 631–3 (1991).
17. Goodman, R. P., Schaap I. A. T., Tardin C. F., Erben C. M., Berry R. M., Schmidt C. F., Tuberfield A.J. Rapid chiral assembly of rigid DNA building blocks for molecular nanofabrication. *Science* **310**, 1661–5 (2005).
18. Aldaye, F. A. & Sleiman, H. F. Modular Access to Structurally Switchable 3D Discrete DNA Assemblies. *J. Am. Chem. Soc.* **129**, 13376–13377 (2007).
19. Erben, C. M., Goodman, R. P. & Turberfield, A. J. A self-assembled DNA bipyramid. *J. Am. Chem. Soc.* **129**, 6992–3 (2007).
20. Mitchell, J. C., Harris, J. R., Malo, J., Bath, J. & Turberfield, A. J. Self-assembly of chiral DNA nanotubes. *J. Am. Chem. Soc.* **126**, 16342–3 (2004).

21. Lund, K., Liu, Y., Lindsay, S. & Yan, H. Self-assembling a molecular pegboard. *J. Am. Chem. Soc.* **127**, 17606–7 (2005).
22. Liu, Y., Ke, Y. & Yan, H. Self-assembly of symmetric finite-size DNA nanoarrays. *J. Am. Chem. Soc.* **127**, 17140–1 (2005).
23. Zheng, J., Birktoft J. J., Chen Y., Wang T., Sha R., Constantinou P.E., Ginell S.L., Mao C., Seeman N. C. From molecular to macroscopic via the rational design of a self-assembled 3D DNA crystal. *Nature* **461**, 74–77 (2009).
24. Andersen, E. S. *et al.* Self-assembly of a nanoscale DNA box with a controllable lid. *Nature* **459**, 73–76 (2009).
25. Douglas, S. M. *et al.* Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **459**, 414–418 (2009).
26. Dietz, H., Douglas, S. M. & Shih, W. M. Folding DNA into Twisted and Curved Nanoscale Shapes. *Science* **325**, 725–730 (2009).
27. Han, D. *et al.* DNA origami with complex curvatures in three-dimensional space. *Science* **332**, 342–346 (2011).
28. Wei, B., Dai, M. & Yin, P. Complex shapes self-assembled from single-stranded DNA tiles. *Nature* **485**, 623–626 (2012).
29. Ke, Y., Ong, L. L., Shih, W. M. & Yin, P. Three-Dimensional Structures Self-Assembled from DNA Bricks. *Science* **338**, 1177–1183 (2012).
30. Ke, Y. *et al.* DNA brick crystals with prescribed depths. *Nat. Chem.* **6**, 994–1002 (2014).

31. Mao, C., LaBean, T. H., Relf, J. H. & Seeman, N. C. Logical computation using algorithmic self-assembly of DNA triple-crossover molecules. *Nature* **407**, 493–6 (2000).
32. Zhang, F., Nangreave, J., Liu, Y. & Yan, H. Structural DNA nanotechnology: state of the art and future perspective. *J. Am. Chem. Soc.* **136**, 11198–11211 (2014).
33. Drew, H. R., McCall, M. J. & Calladine, C. R. Recent Studies of DNA in the Crystal. *Annu. Rev. Cell Biol.* **4**, 1–20 (1988).
34. Drew, H., Takano, T., Tanaka, S., Itakura, K. & Dickerson, R. E. High-salt d(CpGpCpG), a left-handed Z[prime] DNA double helix. *Nature* **286**, 567–573 (1980).
35. Dickerson, R. E. & Drew, H. R. Structure of a B-DNA dodecamer. *J. Mol. Biol.* **149**, 761–786 (1981).
36. Rich, A. & Zhang, S. Z-DNA: the long road to biological function. *Nat Rev Genet* **4**, 566–572 (2003).
37. Lee, J. S., Johnson, D. A. & Morgan, A. R. Complexes formed by (pyrimidine)_n . (purine)_n DNAs on lowering the pH are three-stranded. *Nucleic Acids Res.* **6**, 3073–3091 (1979).
38. Lilley, D. M. Hairpin-loop formation by inverted repeats in supercoiled DNA is a local and transmissible property. *Nucleic Acids Res.* **9**, 1271–1289 (1981).
39. Panayotatos, N. & Wells, R. D. Cruciform structures in supercoiled DNA. *Nature* **289**, 466–470 (1981).

40. Sundquist, W. I. & Klug, A. Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature* **342**, 825–9 (1989).
41. Bacolla, A. & Wells, R. D. Non-B DNA Conformations, Genomic Rearrangements, and Human Disease. *J. Biol. Chem.* **279**, 47411–47414 (2004).
42. Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K. & Neidle, S. Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.* **34**, 5402–5415 (2006).
43. Zhao, J., Bacolla, A., Wang, G. & Vasquez, K. M. Non-B DNA structure-induced genetic instability and evolution. *Cell. Mol. Life Sci.* **67**, 43–62 (2009).
44. Yatsunyk, L. A., Mendoza, O. & Mergny, J.-L. ‘Nano-oddities’: unusual nucleic acid assemblies for DNA-based nanostructures and nanodevices. *Acc. Chem. Res.* **47**, 1836–1844 (2014).
45. Dong, Y., Yang, Z. & Liu, D. DNA nanotechnology based on i-motif structures. *Acc. Chem. Res.* **47**, 1853–1860 (2014).
46. Chakraborty, S., Sharma, S., Maiti, P. K. & Krishnan, Y. The poly dA helix: a new structural motif for high performance DNA-based molecular switches. *Nucleic Acids Res.* **37**, 2810–2817 (2009).
47. Saha, S., Chakraborty, K. & Krishnan, Y. Tunable, colorimetric DNA-based pH sensors mediated by A-motif formation. *Chem. Commun.* **48**, 2513–2515 (2012).
48. Yatsunyk, L. A. *et al.* Guided assembly of tetramolecular G-quadruplexes. *ACS Nano* **7**, 5701–5710 (2013).

49. Zhang, D., Huang, T., Lukeman, P. S. & Paukstelis, P. J. Crystal structure of a DNA/Ba²⁺ G-quadruplex containing a water-mediated C-tetrad. *Nucleic Acids Res.* **42**, 13422–13429 (2014).
50. Muser, S. E. & Paukstelis, P. J. Three-dimensional DNA crystals with pH-responsive noncanonical junctions. *J. Am. Chem. Soc.* **134**, 12557–12564 (2012).
51. Tripathi, S., Zhang, D. & Paukstelis, P. J. An intercalation-locked parallel-stranded DNA tetraplex. *Nucleic Acids Res.* **43**, 1937–1944 (2015).
52. Peyret, N., Seneviratne, P. A., Allawi, H. T. & SantaLucia, J. Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A.A, C.C, G.G, and T.T mismatches. *Biochemistry* **38**, 3468–77 (1999).
53. Tikhomirova, A., Beletskaya, I. V. & Chalikian, T. V. Stability of DNA duplexes containing GG, CC, AA, and TT mismatches. *Biochemistry* **45**, 10563–10571 (2006).
54. Rossetti, G. *et al.* The structural impact of DNA mismatches. *Nucleic Acids Res.* **43**, 4309–4321 (2015).
55. Bochman, M. L., Paeschke, K. & Zakian, V. A. DNA secondary structures: stability and function of G-quadruplex structures. *Nat. Rev. Genet.* **13**, 770–780 (2012).
56. Wei, X., Nangreave, J. & Liu, Y. Uncovering the Self-Assembly of DNA Nanostructures by Thermodynamics and Kinetics. *Acc. Chem. Res.* **47**, 1861–1870 (2014).
57. Song, J. *et al.* Isothermal hybridization kinetics of DNA assembly of two-dimensional DNA origami. *Small* **9**, 2954–2959 (2013).

58. Johnson-Buck, A., Nangreave, J., Jiang, S., Yan, H. & Walter, N. G. Multifactorial Modulation of Binding and Dissociation Kinetics on Two-Dimensional DNA Nanostructures. *Nano Lett.* **13**, 2754–2759 (2013).
59. Paukstelis, P. J., Nowakowski, J., Birktoft, J. J. & Seeman, N. C. Crystal Structure of a Continuous Three-Dimensional DNA Lattice. *Chem. Biol.* **11**, 1119–1126 (2004).
60. Sizov, I., Rahman, M., Gelmont, B., Norton, M. L. & Globus, T. Sub-THz spectroscopic characterization of vibrational modes in artificially designed DNA monocrystal. *Chem. Phys.* **425**, 121–125 (2013).
61. Zhang, W., Brown, E. R., Rahman, M. & Norton, M. L. Observation of terahertz absorption signatures in microliter DNA solutions. *Appl. Phys. Lett.* **102**, 023701 (2013).
62. Paukstelis, P. J. Three-dimensional DNA crystals as molecular sieves. *J. Am. Chem. Soc.* **128**, 6794–5 (2006).
63. Geng, C. & Paukstelis, P. J. DNA Crystals as Vehicles for Biocatalysis. *J. Am. Chem. Soc.* **136**, 7817–7820 (2014).
64. Leontis, N. B. & Westhof, E. The Annotation of RNA Motifs. *Comp. Funct. Genomics* **3**, 518–524 (2002).
65. Greene, K. L., Jones R.L., Li Y., Robinson H., Wang A.H.J., Zon G., Wilson W. D. Solution Structure of a GA Mismatch DNA Sequence, d(CCATGAATGG)₂, Determined by 2D NMR and Structural Refinement Methods. *Biochemistry* **33**, 1053–1062 (1994).

66. Prive, G. G., Helnemann U., Chandrasegran S., Kan L. S., Kopka M. L., Dickerson R.E. Helix geometry, hydration, and G.A mismatch in a B-DNA decamer. *Science* **238**, 498–504 (1987).
67. Hays, F. A. Teegarden A., Jones Z. J. R., Harms M., Raup D., Watson J., Cavaliere E., Ho P. S. How sequence defines structure: A crystallographic map of DNA structure and conformation. *Proc. Natl. Acad. Sci. U. S. A.* **102**, 7157–7162 (2004).
68. Baumgartner, J. *et al.* Nucleation and growth of magnetite from solution. *Nat. Mater.* **12**, 310–314 (2013).
69. Koo, H. S. & Crothers, D. M. Calibration of DNA curvature and a unified description of sequence-directed bending. *Proc. Natl. Acad. Sci. U. S. A.* **85**, 1763–1767 (1988).
70. Hagerman, P. J. Flexibility of DNA. *Annu. Rev. Biophys. Biophys. Chem.* **17**, 265–86 (1988).
71. Haran, T. E., Kahn, J. D. & Crothers, D. M. Sequence Elements Responsible for DNA Curvature. *J. Mol. Biol.* **244**, 135–143 (1994).
72. Kahn, J. D., Yun, E. & Crothers, D. M. Detection of localized DNA flexibility. *Nature* **368**, 163–166 (1994).
73. Lu, X.-J. & Olson, W. K. 3DNA: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures. *Nat. Protoc.* **3**, 1213–1227 (2008).
74. Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 125–132 (2010).

75. Evans, P. R. & Murshudov, G. N. How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **69**, 1204–1214 (2013).
76. Emsley, P., Lohkamp, B., Scott, W. G. & Cowtan, K. Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **66**, 486–501 (2010).
77. Afonine, P. V. *et al.* Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **68**, 352–367 (2012).
78. Joosten, R. P., Long, F., Murshudov, G. N. & Perrakis, A. The PDB_REDO server for macromolecular structure model optimization. *International Union Of Crystallography (IUCr) J* **1**, 213–220 (2014).
79. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
80. Delano, W. The PyMOL Molecular Graphics System, Schrödinger, LLC.
81. Hwang, T. L. & Shaka, A. J. Water Suppression That Works. Excitation Sculpting Using Arbitrary Wave-Forms and Pulsed-Field Gradients. *J. Magn. Reson. A* **112**, 275–279 (1995).
82. Carr, H. Y. & Purcell, E. M. Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments. *Phys. Rev.* **94**, 630–638 (1954).
83. Nath, K., Sarosy, J. W., Hahn, J. & Di Como, C. J. Effects of ethidium bromide and SYBR® Green I on different polymerase chain reaction systems. *J. Biochem. Biophys. Methods* **42**, 15–29 (2000).
84. Roberts, T. M., Kacich, R. & Ptashne, M. A general method for maximizing the expression of a cloned gene. *Proc. Natl. Acad. Sci. U. S. A.* **76**, 760–764 (1979).

85. Price, P. A. The essential role of Ca^{2+} in the activity of bovine pancreatic deoxyribonuclease. *J. Biol. Chem.* **250**, 1981–1986 (1975).
86. Rossetti, G. *et al.* The structural impact of DNA mismatches. *Nucleic Acids Res.* **43**, 4309–4321 (2015).
87. Kan, L. S., Chandrasegaran, S., Pulford, S. M. & Miller, P. S. Detection of a guanine-adenine base pair in a decadeoxyribonucleotide by proton magnetic resonance spectroscopy. *Proc. Natl. Acad. Sci. U. S. A.* **80**, 4263–4265 (1983).
88. Patel, D. J., Kozlowski, S. A., Ikuta, S. & Itakura, K. Deoxyguanosine-deoxyadenosine pairing in the d(C-G-A-G-A-A-T-T-C-G-C-G) duplex: conformation and dynamics at and adjacent to the dG-dA mismatch site. *Biochemistry* **23**, 3207–3217 (1984).
89. Brown, T., Hunter, W. N., Kneale, G. & Kennard, O. Molecular structure of the G-A base pair in DNA and its implications for the mechanism of transversion mutations. *Proc. Natl. Acad. Sci. U. S. A.* **83**, 2402–2406 (1986).
90. Brown, T., Leonard, G. A., Booth, E. D. & Chambers, J. Crystal structure and stability of a DNA duplex containing A(anti) · G(syn) base-pairs. *J. Mol. Biol.* **207**, 455–457 (1989).
91. Hunter, W. N., Brown, T. & Kennard, O. Structural Features and Hydration of d(C-G-C-G-A-A-T-T-A-G-C-G); a Double Helix Containing Two G.A Mispairs. *J. Biomol. Struct. Dyn.* **4**, 173–191 (1986).
92. Leonard, G. A., Booth, E. D. & Brown, T. Structural and thermodynamic studies on the adenine.guanine mismatch in B-DNA. *Nucleic Acids Res.* **18**, 5617–5623 (1990).

93. Nikonowicz, E. P. & Gorenstein, D. G. Two-dimensional proton and phosphorus-31 NMR spectra and restrained molecular dynamics structure of a mismatched GA decamer oligodeoxyribonucleotide duplex. *Biochemistry* **29**, 8845–8858 (1990).
94. Li, Y., Zon, G. & Wilson, W. D. NMR and molecular modeling evidence for a G.A mismatch base pair in a purine-rich DNA duplex. *Proc. Natl. Acad. Sci. U. S. A.* **88**, 26–30 (1991).
95. Li, Y., Zon, G. & Wilson, W. D. Thermodynamics of DNA duplexes with adjacent G-A mismatches. *Biochemistry (Mosc.)* **30**, 7566–7572 (1991).
96. Coimbatore Narayanan, B. *et al.* The Nucleic Acid Database: new features and capabilities. *Nucleic Acids Res.* **42**, D114–D122 (2014).
97. Cheng, J.-W., Chou, S.-H. & Reid, B. R. Base pairing geometry in GA mismatches depends entirely on the neighboring sequence. *J. Mol. Biol.* **228**, 1037–1041 (1992).
98. Chou, S.-H., Cheng, J.-W. & Reid, B. R. Solution structure of [d(ATGAGCGAATA)]₂: Adjacent G-A mismatches stabilized by cross-strand base-stacking and BII phosphate groups. *J. Mol. Biol.* **228**, 138–155 (1992).
99. Chou, S.-H., Zhu, L. & Reid, B. R. The Unusual Structure of the Human Centromere (GGA)₂ Motif: Unpaired Guanosine Residues Stacked Between Sheared G·A Pairs. *J. Mol. Biol.* **244**, 259–268 (1994).
100. Chou, S.-H., Cheng, J.-W., Fedoroff, O. & Reid, B. R. DNA Sequence GCGAATGAGC Containing the Human Centromere Core Sequence GAAT

- Forms a Self-complementary Duplex with Sheared G·A Pairs in Solution. *J. Mol. Biol.* **241**, 467–479 (1994).
101. Chou, S.-H., Chin, K.-H. & Wang, A. H.-J. Unusual DNA duplex and hairpin motifs. *Nucl Acids Res* **31**, 2461–2474 (2003).
102. Grosse-Kunstleve, R. W. & Adams, P. D. Substructure search procedures for macromolecular structures. *Acta Crystallogr. D Biol. Crystallogr.* **59**, 1966–1973 (2003).

