ABSTRACT

Title of Dissertation:	BIOPHYSICAL AND STRUCTURAL CHARACTERIZATION OF THE PARALLEL-STRANDED D(CGA) TRIPLET REPEAT MOTIF
	Emily M. Luteran, Doctor of Philosophy, 2022
Dissertation directed by:	Dr. Paul J. Paukstelis, Associate Professor Department of Chemistry and Biochemistry

Deoxyribonucleic acid (DNA) has been repurposed as a building block in the construction of nanoscale assemblies. The biocompatibility, stability, programmability, structural predictability, and ability to self-assemble inherent to DNA has been leveraged to design various 2D and 3D DNA-based architectures. However, such architectures are commonly restricted by the structural rigidity and stability of the Watson-Crick duplex. Non-canonical DNA interactions can be incorporated to overcome this limitation by retaining the favorable characteristics of DNA while offering structural versatility beyond the constraints of Watson-Crick interactions. Another desirable functional advantage of non-canonical interactions are their sensitivities to the local environment, including cations, salt concentration, or pH, which allow them to undergo predictable structural changes in response to environmental perturbations. The d(CGA) triplet repeat motif is an example of a structurally dynamic non-canonical DNA motif that can transition from parallel-

stranded homo-base paired duplex to anti-parallel unimolecular hairpin in a pH-dependent manner. This dissertation describes the biophysical and structural characterization of the non-canonical d(CGA) repeat motif and related sequence variants.

Thermodynamic parameters obtained from UV absorbance melting curves show that the structural transition resulting from decreasing the pH is accompanied by a significant energetic stabilization as hairpin structures are converted to parallel-stranded duplexes. Additionally, nuclease resistance against double strand-specific nucleases in the parallel-stranded form suggests that this motif may offer unique advantages for cellular applications. CD spectroscopy based kinetic analysis reveals that the time scale of the transition between structural forms is highly dependent upon the direction of the structural change (hairpin to parallel duplex or parallel duplex to hairpin). Biophysical characterization is complimented by the structure determination of four unique d(CGA)-based parallel-stranded duplexes across two crystal structures. Structural analysis confirms the robust structural predictability and defines the specific structural features of d(CGA) triplets in the parallel-stranded form. Finally, we explored 3D DNA crystals containing parallel-stranded d(GGA) triplet repeats as a new platform for drug delivery. Together, the data presented within this dissertation will provide the foundation for the rational incorporation of non-canonical based parallel-stranded interactions, specifically the d(CGA) motif, into the design of DNA-based nanoarchitectures.

BIOPHYSICAL AND STRUCTURAL CHARACTERIZATION OF THE PARALLEL-STRANDED D(CGA) TRIPLET REPEAT MOTIF

by

Emily Michelle Luteran

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
2022

Advisory Committee:
Professor Paul J. Paukstelis, Chair
Professor Philip DeShong
Professor Jason Kahn
Professor Lai-Xi Wang
Professor Helim Aranda-Espinoza, Dean's Representative

Acknowledgements

I would like to express my sincere appreciation for my advisor, Dr. Paul Paukstelis. Without his continuous support and expertise throughout my graduate research this all would not have been possible. Thank you for giving me the opportunity, tools, and confidence to succeed.

Thank you to Dr. Phil DeShong, Dr. Jason Kahn, Dr. Lai-Xi Wang, and Dr. Nicole LaRonde for serving on my dissertation committee and always providing advice and encouragement along the way. I would like to specifically thank Dr. DeShong for the opportunity to work on an exciting project in collaboration with his lab. I am so thankful for his knowledge, mentorship, and kindness. I would also like to thank Dr. Kahn. Not only did he generously allow me to use his instrumentation at my convenience, but his insight and expertise helped to guide the analysis of our thermodynamic data. Thank you to Dr. Helim Aranda-Espinoza for serving as my Dean's Representative.

Thank you to my past lab-mates, Betty and Ron, for teaching me all of the tips and tricks of the lab, the fun conversations, and wonderful friendships. Thank you to my classmates and friends, especially Shannon, Lukasz, Mary, DH, and Ian.

I am also so grateful for my significant other, Daniel. Thank you for being my brainstorming partner, proofreader, figure-checker, and number one supporter. I am so fortunate to have gone through this journey with you.

Finally, my most sincere gratitude to my parents, Sue and Tom, and my sister, Kat. Without their unending love and support I would not be where I am today.

The work presented in this dissertation is adapted from the following research articles:

- Stability of the pH-Dependent Parallel-Stranded d(CGA) Motif
  Emily M. Luteran, Jason D. Kahn, Paul J. Paukstelis
  Biophysical Journal, **2020**
  https://doi.org/10.1016/j.bpj.2020.09.002

- The parallel-stranded d(CGA) duplex is a highly predictable structural motif with two conformationally distinct strands
  Emily M. Luteran, Paul J. Paukstelis
  Acta Crystallographia Section D, **2022**
  https://doi.org/10.1107/S2059798322000304

Table of Contents

List of Tables

List of Figures

List of Abbreviations and Symbols

A	Adenine
bp	Base pair
C	Cytosine
CD	Circular Dichroism
CL	Crosslinked
DNA	Deoxyribonucleic acid
dsDNA	Double-stranded deoxyribonucleic acid
DOX	Doxorubicin
G	Guanine
G4	G-quadruplex
HMW	High molecular weight
hr	Hour
isoC	Isocytosine
I/σI	Signal-to-noise ratio
LbL	Layer-by-layer
min	Minute
MPD	2-methyl-2,4-pentanediol
N	Any nucleobase (A, C, G, T)
NA	Not applicable
NCO	Cobalt Hexammine(III)
NMR	Nuclear magnetic resonance
nt	Nucleotide
P	Phosphate
PDB	Protein data bank
PEG	Polyethylene glycol
ps-duplex	Parallel-stranded duplex
RMSD	Root mean square deviation
RNA	Ribonucleic acid
RT	Room temperature

s	Second
ssDNA	Single-stranded deoxyribonucleic acid
T	Thymine
T_m	Melting temperature
Topo II	Topoisomerase II
U	Uracil
UT	Untreated
UTR	Untranslated region
UV	Ultraviolet
WC	Watson-Crick
Y	Pyrimidine (C or T)
2D	Two-dimensional
3D	Three-dimensional
3′-F	3′-fluorescein labeled DNA
Å	Angstrom
ΔH^o	Enthalpy change at the standard state
ΔG^o_{37}	Formation free energy change at the standard state and 37°C
ΔS^o	Entropy change at the standard state

Chapter 1: Introduction

Chapter 1.1. Canonical DNA Structure

Deoxyribonucleic acid (DNA) is the biomolecule responsible for the storage of genetic information.[1] DNA is an optimal material for this purpose because of structural features that provide inherent stability, rigidity, and programmability. Specifically, DNA polymers are comprised of nucleotide monomeric units that contain a deoxyribose sugar, nitrogenous base, and a phosphate group, where the identity of bases in a linear sequence encodes genetic information.[1] The polymeric DNA chain is formed via phosphodiester bonds between the 3′-hydroxyl group on the deoxyribose sugar of one monomer and the 5′-phosphate group of another monomer. The nature of this linkage gives each DNA strand a 5′ to 3′ directionality.

There are four unique nitrogenous bases found in native DNA including purines, adenine (A) and guanine (G), and pyrimidines, thymine (T) and cytosine (C). The hydrogen bond donor and acceptor pattern presented on each base allows for the formation of complementary Watson-Crick (WC) base pairs between A-T and G-C (Figure 1.1).[1, 2] The A-T base pair is formed via two hydrogen bonds between A_{N6}-T_{O4} and A_{N1}-T_{N3}, while the G-C base pair is formed via three hydrogen bonds G_{N2}-C_{O2}, G_{N1}-C_{N3}, G_{O6}-C_{N4}. The predictable complementary base pairing and base stacking interactions lead to the formation of the iconic double helical DNA structure.

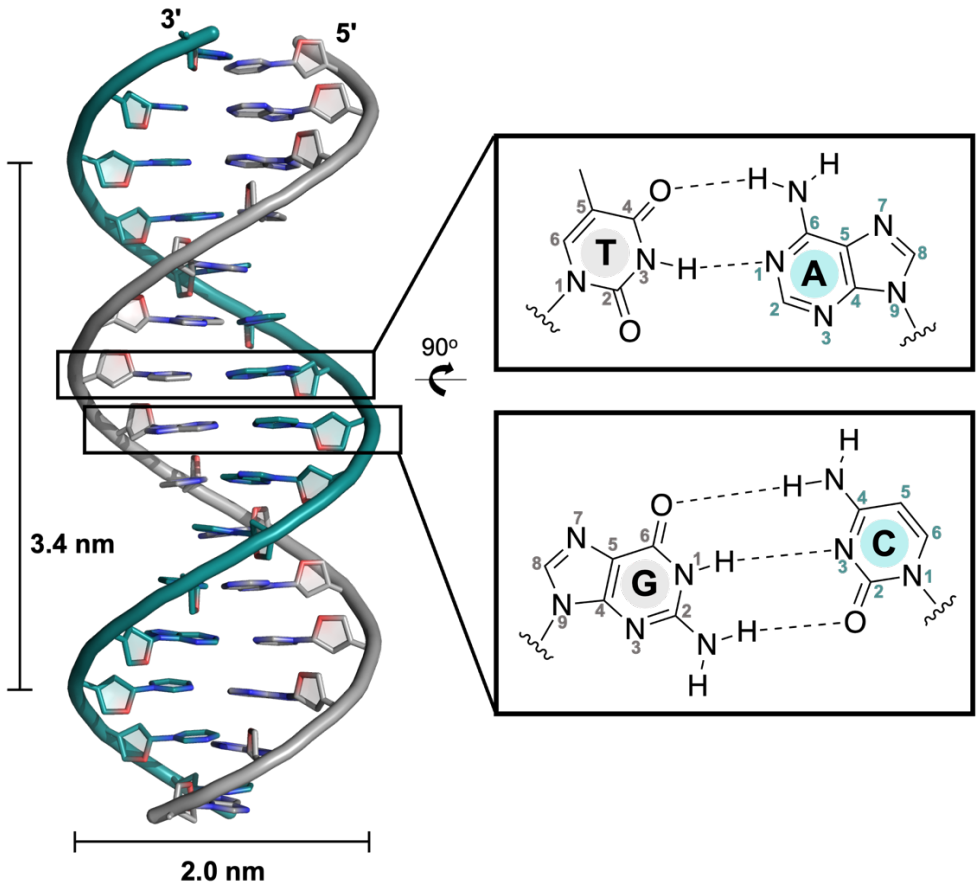

Figure 1.1. Canonical B-form DNA Double Helix. Modeled representation of an ideal Watson-Crick double-stranded DNA duplex. The Watson strand (gray) forms canonical Watson-Crick base pairs with the complementary Crick strand (teal). The top enlarged image highlights the hydrogen bonds involved in the A-T base pair, while the bottom illustrates the G-C base pair. The height of one complete helical turn measured parallel to the helical axis (helical pitch) and diameter are shown on the duplex.

The canonical B-form double helix contains two DNA strands (Watson and Crick) which hybridize through complementary base paring interactions with anti-parallel oriented strands.[2] In this way, the Watson strand oriented in the 5′ to 3′ direction hybridizes with the opposing Crick strand oriented in the 3′ to 5′ direction (Figure 1.1). The helical dimensions of the resulting B-form duplex are highly predictable including a 2 nm diameter and 3.4 nm helical pitch. (Figure 1.1).[3] Though the canonical Watson-

Crick double helix is the commonly recognized form of DNA, likely due to its important role in storing genetic information, there is a wide diversity of accessible DNA structures beyond the iconic double helix.

Chapter 1.2. Non-Canonical DNA Structure

Non-complementary sequences have the potential to form non-canonical DNA structures stabilized by alternative non-Watson-Crick base pairing and other interactions. Each nucleobase has hydrogen bond donor and acceptor sites beyond the canonical WC based pattern that allows access to such non-canonical geometries (Figure 1.2).[4]

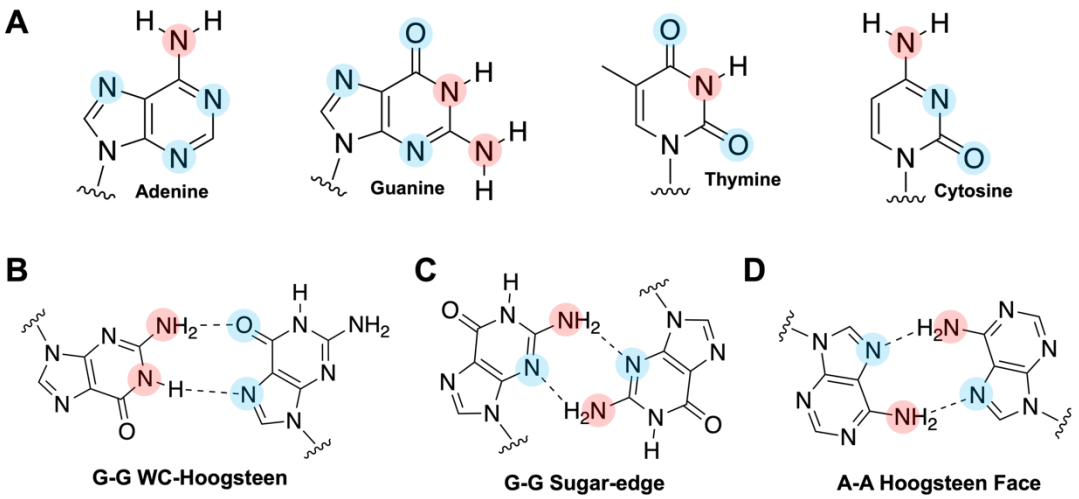

Figure 1.2. Non-canonical base pair hydrogen bond interactions. (A) Each nucleobase has several hydrogen bond donor (red) and acceptor (blue) sites that allow for non-Watson-Crick base-pairing interactions. Select examples of non-canonical base pairing interactions are shown in B-D. **(B)** Two hydrogen bonds (N_1-N_7 and N_2-O_6) are formed between the Watson-Crick edge and Hoogsteen face in this G-G homo-base pair. **(C)** Two hydrogen bonds (N_3-N_2 and N_2-N_3) are formed between sugar edges in this G-G homo-base pair. **(D)** Two hydrogen bonds (N_6-N_7 and N_7-N_6) are formed between Hoogsteen faces of this A-A homo-base pair.

Non-canonical structures stabilized by such alternative base pairing interactions can be controlled by nucleotide sequence composition and environmental factors including pH, the presence and concentration of cations, and temperature. Here, I will provide a brief description of several common non-canonical structures, including G-quadruplexes, i-motifs, and parallel-stranded duplexes, that utilize alternative base pairing interactions.

Chapter 1.2.1. G-Quadruplexes and I-motifs

G-quadruplexes (G4s) are guanine-rich DNA sequences that fold into four-stranded non-canonical structures.[5-7] G4 structures are stabilized by the formation of a square planar network comprised of four guanines (G-quartets). Each guanine within the G-quartet forms non-canonical Hoogsteen to WC-face hydrogen bonds with two adjacent guanines (Figure 1.3A). G-quartets orient in a single plane via π-π stacking interactions on top of one another to form G4s (Figure 1.3B). Cations (mono- or divalent) situated in the center cavity of G-quartets provide further stabilization to G4 structures via coordination to the O6 position of each surrounding guanine (Figure 1.3A).[8, 9] G4 structures form inter- or intra-molecularly and can adopt a diverse range of topologies depending on the polarity of each DNA strand including parallel, antiparallel, or hybrid forms (Figure 1.3B).[10, 11]

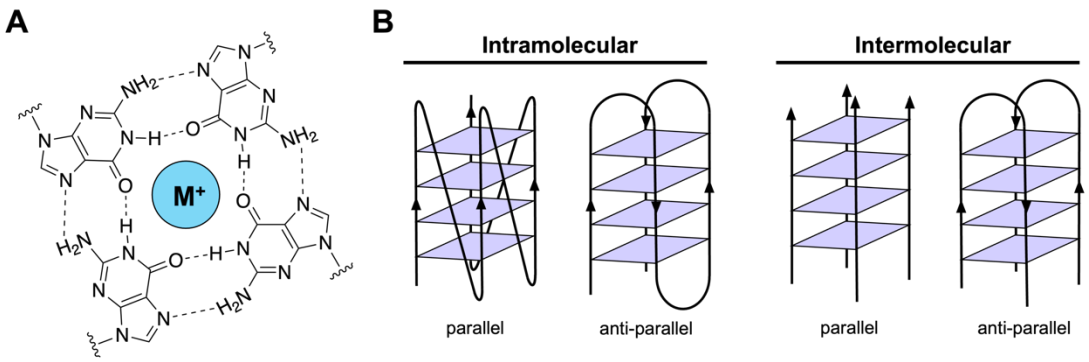

Figure 1.3. G-Quadruplex Structure. (A) Four guanine residues form a square planar network of hydrogen bond interactions to form a G-quartet. Monovalent or divalent cations occupy the center cavity and provide stabilization via coordination to G_{O6}. **(B)** G-quartets (purple sheets) stack on top of one another to form G-quadruplexes. G-quadruplexes can form intra- or intermolecularly with parallel or anti-parallel oriented strands.

Originally characterized by Gehring and colleagues in 1993, the intercalated-motif (i-motif) is a C-rich DNA sequence that folds into a non-canonical quadruplex structure in a pH-dependent manner and is characterized by the presence of the hemi-protonated C-CH+ homo-base pair (Figure 1.4A).[12] Specifically, two parallel-stranded duplexes are held together by hemi-protonated C-CH+ homo-base pairs which subsequently intercalate in an anti-parallel orientation to form the characteristic i-motif structure[12, 13] (Figure 1.4B). As with G4 structures, i-motif structures can form as intramolecular, bimolecular, or tetramolecular assemblies and can adopt a diverse range of intercalation topologies.[14-16]

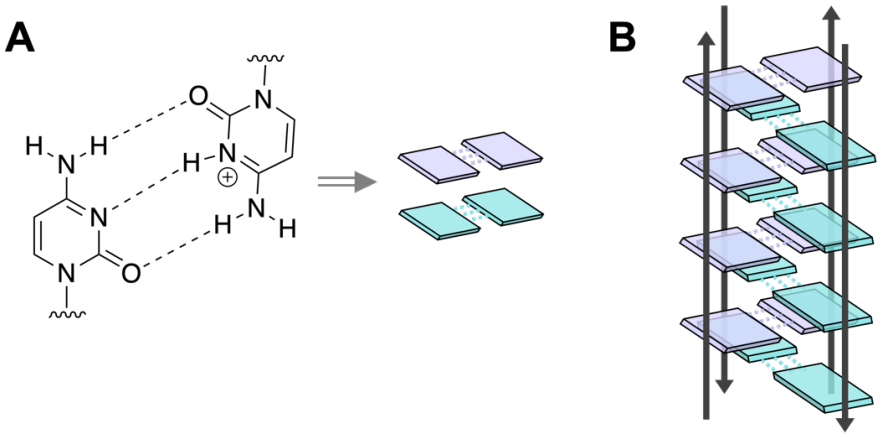

Figure 1.4. I-Motif Structure. (A) The hemi-protonated C-CH+ homo-base pair is the characteristic non-canonical interaction within the i-motif. This base pair relies on protonation of the N3 position of one of the two cytosines in the base pair to form three hydrogen bonds. This is schematically represented as purple or teal rectangles connected with three dashed lines. **(B)** Hemi-protonated C-CH+ homo-base pairs formed between two parallel stranded duplexes are shown as purple or teal rectangles connected with three dashed lines. The base pairs intercalate together in an anti-parallel orientation to form the complete i-motif structure. Shown is a tetramolecular i-motif structure. Similar uni- or bi-molecular structures (not shown) can also form depending on sequence composition and environmental factors.

The formation of G4 and i-motif structures can both be precisely regulated by several factors including DNA sequence composition and length[17-19], loop size[20], pH[21, 22], cation identity and concentration[23-26], temperature[17, 22], or presence of molecular crowding agents.[27-29] The structural tunability and environmental sensitivity of G4 and i-motif structures has implications in biology and DNA nanotechnology, to be expanded upon below in Chapter 1.3.1 and Chapter 1.4, respectively.

Chapter 1.2.2. The d(CGA) triplet repeat motif

The d(CGA) triplet repeat motif is another environmentally sensitive non-canonical motif that can adopt different structural forms in a pH-dependent manner at near-physiological temperature and salt concentration.[30] Neutral pH favors the

formation of a unimolecular anti-parallel hairpin stabilized by canonical G-C base pairs[31, 32], while acidic pH favors the formation of a non-canonical homo-base paired parallel-stranded duplex (ps-duplex) containing the hemi-protonated C-CH+ homo-base pair (Figure 1.5).[33, 34]

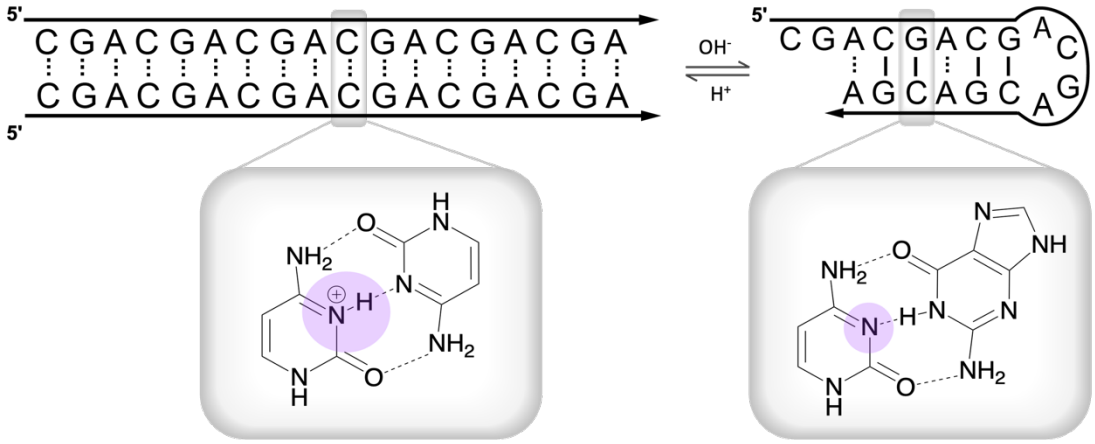

Figure 1.5. The d(CGA) triplet repeat motif. At low pH (left), $(CGA)_n$ sequences adopt a non-canonical homo-base paired parallel-stranded duplex containing the hemi-protonated C-CH+ base pair. The N_3 position (highlighted) of one of the two cytosines must be protonated for the formation of this base pair. Ps-duplex-specific interactions are shown below (Figure 1.6). At neutral pH (right), the cytosine is no longer protonated, favoring the formation of an anti-parallel hairpin containing canonical Watson-Crick G-C base pairs.

At a high DNA concentration and low temperature, changes in salt concentration and pH allow $d(CGA)_4$ to sample four distinct conformations — parallel-stranded duplex, anti-parallel hairpin, anti-parallel B-form duplex, and Z-form duplex.[32] The parallel-stranded duplex is primarily observed in acidic conditions, whereas at higher pH (7.0 and 8.0) and low ionic strength, the structure converts to an anti-parallel hairpin. When ionic strength is increased at pH 8.0, the DNA favors an anti-parallel duplex, whereas very high ionic strength (4 M NaCl) with trace amounts of Ni^{2+} causes a switch in helix handedness to form Z-DNA.

Although the non-canonical d(CGA) motif can adopt distinct structural conformations, the ps-duplex is the predominantly studied form.[33, 35-38] Originally described as **Π**-DNA (because **Π** specifically references parallel-stranded-ness), the $d(CGA)_n$ ps-duplex is stabilized by non-canonical homo-base pair interactions (C-CH+, G-G, and A-A) and inter-strand base stacking interactions[34, 38] (Figure 1.6).

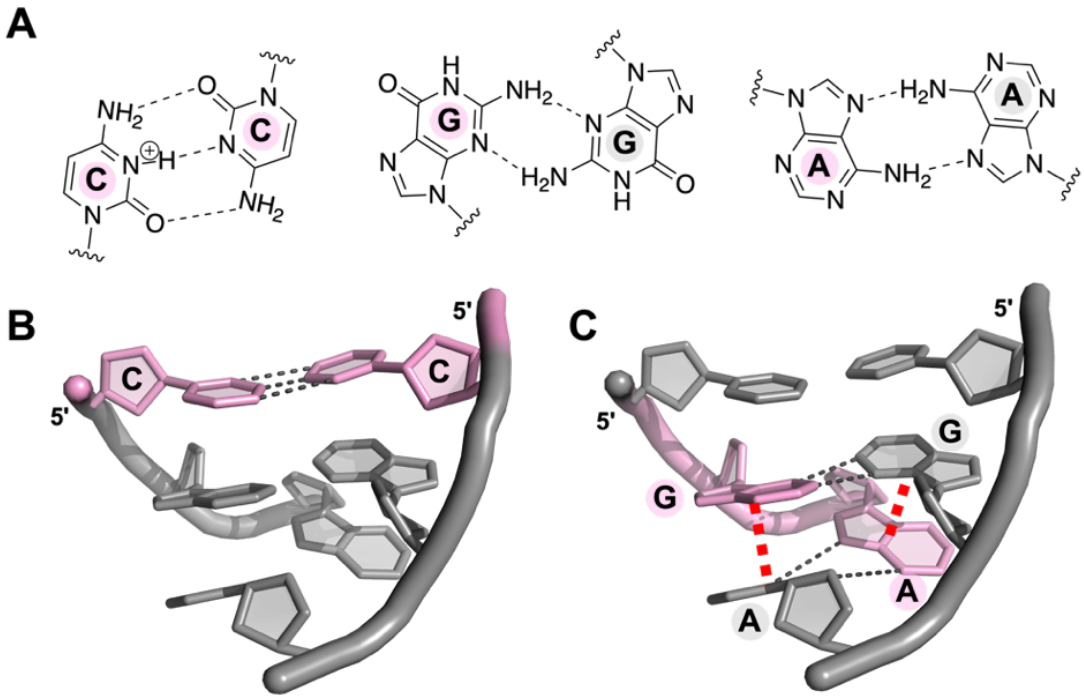

Figure 1.6. The d(CGA) triplet repeat motif in the ps-duplex form. (A) Homo-base pair interactions in the ps-duplex form. C-CH+ homo base pair requires hemi-protonation at the N3 position to form three hydrogen bonds along the Watson-Crick face. The G-G homo base pair forms via two sugar edge hydrogen bonds while the A-A homo base pair forms via two Hoogsteen face hydrogen bonds. **(B)** View of one d(CGA) triplet where the 5′C-CH+ homo base pair is highlighted in pink (PDB 7SB8). **(C)** View of one d(CGA) triplet emphasizing the G/A inter-strand base stacking interactions (PDB 7SB8). The G/A nucleotides in one strand are colored pink, while the G/A nucleotides in the opposing strand are gray. Homo-base pair hydrogen bonding is illustrated by the thin black dashed lines, while the inter-strand G/A stacking interactions are shown by thick red dashed lines.

The C-CH+ homo-base pair requires protonation at the N3 position of one of the two cytosines within the pair to form three hydrogen bonds along the Watson-Crick face.[34] N2-N3 sugar-edge hydrogen bonds stabilize G-G homo-base pairs, while A-A homo-base pairs are formed through N6-N7 Hoogsteen face hydrogen bonds. The GpA dinucleotide step provides significant stabilization to the ps-duplex by the formation of inter-strand G/A base stacking interactions where the G of one strand stacks on top of the A of the opposing strand while maintaining inter-strand homo-base pair interactions (Figure 1.6C).

The structure and stability of the ps-duplex is highly influenced by the 5′-nucleotide of each triplet.[30] Similar G/A-stacking interactions have been observed in ps-duplex structures containing d(GGA) or d(TGA) triplets (Figure 1.7)[35, 37, 39-41], though contiguous repeats of these sequences are unable to form ps-duplexes, unlike d(CGA) triplets.[30] A crystal structure containing both d(CGA) and internal d(TGA) triplets revealed that inter-strand G/A stacking is identical in the d(TGA) triplet and the d(CGA) motif, but there is increased structural asymmetry and duplex bending associated with the C-CH+ base pair.[36] Beyond the additional hydrogen bond interaction within each C-CH+ base pair, the structural details as to why asymmetric d(CGA) duplexes are significantly more stable than symmetric d(TGA) triplets remain unclear and will be addressed in further detail in Chapter 4.

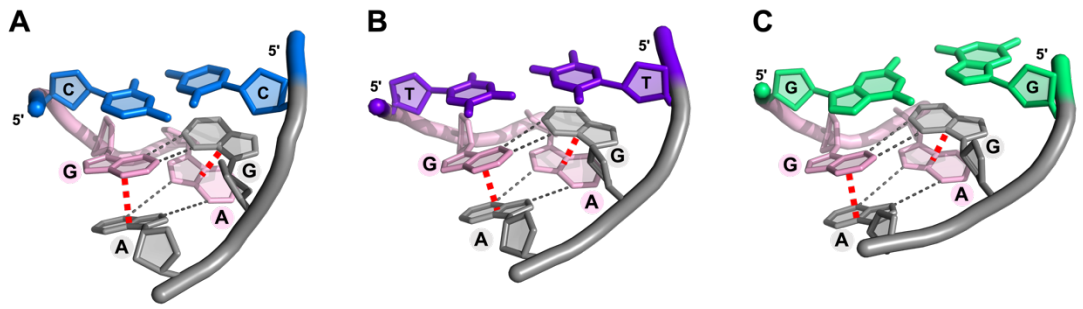

Figure 1.7. G/A stacking interactions are similar in d(CGA), d(TGA), and d(GGA) triplets. Inter-strand G/A stacking interactions are represented by the red dashed line between G of one strand (pink) and A of the opposing strand (gray). Homo-base pair hydrogen bonding is illustrated by the thin black dashed lines. The G/A stacking interactions are identical regardless of the 5′-capping nucleotide. **(A)** C-CH+ base pair is colored blue in the d(CGA) triplet from PDB: 7SB8. **(B)** T-T homo-base pair is colored purple in the d(TGA) triplet from PDB: 4RIM. **(C)** G-G homo-base pair is colored green in the d(GGA) triplet from PDB: 1P1Y.

Additionally, the same cross-strand G/A dinucleotide step seen in d(CGA) and d(TGA) was observed in 5′-d(GGA) sequences forming ps-duplexes both in solution and in the non-canonical motif of a continuously base paired three-dimensional (3D) DNA lattice (Figure 1.7).[40, 42] In these examples, the G-G base pair requires one of the two Gs to adopt a *syn* glycosidic angle to allow Watson-Crick to Hoogsteen face base pairing. The overall structural similarity of the d(CGA), d(TGA), and d(GGA) motifs in the ps-duplex form suggests that they could be used together to expand the sequence diversity of the d(CGA) triplet repeat motif.

Chapter 1.3. Biological relevance of non-canonical DNA structures

Alternative DNA conformations play important roles in the regulation of many biological functions including gene expression and chromosome stability.[43, 44] Interest in understanding the full repertoire of unusual DNA conformations has particularly expanded upon the discovery that non-canonical DNA structures could also be linked

to serious human diseases.[45, 46] Here I will briefly describe the biological implications of the non-canonical structures described in Chapter 1.2.

Chapter 1.3.1. G-quadruplexes and I-motifs

Since the 1960s there has been ample biochemical and structural analysis of the self-assembly of guanines into G4 structures *in vitro*[5, 47-49], whereas direct evidence for the presence of G4 structures *in vivo* is only beginning to be elucidated. Computational analysis used to estimate the position and frequency of G4 structures in the human genome have found that the location of such sequences is not random.[50] The most frequent occurrences were found in telomeric regions, gene promoters, and DNA replication origins.[50-52] The development of antibodies specifically targeting G4 structures lead to a significant breakthrough where computational predictions could be verified *in vivo.*[53, 54] Such studies have provided direct evidence of G4 formation in gene promoters and telomeric regions[53, 55, 56], suggesting that G4 structures are important in regulating gene expression and telomere maintenance. Since then, G4s have been shown to influence gene expression levels in yeast and human cells and affect transcription *in vivo.*[57-60] In particular, there is a significant presence of the G4 motif within several oncogene promotors including, c-MYC[61], pRb[62], Bcl-2[63], hTERT[64], and PDGF-A.[65] This suggests that G4 structures may play a functional role in some forms of cancer.[45]

Several years after recognizing the biological role of G4 structures *in vivo*, researchers began to consider that the i-motif, originally brushed aside as a structural oddity, could also have biological relevance.[21, 66] In 2017, Waller and colleagues used an algorithm initially designed to identify G4-forming sequences to locate potential i-

motif forming sequences in the human genome based on the theory that the complementary sequence of a G4-forming sequence is C-rich.[19] As such, this approach only uses the criterion that i-motif forming sequences must be C-rich and recognizes that other specific sequence requirements could be overlooked. This study in combination with another more specific bioinformatics search[18] revealed that such C-rich sequences are preferentially located at promoter regions, in introns, and in 5′- and 3′-UTRs, indicating that i-motifs may also play a significant role in gene expression.

Similarly to G4 structures, the development of an antibody with high binding affinity and specificity for i-motif structures led to verification of computational predictions *in vivo.*[67] In addition to visualizing i-motif structures in the nuclei of MCF7, U2OS, and HeLa cells, results from this study indicate that i-motif formation is highly cell-cycle dependent. The highest levels are observed in cell-cycle stages undergoing high levels of transcription and cellular growth.[67] This observation agrees with other studies which show that i-motif structures play an important role in transcription regulation of several genes including BCL2[68-70], PDGFRβ[71], and c-MYC.[72, 73]

Chapter 1.3.2. Triplet repeat sequences

In general, triplet (or trinucleotide) repeat sequences consist of three nucleotides consecutively repeated over long stretches of DNA. Expansions of these repeat sequences have been connected to several neurological disorders including Fragile-X syndrome[74-76], Huntington's disease[77, 78], myotonic dystrophy[79] and Friedrich's ataxia.[80, 81] There is evidence indicating that triplet repeat sequences forming non-B DNA structures play an important role in promoting such disease pathologies.[82] One of several recognized mechanisms for the expansion of trinucleotide

repeat sequences associated with hereditary diseases is via the formation of alternative DNA structures (i.e. hairpins, G4s, triplexes) during DNA replication.[83, 84] Specifically, the alternative DNA structures impede and stall replication machinery leading to template strand misalignment. Subsequent resumption of DNA synthesis from the misaligned template results in triplet repeat sequence expansion, disruption of gene expression, and the onset of disease.[83]

Genomic analysis of all triplet repeat sequences, identified as tracts of ≥ 6 tandem triplet repeat units, has revealed an unequal frequency and representation of various triplet repeats sequences in the human genome[85]. Specifically, the $(CAG)_n$ triplet repeat is one of the most frequently identified triplet repeat motifs in the human genome (1055 occurrences), and its expansion has been clearly linked with disease pathologies[86] while the $(CGA)_n$ triplet repeat is one of the least-frequently represented triplets in the genome (16 occurrences) and is not known to be associated with disease.[85] A similarly low frequency and coverage of d(CGA) triplets was seen when a comparable genomic analysis was performed in other eukaryotic organisms.[87] The ability of d(CGA) to adopt distinct structural forms appears to be a trait shared by several other triplet repeat motifs, though d(CGA) is the only triplet known to form perfectly ps-duplex structures.[30, 31, 88, 89]

The uneven representation of triplet repeats in the human genome and association with disease pathologies is largely hypothesized to be related to the ability of a repeat sequence to adopt conformations beyond the traditional double helix and the relative stabilities of such alternative structures.[82, 88, 90] Understanding the structure and dynamics of non-canonical structures formed from repetitive sequences will

provide important information on how these structures can impact biological function and potentially lead to disease states. Therefore, it is important to characterize the structure and stability of triplet repeat sequences, such as $d(CGA)_n$, that can form such non-canonical structures but might not be directly connected to the onset of disease.

Chapter 1.4. Non-canonical DNA structures in nanotechnology

In the 1980s, Dr. Ned Seeman recognized that the inherent programmability, stability, rigidity, self-assembly, and predictability of DNA, evolutionarily optimized for storage of genetic information, could be leveraged in synthetic applications. This realization gave way to the field of DNA nanotechnology where DNA is repurposed as a building block in the creation of nanoscale objects and assemblies.[91] The physical structure of Watson-Crick duplex DNA (2 nm diameter, 3.4 nm helical pitch, and ~150 bp persistence length[92]) offers structural predictability and rigidity on the nanoscale. In addition to desirable structural properties, DNA can self-assemble due to hydrogen bonding and base stacking interactions which are intrinsic to complementary double stranded DNA. Watson-Crick base pairing interactions (A-T and G-C) contribute additional structural programmability and predictability. These features have made the WC duplex a valuable building block in programmed self-assembly of DNA based nanoarchitectures. Several classes of 2D and 3D architectures have been built using the WC double helix including, 2D tile and crystal arrays[93-95], DNA origami[96, 97], 3D objects[98-100], crystals[101], and supramolecular assemblies.[102-104]

The repertoire of nanostructures designed primarily based upon the WC double helix are commonly limited by the structural stability and rigidity of the B-form duplex, along with rapid degradation by nucleases. Again, by turning to biology for inspiration,

researchers have been able to overcome these limitations by incorporating non-canonical DNA interactions into rational nanoarchitecture design. Non-canonical interactions are valuable because they retain the desirable inherent characteristics of DNA, but their unique structural diversity can vastly expand the geometries and topologies of designed architectures. Another valuable advantage of non-canonical interaction-based structures are their sensitivities to the local environment, including cations, salt concentration, or pH.[23, 105-108] This provides a valuable opportunity to expand the functional diversity of DNA architectures, where programmed incorporation of non-canonical motifs will give way to predictable and highly controllable structural changes in response to environmental perturbations (as described in Chapter 1.2).

Previous efforts to design DNA architectures using non-canonical interactions are primarily limited to motifs known to be biologically relevant, including G-quadruplexes[109-112], i-motifs[113-116], and triplex forming strands.[117-119] We are interested in expanding the diversity of functional, non-canonical structures used in DNA nanotechnology by studying biophysical and structural characteristics of the pH-sensitive non-canonical d(CGA) triplet repeat motif.

Chapter 1.5. Motivation and scope of this study

The goal of this research is to characterize the structure, thermodynamic stability, nuclease resistance, and dynamic nature of the non-canonical pH-sensitive d(CGA) triplet repeat motif and related variants. Biophysical and structural characterization of this motif will directly benefit the field of DNA nanotechnology by diversifying the repertoire of functional non-canonical DNA sequence motifs that can

be used in the rational design of DNA architectures. This non-canonical motif will be particularly advantageous to applications that could benefit from pH-triggered structural switching as it can adopt either a ps-duplex or anti-parallel hairpin depending on solution pH. The results of this study could also be used to inform the biological understanding of triplet repeat expansions, specifically, why $(CGA)_n$ is less represented in the genome compared to other similar triplet variations.

In this dissertation, I will discuss thermodynamic parameters obtained from UV melting experiments, nuclease resistance against DNase I and S1 nuclease, circular dichroism-based kinetic studies, and structural analysis of two x-ray crystal structures of d(CGA)-based sequences in the ps-duplex form. Together, our biophysical and structural characteristics of the d(CGA) triplet repeat motif and related variants will provide a foundation for the use of this non-canonical motif in DNA nanotechnology applications. I will also present our efforts to optimize a 3D DNA crystal, containing non-canonical d(GGA) interactions, as a drug delivery vehicle.

Chapter 2: Thermodynamic Stability of d(CGA)-based Triplet Repeat Sequences

*This chapter is derived from the following manuscript:
Luteran, E. M.; Kahn, J. D.; Paukstelis, P. J. (2020). Stability of the pH-Dependent Parallel-Stranded d(CGA) Motif. *Biophys. J.* **119**:1580-1589.

Chapter 2.1. Introduction

Chapter 2.1.1. Overview

In this chapter, I describe our analysis of the thermodynamic stability of oligonucleotides composed of the d(CGA) triplet motif along with several structurally related sequence variants. Our results show that the structural transition resulting from decreasing the pH is accompanied by a significant energetic stabilization as unimolecular hairpin structures are converted to parallel-stranded homo-base paired duplexes (ps-duplexes). In addition, we found that the stability of the ps-duplex form can be tuned by the identity of the 5′-nucleotide within each triplet and the position and frequency of altered triplets within stretches of d(CGA) triplets. This study offers insight into the stability of the d(CGA) triplet repeat motif and provides constraints for using this pH-responsive structural motif for creating DNA-based nanomaterials.

Chapter 2.1.2. Sequence and variant design

We designed several d(CGA)-based repeat sequences to investigate how triplet identity and position within repeat sequences impacts thermodynamic stability in response to pH (Table 2.1). Variant sequences were designed by modifying the $(CGA)_6$ parent sequence to incorporate triplets for which structural information was available (d(TGA) or d(GGA))[36, 37, 40] or by adjusting the $d(CGA)_n$ repeat number (n).

Table 2.1. Sequences of DNA Oligonucleotides.

DNA Name	Sequence
$(CGA)_7$	d(CGA-CGA-CGA-CGA-CGA-CGA-CGA)
$(CGA)_6$	d(CGA-CGA-CGA-CGA-CGA-CGA)
$(CGA)_5$	d(CGA-CGA-CGA-CGA-CGA)
$(CGA)_4$	d(CGA-CGA-CGA-CGA)
$TGA(CGA)_5$	d(TGA-CGA-CGA-CGA-CGA-CGA)
$(CGA)_5TGA$	d(CGA-CGA-CGA-CGA-CGA-TGA)
$(CGATGA)_3$	d(CGA-TGA-CGA-TGA-CGA-TGA)
$(TGA)_6$	d(TGA-TGA-TGA-TGA-TGA-TGA)
$GGA(CGA)_5$	d(GGA-CGA-CGA-CGA-CGA-CGA)
$(GAC)_6$	d(GAC-GAC-GAC-GAC-GAC-GAC)

*Underlined nucleotides differ from the $(CGA)_n$ pattern.

$(CGA)_n$ triplets with various repeat number (n) were created to test how the addition of triplets of the same composition impact stability. $TGA(CGA)_5$, and $GGA(CGA)_5$ variants were synthesized to probe the stabilizing contribution of the 5′-most capping triplet in ps-duplexes while $(CGA)_5TGA$, $TGA(CGA)_5$, $(CGATGA)_3$, and $(TGA)_6$ sequences were designed to test the thermodynamic compatibility of d(CGA) and d(TGA) triplets. We hypothesized that the conservation of the 5′-pyrimidine among d(CGA) and d(TGA) triplets would incur minimal structural penalties, as opposed to incorporation of a bulkier 5′-purine (Figure 1.7). Finally, $(GAC)_6$ allowed a direct comparison with $(CGA)_6$ to examine how a subtle terminal sequence permutation impacts thermodynamic stability.

Chapter 2.2. Results and Discussion

Chapter 2.2.1. d(CGA) sequences adopt unique structures in response to pH

Previous studies demonstrated that $d(CGA)_2$ and $d(CGA)_4$ repeat sequences could adopt an anti-parallel form at pH 7.0 or parallel-stranded duplex form at pH

5.5.[32-34] To determine if this behavior was also found for the parent sequence of this study, $(CGA)_6$, we analyzed the oligonucleotide by circular dichroism (CD) spectroscopy and UV absorbance melting. The CD spectrum for $(CGA)_6$ at pH 5.5 showed a prominent positive band at 265 nm and a negative band at 245 nm, consistent with parallel-stranded homo-duplex formation (Figure 2.1A).[32, 120] A two-state UV absorbance melting transition (in which only two populations of molecules exist: folded or unfolded) was observed for $(CGA)_6$ at pH 5.5, with a concentration-dependent melting temperature (T_m) indicating a reaction molecularity greater than one (Figure 2.1B). This observation is consistent with the parallel bimolecular complexes seen in crystal and solution structures containing the d(CGA) motif.[32, 35, 120, 121] At pH 7.0, the measured T_m was independent of concentration, indicating a unimolecular structure, and the CD spectrum was characteristic of anti-parallel strands (Figure 2.1B). These results are consistent with $(CGA)_6$ forming anti-parallel hairpin structures at neutral pH and parallel-stranded homo-duplexes under acidic conditions.

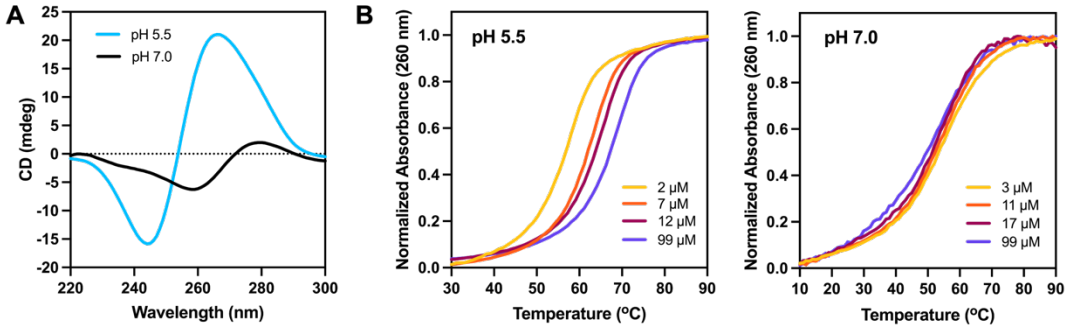

Figure 2.1. $(CGA)_6$ adopts a parallel-stranded duplex or anti-parallel hairpin in response to pH. (A) Shown are CD spectra of 10 µM $(CGA)_6$ at pH 5.5 and pH 7.0. At pH 5.5, the prominent positive band at 265 nm and negative band at 245 nm are consistent with parallel-stranded duplex formation. At pH 7.0, the positive band at 280 nm and negative band at 260 nm are characteristic of anti-parallel strands[120]. **(B)** Normalized temperature versus absorbance curves for $(CGA)_6$ show concentration-

dependent T_m at pH 5.5 and concentration-independent T_m at pH 7.0, suggesting bi-/multimolecular and largely unimolecular structure formation, respectively.

The pH-dependent structural forms of d(CGA)-based variant sequences were also analyzed by CD and UV thermal melting. All variants except $(TGA)_6$ and $(CGATGA)_3$ had CD spectra consistent with the pH-dependent structures seen in $(CGA)_6$ (Figure 2.2). Interestingly, CD spectra for $(CGATGA)_3$ and $(TGA)_6$ had bands characteristics of parallel-oriented strands at both pH 5.5 and 7.0, but the intensity of the signal was significantly weaker for $(TGA)_6$. This suggests a decreased pH sensitivity with increasing d(TGA) triplet content, likely due to the absence of pH-dependent C-CH+ homo-base pairs.

All variants except $(TGA)_6$ had concentration-dependent, two-state melting curves at pH 5.5, indicating bimolecular interactions (Figure 2.3). $(CGA)_6$ was the only oligomer that exhibited clear two-state melting at pH 7.0, with all other sequence variants having an apparent minor lower-temperature melting transition between 20°C and 40°C. The multistate transitions at pH 7.0 may reflect populations of hairpin and anti-parallel hairpin duplex structures, which are also hypothesized to form in other triplet repeat sequences.[31]

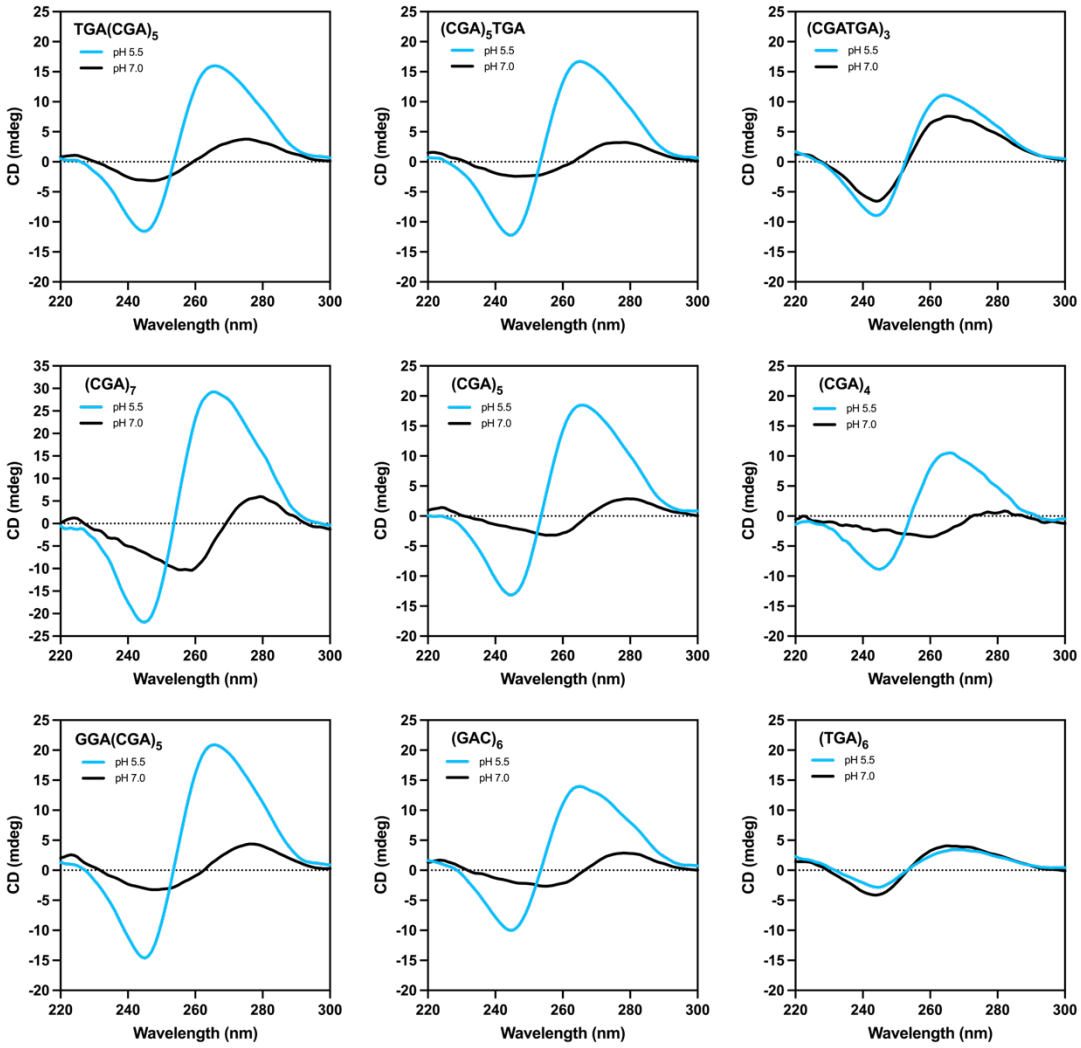

Figure 2.2. CD spectra of d(CGA)-based variant sequences. Shown are the CD spectra of d(CGA)-based variant sequences tested at pH 5.5 (blue) and pH 7.0 (black). The prominent positive band at ~265 nm and the negative band at ~245 nm are characteristic of parallel-stranded duplex formation.[32, 120] As the number of d(CGA) triplets is decreased, the intensity of the bands associated with parallel-stranded duplexes also decreases. The positive band at ~280 nm and negative band at ~260 nm are characteristic of anti-parallel oriented strands.[32, 120] $(CGATGA)_3$ and $(TGA)_6$ are the only variants that do not exhibit pH dependent structural changes, suggesting that pH sensitivity can be tuned by the incorporation of d(TGA) triplets.

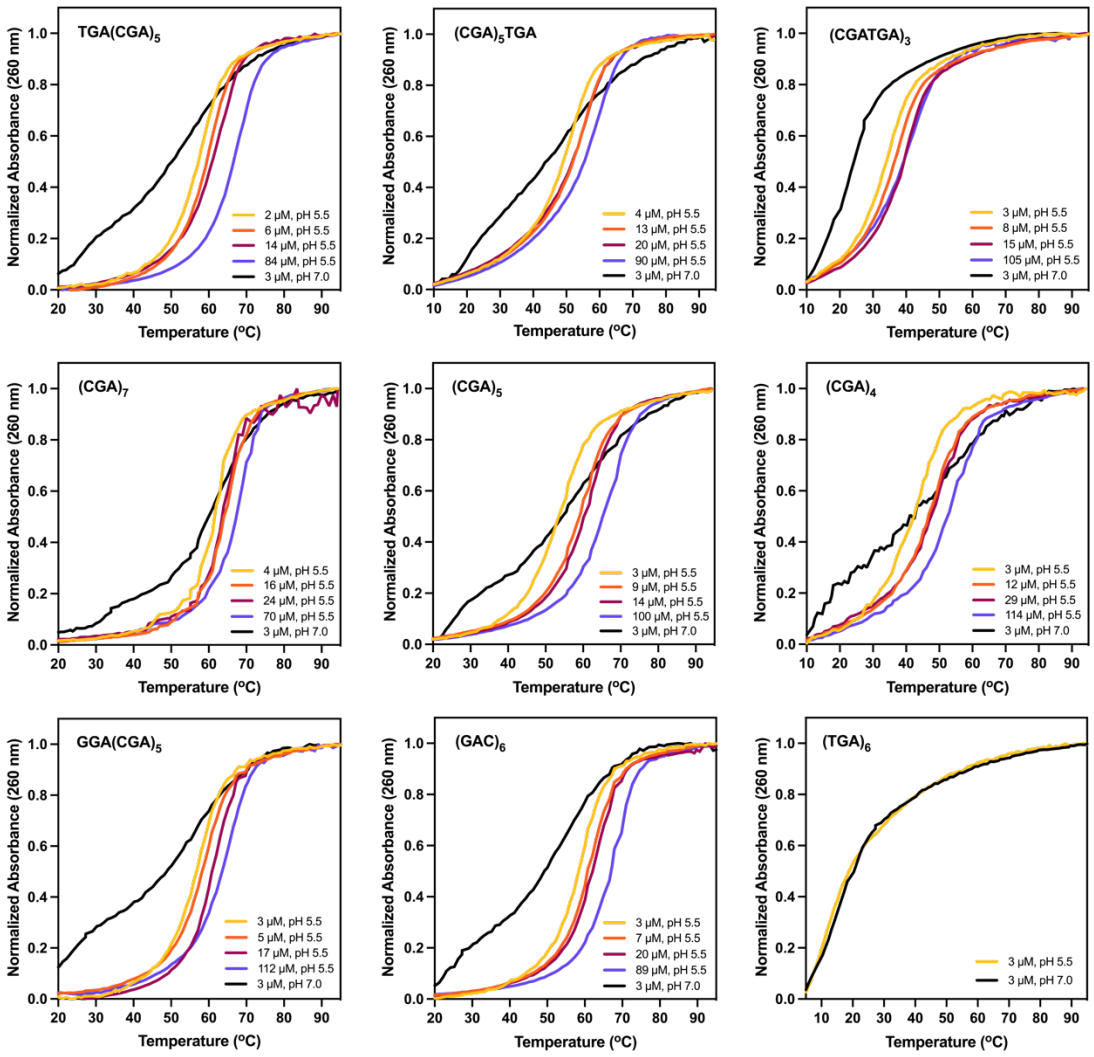

Figure 2.3. Normalized temperature versus absorbance curves. UV absorbance melting curves for $(CGA)_n$ variants at pH 5.5 (yellow, orange, magenta, purple) and pH 7.0 (black). All $(CGA)_n$ variants, except $(TGA)_6$, have concentration-dependent, two-state melting curves at pH 5.5, indicating bimolecular interactions. At pH 7.0, all variants exhibit non-two-state melting with apparent minor lower-temperature melting transitions between 20-40°C.

Chapter 2.2.2. Stability of the ps-duplex and anti-parallel hairpin forms

$(CGA)_6$ and variants were analyzed by UV melting to establish thermodynamic parameters for the conformations observed at each pH (Table 2.2). Thermodynamic parameters for parallel duplex (pH 5.5) or hairpin (pH 7.0) formation were extracted

from UV melting curves fit to two-state bimolecular or unimolecular models, respectively. For each data set, the entire absorbance versus temperature curve was fit with the van't Hoff equation assuming constant enthalpy (ΔH^o) and entropy (ΔS^o) changes, as described in Chapter 2.4.4.

Table 2.2. Thermodynamic parameters for $(CGA)_6$ and variants at pH 5.5 and 7.0.

Thermodynamic Parameters, pH 5.5*				
Sequence	**T_m (^{o}C)****	**ΔG^o_{37} (kcal/mol)**	**ΔH^o (kcal/mol)**	**ΔS^o (e.u.)**
$(CGA)_7$	61	-15.6 ± 0.9	-107.9 ± 7.2	-298 ± 21
$(CGA)_6$	60	-14.2 ± 0.7	-92.1 ± 3.1	-251 ± 9
$(CGA)_5$	54	-11.8 ± 0.4	-75.8 ± 1.4	-207 ± 4
$(CGA)_4$	43	-8.9 ± 0.2	-60.9 ± 3.5	-168 ± 11
$TGA(CGA)_5$	58	-13.4 ± 0.5	-85.6 ± 2.4	-233 ± 7
$(CGA)_5TGA$	49	-10.3 ± 0.3	-65.8 ± 2.2	-179 ± 7
$(CGATGA)_3$	32	-6.9 ± 0.3	-59.7 ± 3.2	-171 ± 10
$(TGA)_6$	NA	NA	NA	NA
$GGA(CGA)_5$	56	-13.2 ± 0.7	-91.2 ± 2.6	-252 ± 8
$(GAC)_6$	58	-13.3 ± 0.4	-85.8 ± 2.3	-234 ± 7
Thermodynamic Parameters, pH 7.0***				
Sequence	**T_m (^{o}C)****	**ΔG^o_{37} (kcal/mol)**	**ΔH^o (kcal/mol)**	**ΔS^o (e.u.)**
$(CGA)_6$	55	-2.0 ± 0.2	-33.5 ± 1.1	-102 ± 3

*Thermodynamic parameters obtained at pH 5.5 are reported as averages of van't Hoff analysis curve fitting results assuming bimolecular duplex formation, with oligomer concentrations between 1 and 120 x 10^{-6} M. Errors in ΔH^o and ΔS^o are reported as one standard deviation from at least four concentrations, and error in ΔG^o is propagated as previously described.[122-124]

**T_m values are calculated for oligomer concentration of 3 x 10^{-6} M.

***Thermodynamic parameters and T_m for pH 7.0 are reported as the average of four van't Hoff curve fit data sets, assuming unimolecular hairpin formations, with oligomer concentrations between 1 and 100 x 10^{-6} M.

Although $(CGA)_6$ exhibited apparent two-state melting profiles at both pH 5.5 and 7.0 and had similar T_ms in this concentration range, detailed examination of the pH 7.0 melting curves (Figure 2.1B) showed that there are significant underlying differences in the stabilities of the two structures. The higher baseline slope of the anti-parallel form at pH 7.0 suggests that its structure may be changing as a function of

temperature even at low temperature. The decrease in the slope at T_m as the concentration increases (i.e., an apparent decrease in the magnitudes of ΔH^o and ΔS^o) suggests some bimolecular or multimolecular behavior at pH 7.0. Comparing the thermodynamic parameters of $(CGA)_6$ determined at pH 5.5 to pH 7.0, the standard state formation free energy (ΔG^o_{37}) and ΔH^o were destabilized by 12.2 and 58.6 kcal/mol, respectively; the similar T_ms arise from the difference between unimolecular interactions at very high effective local concentration versus much lower concentrations of bimolecular interactions. This suggests that the parallel-stranded duplex form containing intermolecular G/A stacking and hydrogen bonding interactions provides substantial structural stabilization versus the anti-parallel hairpin.

Chapter 2.2.3. The identity of the 5′-nucleotide impacts ps-duplex stability

We compared ΔG^o_{37} for 18-nt variants with different 5′-terminal capping triplets for which structural information was available (d(CGA), d(TGA), or d(GGA))[36, 37, 40] to the 15-nt $(CGA)_5$ to understand how the identity of the 5′-terminal triplet contributed to thermodynamic stability of the parallel-stranded duplex (Table 2.3). The 5′-d(GGA) triplet ($\Delta\Delta G^o_{37}$ = –1.4 kcal/mol with respect to $(CGA)_5$) and the 5′-d(TGA) triplet (–1.6 kcal/mol with respect to $(CGA)_5$) showed similar contributions, both being less stabilizing than the 5′-d(CGA) triplet, estimated to contribute –2.4 kcal/mol. Assuming each 5′-triplet contains similar G/A inter-strand stacking interactions (Figure 1.7), the differences in ΔG^o_{37} between d(CGA), d(GGA), and d(TGA) triplets can be attributed primarily to the identity of the terminal 5′-nucleotide.

Table 2.3. $\Delta\Delta G^{o}_{37}$ values for $(CGA)_6$ and variants compared to $(CGA)_5$ at pH 5.5.

*Thermodynamic Parameters, pH 5.5**		
Sequence	**ΔG^{o}_{37} (kcal/mol)**	**$\Delta\Delta G^{o}_{37}$ (kcal/mol)****
$(CGA)_7$	-15.6 ± 0.9	-3.8 ± 1.0
$(CGA)_6$	-14.2 ± 0.7	-2.4 ± 0.8
$(CGA)_5$	-11.8 ± 0.4	0
$(CGA)_4$	-8.9 ± 0.2	+2.9 ± 0.4
$TGA(CGA)_5$	-13.4 ± 0.5	-1.6 ± 0.7
$(CGA)_5TGA$	-10.3 ± 0.3	+1.4 ± 0.5
$(CGATGA)_3$	-6.9 ± 0.3	+4.9 ± 0.5
$(TGA)_6$	NA	NA
$GGA(CGA)_5$	-13.2 ± 0.7	-1.4 ± 0.8
$(GAC)_6$	-13.3 ± 0.4	-1.5 ± 0.6

*Thermodynamic parameters obtained at pH 5.5 are reported as averages of van't Hoff analysis curve fitting results assuming bimolecular duplex formation, with oligomer concentrations between 1 and 120 x 10^{-6} M.

**$\Delta\Delta G^{o}_{37}$ for each sequence calculated with respect to $(CGA)_5$. Negative values represent a stabilized free energy of formation, whereas positive values represent destabilized free energy of formation.

The small change in stability between the 5′-terminal d(GGA) and d(TGA) suggest that the 5′-G and 5′-T are thermodynamically comparable in the parallel duplex. Interestingly, the entropy of formation for $GGA(CGA)_5$ and $(CGA)_6$ is considerably lower than for $TGA(CGA)_5$ (Table 2.2), which could result from conformational variability introduced by a G residue in the *syn* conformation as seen in crystal structures containing the parallel-stranded d(GGA) motif (Figure 1.7). The destabilization caused by the 5′-T can also be attributed to enthalpic destabilization (–92.1 to –85.6 kcal/mol), reflecting the loss of one hydrogen bond by replacing the hemi-protonated C-CH+ with the T-T base pair. Additional electronic effects from the cationic C-CH+ base pair that are not found in the T-T or G-G base pairs may also contribute.

Though the 5′-d(GGA) triplet appears compatible with 5′-d(CGA) in the parallel-stranded duplex, incurring only a slight thermodynamic destabilization, our preliminary modeling suggests that longer stretches of internal d(GGA) triplets would not be well accommodated because of structural clashes that would arise from an internal G_{syn} residue.[40, 42] CD data further suggest that tandem d(GGA) triplets are not pH sensitive and do not form the expected structures at either pH (Figure 2.4). Together, these suggest that oligonucleotides solely comprised of d(GGA) triplets are most likely unstructured or weakly associated under the conditions examined.

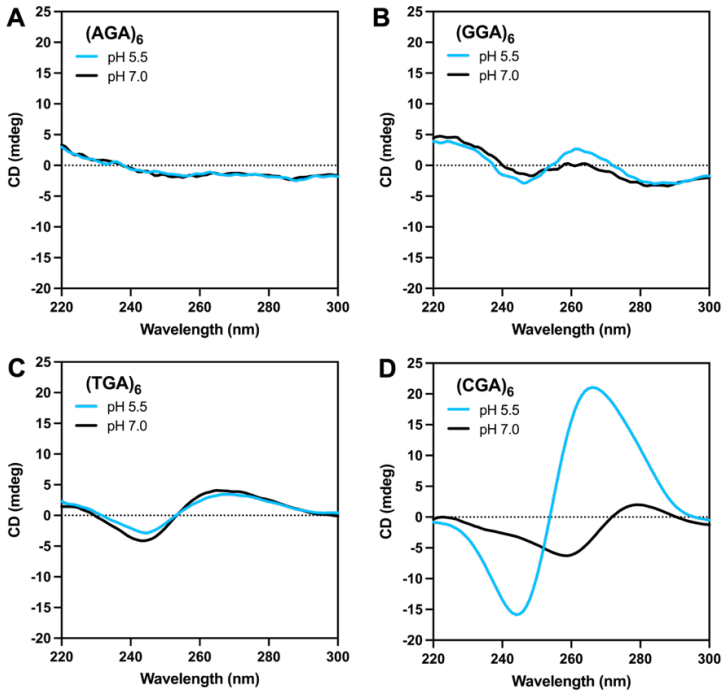

Figure 2.4. CD spectra of $(NGA)_6$ sequences. (A) $(AGA)_6$ does not have CD bands characteristic of parallel- or anti-parallel oriented strands and is not pH dependent. The lack of CD signal could suggest a lack of structure at each pH. **(B)** $(GGA)_6$ also does not have CD bands characteristic of parallel or anti-parallel oriented strands and is not pH dependent. **(C)** $(TGA)_6$ has weak CD signal characteristic of the parallel-stranded form at each pH, suggesting a lack of pH-dependent structural sensitivity. **(D)** $(CGA)_6$ is the only $(NGA)_6$ sequence that exhibits a pH-dependent structural change. At pH 5.5 the CD signal is characteristic of parallel-oriented strands, while at pH 7.0 the CD signal is characteristic of anti-parallel oriented strands.

Notably, the ΔG^{o}_{37} for $(GAC)_6$ (–13.3 kcal/mol) at pH 5.5 was different than $(CGA)_6$ (–14.2 kcal/mol), though d(GAC) is a simple permutation of d(CGA). The structural basis for this difference is not immediately clear as both sequences have CD spectra characteristic of parallel-oriented strands (Figure 2.5A). The variation in thermodynamic stability suggests that the identity of the homo-base pair interactions at the 5′- or 3′- terminus could play an important role in parallel-stranded duplex nucleation and stabilization. Nucleotides providing strong intermolecular interactions could help to lock the duplex in place to avoid premature end fraying, contributing to enhanced overall thermodynamic stability.

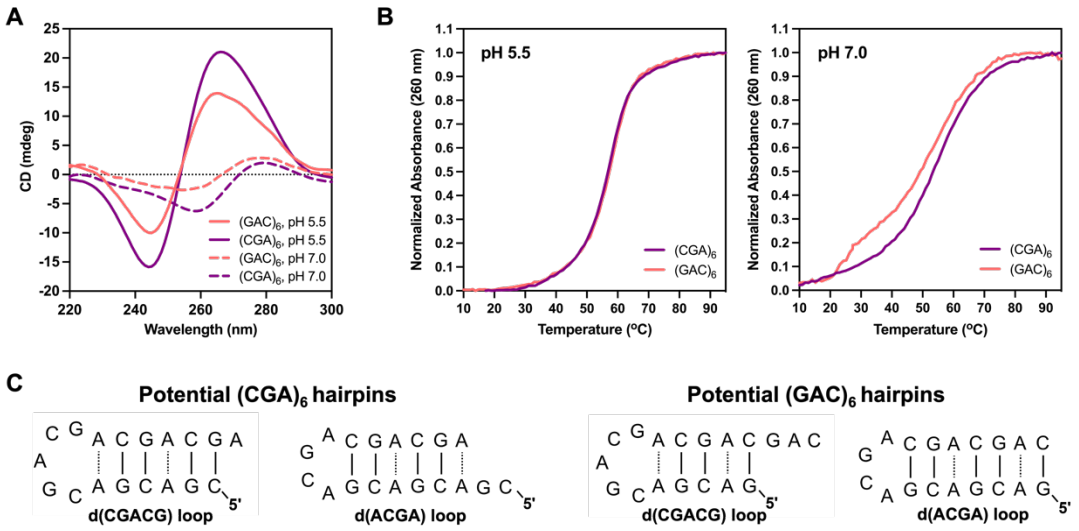

Figure 2.5. $(CGA)_6$ and the sequence permutation $(GAC)_6$ (A) CD spectra of $(CGA)_6$ and $(GAC)_6$ at pH 5.5 and pH 7.0. $(GAC)_6$ exhibits a structural pH-dependence that is similar to $(CGA)_6$, but the intensity of the CD bands corresponding the parallel-stranded duplex form are less intense. This could suggest that the 5′ or 3′ terminal homo-base pairs play an important role in maintaining the parallel-stranded duplex structure. **(B)** UV melting curves comparing 3 μM $(CGA)_6$ and $(GAC)_6$ at pH 5.5 (left) and pH 7.0 (right). Different melting curves at pH 7.0 suggest that the anti-parallel hairpin forms are not identical, though the two sequences are permutations. The multi-state transition in $(GAC)_6$ at pH 7.0 suggests the presence of multiple populations. **(C)** Potential hairpin forms for $(CGA)_6$ and $(GAC)_6$. Solid lines represent canonical Watson-Crick base pairs, while dashed lines represent non-canonical base pairing. The multi-state melting transition for $(GAC)_6$ at pH 7.0 could result from the dynamic interconversion between each hairpin shown.

Additionally, it was unclear why $(GAC)_6$ behaved differently than $(CGA)_6$ at pH 7.0 (Figure 2.5B). We suspected that the difference could arise from variations in hairpin loop and stem size, as well as the likelihood of interconversion between different hairpin forms (Figure 2.5C). The $(CGA)_6$ hairpin could form a 6-bp stem where a d(CGACG) loop is capped by a A-A homo-base pair or a d(ACGA) loop capped by a G-C base pair. The two-state transition seen at pH 7.0 could represent one of these two distinct structures. The non-two state melt and slight destabilization seen

for $(GAC)_6$ could arise from dynamic interconversion between two hairpin forms with the same loop compositions as $(CGA)_6$, but contain distinct hairpin stems – a d(CGACG) loop only accommodating a 5-bp stem or a d(ACGA) loop accommodating a 7-bp stem. There is also the possibility that the minor low temperature melting transitions could be a result of higher order multi-molecular interactions.

Chapter 2.2.4. d(TGA) triplet frequency and position impacts duplex stability

The identity of the 5′-pyrimidine in the d(CGA) triplet motif leads to subtle but significant structural changes in the parallel-stranded duplex, evident in crystal structures containing d(CGA) and d(TGA) repeats.[36, 37] The C-CH+ base pair induces a structural asymmetry, likely resulting from a hydrogen bond between N4 of one cytosine to the nonbridging phosphate oxygen of the previous interchain adenosine[36], discussed further in Chapter 4. In contrast, the T-T base pair retains duplex symmetry at the expense of stacking interactions. To determine if the subtle structural differences between d(TGA) and d(CGA) triplets corresponded to differences in thermodynamic stability, we tested several sequence variants that incorporate d(TGA) triplets. Our results showed that the frequency and position of d(TGA) triplets within $(YGA)_6$ oligomers had a significant impact on the thermodynamic stability of the parallel-stranded duplex (Figure 2.6).

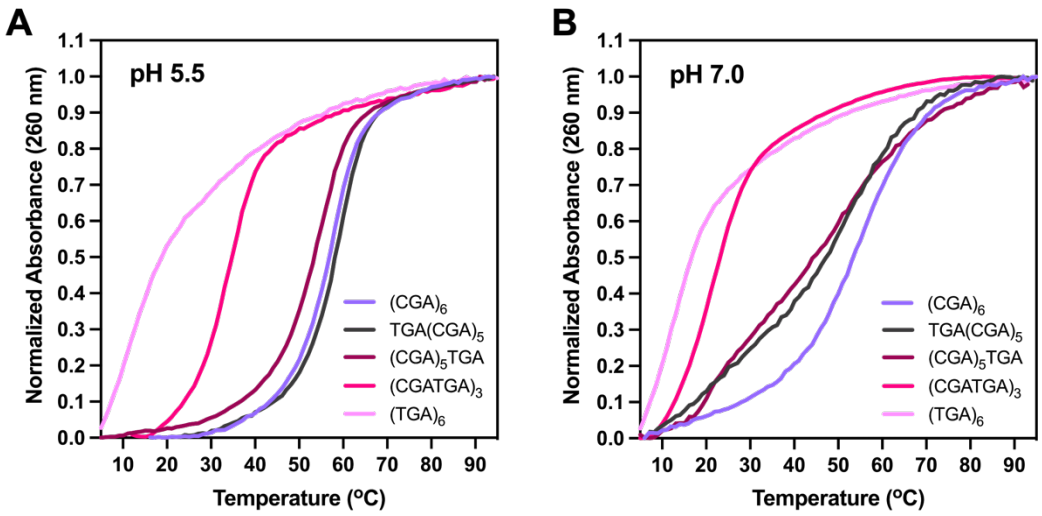

Figure 2.6. UV melting curves for $(YGA)_6$ oligomers. The addition of d(TGA) triplets decreases the T_m of $(YGA)_6$ oligomers. Shown are normalized temperature versus absorbance curves for $(YGA)_6$ oligomers with d(TGA) triplets (n = 6, 3, 1, and 0). **(A)** At pH 5.5, all variants except $(TGA)_6$ exhibit two-state melting. **(B)** At pH 7.0, $(CGA)_6$ is the only variant to retain two-state melting. $(CGA)_5TGA$ and $TGA(CGA)_5$ have broad melting curves with multiple transitions, whereas $(TGA)_6$ and $(CGATGA)_3$ do not have clear melting transitions, suggesting a lack of stable structure formation.

At pH 5.5, $(CGA)_6$ had a clear two-state melting transition, indicating structure formation when the sequence was completely comprised of d(CGA) triplets. In contrast, $(TGA)_6$ had no evident melting transition at pH 5.5 or 7.0, suggesting that sequences comprised solely of d(TGA) triplets are unable to adopt stable structures at either pH. This indicated that the interactions formed by the T-T base pair were either not strong enough to nucleate stable structure formation or were transient.

We tested the destabilizing effects associated with position and number of d(TGA) triplets in a given $(YGA)_6$ sequence using the sequence variants $TGA(CGA)_5$, $(CGA)_5TGA$, $(CGATGA)_3$, and $(TGA)_6$. As described above (Chapter 2.2.3), in the parallel duplex form, the addition of a d(TGA) triplet to the 5′-end stabilized the free energy of formation by –1.6 ± 0.7 kcal/mol compared with $(CGA)_5$ (Table 2.2). When

the d(TGA) triplet was placed at the 3′-end (variant sequence $(CGA)_5TGA$), we observed a corresponding destabilization of free energy of duplex formation by 1.4 ± 0.7 kcal/mol compared with $(CGA)_5$. This indicated that the position of the d(TGA) triplet significantly impacted the thermodynamic stability of the parallel-stranded duplex, likely due to flexibility associated with the T-T base pair. When the T-T homo-base pair was positioned internally within the 3′ triplet, there was a significant impact on duplex stability, presumably through disruption of surrounding 5′- and 3′- G-G and A-A homo-base pairs (Figure 2.7). This contrasts with positioning the d(TGA) triplet at the 5′-end, where the T-T homo-base pair would only disrupt 3′- G-G and A-A homo-base pairs. Correspondingly, more substantial destabilization (7.3 ± 0.8 kcal/mol compared to $(CGA)_6$) was observed when multiple d(TGA) triplets were added as in $(CGATGA)_3$ where there are multiple sites for T-T induced disruption of surrounding base pairs.

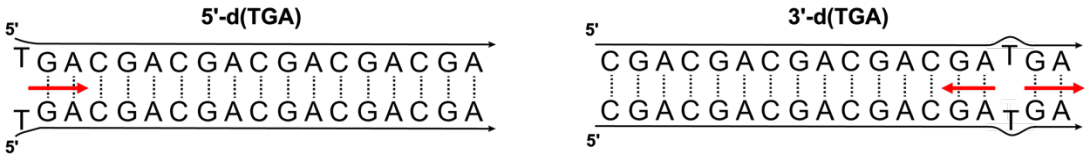

Figure 2.7. d(TGA) position affects the surrounding base pair interactions. When the d(TGA) triplet is positioned at the 5′-end (left), flexibility of the T-T homo-base pair could only disrupt 3′-down strand base pairs, as indicated by the red arrow. This is in contrast with positioning the T-T homo-base pair internally by incorporating the d(TGA) triplet at the 3′-end where flexibility associated with the T-T base pair could disrupt 5′-up-strand or 3′-down-strand base pairs, as indicated by the red arrows.

Although the d(TGA) triplets decrease thermodynamic stability, they have the potential to modulate pH sensitivity. CD results suggested that the presence of increasing d(TGA) triplets corresponds to decreased pH sensitivity (Figure 2.8).

d(TGA) triplets can be combined with other strong parallel-strand inducing triplets, such as d(GGA) or d(CGA), potentially providing a unique system for tuning structure formation over specific pH ranges. Information on the thermodynamic stability of sequences containing tandem d(YGA) triplets can be combined with structural data (Chapter 4) to provide valuable insight in optimizing parallel-stranded duplexes for 3D DNA crystal design or other nanoscale architectures.

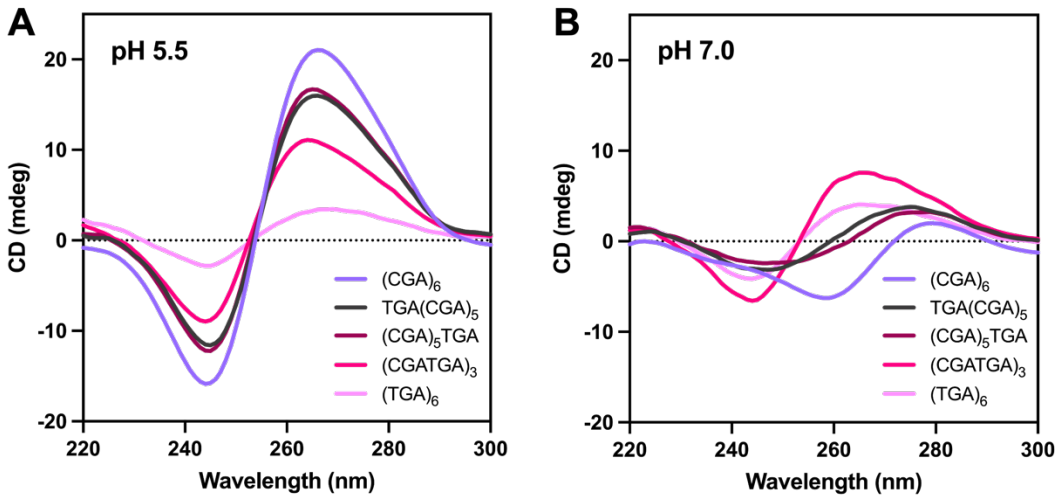

Figure 2.8. CD spectra of $(YGA)_6$ oligomers. (A) At pH 5.5, all oligomers have CD spectra with bands characteristic of parallel-oriented strands. As the number of d(TGA) triplets increase within a sequence, the corresponding intensity of the CD bands decrease. **(B)** At pH 7.0, the CD spectra does not drastically change for $(TGA)_6$ and $(CGATGA)_3$, compared to pH 5.5, suggesting a decrease in pH sensitivity as the number of d(TGA) triplets increases within a sequence.

Chapter 2.2.5. Anti-parallel homogeneity correlates with repeat length

At pH 7.0, there were interesting differences in the melting profiles for $(CGA)_n$ (n = 3-7) (Figure 2.9A). When n = 4 or 5, the melting profile was broad with multiple transitions, whereas when $n \geq 6$, the melting transition was still broad but became more two-state. The multi-transition melting profile seen in shorter $(CGA)_n$ sequences ($n \leq$

5) suggested that a temperature-dependent equilibrium between hairpin and duplex or higher-order forms may exist. In contrast, the two-state melting transition for longer repeat sequences (n ≥ 6) suggested they exist predominately in the hairpin conformation. This trend is similar to the length-dependent conformational equilibrium seen in other triplet repeat sequences[31]. CD spectra at pH 7.0 confirmed the presence of anti-parallel oriented strands (negative band at 260 nm and positive band at 280 nm) when n = 5-7 (Figure 2.9B). The intensity of the bands decreased with repeat number and was lost when n ≤ 4 suggesting that the hairpin structure cannot form at low repeat number. In combination, this data further supports the importance of sequence length in rational structure design.

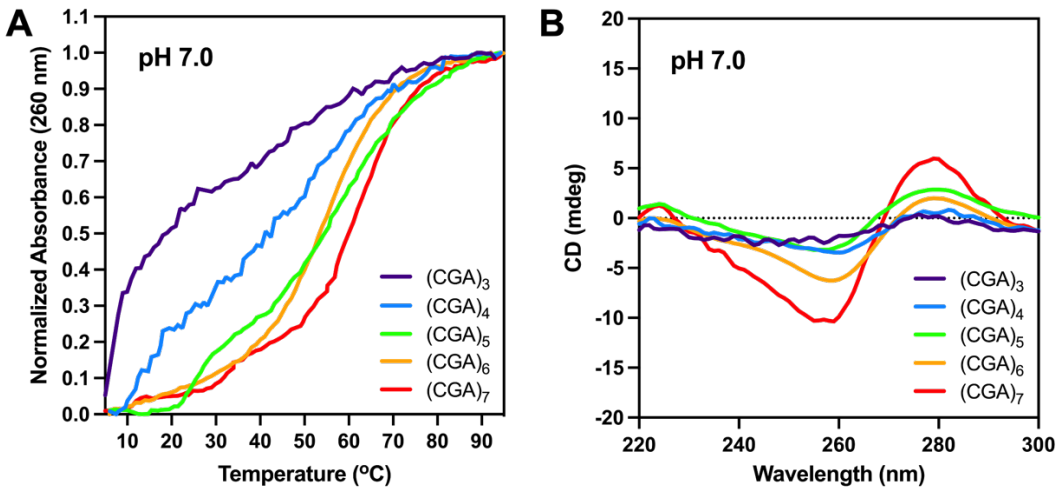

Figure 2.9. UV melting curves and CD spectra for $(CGA)_n$ sequences at pH 7.0. **(A)** Normalized temperature versus absorbance curves for 3 μM $(CGA)_n$ (n = 3-7) variants at pH 7.0. $(CGA)_3$ does not have a clear melting transition. $(CGA)_4$ and $(CGA)_5$ melting transitions are broad and multi-state. $(CGA)_6$ and $(CGA)_7$ melting transitions are still relatively broad but become more two-state. **(B)** CD spectra of $(CGA)_n$ (n = 3-7) at pH 7.0. The CD signal intensity corresponding to anti-parallel oriented strands decreases with repeat number and is lost when n ≤ 4.

Chapter 2.2.6. ΔG^{o}_{37} estimates used to predict the stability of new ps-duplexes

To estimate the energetic contribution of each parallel-stranded triplet unit, we compared thermodynamic parameters of $d(CGA)_n$ oligonucleotides with a differing number of repeat units (n). Strong apparent enthalpy-entropy compensation was observed for the set of $(CGA)_n$ variants tested (Figure 2.10). The positive correlation seen in the ΔH^o versus ΔS^o plot demonstrated the strong enthalpy-entropy compensation, whereas ΔG^{o}_{37} versus ΔH^o plot confirmed that the apparent compensation was not an artifact of experimental error.

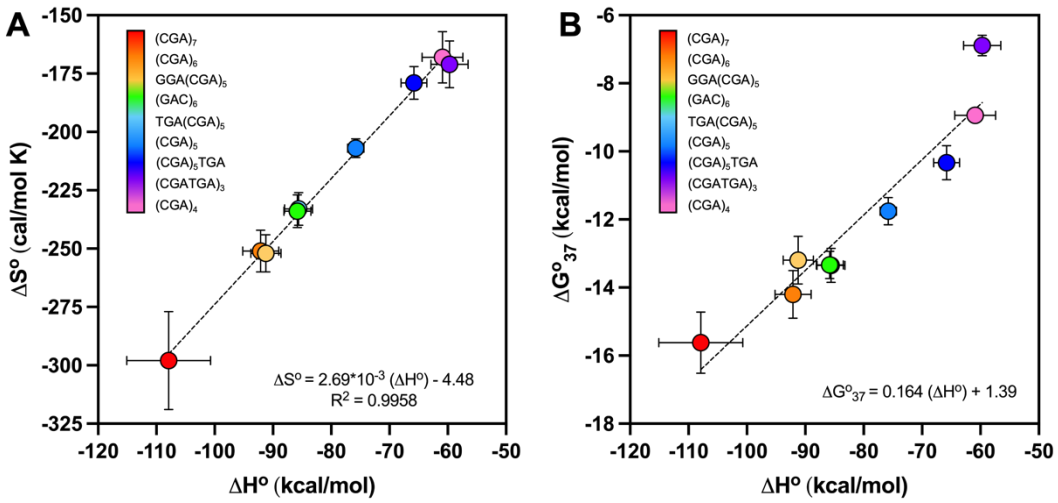

Figure 2.10. Apparent entropy-enthalpy compensation of all $(CGA)_n$ variants. (A) The positive correlation between ΔH^o and ΔS^o indicates a strong enthalpy-entropy compensation for all oligonucleotides tested. **(B)** The ΔG^{o}_{37} versus ΔH^o plot suggests that compensation is due to the underlying physical reality, as opposed to experimental error.

As a simple predictor of the ΔG^{o}_{37} for $(CGA)_n$ sequences, the number of d(CGA) units were plotted against the respective experimental ΔG^{o}_{37} value (Figure 2.11). Interestingly, the y-intercept (-0.250 kcal/mol) of the linear fit, which corresponds to the ΔG^{o}_{37} for a n = 1 sequence, could represent the gain in stability upon

formation of the first triplet unit slightly overbalancing the entropic penalty of bringing the strands together. The linear relationship can be used to predict the ΔG^{o}_{37} for a $(CGA)_n$ triplet of any length and can be combined with other estimates (5′-d(GGA), 5′-d(TGA), and 3′-d(TGA)) to predict ΔG^{o}_{37} of various sequences (Figure 2.11). Importantly, these results were internally consistent and provide a baseline for understanding the energetic contributions of these triplets.

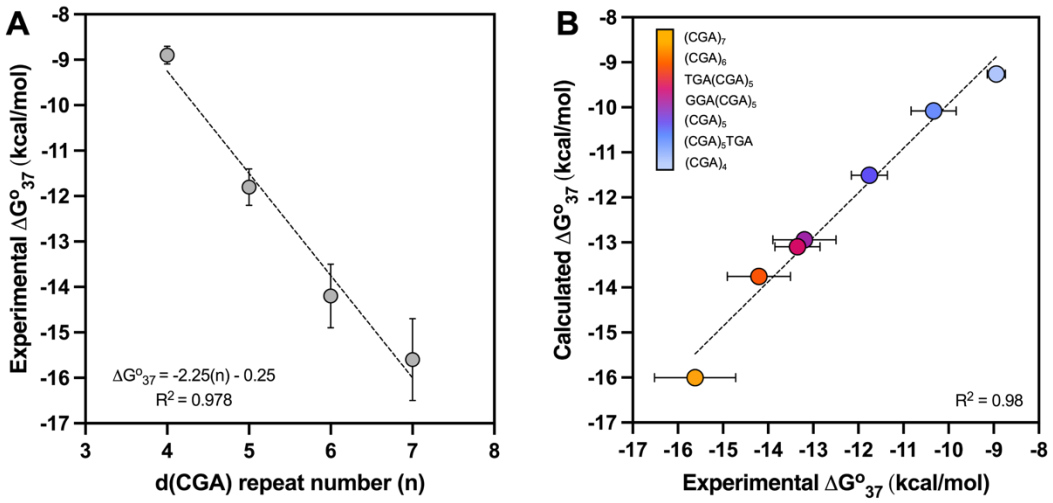

Figure 2.11. ΔG^{o}_{37} values can be used to estimate the stability of new triplet repeat sequences. (A) Linear fit of $(CGA)_n$ sequences versus experimentally obtained ΔG^{o}_{37} values. **(B)** Experimental ΔG^{o}_{37} is highly correlated to calculated ΔG^{o}_{37} for all sequences, assuming additive contributions from each triplet. Estimates for d(CGA) were obtained from the fit in **A**, while estimates for 5′-d(GGA), 5′-d(TGA), and 3′-d(TGA) were obtained from Table 1. All sequences were included except $(TGA)_6$, $(GAC)_6$, and $(CGATGA)_3$.

Chapter 2.3. Summary and Conclusions

Here, I have described our systematic approach to determine the thermodynamic stability of the d(CGA) triplet repeat motif and variants in solution. CD spectroscopy confirmed the formation of the ps-duplex at pH 5.5 or anti-parallel form at pH 7.0 for all variants tested except $(CGATGA)_3$ and $(TGA)_6$. This indicates that pH

sensitivity decreases as d(TGA) triplets are added and highlights potential approaches for tuning the pH-dependency of $(YGA)_6$ sequences. In parallel, UV absorbance thermal melting experiments showed that $(CGA)_6$ and variants undergo pH-induced structural switching that coincides with a significant thermodynamic destabilization. This trend is likely linked to the loss of strong inter-strand G/A base stacking and homo-base pair interactions as pH is increased from 5.5 to 7.0. Additionally, changes in thermodynamic parameters among variants indicate that the identity of the 5′-nucleobase within each triplet and the position and frequency of such triplets within stretches of d(CGA) triplets can tune ps-duplex stability. The thermodynamic stability data presented here can be used to strategically design and optimize sequences for nanotechnology applications that could benefit from pH-triggered structural switching. In the next chapter, I will discuss the nuclease sensitivity of the d(CGA) motif, which will demonstrate the unique advantage this motif offers for applications in cellular environments.

Chapter 2.4. Experimental procedures

Chapter 2.4.1. Oligonucleotide synthesis and purification

DNA oligonucleotides were synthesized on a 1 μmol scale using standard phosphoramidite chemistry on an Expedite 8909 Nucleic Acid Synthesizer (PerSeptive Biosystems, Framingham, MA) with reagents from Glen Research (Sterling, VA). Oligonucleotides were purified by denaturing 20% (19:1) acrylamide/bis-acrylamide and 7 M urea gel electrophoresis. DNA bands were identified and excised using UV

shadowing. DNA was electroeluted from gel slices, ethanol precipitated, and dialyzed against deionized water.

Chapter 2.4.2. Circular dichroism (CD) spectroscopy

Circular dichroism (CD) spectra were obtained using a Jasco J-810 spectropolarimeter fitted with a thermostatted cell holder (Jasco, Easton, MD). Samples were prepared using 10 μM DNA in the same buffers used for UV melting experiments: 20 mM MES, 100 mM sodium chloride (pH 5.5) or 20 mM sodium cacodylate, 100 mM sodium chloride (pH 7.0). Samples were incubated at 4°C overnight before data collection. Data were collected at room temperature using a 1 mm pathlength cuvette at wavelengths from 220 to 300 nm and represented as the average of three individual scans.

Chapter 2.4.3. UV Absorbance Thermal Melting Procedures

UV melting curves were obtained on a Cary Bio 100 UV-visible spectrophotometer (Varian/Agilent, Palo Alto, CA) equipped with a 12-cell sample changer and Peltier heating/cooling system. DNA samples were diluted to working concentrations in either 20 mM MES (pH 5.5) or 20 mM sodium cacodylate (pH 7.0), both supplemented with 100 mM sodium chloride. Sample absorbance data were obtained at 260 nm as the temperature was ramped from 4°C to 95°C at 1°C/min. In addition, data were collected at 260 nm for DNA renaturation from 95°C to 4°C at 1°C/min to assess reversibility. Self-masking cuvettes with 1-cm path lengths were used for 1.4 – 20 μM samples, whereas a 2-mm path length cuvette was used for 50 –

120 μM samples. Final working oligonucleotide concentrations were calculated from the absorbance at 85ºC.

Chapter 2.4.4. Melting Curve Data Analysis

The measured absorbance values for samples at a high concentration were corrected for deviations from linearity with respect to DNA concentration. A calibration curve was obtained and data falling above the linear range were corrected accordingly. We verified that this procedure yielded absorbance versus wavelength curves that were superimposable at all concentrations for samples that exhibited unimolecular behavior. Thermodynamic parameters for parallel-stranded duplex formation at pH 5.5 were obtained from the fit of each temperature versus absorbance melting curve to the bimolecular van't Hoff expression, as described by Petersheim and Turner.[125] Similarly, the van't Hoff expression for unimolecular hairpin formation was used to fit the melting curves and extract thermodynamic parameters for data at pH 7.0, as described by Siegfried and Bevilacqua.[126] The thermodynamic parameters obtained from the fits to the melting curves are reported as the average ± SD of the results from independent fits from experiments with four to six independently prepared samples at concentrations ranging from 2 to 120 μM.

Chapter 3: Stability of d(CGA)-based sequences in *in vivo-like* conditions

*This chapter is derived from the following manuscript:
Luteran, E. M.; Kahn, J. D.; Paukstelis, P. J. (2020). Stability of the pH-Dependent Parallel-Stranded d(CGA) Motif. *Biophys. J.* **119**:1580-1589.

Chapter 3.1. Introduction

Chapter 3.1.1. Overview

With increasing interest in using DNA nanotechnology in cellular applications, we assessed the stability of the DNA structures formed by d(CGA)-based sequences in conditions that mimic cellular environments. To do this, we qualitatively examined the extent to which d(CGA)-based oligonucleotides resist degradation by nucleases with preferences for double-stranded B-DNA (DNase I) or unpaired, single-stranded DNA (S1 nuclease). In this chapter, I describe the unique differences in nuclease digestion pattern observed from sequence variations, providing evidence for subtle structural differences among variants. In addition, I describe an increased resistance against double strand-specific nucleases in the ps-duplex form relative to the anti-parallel form, illustrating the unique advantage the ps-duplex form may offer in DNA nanotechnology applications used in cellular conditions. While investigating nuclease stability, we observed the formation of unexpected high molecular weight (HMW) complexes. At the end of this chapter, I will briefly describe our efforts to understand this unexpected HMW structure formation.

Chapter 3.1.2. Sequence and variant design

We chose a subset of the sequences originally designed to probe thermodynamic stability (Chapter 2) in our assessment of nuclease sensitivity. Specifically, we focused our analysis on d(CGA)-based repeat sequences that incorporated d(TGA) triplets (Table 3.1). This subset of sequences allowed us to directly explore the importance of pyrimidine identity and position within repeat sequences. Differences in nuclease degradation pattern across variants allowed us to make conclusions about the effect of subtle triplet mutations on the structure formed at each pH. Considering distinct changes in thermodynamic stability among d(YGA) variants (described in Chapter 2), we expected to observe corresponding differences in nuclease degradation patterns among variants incorporating d(TGA) triplets.

Table 3.1. DNA oligonucleotides used to assess nuclease stability.

DNA Name	Sequence
$(CGA)_6$	d(CGA-CGA-CGA-CGA-CGA-CGA)
$TGA(CGA)_5$	d(<u>T</u>GA-CGA-CGA-CGA-CGA-CGA)
$(CGA)_5TGA$	d(CGA-CGA-CGA-CGA-CGA-<u>T</u>GA)
$(CGATGA)_3$	d(CGA-<u>T</u>GA-CGA-<u>T</u>GA-CGA-<u>T</u>GA)
$(TGA)_6$	d(<u>T</u>GA-<u>T</u>GA-<u>T</u>GA-<u>T</u>GA-<u>T</u>GA-<u>T</u>GA)

*Underlined nucleotides differ from the $(CGA)_n$ pattern.

Chapter 3.2. Results and Discussion

Chapter 3.2.1. Increased DNase I resistance observed in the ps-duplex form

DNase I is an endonuclease found in serum that nonspecifically digests Watson-Crick double-stranded DNA (dsDNA).[127, 128] In the presence of Ca^{2+} and Mg^{2+}, phosphodiester bonds within each strand are independently hydrolyzed in a random

fashion.[129] The digestion results in a mixture of products including mono- and oligonucleotides depending on the duration of the reaction (Figure 3.1A). We hypothesized that the DNase I foot-printing pattern would be different depending on which structural form (ps-duplex or anti-parallel hairpin) the d(CGA)-based sequences adopted. We expected the non-canonical ps-duplex form to resist DNase I digestion because of the lack of recognizable regions, whereas the WC base pairs in the hairpin stem of the anti-parallel hairpin would be recognized and cleaved by DNase I (Figure 3.1B).

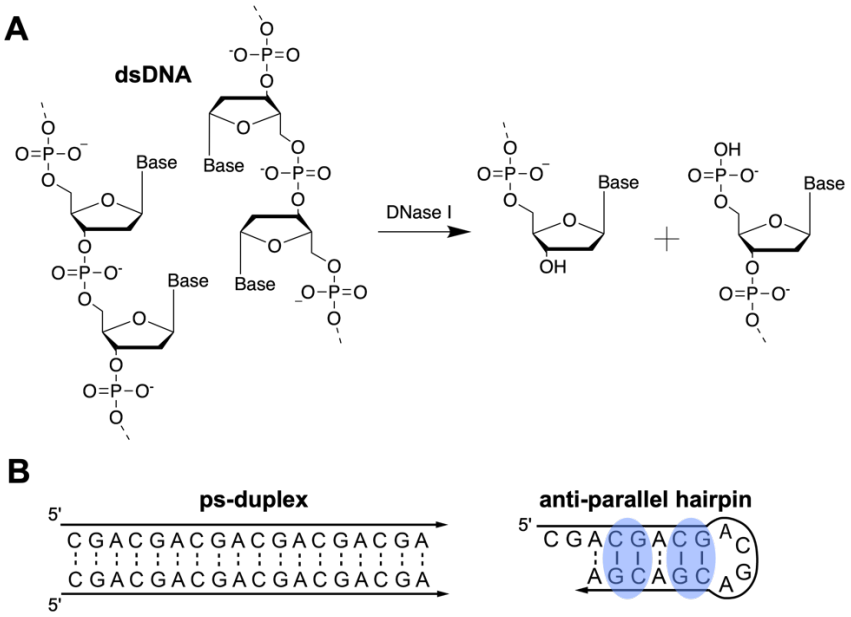

Figure 3.1. Hypothesized DNase I cleavage sites. (A) DNase I recognizes the minor groove of double-stranded DNA and independently hydrolyzes the phosphodiester backbone of each strand to produce products containing a 3′-OH and 5′-phosphate. **(B)** Predicted DNase I cleavage sites for each structural form of d(CGA)-based sequences. The ps-duplex (left) does not have any DNase I recognizable regions, therefore we predicted it would resist digestion. The anti-parallel hairpin (right) does contain WC-base paired regions (blue) that could be recognized and hydrolyzed by DNase I.

CD was used to verify that the ps-duplex or anti-parallel hairpin structures were maintained in the solution conditions present in DNase I assay buffers. Magnesium

chloride (up to 10 mM) and calcium chloride do not alter the expected structures at either pH, as evident by minimal changes in CD spectra at each pH (Figure 3.2A,B). Control experiments confirmed the activity of DNase I for single-stranded (ssDNA) or dsDNA substrates at each pH tested. As expected, we observed DNase I cleavage for dsDNA substrates at pH 5.5 and 7.5 and reduced activity for ssDNA substrates (Figure 3.2C).

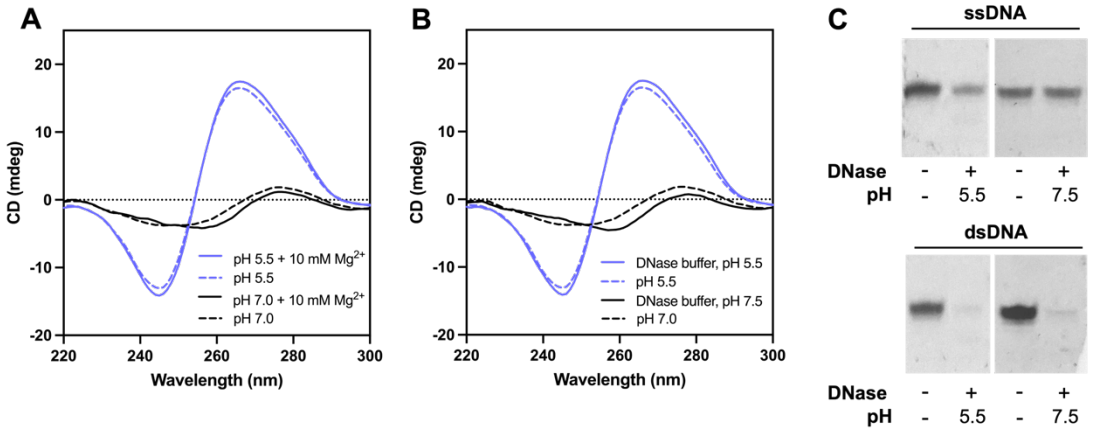

Figure 3.2. DNase I assay control experiments. (A) CD spectra of $GGA(CGA)_5$ in CD buffer at pH 5.5 (purple, dashed) or pH 7.0 (black, dashed) compared to samples supplemented with 10 mM Mg^{2+} (solid lines). The addition of 10 mM Mg^{2+} does not alter the expected structures at either pH. **(B)** CD spectra of $GGA(CGA)_5$ in DNase I buffer containing 0.5 mM calcium chloride and 2.5 mM magnesium chloride. DNase I buffer conditions do not alter the expected structures at pH 5.5 (purple) or pH 7.0 (black). **(C)** ssDNA and dsDNA exposed to DNase I at pH 5.5 and 7.5. DNase I hydrolyzed dsDNA at pH 5.5 and 7.5. DNase I had reduced activity for ssDNA substrates at pH 5.5 and 7.5. Mg^{2+} was added and incubated with d(GGACAGCTGGGAG) to form dsDNA substrate. The same sequence was used without Mg^{2+} as the ssDNA substrate.

Each variant tested was incubated with DNase I at pH 5.5 and pH 7.5 (Figure 3.3). At pH 7.5, $(CGA)_6$, $TGA(CGA)_5$, and $(CGA)_5TGA$ showed clear cleavage patterns, suggesting that the structure formed at pH 7.5 had DNase I recognizable

regions. This supported previous CD and UV thermal melting analysis (Chapter 2) indicating the formation of an anti-parallel hairpin at near-neutral pH. The difference in DNase I cleavage pattern among variants at pH 7.5 suggests that the anti-parallel hairpin formed for each variant has subtle structural differences. Importantly, these same variants showed clear DNase I resistance when the reaction was repeated at pH 5.5 (Figure 3.3). This indicates that the parallel-stranded duplex formed by $(CGA)_6$, $TGA(CGA)_5$, $(CGA)_5TGA$, and $(CGATGA)_3$ was not recognized by DNase I. $(CGATGA)_3$ appeared to have slight protection at pH 5.5 compared to pH 7.5, whereas the cleavage of $(TGA)_6$ was identical at both pHs. These results are consistent with our CD and UV melting observations (Chapter 2) showing no pH-dependent structural change for $(TGA)_6$.

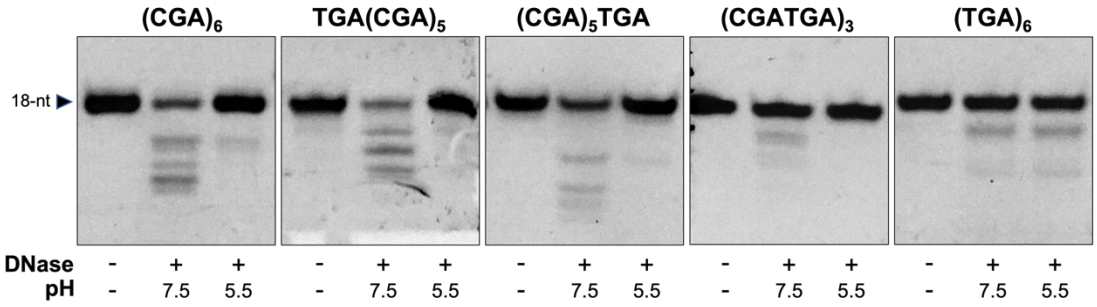

Figure 3.3. The ps-duplex form exhibits enhanced DNase I resistance. $(YGA)_6$ variants exposed to DNase I at pH 7.5 or 5.5. DNase I reactions were quenched, and reaction products were analyzed via denaturing gel electrophoresis. DNase I preferentially hydrolyzed the anti-parallel hairpin structure formed at pH 7.5, while the ps-duplex formed at pH 5.5 was resistant to hydrolysis.

Increased resistance against double strand specific nucleases in the ps-duplex form relative to the anti-parallel hairpin suggests that the parallel motifs may offer unique advantages for the design, assembly, and delivery of nanostructures in cellular environments.

Chapter 3.2.2. S1 nuclease reaction products reveal preferential digestion sites

S1 nuclease is an endonuclease that preferentially degrades single-stranded DNA (ssDNA).[130, 131] In the presence of Zn^{2+}, phosphodiester bonds of ssDNA are hydrolyzed to produce mono- or oligonucleotides with a 5′-phosphoryl group (Figure 3.4A). We hypothesized that the ps-duplex would not be digested by S1 nuclease because it was thought to form a fully double-stranded duplex that does not have S1 accessible single-stranded regions. In contrast, the terminal end overhangs and loop region of the anti-parallel hairpin was expected to be hydrolyzed by S1 nuclease (Figure 3.4B).

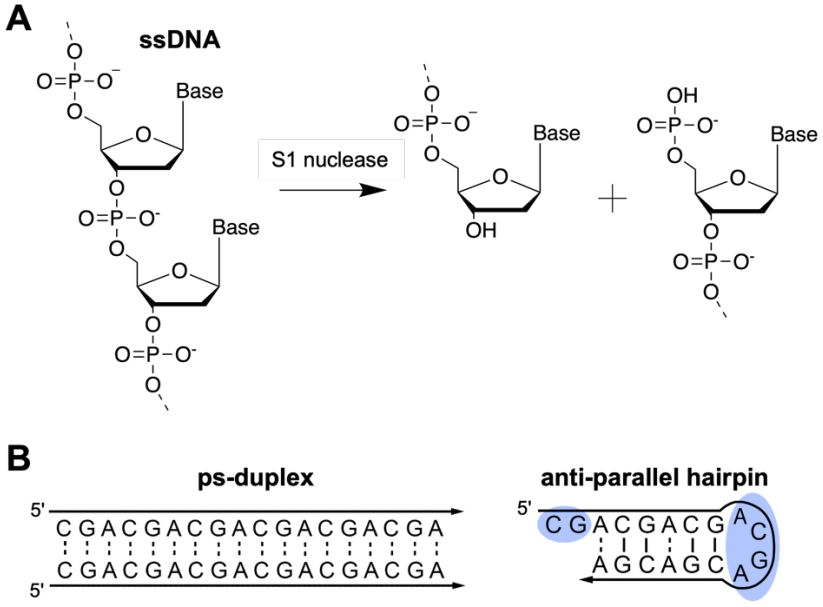

Figure 3.4. Hypothesized S1 nuclease cleavage sites. (A) S1 nuclease preferentially hydrolyzes the phosphodiester backbone of ssDNA to produce products containing a 3′-OH and 5′-phosphate. **(B)** Predicted S1 nuclease cleavage sites. The ps-duplex (left) does not have any S1 nuclease recognizable regions, therefore resisting cleavage. The anti-parallel hairpin (right) does have single-stranded nucleotides in the loop region and terminal end overhang that could be recognized and hydrolyzed by S1 nuclease (highlighted in blue).

CD was used to verify that the ps-duplex or anti-parallel hairpin structures were maintained in the presence of S1 nuclease assay buffer. The lack of change in CD spectra indicated that zinc sulfate and increased concentration of sodium chloride in S1 nuclease reaction buffer did not significantly alter the expected structures at either pH (Figure 3.5A). Control experiments confirmed the activity of S1 nuclease for ssDNA substrates at each pH tested. As expected, S1 nuclease functioned optimally at low pH (Figure 3.5B).

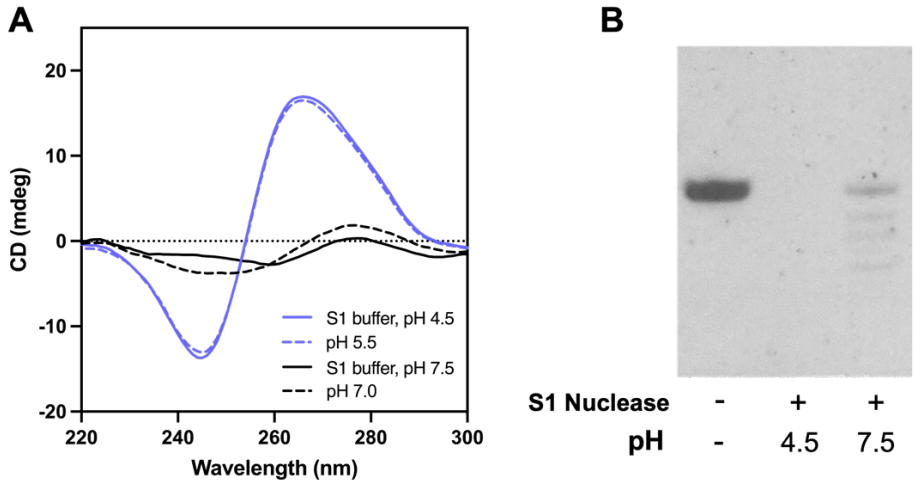

Figure 3.5. S1 nuclease assay control experiments. (A) CD spectra of GGA$(CGA)_5$ in CD buffer at pH 5.5 (purple, dashed) or pH 7.0 (black, dashed) compared to samples prepared in S1 nuclease assay buffer (solid lines). S1 nuclease buffer did not alter the expected structures at either pH. **(B)** S1 nuclease hydrolyzed ssDNA at pH 4.5 and 7.5. S1 nuclease exhibited slightly reduced activity at pH 7.5 than at pH 4.5.

$(YGA)_6$ sequences were exposed to S1 nuclease to test d(YGA) triplet repeat susceptibility to single strand-specific nucleases. At pH 7.5, S1 nuclease cleavage products for $(CGA)_6$, TGA$(CGA)_5$, $(CGA)_5$TGA, and $(CGATGA)_3$ suggest that single-stranded regions are primarily associated with hairpin ends (Figure 3.6). Intense bands visible at ≥14 nucleotides (nt) are consistent with the degradation of terminal non-base-

paired nucleotides or single-stranded triplet overhands at the hairpin terminus. Distinct differences in S1 nuclease degradation pattern among all variants at pH 7.5 suggests that hairpins formed were not structurally identical. Weak product bands for $(CGATGA)_3$ and $(TGA)_6$ indicated that the structure formed by these variants at pH 7.5 present substantial S1 nuclease recognizable single-stranded regions, consistent with CD and thermal melting data presented in Chapter 2.

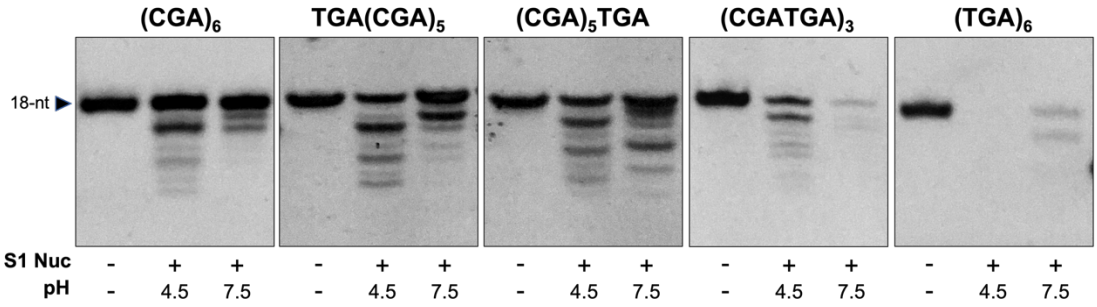

Figure 3.6. $(YGA)_6$ variants exposed to S1 nuclease at pH 7.5 or 4.5. At pH 7.5, intense bands for degradation products ≥14 nt suggest that non-paired hairpin overhangs or single stranded triplet overhangs were degraded by S1 nuclease. At pH 4.5, intense 18-nt bands indicate that the reaction mixture contained a large population of perfectly aligned ps-duplexes that were not recognized by S1 nuclease. However, bands at 15, 12, and 9-nt indicate that the ps-duplex form can undergo frame-shifting to expose single-stranded d(CGA) triplets to S1 nuclease digestion.

$TGA(CGA)_5$ and $(CGA)_5TGA$ were exposed to S1 nuclease for 2 days at room temperature to probe the stability of d(CGA)-based sequences against S1 nuclease over an extended reaction period (Figure 3.7). Each oligonucleotide was almost completely digested at pH 7.5 when exposed to S1 nuclease for the extended reaction time. This result suggested that the anti-parallel hairpin is dynamic enough to be susceptible to S1 nuclease digestion over time. In contrast, the ps-duplex form resisted S1 nuclease digestion over 2 days, illustrating the stability and utility of this structure at room temperature.

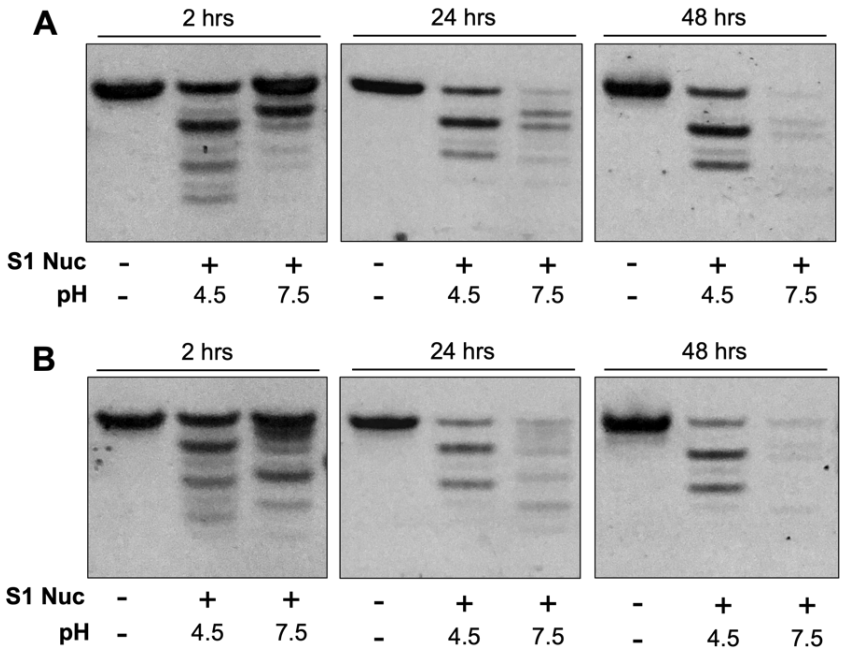

Figure 3.7. Extended S1 nuclease digestion. The product bands for **(A)** TGA$(CGA)_5$ and **(B)** $(CGA)_5$TGA at pH 7.5 were almost completely digested when exposed to S1 nuclease for 48 hrs, suggesting that the structure formed at this pH is dynamic or not resistant to S1 digestion over time. In contrast, the intensity of the 18, 15, and 12-nt bands corresponding to digestion by S1 nuclease at pH 4.5 did not significantly decrease over two days of exposure. This suggests that the ps-duplex formed is stable and resistant to S1 degradation at room temperature.

$(CGA)_6$, TGA$(CGA)_5$, $(CGA)_5$TGA, and $(CGATGA)_3$ were only partially digested by S1 nuclease at its optimal pH of 4.5 (Figure 3.6). CD experiments indicated that d(CGA)-based oligonucleotides form ps-duplexes in conditions similar to S1 nuclease reaction conditions (Figure 2.2, Figure 3.5A), explaining the lack of complete S1 digestion. The intense 15-nt band present in the degradation of $(CGA)_6$ is consistent with the hypothesis that a small population of the ps-duplexes were frame-shifted by one d(CGA) triplet, resulting in 3-nt S1 nuclease-accessible single-stranded overhangs (Figure 3.8). The intense unreacted band remaining at 18-nt suggests that most oligomers in the reaction populated the perfectly aligned ps-duplex, which resisted S1

nuclease digestion. The low molecular weight degradation products (15-nt and 12-nt) could result from frame-shifted duplexes containing one or more S1 nuclease-accessible triplet overhang. Apparently, once duplexes were formed, frameshifting was extremely slow at room temperature, or else equilibration among frames would have exposed all of the sequence to S1 digestion over the extended reaction period.

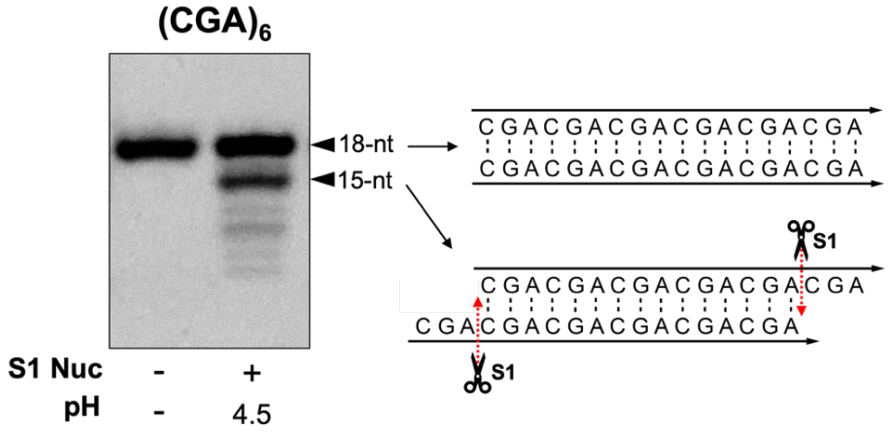

Figure 3.8. Ps-duplex frameshifting revealed by S1 nuclease digestion of $(CGA)_6$ at pH 4.5. The intense band at 18-nt indicates that most oligomers populated the perfectly aligned 18-nt ps-duplex, resisting S1 nuclease digestion. The intense 15-nt band suggests that a small population of the ps-duplexes were frame-shifted by one d(CGA) triplet resulting in S1 nuclease digestion.

$TGA(CGA)_5$ and $(CGA)_5TGA$ showed S1 nuclease degradation patterns similar to each other, consistent with frame-shifted triplet cleavage, as intense bands persisted at 15, 12, and 9-nt. Each intense band was coupled with a faint band, where the position and intensity were used to gain insight into the specific S1 cleavage site. The cleavage of a 5′-triplet produced a fast migrating 15-nt product containing a terminal 5′-phosphate, whereas the cleavage of a 3′-triplet produced a slower migrating 15-nt product that did not have a 5′-phosphate (Figure 3.9). The faster migrating 15-nt band was very intense in $TGA(CGA)_5$, suggesting that the 5′-d(TGA) triplet was

preferentially digested. The opposite was seen in $(CGA)_5TGA$, in which the slower migrating 15-nt band was more intense, suggesting that the 3′-d(TGA) was cleaved. The difference in S1 cleavage pattern observed between $TGA(CGA)_5$ and $(CGA)_5TGA$ supports our initial hypothesis drawn from thermodynamic parameters that the T-T homo-base pair is associated with inherent flexibility leading to the disruption of surrounding base pair interactions (Figure 2.7).

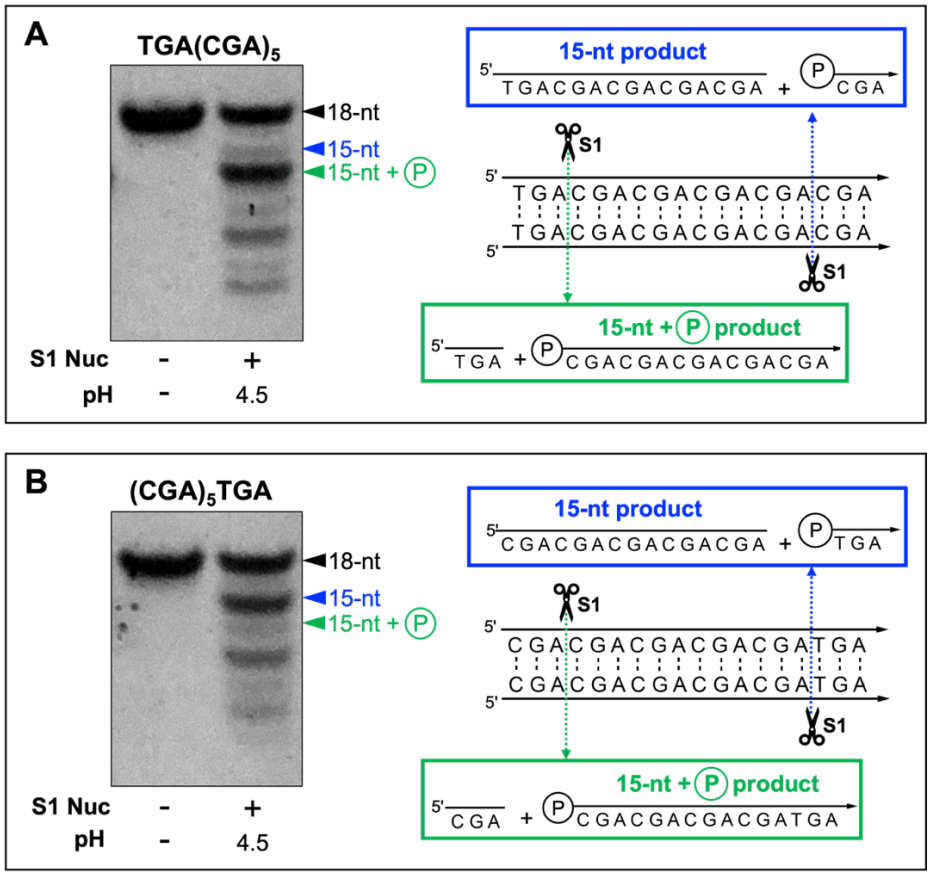

Figure 3.9. S1 nuclease cleavage of $TGA(CGA)_5$ and $(CGA)_5TGA$ ps-duplexes. The position and intensity of S1 nuclease cleavage products of **(A)** $TGA(CGA)_5$ and **(B)** $(CGA)_5TGA$ were used to gain insight into the preferred S1 cleavage site. The 15-nt product (blue) represents the S1 nuclease cleavage product that does not have a terminal 5′-phosphate group (P). The 15-nt + P (green) represents the S1 nuclease cleavage product that does contain a terminal 5′-phosphate, leading to increased gel mobility. In the $TGA(CGA)_5$ reaction, the 15-nt + P product is the most intense band, while the 15-nt band is most intense for $(CGA)_5TGA$, suggesting that the d(TGA)

triplet is preferentially cleaved in each reaction. 3-nt products migrated too quickly through the gel to be resolved.

At pH 4.5, $(TGA)_6$ was almost completely digested by S1 nuclease by the end of the two-hour reaction period, indicating lack of structure formation at each pH (Figure 3.6). This further confirmed results from CD (Chapter 2, Figure 2.2) and UV melting experiments (Chapter 2, Figure 2.3) that suggest $(TGA)_6$ does not form stable structures at acidic or neutral pH.

Chapter 3.2.3. High molecular weight structure formation

Unexpected high molecular weight (HMW) products became apparent over time for d(CGA)-based sequences, except $(TGA)_6$, and persisted in 8 M urea denaturing gels (Figure 3.10A). We found that the number and intensity of HMW products is unique for each variant tested. Preliminary experiments suggested that these HMW structures appeared more quickly at lower pHs, at higher temperature, and in sequences containing higher $d(CGA)_n$ repeat number (n) (Figure 3.10). The formation of the stable HMW structures likely relies on cytosine reactivity as HMW bands were never observed for $(TGA)_6$. We cannot confirm the identity or mechanism of HMW band formation, but we suspect that stable aggregates or cross-links are formed from reactions with depurination products (abasic sites) of the d(CGA) triplet repeats over time (Figure 3.11).

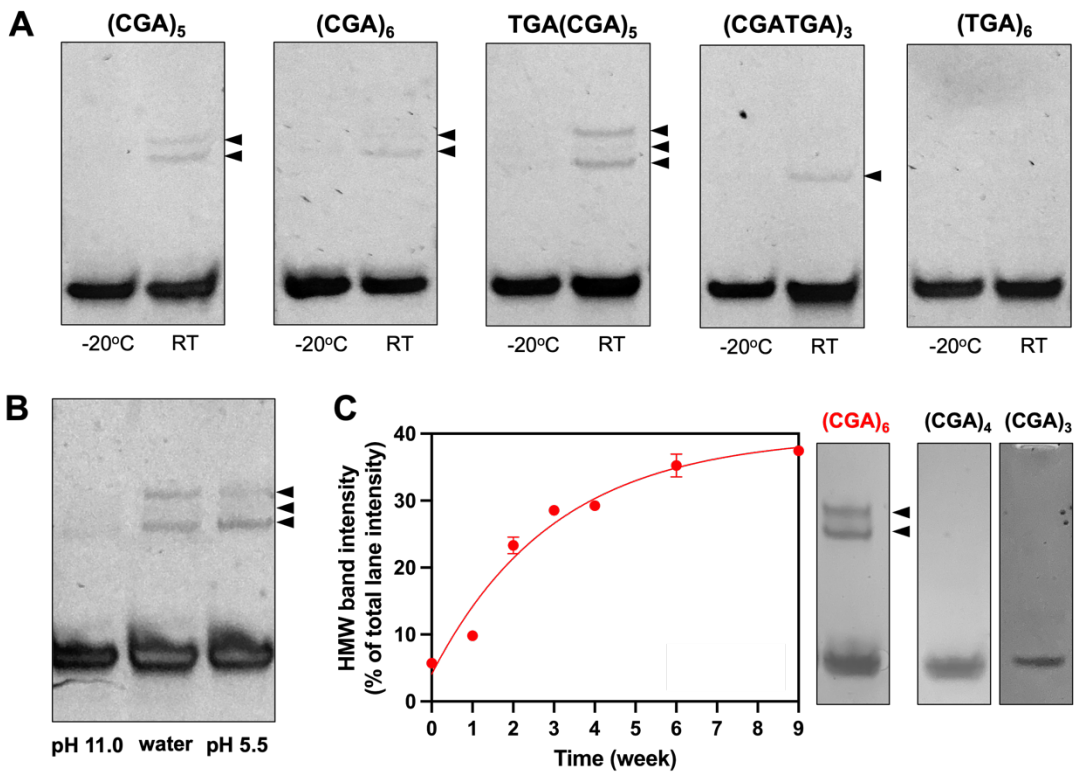

Figure 3.10. HMW structures form over time for d(CGA)-based oligonucleotides. High molecular weight (HMW) species are indicated in each gel by black arrows. **(A)** $d(CGA)_6$ and variants stored at -20°C or room temperature (RT) for 3 weeks and analyzed on 8 M denaturing gel. HMW bands formation is favored when stored at higher temperature. $(TGA)_6$ does not have HMW band formation when stored at either temperature. **(B)** HMW band formation is favored when $TGA(CGA)_5$ (shown) and other related d(CGA)-based variants (not shown) are stored in water or at pH 5.5, compared to pH 11.0. Samples were stored in the respective pH condition for 8 days at RT prior to analysis on 8 M denaturing gel. **(C)** HMW band formation is favored in sequences containing higher d(CGA) triplet repeat number. $(CGA)_n$ (n = 6, 4, 3) sequences were prepared in water and stored at RT for up to 9 weeks. For $(CGA)_6$ (red data points), HMW band intensity increased over time, while $(CGA)_4$ and $(CGA)_3$ (not plotted) did not form HMW bands over the incubation time course. An aliquot of each sample was frozen at each time point and stored at -20°C until gel analysis. The gel lane representative of the 9-week time point is shown for each sequence.

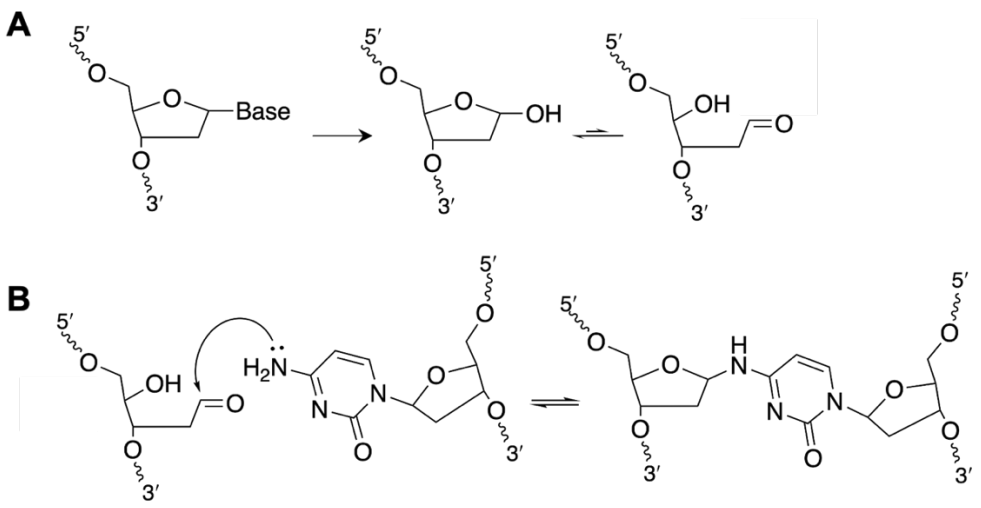

Figure 3.11. Speculative mechanism for HMW product formation. High molecular weight products could be a result of crosslinking interactions that form between abasic sites on one strand and intact cytosines of the opposing strand. **(A)** The formation and structure of an abasic site is shown for one nucleotide within a DNA strand. The loss of a nucleobase results in the formation of an abasic site containing a 1′-OH. The abasic site can also adopt the tautomeric form containing a ring-opened aldehyde. **(B)** The speculative mechanism resulting in crosslinked inter-strand HMW products is shown. The ring-opened aldehyde tautomer of the abasic site can react with the N4 position of a cytosine from the opposing intact strand. This interaction generates inter-strand covalent crosslinks. These higher molecular weight products could correspond to the slowly migrating bands observed in the denaturing gels shown in Figure 3.10.

Chapter 3.3. Summary and Conclusions

In this chapter, I have described our approach to evaluate the stability of the pH-dependent structures formed by d(CGA)-based sequences in the presence of DNase I and S1 nuclease to gauge the stability of this motif in conditions that mimic cellular environments. DNase I reactivity revealed that the ps-duplex form showed increased resistance against double-strand specific nucleases relative to the anti-parallel hairpin, while S1 nuclease foot-printing indicated that the ps-duplex form undergoes triplet frameshifting. Additionally, the S1 nuclease product band intensity and pattern were

used to distinguish the preferred S1-cleavage sites among variants. Together, the nuclease sensitivity data presented here demonstrates that the parallel-stranded motif may offer unique advantages for the design, assembly, and delivery of nanostructures used in cellular environments. In the next chapter, I will describe two crystal structures of d(CGA)-based triplet repeat sequences that have allowed us to characterize and establish structural parameters of d(CGA) triplets in the ps-duplex form.

Chapter 3.4. Experimental procedures

Chapter 3.4.1. Oligonucleotide synthesis and purification

DNA oligonucleotides were synthesized on a 1 μmol scale using standard phosphoramidite chemistry on an Expedite 8909 Nucleic Acid Synthesizer (PerSeptive Biosystems, Framingham, MA) with reagents from Glen Research (Sterling, VA). Oligonucleotides were purified by denaturing 20% (19:1) acrylamide/bis-acrylamide and 7 M urea gel electrophoresis. DNA bands were identified and excised using UV shadowing. DNA was electroeluted from gel slices, ethanol precipitated, and dialyzed against deionized water.

Chapter 3.4.2. DNase I stability assay

DNA was incubated overnight at 4°C with 50 mM magnesium formate to induce structure formation, as previously described [132]. The DNA was diluted to 40 μM and was incubated with 7 U/mL DNase I (Thermo Fisher Scientific), 10 mM MES (pH 5.5) or 10 mM Tris (pH 7.5), 2.5 mM magnesium chloride, and 0.5 mM calcium chloride at 37°C for 10 min. The reaction was quenched by the addition of 5 mM EDTA and incubated at 75°C for 10 min. Samples were analyzed by 20% denaturing (19:1)

acrylamide/bis-acrylamide and 7 M urea gel electrophoresis and stained with SYBR Gold (Thermo Fisher Scientific).

Chapter 3.4.3. S1 nuclease stability assay

20 μM DNA was incubated with 200 U/mL S1 nuclease (Thermo Fisher Scientific, Waltham, MA) in 40 mM sodium acetate (pH 4.5) or 40 mM Tris HCl (pH 7.5), 300 mM sodium chloride, and 2 mM zinc sulfate at 22°C for 2 hrs. The reaction was quenched with 30 mM EDTA and incubated at 70°C for 10 min. Samples were analyzed by 20% denaturing (19:1) acrylamide/bis-acrylamide and 7 M urea gel electrophoresis and stained with SYBR Gold (Thermo Fisher Scientific).

Chapter 3.4.4. Circular dichroism (CD) spectroscopy

Circular dichroism (CD) spectra were obtained using a Jasco J-810 spectropolarimeter fitted with a thermostatted cell holder (Jasco, Easton, MD). Samples were prepared using 10 μM DNA in 20 mM MES (pH 5.5) or 20 mM sodium cacodylate (pH 7.0), with 100 mM sodium chloride. Samples were incubated at 4°C overnight before data collection. Data were collected at room temperature using a 1 mm pathlength cuvette at wavelengths from 220 to 300 nm and represented as the average of three individual scans.

Chapter 4: Structural analysis of d(CGA)-based parallel-stranded duplexes

*This chapter is derived from the following manuscript:
Luteran, E. M.; Paukstelis, P. J. (2022). The parallel-stranded d(CGA) duplex is a highly predictable structural motif with two conformationally distinct strands. *Acta. Cryst. D78,* https://doi.org/10.1107/S2059798322000304.

Chapter 4.1. Introduction

Chapter 4.1.1. Overview

In this chapter I describe two crystal structures of oligonucleotides containing multiple tandem d(CGA) triplet repeats in the ps-duplex form. These structures are the longest ps-duplexes to be solved comprised solely of such triplets. The structure determination of four unique ps-duplexes across these two different crystal structures has allowed us to thoroughly characterize and define the structural features of d(CGA) triplets and the ps-duplexes they form. Despite crystallization and molecular packing differences, the resulting ps-duplex structures have strikingly low RMSD values, demonstrating the robust structural uniformity of the d(CGA) triplet repeat motif in the ps-duplex form. Also, we note that each ps-duplex contains two unique d(CGA) triplet conformations (i.e. loose or rigid) based on differences in backbone torsion angles, hydrogen bonding distances, and base stacking interactions. Within each strand, d(CGA) triplets only adopt one of the two unique conformations. Thus, the resulting ps-duplex contains one strand comprised solely of d(CGA) triplets in the "loose" conformation while the opposing strand is completely comprised of d(CGA) triplets in the "rigid" conformation. This illustrates that ps-duplexes containing d(CGA) triplets

are not structurally symmetrical and the apparent structural asymmetry is propagated separately based on the d(CGA) triplet conformations within each strand.

Chapter 4.1.2. X-ray crystallography-based approach

d(CGA)-based sequences have been primarily observed in the ps-duplex form either in the context of other complex secondary structures[35, 36, 133] or in isolation with few (≤ 4) triplet repeat units.[33, 120, 134] Because d(CGA) is known adopt different structural forms in a pH-dependent manner, we intended to use x-ray crystallography to capture and characterize both structural conformations (ps-duplex and anti-parallel hairpin). Towards this goal, we synthesized and purified several d(CGA)-based sequences for crystallization containing ≥ 4 triplet repeat units. The resulting DNA library contained d(CGA)-based sequences with various triplet repeat number (4 – 7) and/or mutated triplets. Triplets were considered mutated if they contained any deviation from the d(CGA) pattern (Figure 4.1). All sequences except $(CGATGA)_3$, $(TGA)_6$, and $(TGA)_5$ could adopt the ps-duplex and anti-parallel hairpin forms in solution based on CD analysis. Our goal was to find sequences that could also be captured in both forms via crystallography.

Table 4.1. d(CGA)-based sequences tested for crystallization.

DNA Name	Sequence (5′ to 3′)
$(CGA)_6$*	d(CGA-CGA-CGA-CGA-CGA-CGA)
$(CGA)_5$*	d(CGA-CGA-CGA-CGA-CGA)
TGA$(CGA)_5$	d(TGA-CGA-CGA-CGA-CGA-CGA)
$(CGATGA)_3$	d(CGA-TGA-CGA-TGA-CGA-TGA)
$(TGA)_6$	d(TGA-TGA-TGA-TGA-TGA-TGA)
$(TGA)_5$	d(TGA-TGA-TGA-TGA-TGA)
GGA$(CGA)_5$	d(GGA-CGA-CGA-CGA-CGA-CGA)
$(CGA)_7$	d(CGA-CGA-CGA-CGA-CGA-CGA-CGA)
$(CGA)_5$TGA*	d(CGA-CGA-CGA-CGA-CGA-TGA)
$(GAC)_6$	d(GA-CGA-CGA-CGA-CGA-CGA-C)
GGA$(CGA)_4$GGA	d(GGA-CGA-CGA-CGA-CGA-GGA)
TGA$(CGA)_4$TGA	d(TGA-CGA-CGA-CGA-CGA-TGA)
$(TGA)_2(CGA)_4$*	d(TGA-TGA-CGA-CGA-CGA-CGA)
$(CGA)_4$	d(CGA-CGA-CGA-CGA)
GA$(CGA)_4$*	d(GA-CGA-CGA-CGA-CGA)
GA$(CGA)_5$*	d(GA-CGA-CGA-CGA-CGA-CGA)
GA$(CGA)_6$*	d(GA-CGA-CGA-CGA-CGA-CGA-CGA)
A$(CGA)_5$	d(A-CGA-CGA-CGA-CGA-CGA)

*Unmodified and brominated sequences were synthesized. Underlined nucleotides represent positions where 5-U-Br or 5-C-Br were incorporated to generate brominated sequences. Up to two brominated sites were incorporated per sequence.
**Mutated triplets are designated by red text.

Following synthesis and purification, each DNA sample was screened against two 96-condition crystallization screens containing various cations, polyamines, and precipitants prepared over a range of pHs. Each screened DNA sample was visually assessed for crystal formation using light microscopy after several days of incubation. Any samples that contained solid, birefringent material were flagged as “promising hits”. Conditions that produced promising hits were pursued for further optimization. To optimize a promising hit condition, the concentrations of components within the condition were individually adjusted and tested in larger drop volumes to tune the solution condition to favor single crystal formation. Though several DNA sequences resulted in initial promising hits, many of the samples failed to produce diffraction

quality crystals during optimization. In such cases, optimization produced crystals that were either too small, clustered, rounded, or did not diffract (Figure 4.1). Many iterations of sequence and condition optimization were required to obtain diffraction data. $(CGA)_5TGA$ and $GA(CGA)_5$ were the only two sequences that produced high quality diffraction data that ultimately resulted in the successful structure determination described in this chapter.

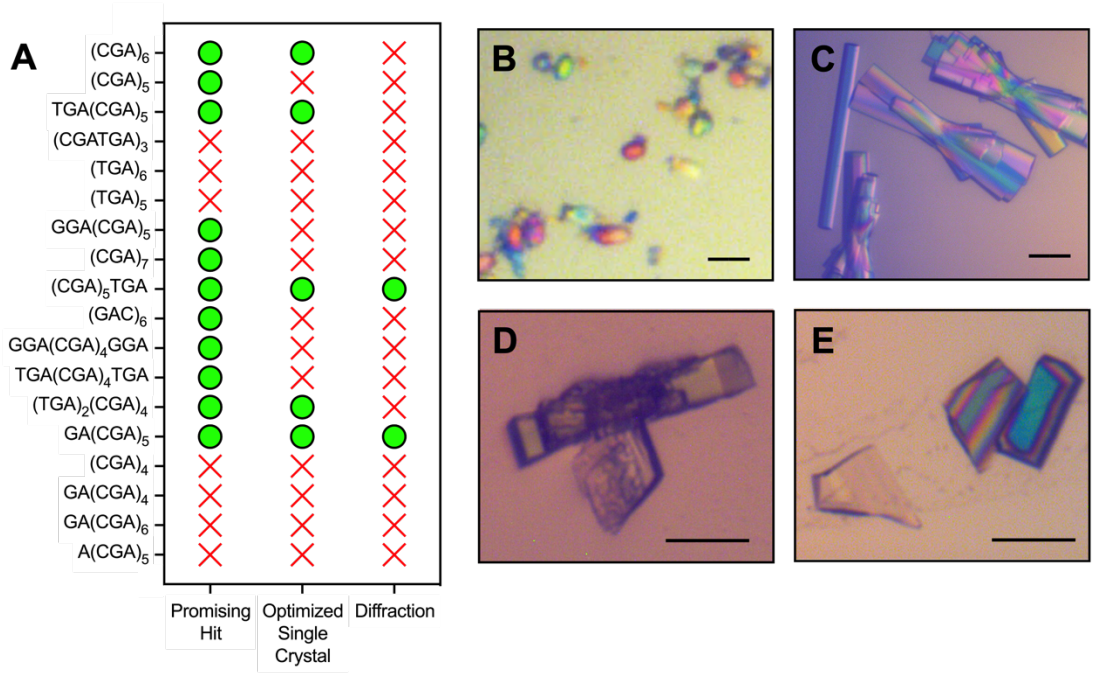

Figure 4.1. Crystallization of d(CGA)-based sequences. (A) Schematic representation of successful (green circle) or unsuccessful (red X) crystallization steps for each DNA tested. Any birefringent solid material visible after the initial 96-condition screening was considered a successful promising hit and pursued further in condition optimization. Any large, birefringent, single crystal was considered a successful optimization. Such crystals were then frozen and sent for data collection. Diffraction data extending to resolutions ≤ 3.0 Å without significant ice rings or twinning were collected and considered successful. Images of optimized crystals that did not result in diffraction for **(B)** $(CGA)_6$ and **(C)** $(TGA)_2(CGA)_4$. Images of optimized crystals that produced diffraction data for **(D)** $GA(CGA)_5$ and **(E)** $(CGA)_5TGA$. All scale bars are 70 μm.

Chapter 4.2. Results and discussion

Chapter 4.2.1. Overview of structure determination

CD spectroscopy was used to verify that $(CGA)_5TGA$ and $GA(CGA)_5$ could adopt each expected structural form in solution (Figure 4.2). The positive band at 265 nm and negative band at 245 nm characteristic of the ps-duplex form are present for both sequences at pH 5.5. The presence of a weak positive band at 280 nm and weak negative band at 260 nm verified the formation of the anti-parallel hairpin form for both sequences at pH 7.0. Together, this confirmed that $(CGA)_5TGA$ and $GA(CGA)_5$ adopted the ps-duplex form at low pH and anti-parallel hairpin form at neutral pH in solution. This CD data was used to inform the selection of the initial search model used in molecular replacement for each diffraction data set.

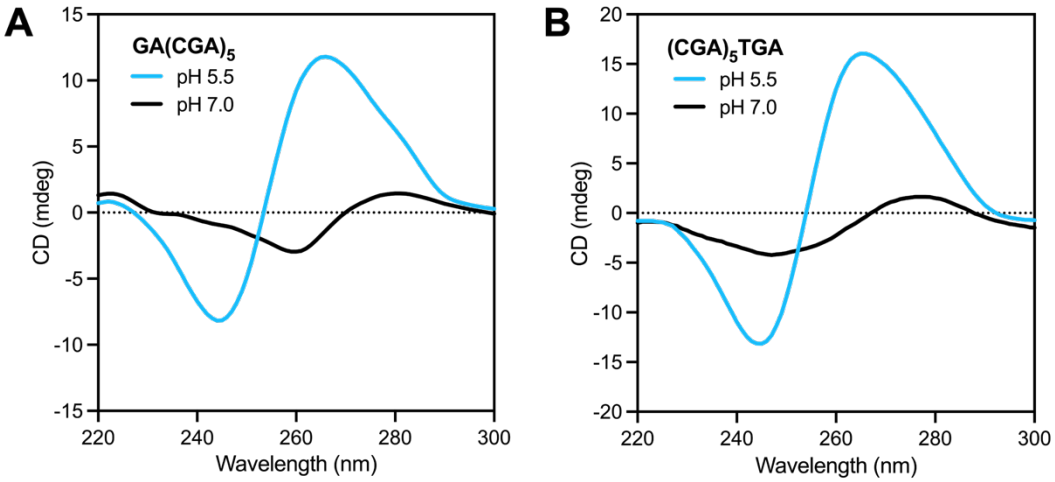

Figure 4.2. CD spectra of $GA(CGA)_5$ and $(CGA)_5TGA$. The positive band at 265 nm and negative band at 245 nm are characteristic of the ps-duplex form[30, 32, 34]. The anti-parallel form has a weak positive band at 280 nm and weak negative band at 260 nm[30, 32, 34]. **(A)** CD spectrum for $GA(CGA)_5$ at pH 5.5 (blue) or pH 7.0 (black). **(B)** CD spectrum for $(CGA)_5TGA$ at pH 5.5 (blue) or pH 7.0 (black). All DNA samples were diluted to 10 µM in buffer containing 100 mM NaCl and 20 mM 20 mM MES (pH 5.5) or 20 mM sodium cacodylate (pH 7.0) as described in Chapter 4.4.2.

Using the d(CGA)-based ps-duplex region of a previously solved structure (PDB: 1IXJ)[35] as a search model for molecular replacement, we determined the crystal structure of $(CGA)_5TGA$ at pH 5.5 in the ps-duplex form at 2.30 Å (Table 4.2). Despite being at a pH that strongly favors the hairpin form in solution (Figure 4.2), we determined the crystal structure of $GA(CGA)_5$ at pH 7.4 also in the ps-duplex form at 1.32 Å (Table 4.2). Two d(CGA) triplets from the $(CGA)_5TGA$ ps-duplex structure were used as the search model in the molecular replacement for the $GA(CGA)_5$ dataset.

Table 4.2. Data collection and refinement statistics.

	$GA(CGA)_5$	**$(CGA)_5TGA$**
PDB ID	7SB8	7T6Y
Sequence	d(GACGACGACGACGACGA)	d(CGACGACGACGACGATGA)
Data Collection Statistics		
Wavelength (Å)	0.979	0.979
Space group	P 1 21 1	C 1 2 1
a, b, c (Å)	19.68, 30.42, 180.82	84.50, 32.35, 32.26
α,β,γ (°)	90, 90.4, 90	90, 91.03, 90
Resolution range (Å)	90.41 - 1.32 (1.36 - 1.32)	32.25 - 2.30 (2.38 - 2.30)
Multiplicity	4.5 (4.4)	1.8 (1.8)
Completeness (%)	94.7 (84.3)	87.2 (88.6)
$I/\sigma I$	9.9 (2.0)	5.7 (5.0)
R_{pim}*	0.049 (0.350)	0.109 (0.288)
$CC_{1/2}$**	0.995 (0.820)	0.940 (0.395)
Refinement Statistics		
No. Reflections	48419 (4340)	3487 (350)
R_{work}	0.171 (0.191)	0.204 (0.225)
R_{free}	0.214 (0.292)	0.242 (0.271)
No. DNA atoms	2100	740
No. ligand atoms	101	10
No. solvent atoms	480	119
RMSD, bond lengths (Å)	0.008	0.009
RMSD, bond angles (°)	0.94	1.00
Average B-factor	18.69	13.22
DNA	16.06	12.47
ligands	33.81	14.80
solvent	27.01	17.97

*Precision-indicating merging R factor.[135] $R_{pim} = \sum_{hkl}\left[\frac{1}{N-1}\right]^{1/2}\sum_i|I_i(hkl) - \overline{I(hkl)}|/\sum_{hkl}\sum_i I_i(hkl)$ where $I_i(hkl)$ is the i_{th} observation of reflection *hkl*.

**Correlation coefficient between reflection intensities from the data set randomly split into two halves.

***Values in parentheses correspond to the high-resolution shell.

Several structures that rely on the C-CH+ hemi-protonation have also crystallized as ps-duplexes at above-neutral pH, suggesting that factors beyond pH influence this structural preference.[37, 136] The high local concentration of DNA and the presence of crowding agents (such polyethylene glycol) have been demonstrated to increase the observed pH of the structural transition in C-CH+ mediated structures.[29, 32, 137] CD measurements of d(CGA)-based repeat sequences are consistent with these observations as the presence of crowding agents shifts the favorability range of the ps-duplex to higher pH (Figure 4.3). Specifically, the addition of 30% PEG2000 increased the pH of the structural transition by 0.33 ± 0.06 and 0.32 ± 0.13 pH units for $(CGA)_5TGA$ and $(CGA)_5$, respectively (Figure 4.3).

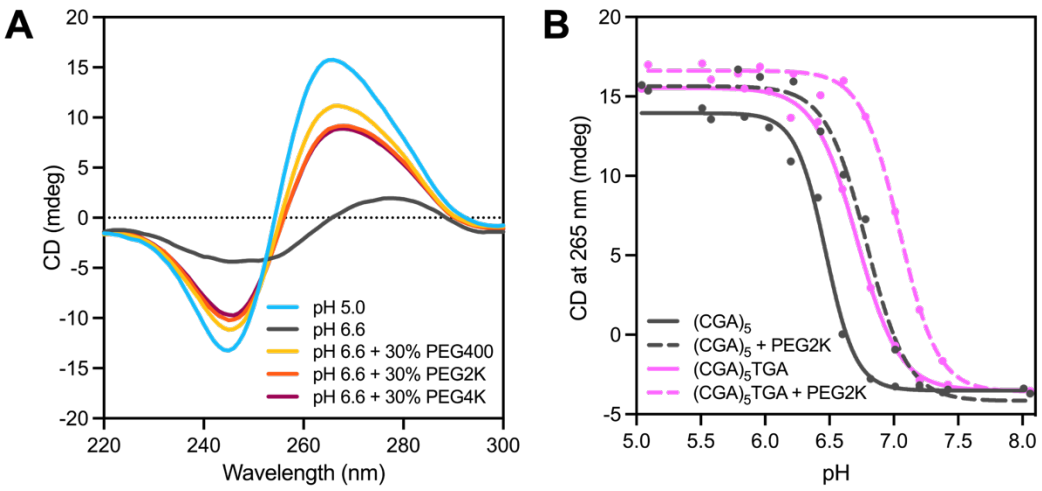

Figure 4.3. Molecular crowding agents impact the pH of the structural transition in d(CGA)-based sequences. (A) $(CGA)_5$ forms a ps-duplex at pH 5.0 (light blue) and anti-parallel hairpin at pH 6.6 (gray) without molecular crowding agents. The formation of the ps-duplex form at pH 6.6 is favored in the presence of 30% PEG400 (yellow), PEG2000 (orange), and PEG4000 (maroon). **(B)** Crowding agents increase the pH of the structural transition from ps-duplex to anti-parallel hairpin form. The transition was measured as the loss of characteristic ps-duplex signal at 265 nm in native conditions (solid lines) or in the presence of 30% PEG2000 (dashed lines) for $(CGA)_5$ (gray) and $(CGA)_5TGA$ (pink).

Thermodynamic measurements (described in Chapter 2) have demonstrated a significantly greater stability in the ps-duplex over the anti-parallel hairpin form.[30] Therefore, it is not surprising that the significantly more stable ps-duplex form is dominant in crowded crystallization conditions where structural stability is advantageous. It may also thus be possible for d(CGA) ps-duplexes to form in crowded cellular environments, similarly to other C-CH+ mediated DNA structures.[67, 138, 139] Despite testing multiple constructs of d(CGA)-derived oligonucleotides, we were unable to determine a structure in the hairpin form.

Chapter 4.2.2. Crystal packing

In the $GA(CGA)_5$ crystal structure, six strands form three parallel-stranded homo-duplexes (Duplex 1, 2, and 3) in the asymmetric unit (Figure 4.4A). Duplex 2 is coaxially stacked between duplexes 1 and 3 though 3′ to 5′ end stacking of the terminal G1-G1 and A17-A17 base pairs. This arrangement results in a junction of three tandem sets of inter-strand G/A stacking interactions at each duplex intersection to stabilize the crystal lattice (Figure 4.4B). This packing arrangement forms columns of alternating ps-duplexes propagating throughout the crystal along the c-axis. Interestingly, this is the first instance of 3′ to 5′ end stacking in this class of ps-duplexes; other ps-duplexes containing the d(CGA) motif stack in the 3′-3′ or 5′-5′ orientation.[36] This difference is likely due to the lack of 5′-C. The exposed 5′-G allows for preferential formation of inter-duplex G/A stacking interactions with the 3′-A of another duplex that directly mimic the internal inter-strand G/A stacking interactions. Symmetry related duplexes can be used to extrapolate the ps-duplex structure beyond 17-nt, but the absence of the

5′-C in this sequence disrupts the internal consistency of the d(CGA) repeating unit. In the $(CGA)_5TGA$ structure, two strands form one homo-duplex (Duplex 4) in the asymmetric unit (Figure 4.4C). The duplex is stacked with crystallography identical duplexes via 5′-5′ stacking of C1-C1 base pairs and 3′-3′ stacking of A18-A18 base pairs.

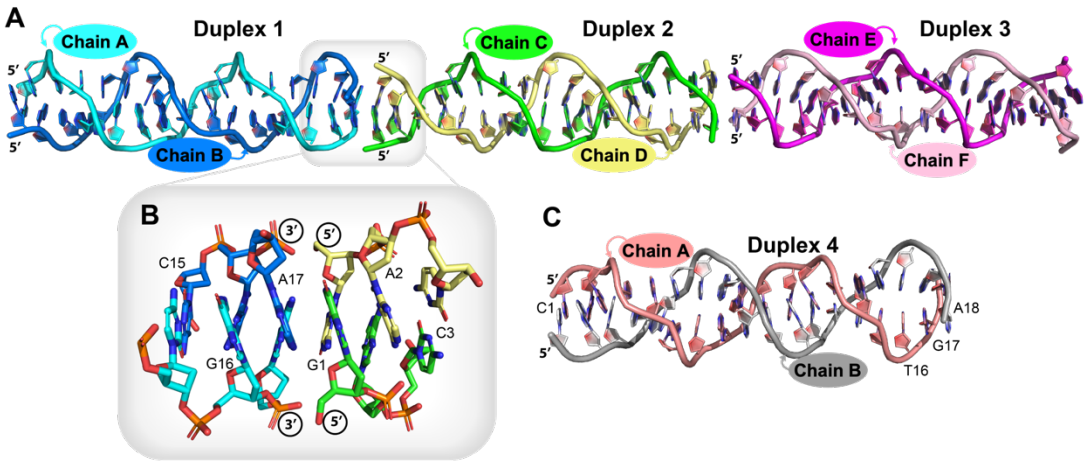

Figure 4.4. Overview of the d(CGA)-based parallel-stranded homo-duplexes. **(A)** The asymmetric unit for $GA(CGA)_5$. The individual chains within each duplex (1-3) are labeled and colored accordingly. Duplex 1: chain A (cyan), chain B (blue); Duplex 2: chain C (green), chain D (yellow); Duplex 3: chain E (magenta), chain F (light pink). **(B)** 3′ to 5′ end stacking of duplex 1 and 2. The 3′ A17-A17 base pair of duplex 1 forms stacking interactions with the 5′ G1-G1 base pair of duplex 2 to form three tandem G/A stacking interactions. **(C)** The $(CGA)_5TGA$ asymmetric unit. Each chain within duplex 4 is labeled and colored accordingly: chain A (salmon), chain B (gray). PyMOL graphics software was used for all figures.[140]

$GA(CGA)_5$ and $(CGA)_5TGA$ crystals both grew in the presence of divalent cations where they primarily mediate inter-duplex crystal packing interactions (Figures 4.5 and 4.6). When possible, anomalous difference maps and coordination distances were used to verify cation identity and placement (Figures 4.5D-F and 4.6C-D).

$GA(CGA)_5$ (duplexes 1-3) crystallized in the presence of cobalt hexammine(III) (NCO) and strontium (Sr^{2+}). Specifically, NCO is positioned in multiple conformations between the Hoogsteen faces of guanines from two different duplexes, where the GN_7-GN_7 and GO_6-GO_6 distances are 8.7 ± 0.3 Å and 7.6 ± 0.1 Å, respectively (Figure 4.5A). Sr^{2+} mediates the remaining inter-duplex guanine positions in two distinct modes. The first set of Sr^{2+} mediated interactions are similar to the NCO positions but the GN_7-GN_7 and GO_6-GO_6 distances are shorter (7.9 ± 0.1 Å and 6.8 ± 0.1 Å, respectively) (Figure 4.5B). The remaining Sr^{2+} cations are similarly positioned between two guanines from separate ps-duplexes, but the Hoogsteen faces are positioned such that GN_7-GO_6 are oriented together (9.31 ± 0.03 Å) (Figure 4.5C).

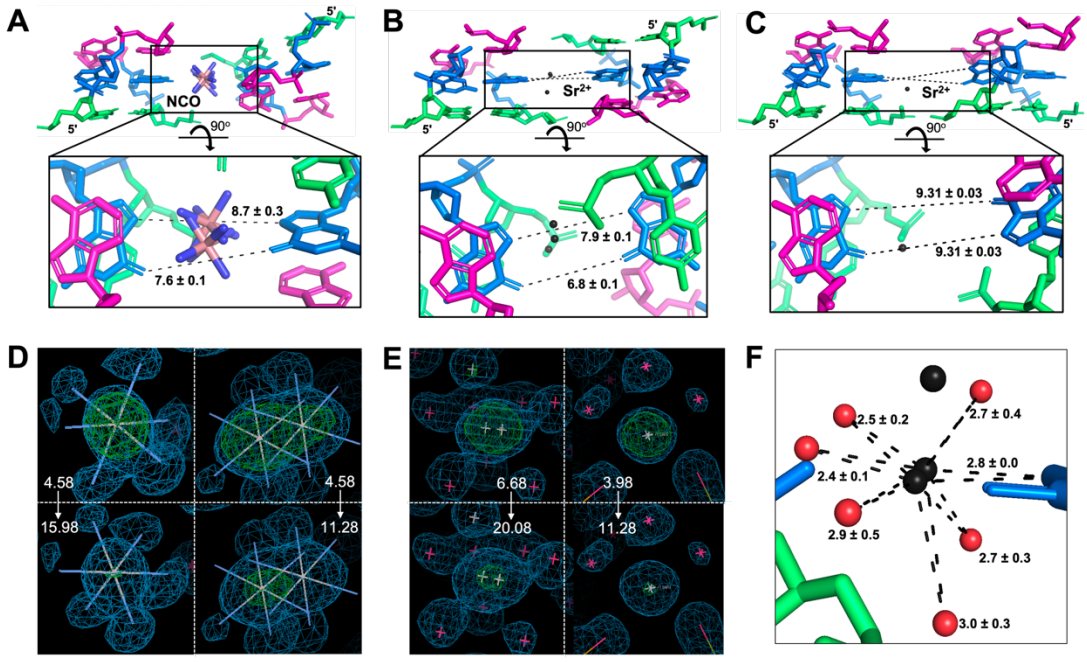

Figure 4.5. Cation mediated crystal packing interactions observed in the GA(CGA)$_5$ crystal structure. Cations **(A)** cobalt hexammine(III) (NCO) and **(B, C)** strontium (Sr^{2+}) are always found positioned between two guanines from different duplexes. d(CGA) triplets surrounding each cation are shown and colored as follows: cytosine (green), guanine (blue), adenine (magenta). The averaged bond distances between cation mediated guanines are shown. **(D)** Anomalous difference maps are shown at high and low σ values for two representative NCO cations. The $2mF_o$-DF_c electron density map is contoured to 1 σ. **(E)** Anomalous difference maps are shown at high and low σ values for two representative Sr^{2+} cations. The $2mF_o$-DF_c electron density map is contoured to 1 σ. **(F)** Average coordination distances are shown for a representative Sr^{2+} cation.

In the $(CGA)_5TGA$ structure, barium (Ba^{2+}) mediates inter-duplex packing in two distinct environments. One mode is almost identical to the first set of Sr^{2+} mediated interactions in the GA$(CGA)_5$ structure (Figure 4.5B), where GN_7-GN_7 and GO_6-GO_6 distances are 7.8 ± 0.1 Å and 6.9 ± 0.0 Å, respectively (Figure 4.6A). The remaining Ba^{2+} cations are positioned between the Hoogsteen face of one guanine and the phosphate oxygen of the opposing duplex guanosine where the GN_7-PO_2 and GN_6-PO_2 distances are 9.4 ± 1.3 Å and 8.9 ± 0.9 Å (Figure 4.6B). Despite the presence of

different cations with unique packing interactions, the resulting ps-duplex structures were highly uniform.

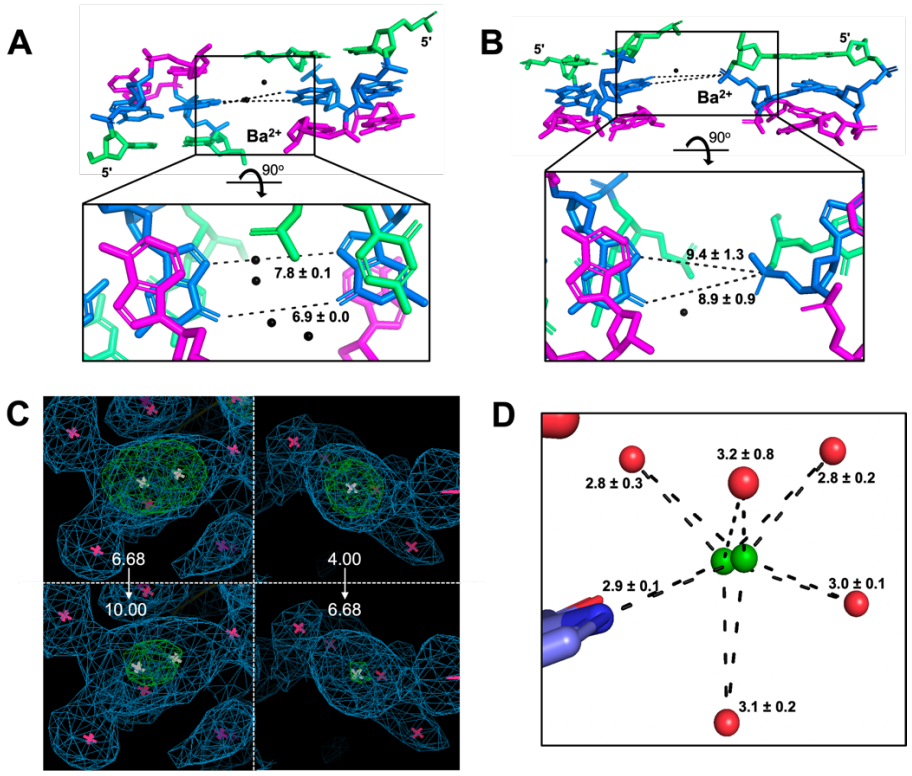

Figure 4.6. Cation mediated crystal packing interactions observed in the $(CGA)_5TGA$ crystal structure. Barium (Ba^{2+}) cations are always found positioned between guanines from different duplexes. **(A and B)** Two types of unique cation interactions are shown. d(CGA) triplets surrounding each cation are shown and colored as follows: cytosine (green), guanine (blue), adenine (magenta). The averaged bond distances between cation mediated guanines are shown. **(C)** Anomalous difference maps are shown at high and low σ values for two representative Ba^{2+} cations. The $2mF_o$-DF_c electron density map is contoured to 1 σ. **(D)** Average coordination distances are shown for Ba^{2+}.

Chapter 4.2.3. d(CGA) ps-duplexes are highly uniform

Though these structures were solved from individual crystals with different DNA sequences, solution conditions, and crystal packing arrangements, the resulting ps-duplex structures are nearly identical over the length of tandem d(CGA) repeats

(Figure 4.7A). The three duplexes from the GA(CGA)$_5$ structure have RMSD values between 0.421 Å and 0.451 Å for 700 atoms fit (Figure 4.7B).

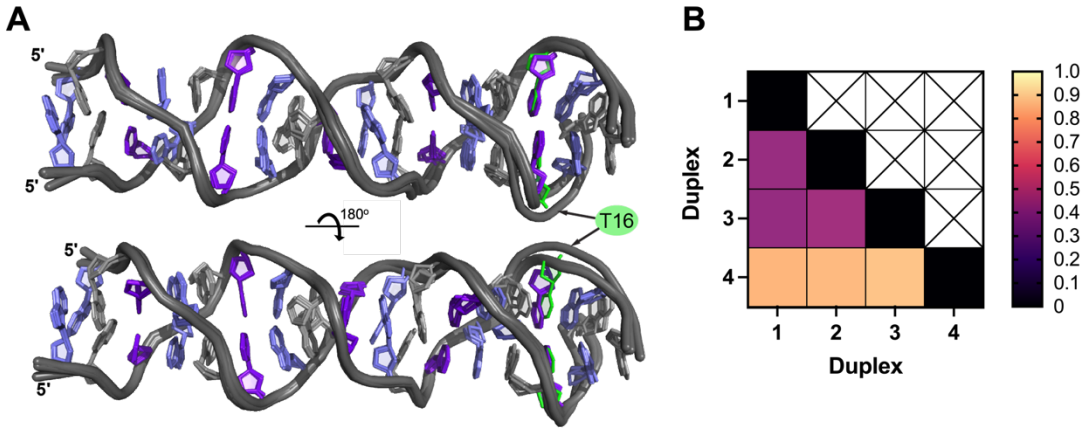

Figure 4.7. d(CGA) triplets in the ps-duplex form are structurally isomorphous. **(A)** Overlay of duplexes 1-4 illustrates the robust structural uniformity of the ps-duplex form across different sequences, solution conditions, and crystal packing arrangements. Structural deviations are primarily observed surrounding the C16T substitution position in (CGA)$_5$TGA. Nucleotides are colored as follows: C (purple), G (light purple), A (gray), T (green). The 5' C-CH+ homo-base pair was omitted from (CGA)$_5$TGA in this overlay for simplicity. **(B)** RMSD values from pair-wise alignment of duplexes 1-4. All ps-duplexes are highly similar with RMSD values below 1.0.

Despite being crystallized in different conditions and containing the C16T substitution, duplex 4 is also highly similar to duplexes 1-3 (respective RMSD values: 0.846 Å, 0.855 Å, 0.877 Å for 698 atoms fit) (Figure 4.7). The structural deviations associated with duplex 4 are primarily observed near the substitution site. Weaker electron density and correspondingly higher B-factors observed from A12 to A18 in duplex 4 may also contribute to increased RMSD values, though the overall ps-duplex structure is maintained. Beyond this, additional structural deviations among duplexes 1-4 arise from subtle differences in the phosphate backbones, likely resulting from solvent interactions that influence crystal packing (Figure 4.8).

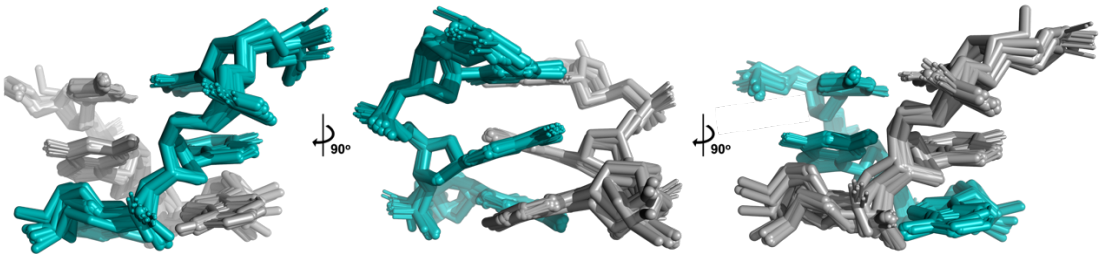

Figure 4.8. Overlay of all d(CGA) triplets from duplexes 1-4. Overlay of all d(CGA) triplets rotated 90º to show the difference in deviation along the phosphate backbone from each strand (colored teal or gray). Minimal overall deviations demonstrate the high structural predictability of the ps-duplex form of the d(CGA) triplet.

We also compared the structures of isolated d(CGA) base-paired triplets from all three duplexes to triplets within each duplex (intra-duplex) or from other duplexes (inter-duplex). Not surprisingly, comparison of all individual d(CGA) triplets results in high similarity as evident by the low RMSD of the full duplexes. Individual d(CGA) triplets at different positions in the same duplex (intra-duplex d(CGA) triplets) are almost identical (0.122 Å to 0.557 Å for 124 atoms fit) (Figure 4.9), indicating that there are no unique position-specific structural features along the helical length. The 3′-end d(CGA) triplet of duplexes 1-3 and 3′-most d(CGA) triplet of duplex 4 are the sources of the largest deviations among intra-duplex triplets (RMSD ranging from 0.634 Å to 0.881 Å for 124 atoms fit). This position in duplexes 1-3 is likely associated with greater deviations due to duplex end flexibility or crystal contact interactions, while deviations in duplex 4 are likely influenced by the structural changes induced by the adjacent d(TGA) triplet.

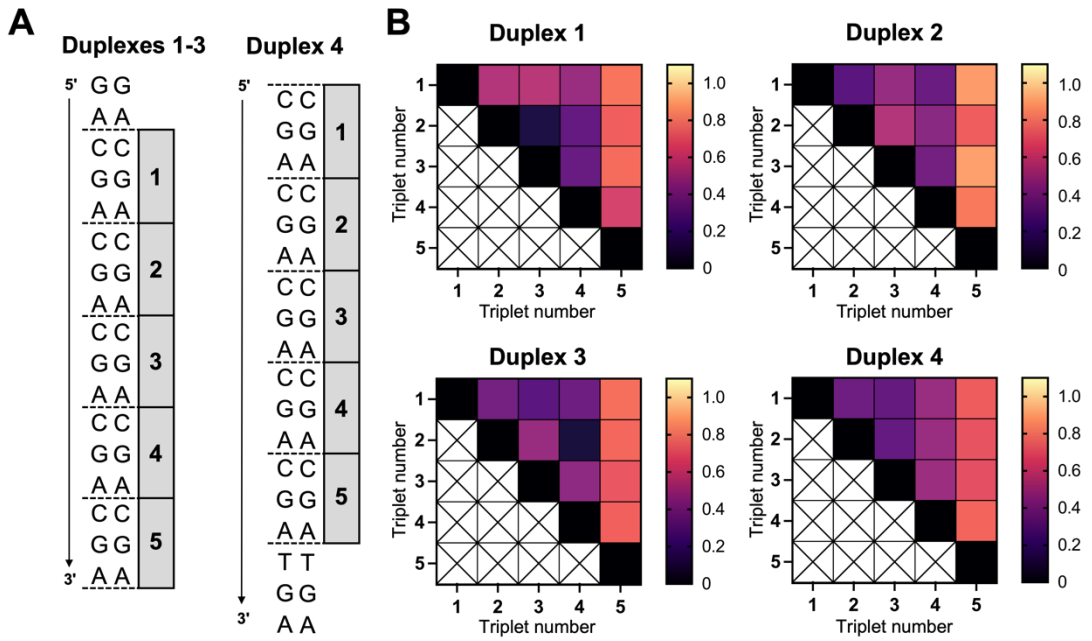

Figure 4.9. Individual intra-duplex d(CGA) triplets are highly uniform. (A) Individual d(CGA) triplets are numbered 1-5 starting at the first complete d(CGA) triplet within each duplex. Specific triplet numbering is shown for duplexes 1-3 (left) and duplex 4 (right). **(B)** RMSD values obtained from the alignment of homo-base paired d(CGA) triplets within the same duplex (intra-duplex) using the triplet numbering scheme from (A). Alignments containing the 3′-d(CGA) triplet (position 5) from each duplex always result in the highest RMSD value, indicating that this position is associated with structural flexibility.

Similarly low RMSD values were observed when comparing individual d(CGA) triplets from different duplexes (inter-duplex) (Figure 4.10). In the inter-duplex comparison, d(CGA) triplets from within duplex 3 exhibited the largest range of RMSD values when compared to triplets from other duplexes (0.459 Å to 1.043 Å for 124 atoms fit). Overall, the low RMSD values in the comparison of tandem and individual d(CGA) triplets illustrates that the ps-duplex form of d(CGA) repeat containing sequences are structurally isomorphous, even in different environmental contexts.

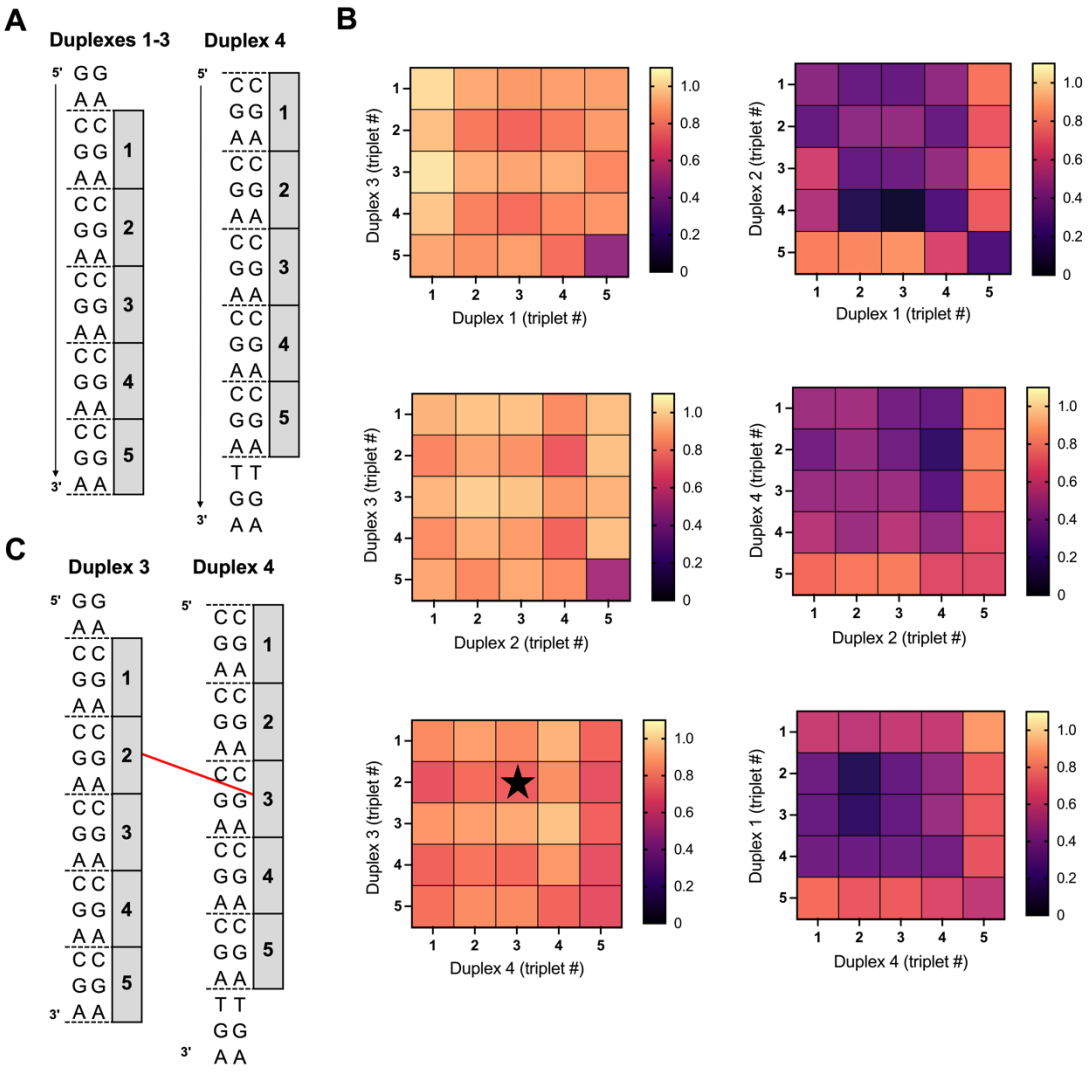

Figure 4.10. Individual inter-duplex d(CGA) triplets are highly uniform. (A) Individual d(CGA) triplets are numbered 1-5 starting at the first complete d(CGA) triplet within each duplex. Specific duplex numbering is shown for duplexes 1-3 (left) and duplex 4 (right). **(B)** RMSD values obtained from the alignment of d(CGA) triplets from different duplexes (inter-duplex). d(CGA) triplets from duplex 3 have the highest RMSD values. **(C)** Schematic representation of one set of d(CGA) triplets compared within duplexes 3 and 4 indicated by the star in B.

Chapter 4.2.4. d(CGA) helical and base parameters

The high degree of similarity of the four unique ps-duplexes obtained from our crystal structures has allowed us to establish helical (Figure 4.11) and base-pair parameters (Figure 4.12) for this motif. Like other d(CGA) and d(TGA) homo-duplex structures[36, 37], the ps-duplexes form right-handed helices that lack distinct major and minor grooves. The d(CGA) ps-duplex form requires 9.0 ± 0.1 base pairs to complete one helical turn resulting in an average helical pitch of 32.2 Å ± 0.5 Å. The decreased helical pitch of the ps-duplex form, compared to B-DNA, is primarily a result of the large helical rise (5.0 Å ± 0.1 Å) and twist (84^{o} ± 1^{o}) associated with the inter-strand G/A base step. As previously observed[37], there is a notable difference in base pair parameters between purine and pyrimidines in the ps-duplex. Purine nucleotides adopt larger shear, propeller, stretch, and buckle angles to accommodate the hydrogen bonds while maintaining the duplex-stabilizing inter-strand G/A base stacking interactions. We quantified the range of helical and base-pair parameters using 3DNA v2.4[141] along each nucleotide position to highlight the periodic fluctuation of each parameter along individual d(CGA) triplets throughout the entire duplex (Figure 4.11 and 4.12).

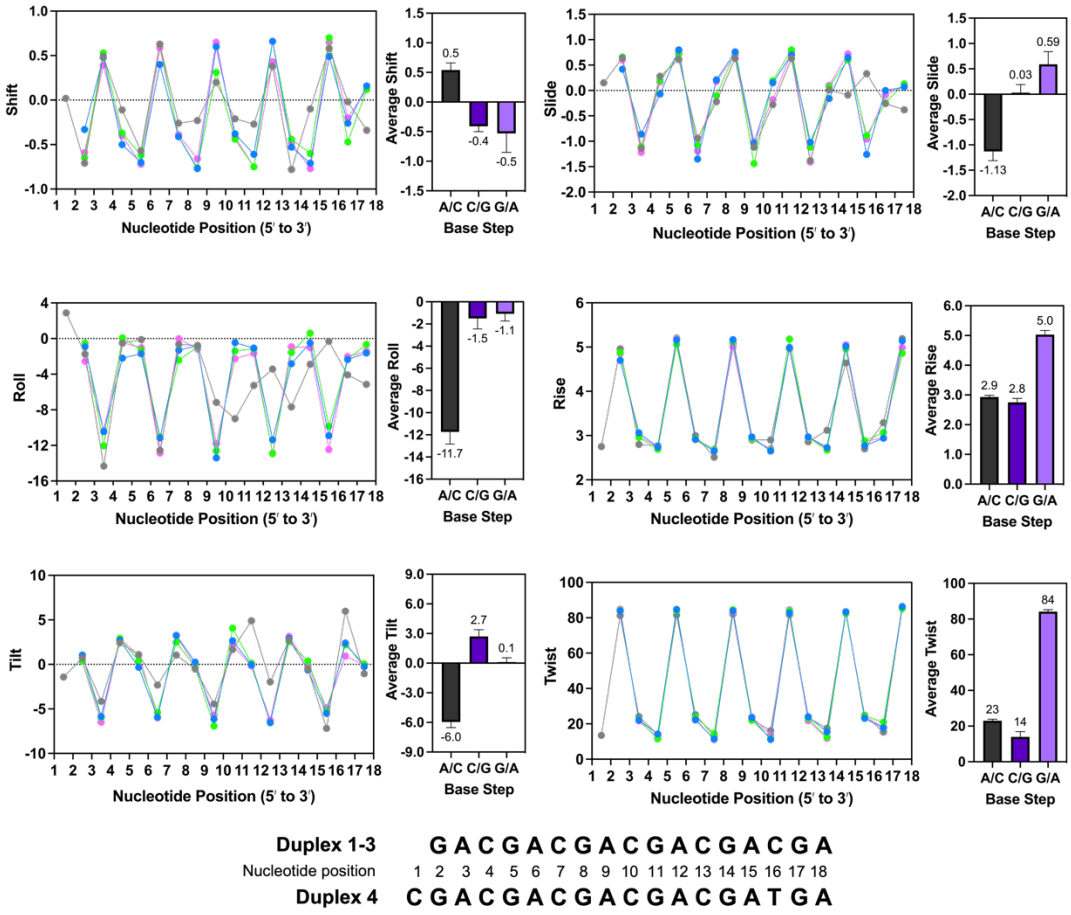

Figure 4.11. Helical parameters. Helical parameters were obtained from 3DNA v2.4[141] and are shown for each base step along duplexes 1 (blue), 2 (green), 3 (pink), and 4 (gray). Bar graphs represent the average and standard deviation of each parameter for A/C, C/G, and G/A base steps in duplexes 1-4. Nucleotide position for each duplex is shown.

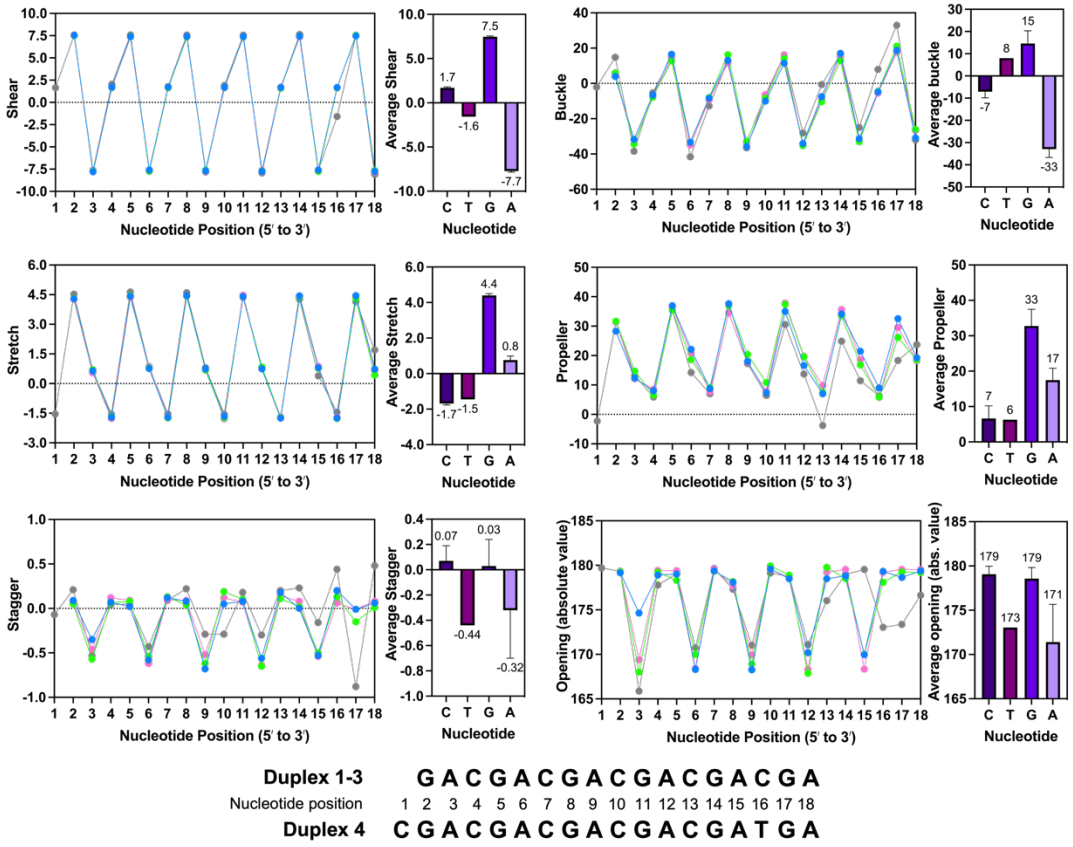

Figure 4.12. Simple base-pair parameters. Simple base-pair parameters were obtained from 3DNA v2.4[141] and are shown for each base pair along duplexes 1 (blue), 2 (green), 3 (pink), and 4 (gray). Bar graphs represent the average and standard deviation of each parameter for each nucleotide in duplexes 1-4. Nucleotide position for each duplex is shown.

We observed distinct ranges of phosphate backbone torsion angles (α, β, γ) among individual strands within each duplex (Figure 4.13). Notably, one strand in each duplex, referred to as the "loose" strand, adopts a wide range of α, β, and γ torsion angles (246° ± 66°, 180° ± 36°, and 91° ± 66°, respectively), while the opposing strand of the duplex, referred to as the "rigid" strand, has a much narrower range (292° ± 5°, 164° ± 14°, and 59° ± 32°, respectively) (Figure 4.13). The ability to adopt a wide range of torsion angles suggests that one strand is generally more flexible than its partner

strand and led us to adopt the nomenclature of loose and rigid, respectively. This also indicates that each strand within the ps-duplex is conformationally unique.

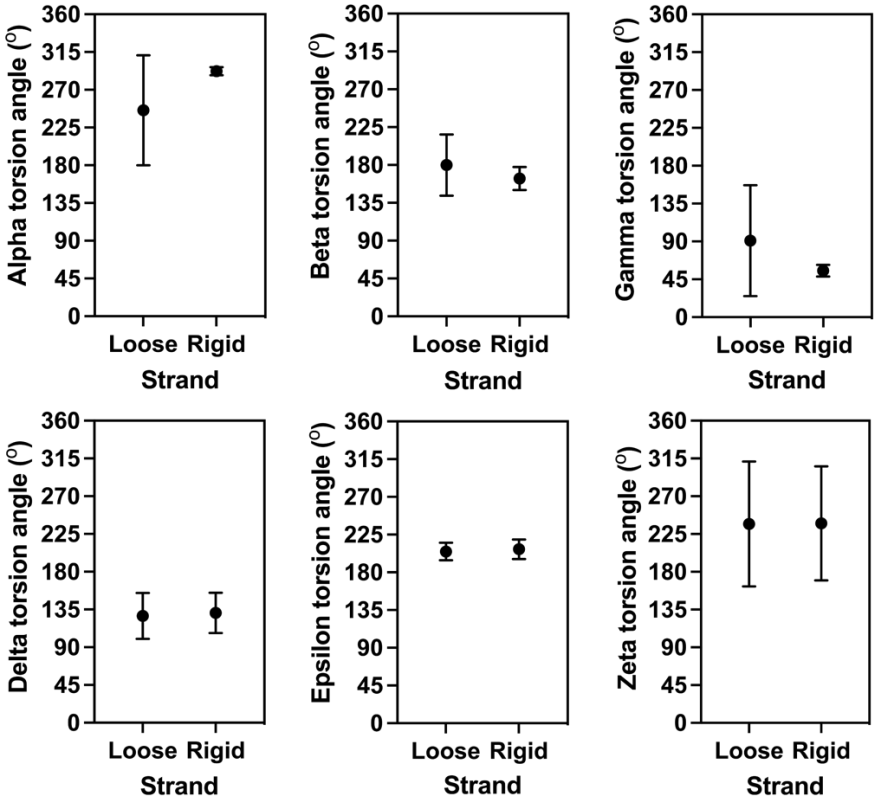

Figure 4.13. Torsion angles. Torsion angles (α, β, γ, δ, ε, and ζ) obtained from 3DNA v2.4[141]. Torsion angles were calculated for all nucleotides in duplexes 1-4, excluding the TGA triplet, in each strand (loose or rigid). There is a difference in the range of α, β, and γ angles depending on the strand, indicating that each strand has unique structural character. Plots illustrate the average and standard deviation of each measured angle.

Chapter 4.2.5. Structural asymmetry is induced by the C-CH+ base pair

We assessed the overall symmetry of the ps-duplex form to determine how structural asymmetries are propagated through the ps-duplex. The linearity of each duplex along base pair units was measured by connecting the mid-point of the hydrogen bonding partners of each homo-base pair (Figure 4.14A). Each d(CGA) triplet

exhibited a similar bending pattern centered around the largest deviation from linearity (25.0° ± 3.9°) at the C-CH+ base pair (Figure 4.14B). The G-G centric angle does not propagate significant deviations from linearity, while the magnitude of the A-A centric deviation is highly dependent upon the identity of the following nucleotide (C or T). When a C-CH+ base pair is present, the adjacent 5′-A-A centric angle adopts a deviation (20.8° ± 4.3°) similar in magnitude to the C-CH+ centric angle. Alternatively, the A-A centric deviation is smaller (9.0°) when followed by a T-T base pair (Figure 4.14B).

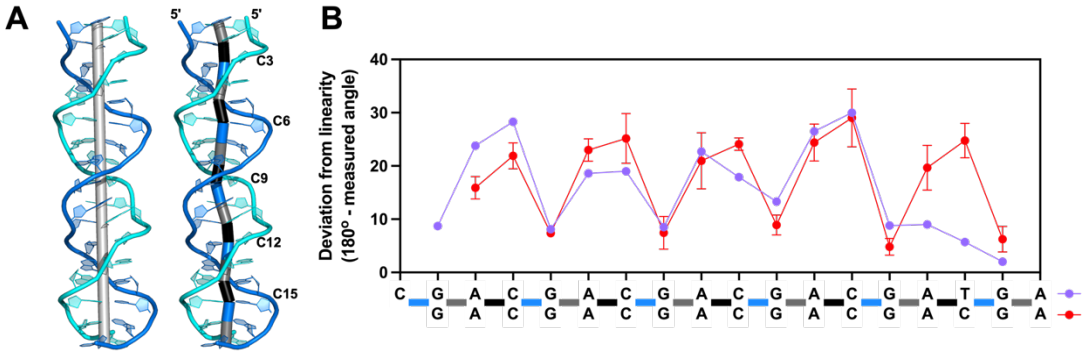

Figure 4.14. Parallel-stranded homo-duplex asymmetry. (A) Deviations from linearity along base pairs of ps-duplex 1. Light gray cylinder represents the helical axis (left). The deviation from linearity of a base pair is measured as the angle between the two cylinders adjacent to the base pair of interest (right). Individual cylinders were created by connecting points placed at the midpoint of the hydrogen bonding partners of each base pair. The resulting cylinders are colored based on the identity of the base pairs they connect; G-A (gray), A-C (black), C-G (blue). **(B)** Deviation from linearity of each base pair along the $(CGA)_5TGA$ (purple) or $GA(CGA)_5$ (red) sequence. Colored bars along the sequence correspond to the same cylinders connecting base pairs from A. The angles measured for $GA(CGA)_5$ are represented as the average of duplexes 1-3 and $(CGA)_5TGA$ is from duplex 4.

Structural overlays of the A/C and A/T base steps within the $(CGA)_5TGA$ structure indicate that this deviation coincides with the extension of one cytidine from the helical axis to align the Watson-Crick faces for the formation of the hemi-

protonated C-CH+ base pair (Figure 4.15), as previously noted.[36] There is also a slight displacement of the adjacent adenosine on the same strand which could be required to accommodate the cytosine deviation. This contrasts with the T-T base pair which makes interactions in a perfectly symmetrical manner; therefore, the adjacent adenosine also remains unbent. We conclude that the A-A base pair provides structural flexibility to accommodate deviations from linearity induced by the C-CH+ base pair.

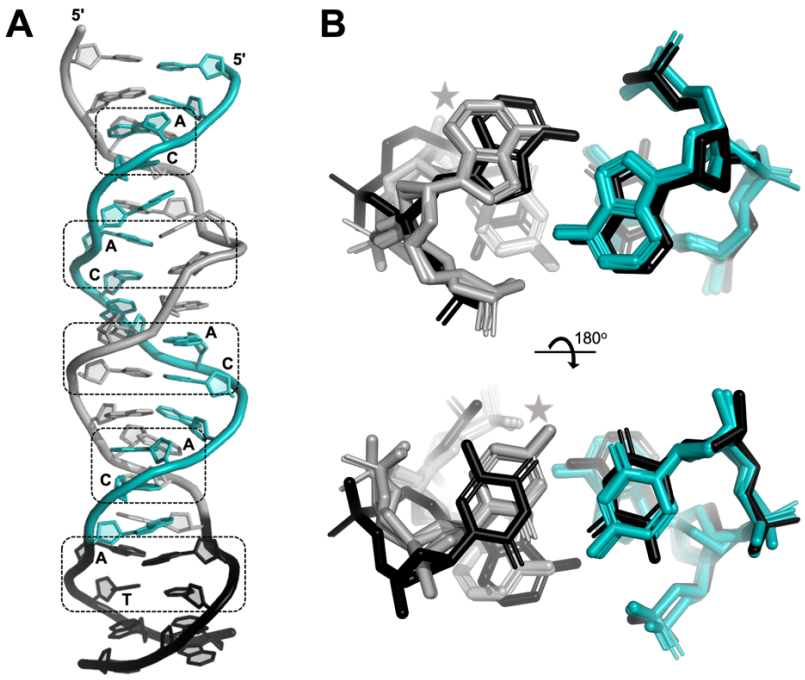

Figure 4.15. A/(C or T) di-nucleotide step within the $(CGA)_5TGA$ duplex. (A) Loose and rigid strands within the $(CGA)_5TGA$ structure are individually colored gray or teal, respectively. The 3′-TGA triplet and 5′-adjacent A-A homo-base pair are colored black. All A/(C or T) base steps within this structure are boxed. **(B)** Overlay of A/(C or T) base steps from (A), rotated 180° to highlight the position of the A-A and C-C base pairs. Nucleotides within the rigid strand (teal) do not show significant structural deviation when compared to the black strand. However, nucleotides within the loose strand (gray, starred) extend away from the helical axis.

Chapter 4.2.6. Each strand has unique structural character

Backbone torsion angle analysis and the duplex asymmetry suggested that the two strands of the ps-duplex have unique structural characteristics. These differences are correlated with two distinct hydrogen bond interactions that form within the A/C step between d(CGA) triplets (Figure 4.16A). The first hydrogen bond is between cytosine N4 (C-N4) and a non-bridging phosphate oxygen (O2P) of the previous adenosine within the same strand. There is no bond equivalent to the C-N4 to O2P bond in the T-T base pair, further suggesting that this bond could be influential in controlling the relative position of the C-CH+ and A-A base pairs. The second hydrogen bond is between the same non-bridging phosphate oxygen and the adenosine N6 (A-N6) of the opposing strand. Interestingly, depending on the strand within each duplex (loose or rigid), there are unique differences in A-N6 to O2P and C-N4 to O2P bond lengths (Figure 4.16B). In the rigid strand, all A-N6 to O2P and C-N4 to O2P bonds distances remain between 2.8 to 3.1 Å. However, the average A-N6 to O2P and C-N4 to O2P distances within the loose strand are increased to 4.1 ± 0.4 Å and 3.5 ± 0.1 Å, respectively. The cytosine that is displaced from the helical axis is always on the loose strand, where the increased bond lengths and wider range of torsion angles within the loose strand coincide with this displacement.

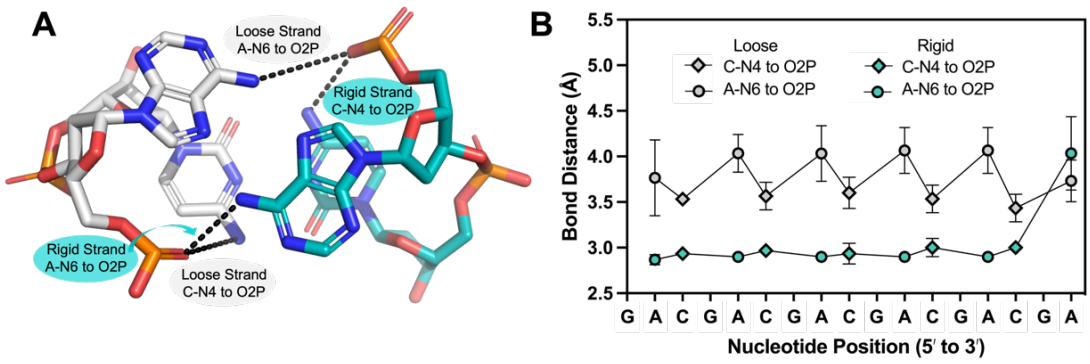

Figure 4.16. Base-to-backbone bond distances distinguish loose and rigid strands **(A)** The A/C step highlighting the A-N6 to O2P and C-N4 to O2P interactions within loose (gray) and rigid (teal) strands. Chain A (duplexes 1 and 4), chain C (duplex 2), and chain E (duplex 3) have been characterized as loose strands. Chain B (duplexes 1 and 4), chain D (duplex 2), and chain F (duplex 3) have been characterized as rigid strands. **(B)** Loose (gray) and rigid (teal) strand bond distances represented along the $GA(CGA)_5$ sequence. A-N6 to O2P distances are plotted as circles and C-N4 to O2P distances are plotted as diamonds. Each data point represents the average and standard deviation of distances measured from duplexes 1-3. Loose strand bond distances alternate between 3.5 ± 0.1 Å and 4.1 ± 0.1 Å depending on the identity of the nucleotide involved in the interaction while rigid strand bond distances remain between 2.8 to 3.1 Å, regardless of the interaction.

Accompanying the differences in hydrogen bonding are distinct differences in base stacking interactions between loose and rigid strands (Figure 4.17A). Base pair overlap areas (excluding exo-cyclic groups) calculated for each duplex using 3DNA v2.4[141] indicate that intra-rigid strand A/C and C/G steps maintain similar overlap areas of 2.4 ± 0.5 $Å^2$ and 2.3 ± 0.4 $Å^2$, respectively (Figure 4.17B). The inter-strand G/A stacking interaction adjacent to the A/C step on the rigid strand also has a similar overlap area of 2.8 ± 0.9 $Å^2$ (Figure 4.17B). However, the stacking areas of the loose strand are more variable. The A/C step on the loose strand has the lowest base overlap area (0.4 ± 0.1 $Å^2$), while the G/A (inter-strand) and C/G (intra-strand) stacking interactions surrounding the A/C step have the highest stacking overlap area (4.6 ± 0.4

Å^2 and 4.0 ± 0.2 Å^2, respectively; Figure 4.17B). The large stacking interactions surrounding the bent A/C step within the loose strand contributes additional stabilization that may compensate for the increased base-to-phosphate hydrogen bond distances.

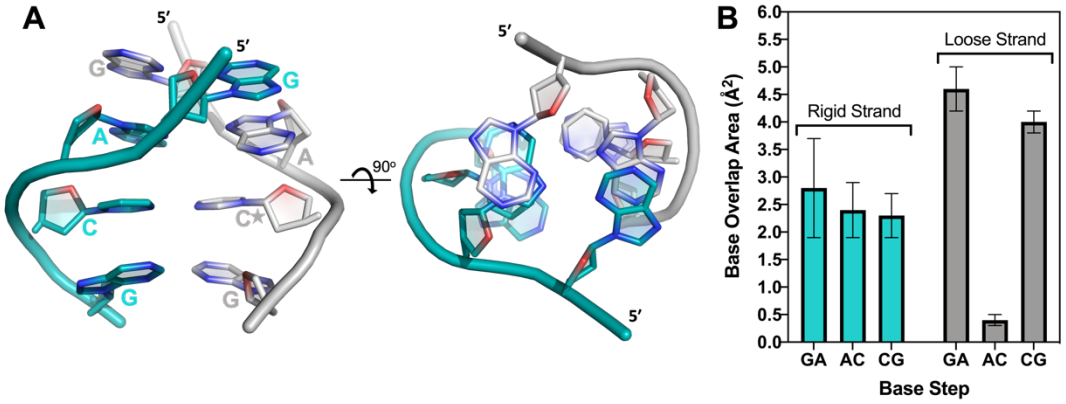

Figure 4.17. Base overlap areas are used to distinguish loose and rigid strands. (A) Base overlap areas are different for loose and rigid strands. View of all unique base stacking interactions (inter-strand G/A, intra-strand A/C, and intra-strand C/G) that contribute to d(CGA) triplet stabilization. 90º rotation illustrates difference in stacking overlap area between strands. The rigid strand (teal) maintains consistent stacking overlap areas, while the loose strand (gray) is highly variable. The star denotes the cytosine that is extended from the helical axis. **(B)** Base stack overlap areas are represented as the average and standard deviation of overlap areas from d(CGA) triplets from duplexes 1-4 and are shown for G/A, A/C, and C/G base steps.

The overall structural asymmetry and accompanying differences in hydrogen bonding and base stacking interactions among strands are observed throughout each ps-duplex studied. Though it would be conceivable to expect the structural asymmetry to be propagated on a per-triplet basis, we observed the propagation on a per-strand basis over the entire length of the d(CGA) repeats. Thus, each ps-duplex is composed of two structurally unique strands where all triplets within a strand adopt either the loose or rigid conformation. The structural homogeneity of triplets within strands

implies that duplexation of tandem d(CGA) triplets could occur in a cooperative manner. Further, the distinct conformations of each strand could play separate roles in accommodating the structural asymmetry. The rigid strand is the structural scaffold strand that maintains consistent hydrogen bonding and stacking interactions, while the loose strand provides structural flexibility to stabilize and accommodate deviations from linearity induced by the C-CH+ base pair.

Chapter 4.2.7. d(YGA) triplets are structurally compatible, but not identical

It was previously hypothesized that d(TGA) triplets could be useful discriminators in the programmable pairing of long stretches of d(CGA) triplets based on slight structural and thermodynamic deviations incurred by the 5′-nucleotide.[30] Previous crystal structures have reported differences in d(CGA)[36] and d(TGA)[37] triplets from separate sequence contexts but have not yet examined the structural compatibility when d(YGA) triplets are present within the same sequence. The crystal structure of $(CGA)_5TGA$ has allowed us to evaluate the structural compatibility of d(CGA) and d(TGA) triplets within the same DNA sequence. We observed that the incorporation of a 3′-d(TGA) triplet significantly alters the bond distances of loose and rigid strands in upstream d(CGA) triplets. Within the d(CGA) triplet directly adjacent to the d(TGA) triplet, the rigid strand C-N4 to OP2 and A-N6 to OP2 bond distances increased from an average of 3.0 Å to 4.4 Å and 4.9 Å, respectively, while the loose strand A-N6 to OP2 distance decreased from 4.0 Å to 3.5 Å (Figure 4.18A). The C1′-C1′ distance for the T-T homo base pair is 1.4 Å wider than the C-CH+ homo-base pair, therefore, upstream swelling of the rigid strand could be required to accommodate the wider T-T homo-base pair (Figure 4.18B).

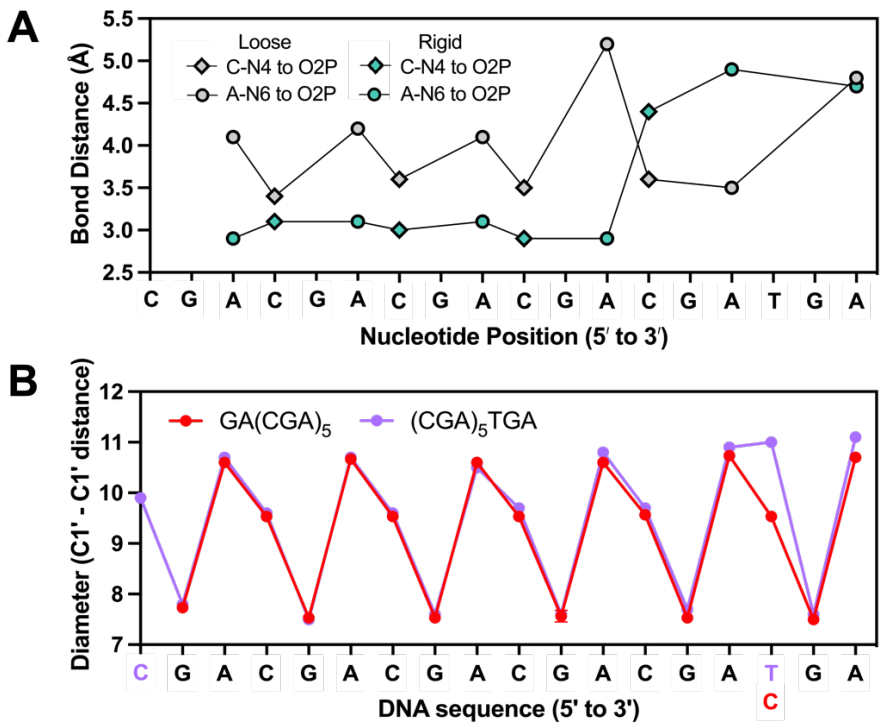

Figure 4.18. Base-to-backbone bond distances and duplex diameter of $(CGA)_5TGA$. (A) Loose (gray) and rigid (teal) strand nucleotide to backbone bond distances plotted along the $(CGA)_5TGA$ sequence. A-N6 to O2P distances are plotted as circles and C-N4 to O2P distances are plotted as diamonds. The rigid strand bond distances significantly increase in the d(CGA) triplet directly upstream of the d(TGA) triplet. **(B)** Duplex diameter as measured by the C1′-to-C1′ distance of each base pair along sequences $GA(CGA)_5$ (red) and $(CGA)_5TGA$ (purple). $GA(CGA)_5$ data points are represented as the average and standard deviation of measurements from duplexes 1-3, while $(CGA)_5TGA$ data points are from duplex 4. There is a significant difference in diameter of C-C vs T-T homo-base pairs.

Increased base overlap areas of the G/A steps adjacent to the d(TGA) triplet could also contribute additional stabilization to compensate for the extended rigid strand bond distances (Figure 4.19). Enthalpic destabilization (described in Chapter 2) observed in sequences containing d(TGA) triplets was previously attributed to the loss of one hydrogen bond from replacing the hemi-protonated C-CH+ with a T-T base pair.[30] The structure described here further suggests this destabilization could also be

due to the loss of the C-N4 to O2P hydrogen bond and swelling of adjacent d(CGA) triplets that coincide with the addition of a T-T base pair. The incorporation of a d(TGA) triplet at the 3′-end of a long stretch of d(CGA) triplets does not disrupt the overall ps-duplex structure but induces slight structural changes in the adjacent d(CGA) triplet. We observed a high range of B-factors (9 – 40 Å^2) for the atoms in the T-T and surrounding homo-base pairs (A15 to A18), indicating that the atoms within the d(TGA) triplet are associated with flexibility within the crystal. This supports conclusions drawn from thermodynamic (Chapter 2) and nuclease sensitivity experiments (Chapter 3) claiming that the T-T homo base pair is associated with inherent flexibility that disrupts surrounding base pair interactions.

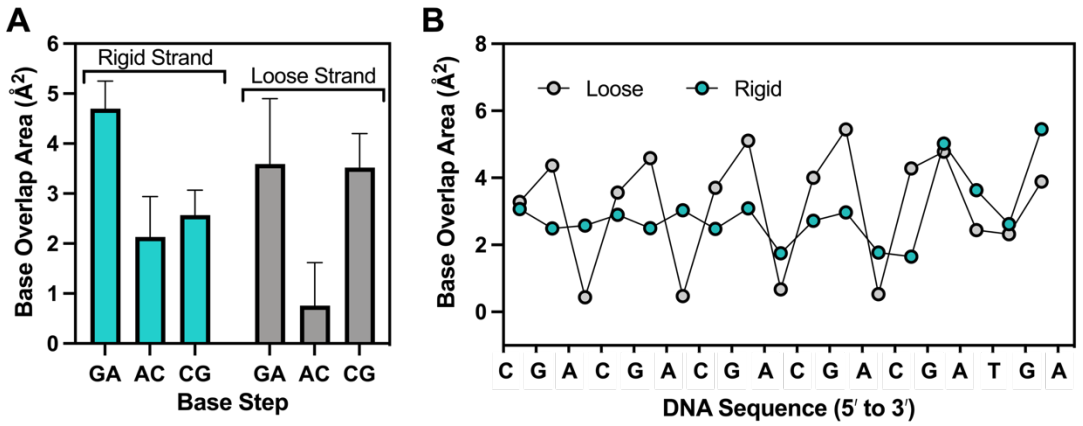

Figure 4.19. Base overlap areas distinguish loose and rigid strands in (CGA)$_5$TGA. **(A)** Base stack overlap areas are represented as averages of overlap areas from d(CGA) triplets in duplex 4 and are shown for each respective base step. **(B)** Intra-strand (A/C or C/G) and inter-strand (G/A) base step overlap areas plotted for each step along the (CGA)$_5$TGA sequence. Teal points represent base overlaps between nucleotides in rigid strands and gray points represent base overlap between nucleotides in loose strands. In the (G/A) inter-strand base step, teal points represent the base overlap of a rigid strand guanine on top of a loose strand adenine in the 5′ to 3′ direction. Gray points represent the reverse; a loose strand guanine stacking on top of a rigid strand adenine in the 5′ to 3′ direction.

Though d(CGA) and d(TGA) triplets are not structurally identical within the ps-duplex, they could be used to control rigid and loose strands. Interestingly, 5-Br-UGA triplets have been shown to offer increased stability to the ps-duplex via the formation of a halogen bond with the phosphate oxygen of an adjacent adenosine.[37] This indicates the valuable potential of the 5-Br-UGA triplet in the rational design of d(CGA) containing ps-duplexes. To fully evaluate the potential use as discriminator triplets, structural analysis of d(CGA)-based repeat sequences containing internal d(TGA) triplets (and 5-Br-UGA triplets) are needed to understand the effect of their incorporation on downstream d(CGA) triplets.

Chapter 4.3. Summary and conclusions

The crystal structures described here have allowed us to characterize the d(CGA) triplet repeat motif in the ps-duplex form and establish structural features for its use as a building block in DNA nanotechnology applications. The generalized helical and base parameters established by these structures will serve as constraints for the incorporation of d(CGA)-based triplets into rational structure design. Particularly, requiring an integer number of base pairs per turn (9.0 ± 0.1 base pairs) simplifies its use from a design perspective, as incorporation of three d(CGA) repeats completes exactly one helical turn.

Consistent with previous d(CGA) base-paired triplets, we observed a structural asymmetry that is propagated throughout each duplex. Our crystal structures containing multiple tandem d(CGA) triplets have demonstrated that this asymmetry is propagated on a per-strand basis, where all triplets within each strand adopt a specific conformation and play a unique role in accommodating the asymmetry. Specifically, the rigid strand

serves as the structural scaffold that maintains hydrogen bonding and stacking interactions, while the loose strand provides structural flexibility to stabilize and accommodate deviations from linearity induced by the C-CH+ base pair. The distinct structural character between triplets within rigid and loose strands could be a useful tool in the structural programming of DNA based architectures where specific control of each strand within the duplex may be desirable.

The structural similarity of tandem and individual d(CGA) base-paired triplets obtained from different solution environments demonstrates that the ps-duplex form is a robust and highly predictable structure. These desirable structural characteristics in combination with thermodynamic stability (Chapter 2), and nuclease resistance (Chapter 3) strongly suggests that the d(CGA) motif can be used to reliably integrate the ps-duplex form into nanostructures. This motif has the added benefit of allowing conditional control of the ps-duplex form in solution through mild fluctuations in pH or the addition of crowding agents. In the next chapter, I will discuss the kinetics of the pH-dependent structural transition between hairpin and ps-duplex form.

Chapter 4.4. Experimental procedures

Chapter 4.4.1. Oligonucleotide synthesis and purification

DNA oligonucleotides were synthesized on a 1 μmol scale using standard phosphoramidite chemistry on an Expedite 8909 Nucleic Acid Synthesizer (PerSeptive Biosystems, Framingham, MA) with reagents from Glen Research (Sterling, VA). Oligonucleotides were purified by denaturing 20% (19:1) acrylamide/bis-acrylamide and 7 M urea gel electrophoresis. DNA bands were identified and excised using UV

shadowing. DNA was electroeluted from gel slices, ethanol precipitated, and dialyzed against deionized water.

Chapter 4.4.2. Circular Dichroism (CD) Spectroscopy

Circular dichroism (CD) spectra were obtained using a Jasco J-810 spectropolarimeter fitted with a thermostatted cell holder (Jasco, Easton, MD). Samples were prepared using 10 μM DNA in 20 mM MES (pH 5.5) or 20 mM sodium cacodylate (pH 7.0), with 100 mM sodium chloride. Samples were incubated at 4°C overnight before data collection. Data were collected at room temperature using a 1 mm pathlength cuvette at wavelengths from 220 to 300 nm and represented as the average of three individual scans.

Chapter 4.4.3. Oligonucleotide crystallization

$(CGA)_5TGA$ was crystallized by mixing 2 μL of 200 μM DNA solution with 2 μL of crystallization solution (20% 2-methyl-2,4-pentanediol (MPD), 120 mM barium chloride, 30 mM sodium cacodylate, pH 5.5). $GA(CGA)_5$ was crystallized by mixing 1 μL of 125 μM DNA solution with 2 μL of crystallization solution (8% PEG400, 96 mM strontium chloride, 32 mM lithium chloride, 8 mM hexamminecobalt(III) chloride, 24 mM sodium cacodylate, pH 7.4). Crystallization was performed in sitting drops, equilibrated against 300 μL of 30% MPD or PEG400 (for $(CGA)_5TGA$ and $GA(CGA)_5$, respectively) in the well reservoir, and incubated at 22°C. Crystals were observed within 7 days of plating.

Chapter 4.4.4. Data collection, processing, and structure determination

Crystals were removed from drops with nylon cryo-loops, immediately dipped in the respective crystallization condition supplemented with 30% MPD or PEG400 and plunged into liquid nitrogen. Diffraction data were collected at the Advanced Photon Source (APS), Argonne National Laboratory. $(CGA)_5TGA$ data was collected on the 24-ID-E beamline and $GA(CGA)_5$ data was collected on the 24-ID-C beamline.

Data processing for $(CGA)_5TGA$ was carried out with iMosflm[142] and $GA(CGA)_5$ was carried out with XDS[143] and Aimless.[144] Initial phases were obtained by molecular replacement using Phaser.[145] The parallel stranded homo-duplex d(CGA) triplet region from PDB id: 1IXJ[35] was used as the search model for $(CGA)_5TGA$, and two tandem d(CGA) units from the refined $(CGA)_5TGA$ structure were used as the search model for $GA(CGA)_5$. Model building and refinement was carried out in Phenix[146] and Coot[147], respectively, for both datasets. Data collection and refinement statistics are given in Table 4.2.

Chapter 5: Kinetic analysis of the pH-induced structural transition

Chapter 5.1. Introduction

Chapter 5.1.1. Overview

In this chapter, I will describe our analysis of the pH-dependent structural transition observed in oligonucleotides containing the d(CGA) triplet repeat motif. To do this, we initially obtained the pH range over which the ps-duplex to anti-parallel hairpin transition occurs (transition point pH range) for several d(CGA)-based sequences containing d(CGA) and/or d(TGA) triplets. We demonstrated that d(CGA) repeat number (n) and d(TGA) composition can be used to tune the pH range and sensitivity of the transition between structural forms. Using the transition point pH ranges as a guide, we collected kinetic information on the transition between each structural form using time-dependent CD spectroscopy. We note that the transitions between anti-parallel hairpin and ps-duplex are not completed within the same time scale in each direction (hairpin to ps-duplex vs. ps-duplex to hairpin). Specifically, the transition from ps-duplex to anti-parallel hairpin was completed within seconds, while the anti-parallel hairpin to ps-duplex transition was completed within an hour. We also demonstrated that the structural transition is reversible over several cycles of pH fluctuation. This work offers insight into the dynamic structural versatility of the d(CGA) triplet repeat motif and provides further constraints for inclusion of this motif in the rational design of pH-controllable nanodevices.

Chapter 5.1.2. Sequence and variant design

We designed two classes of d(CGA)-based repeat sequences to characterize the pH range of the structural transition (Table 5.1). The first subset of sequences was designed to examine the extent to which repeat number (n) in $d(CGA)_n$ sequences influenced the transition point pH range. These sequences were also used in kinetic analysis of the transition between structural forms. The second class contained d(CGA)-based sequences which incorporated d(TGA) triplets. This allowed us to identify the extent to which pyrimidine identity and frequency impacted the transition point pH range. Due to distinct differences in thermodynamic stability (Chapter 2), CD spectra (Chapter 2), and nuclease sensitivity (Chapter 3) we expected to observe corresponding changes in the transition point pH range for sequences containing a higher number of d(TGA) triplets.

Table 5.1. Two classes of d(CGA)-based oligonucleotides used to study the pH-dependent structural transition

Class 1: $d(CGA)_n$ oligonucleotides	
DNA Name	**Sequence (5′ to 3′)**
$(CGA)_7$	d(CGA-CGA-CGA-CGA-CGA-CGA-CGA)
$(CGA)_6$	d(CGA-CGA-CGA-CGA-CGA-CGA)
$(CGA)_5$	d(CGA-CGA-CGA-CGA-CGA)
$(CGA)_4$	d(CGA-CGA-CGA-CGA)
$(CGA)_3$	d(CGA-CGA-CGA)
Class 2: $(YGA)_6$ oligonucleotides	
DNA Name	**Sequence (5′ to 3′)**
$(CGA)_5TGA$	d(CGA-CGA-CGA-CGA-CGA-<u>T</u>GA)
$(TGA)_2(CGA)_4$	d(<u>T</u>GA-<u>T</u>GA-CGA-CGA-CGA-CGA)
$(CGATGA)_3$	d(CGA-<u>T</u>GA-CGA-<u>T</u>GA-CGA-<u>T</u>GA)
$(TGA)_6$	d(<u>T</u>GA-<u>T</u>GA-<u>T</u>GA-<u>T</u>GA-<u>T</u>GA-<u>T</u>GA)

*Underlined nucleotides differ from the $(CGA)_n$ pattern.

Chapter 5.2. Results and discussion

Chapter 5.2.1. The pH-induced structural transition

The anti-parallel hairpin and ps-duplex structures adopted by d(CGA)-based sequences have been differentiated by CD spectroscopy (Figure 5.1A)[30], but the pH range at which this structural transition occurs has not been thoroughly studied. CD spectroscopy is a useful approach used to detect the structural transition due to characteristic differences in CD signal between the ps-duplex and anti-parallel hairpin forms. The ps-duplex has a prominent positive band at 265 nm and negative band at 245 nm (Figure 5.1B), while in the anti-parallel hairpin form the peak intensities significantly decrease and shift to 280 nm and 260 nm (Figure 5.1C).[32, 120]

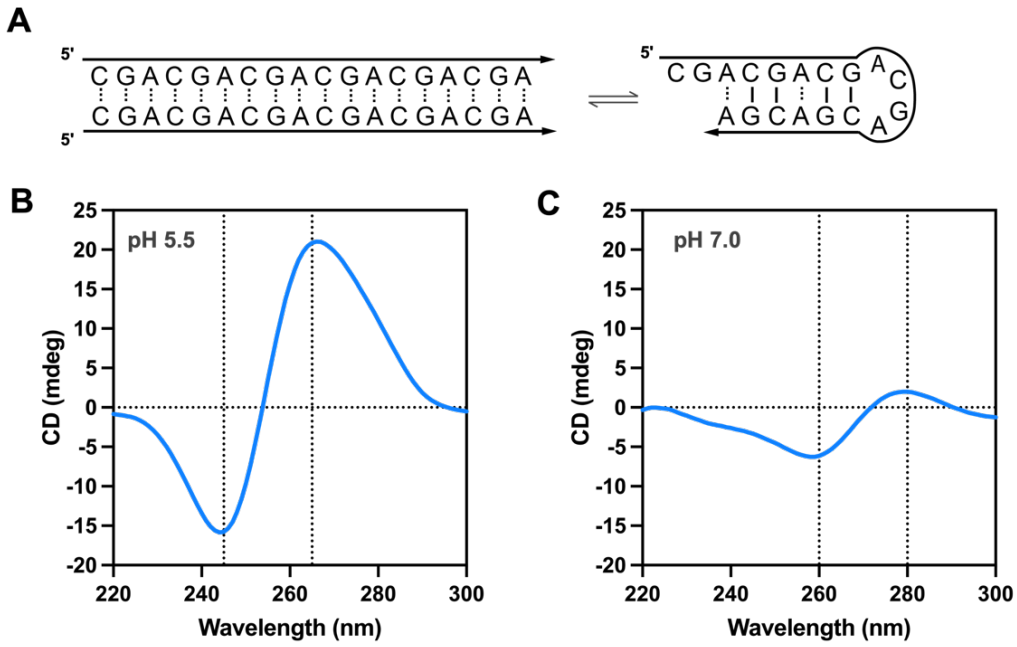

Figure 5.1. CD spectroscopy is used to distinguish the distinct conformations of $(CGA)_6$. (A) Schematic representation of the ps-duplex and anti-parallel hairpin formed by $(CGA)_6$ at pH 5.5 and 7.0, respectfully. **(B)** CD spectra of the ps-duplex form contains characteristic strong negative band at 245 nm and strong positive band at 265 nm. **(C)** CD spectra of the anti-parallel hairpin form contains the characteristic weak negative band at 260 nm and weak positive band at 280 nm.

To obtain the transition point pH range for all sequences, we monitored the intensity of the CD signal at 265 nm as pH was adjusted. CD signal at this wavelength is characteristically intense in the ps-duplex form, while it is approximately zero in the anti-parallel hairpin form (Figure 5.1). Therefore, any signal at this wavelength will directly report on the presence of the ps-duplex form. DNA samples were equilibrated in buffers ranging from pH 5.0 to 8.0 and CD signal at 265 nm was plotted as a function of pH to obtain the transition point pH ranges (described in Chapter 5.4.2). The transition point pH was specifically defined as the pH at which 50% of the sample population adopted the ps-duplex form in reference to maximum CD signal for a given sample.

In the first class of $(CGA)_n$ sequences with repeat number (n) from 3 to 7, we observed that the overall CD signal at 265 nm was proportional to the number of d(CGA) triplets in the sequence (Figure 5.2). There was a corresponding decrease in CD signal at 265 nm as d(CGA) repeat number decreased. This is likely due to the reduction of ps-duplex character in solution as ps-duplex-forming triplets are removed in sequences with lower repeat number. Additionally, we observed well-defined, sharp structural transitions as pH was adjusted for $(CGA)_n$ oligonucleotides where n = 4 to 7 (Figure 5.2). Each dataset was fit to a sigmoidal equation (described in Chapter 5.4.2) and the transition point pH range was calculated to be within pH 6.3 to 6.7 (Figure 5.2).

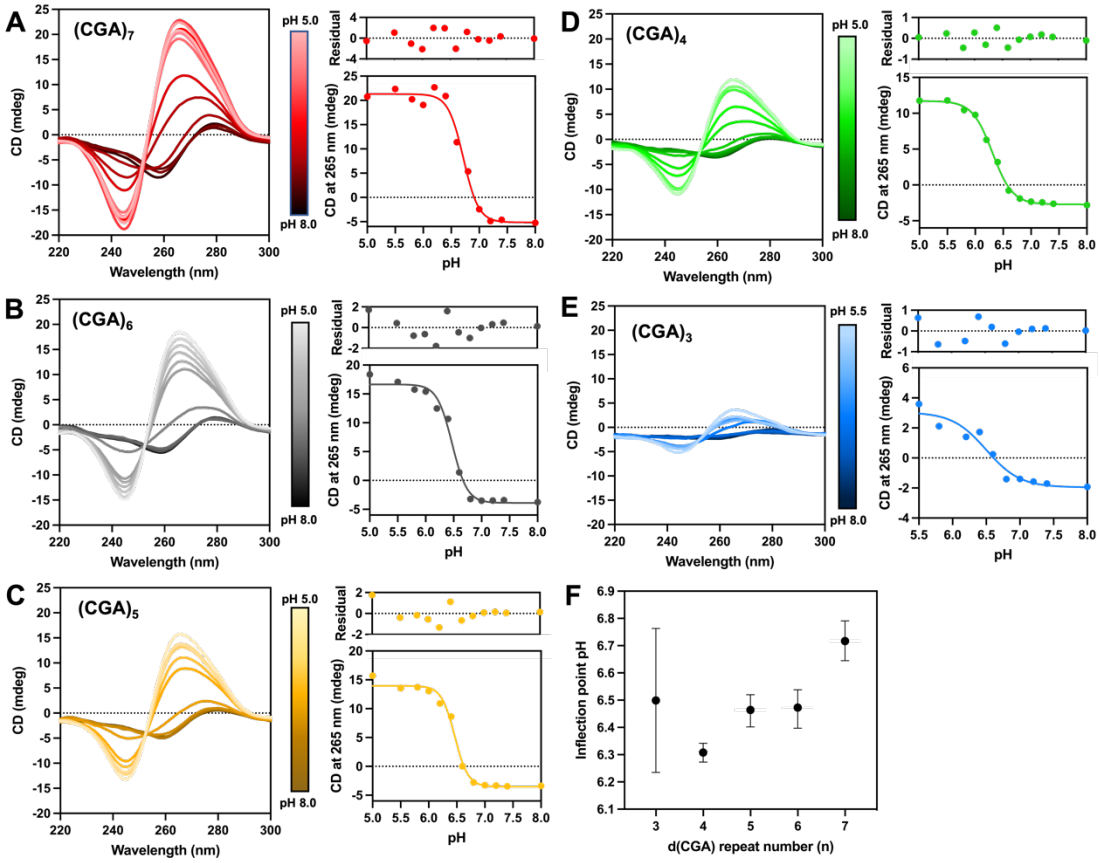

Figure 5.2. CD spectra of d(CGA)$_n$ sequences as a function of pH. CD spectra (left) for each (CGA)$_n$ sequence is shown where repeat number (n) equals **(A)** 7, **(B)** 6, **(C)** 5, **(D)** 4, or **(E)** 3. For each sequence, the CD signal at 265 nm was plotted versus pH and fit to the equation described in Chapter 5.4.2 to obtain the transition point pH (right). Residuals of each fit are shown. **(F)** Transition point pH plotted versus d(CGA) repeat number. The large error range associated with (CGA)$_3$ is likely due to the broadened transition and inadequate fit.

We note that as the number of d(CGA) triplets was increased there was a slight corresponding increase in transition point pH (Figure 5.2F). This is likely attributed to the additional hemi-protonated C-CH+ base pair added with each additional d(CGA) triplet unit. Conversion from ps-duplex to anti-parallel hairpin is assumed to be induced by the disruption of all C-CH+ base pairs. Therefore, it is reasonable that the pH of the structural transition occurred at slightly higher pH as additional C-CH+ base pairs

requiring deprotonation were present. $(CGA)_3$ also exhibited a pH-dependent structural transition, but the slope of the transition was broader than sequences containing ≥ 4 repeat units. The CD spectra for $(CGA)_3$ at neutral pH also lacked the expected bands for the anti-parallel hairpin form and appeared to be unstructured. This indicates that three d(CGA) triplet repeat units may not be sufficient to form the hairpin structure. Overall, we note that $(CGA)_n$ triplet repeat unit number (n) is a useful parameter for fine-tuning the specific pH at which the structural transition from ps-duplex to anti-parallel hairpin occurs.

Dual wavelength parametric tests (Wallimann plots[148]) were constructed to confirm the hypothesized two-state model, in which only two populations exist (ps-duplex or anti-parallel hairpin). For $(CGA)_n$ sequences where $n \geq 4$, the CD intensity at one characteristic wavelength was plotted against the CD intensity at another characteristic wavelength. This was performed for three representative wavelengths including 265 nm, 245 nm, and 260 nm. The linearity of each plot confirmed the cooperative, two-state pH-dependent structural transition between ps-duplex and anti-parallel hairpin for sequences where $n \geq 4$ (Figure 5.3).

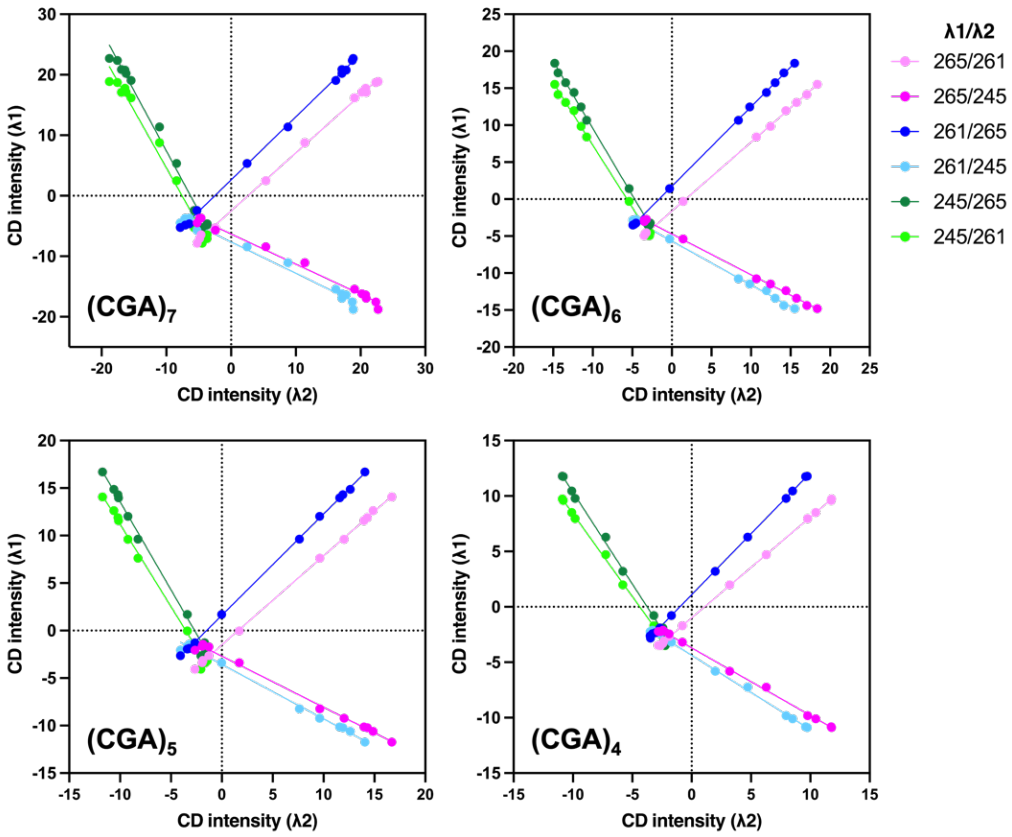

Figure 5.3. Wallimann plots for $(CGA)_n$ where n = 4 – 7. Three characteristic wavelengths (265 nm, 260 nm, and 245 nm) were used for Wallimann analysis[148]. Linear regression was performed and R^2 values ≥ 0.9845 for each pair support the two-state model in $(CGA)_n$ sequences where n = 4 – 7.

Similar CD experiments were performed for the second class of d(CGA)-based sequences containing d(TGA) triplets. We observed that the intensity of the CD signal at 265 nm decreased as the number of incorporated d(TGA) triplets increased, suggesting that the ps-duplex form was disrupted or destabilized as d(TGA) triplets were introduced (Figure 5.4). This result is consistent with previous results (Chapter 2) indicating that the incorporation of d(TGA) triplets decreases the thermodynamic stability of the ps-duplex form.

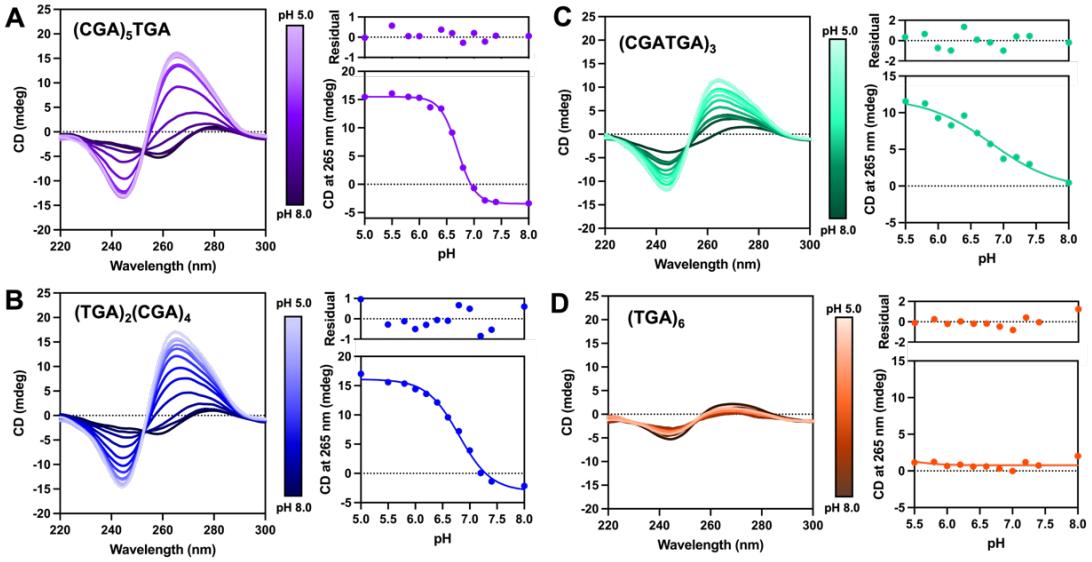

Figure 5.4. CD spectra of d(YGA)$_6$ sequences as a function of pH. CD spectra (left) for (YGA)$_6$ sequences at pH 5.0 to 8.0 is shown for **(A)** (CGA)$_5$TGA, **(B)** (TGA)$_2$(CGA)$_4$, **(C)** (CGATGA)$_3$, and **(D)** (TGA)$_6$. For each sequence, the CD signal at 265 nm was plotted versus pH and fit to the equation described in Chapter 5.4.2. to obtain the transition point pH (right). Residuals of each fit are shown.

The slope of the pH-dependent transition point was broadened and became less cooperative as the number of d(TGA) triplets increased within a d(CGA)-based sequence. Specifically, we observed sharp, two-state transitions for sequences containing ≤ 2 d(TGA) triplets with the transition point pH calculated to be between pH 6.7 and 6.8. We observed a significant broadening of the transition point slope and decrease in linearity in Wallimann plots as the d(TGA) composition within a d(CGA)-based sequence increased beyond 50%. The CD spectra for (TGA)$_6$ did not change over the pH range tested. These results indicated that there is a threshold (≤ 50% sequence composition) to the number of d(TGA) triplets that can be incorporated within a sequence in order to retain the ability to cooperatively transition between ps-duplex and anti-parallel hairpin forms.

Chapter 5.2.2. The ps-duplex to anti-parallel hairpin transition

We performed time course experiments to assess the rate of the pH-induced structural transition from ps-duplex to anti-parallel hairpin. $(CGA)_6$ was equilibrated in CD buffer at pH 6.0 to initially populate the ps-duplex form (Figure 5.5A). The pH of the solution was manually adjusted to pH 7.0 and the CD signal corresponding to the ps-duplex form at 265 nm was monitored for 20 minutes as described in Chapter 5.4.3. Following the pH jump, we observed that the CD signal dropped to zero immediately after the 15-second mixing period (Figure 5.5B). Following the complete 20-minute reaction period, CD spectra had signals characteristic of the anti-parallel hairpin form at 260 nm and 280 nm (Figure 5.5A). Due to limitations in instrumentation, we could not collect faster initial time points. Although we could not fit these data to obtain the rate of the transition from pH 6.0 to 7.0, we qualitatively concluded that the ps-duplex to anti-parallel hairpin structural transition is rapid and completed within 15 seconds of the pH increase.

Additional experiments using stopped-flow instrumentation coupled to the CD spectrometer could be performed to access the millisecond time scale data that was inaccessible in our original experimental setup. This would generate a more complete dataset that could be fit to obtain the quantitative rate of the ps-duplex to anti-parallel hairpin structural transition. Stopped-flow instrumentation specifically allows for accurate automated mixing and rapid CD measurements. The proposed experimental set up would improve upon the current approach as it would not be limited by manual pH adjustment, mixing, and delayed scan start time.

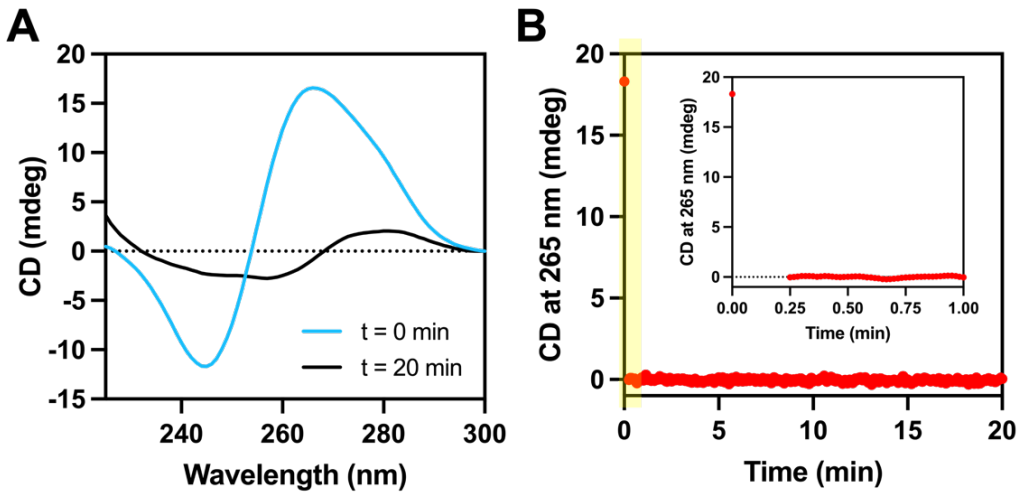

Figure 5.5. The structural transition from ps-duplex to anti-parallel hairpin is rapid. (A) CD spectra at t = 0 min (blue) confirmed that $(CGA)_6$ populated the ps-duplex form prior to pH adjustment. Full CD spectra collected after the time course at t = 20 min (black) has signal characteristic of the anti-parallel hairpin form. **(B)** Time course measurement of CD signal at 265 nm. Immediate decrease in signal indicates that the transition is rapid. **(B, inset)** Highlighted section of the time course zoomed in to show that the transition occurred within 15 seconds of the pH increase. Due to limitations in instrumentation, we could not collect more rapid initial timepoints.

Chapter 5.2.3. The anti-parallel hairpin to ps-duplex transition

We performed a similar time course experiment to assess the rate of ps-duplex formation. $(CGA)_6$ was equilibrated in CD buffer at pH 7.0 to initially populate the anti-parallel hairpin form (Figure 5.6). The pH was manually adjusted to 6.0 and CD signal was monitored at 265 nm for 40 minutes. We observed that the formation of the ps-duplex form is comprised of at least two phases (Figure 5.6). Due to the same instrumentational limitations described in 5.2.2., we could not obtain initial time points faster than 15 seconds. As such, we could not confidently obtain kinetic rates for this structural transition. Stopped-flow CD would be a useful approach to overcome this issue, as described in Chapter 5.2.4.

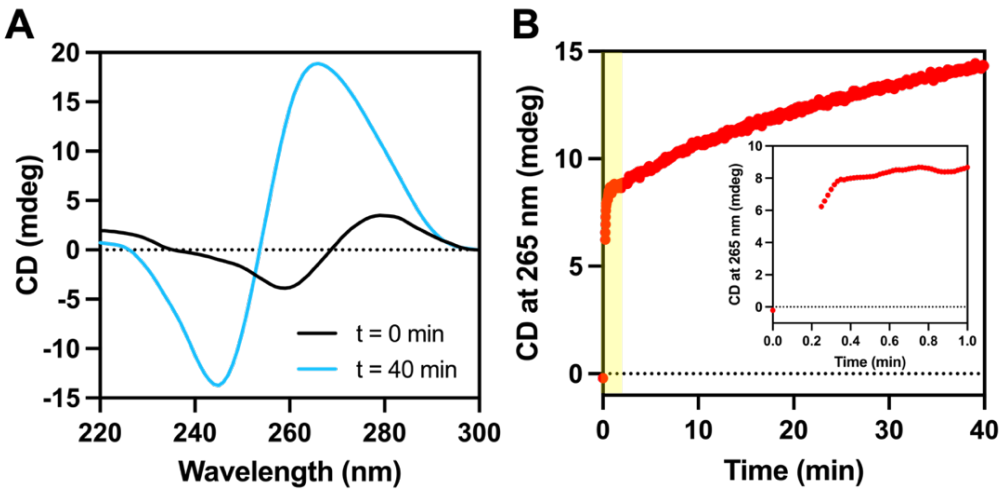

Figure 5.6. Time course of the structural transition from anti-parallel hairpin to ps-duplex. (A) CD spectra at t = 0 min (black) confirmed that $(CGA)_6$ populated the anti-parallel hairpin form prior to pH adjustment. CD spectra collected after the 40 min time course (blue) has signal characteristic of the ps-duplex form. **(B)** Time course measurement of CD signal at 265 nm after pH is adjusted from 7.0 to 6.0. Immediate sharp fast phase is followed by a slow second phase. **(B, inset)** Highlighted region of the time course measurement zoomed in to highlight the fast phase. Due to limitations in instrumentation, we could not collect more rapid initial time points.

Qualitatively, we observed that the formation of the ps-duplex was apparently bi-phasic. The kinetic profile contained an initial fast phase (k_{fast}) that occurred within 30 seconds of the pH decrease, followed by a second slow phase (k_{slow}). We hypothesized that the fast phase was associated with the dissociation of the anti-paralle hairpin followed by initial hybridization of two strands to form a ps-duplex. Because the $(CGA)_6$ sequence is comprised of several identical d(CGA) triplets, the initial association of the two strands might not immediately produce a fully registered 18 base-paired duplex (Figure 5.7). The lower signal at 265 nm is consistent with the number of d(CGA) homo-base pairs present (Figure 5.2). The initial association could result in a frameshifted duplex comprised of one or more single-stranded d(CGA) triplet overhangs on each strand (similarly described in Chapter 3). If this were the case, the

second slower phase could reflect the dynamic correction of frameshifted duplexes to form fully registered ps-duplexes (Figure 5.7).

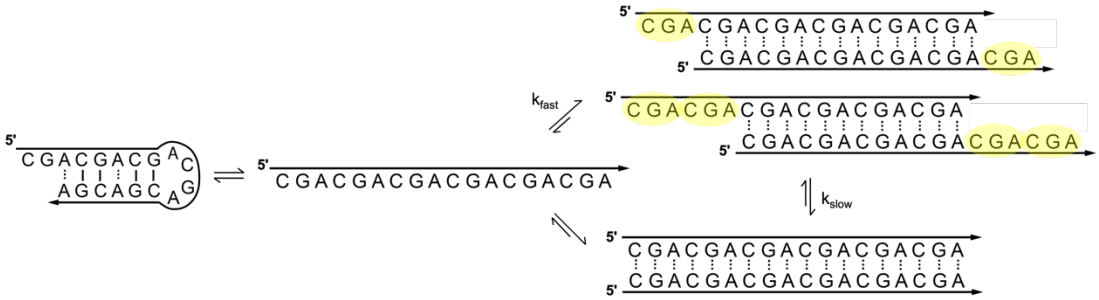

Figure 5.7. Hypothesized mechanism for ps-duplex association. The anti-parallel hairpin initially dissociates into single stranded DNA following the decrease in pH. The single strands hybridize to either form a frameshifted or perfectly registered duplex. We hypothesize that the fast phase (k_{fast}) observed in the time course experiments is due to the initial association of single strands to form frameshifted duplexes. Two potential frameshifted duplexes are shown comprised of either a 15-bp or 12-bp duplex region. The single stranded d(CGA) overhangs of each frameshifted duplex are highlighted in yellow. Dynamic triplet re-registration over time (k_{slow}) could lead to the formation of the fully aligned 18-bp ps-duplex.

To test this, we performed similar time course experiments with $d(CGA)_n$ sequences containing different d(CGA) repeat number (n = 4 to 7). We hypothesized that decreasing repeat number would reduce the population of frameshifted duplexes, correspondingly altering the slow phase. Each $d(CGA)_n$ variant had unique time course profiles (Figure 5.8A). Sequences containing low repeat number, $(CGA)_4$ and $(CGA)_5$, appeared to have a fast phase followed by a slow phase that plateaued within 20 minutes. CD spectra of $(CGA)_4$ and $(CGA)_5$ at the end of the 40-minute time course were superimposable with CD spectra of samples that had been equilibrated in pH 5.5 buffer overnight (Figure 5.8B). This indicated that $(CGA)_4$ and $(CGA)_5$ fully equilibrated to the ps-duplex form following the pH decrease within 40-minutes.

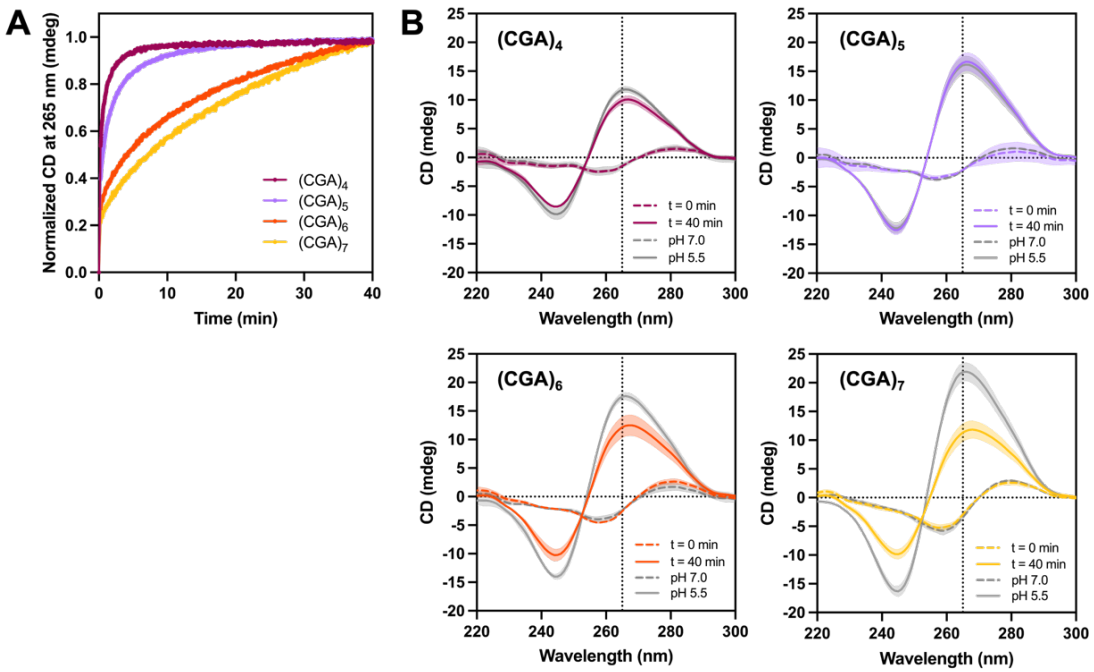

Figure 5.8. Time course of the structural transition from anti-parallel hairpin to ps-duplex for $(CGA)_n$ (n = 4 – 7) oligonucleotides. (A) Time course measurement of CD signal at 265 nm after pH was adjusted from 7.0 to 6.0 for $(CGA)_4$ (magenta), $(CGA)_5$ (purple), $(CGA)_6$ (orange), and $(CGA)_7$ (yellow). Data was normalized to the intensity of the ps-duplex at the end of the time course and represents the average of three individual trials. Each variant has a unique rate profile containing an initial fast phase followed by a second slow phase. The slow phase plateaus within 20 minutes for $(CGA)_4$ and $(CGA)_5$. The second phase does not plateau within the 40-minute time course for $(CGA)_6$ or $(CGA)_7$. **(B)** CD spectra of samples at t = 0 min (colored, dashed lines) and t = 40 min (colored, solid lines) overlaid with samples equilibrated at pH 7.0 (gray, dashed lines) or pH 5.5 (gray, solid lines). All spectra are shown as the average and standard deviation of three individual samples. $(CGA)_4$ and $(CGA)_5$ spectra at t = 40 min superimpose with the equilibrated samples, while $(CGA)_6$ and $(CGA)_7$ do not.

However, this was not observed in sequences containing higher repeat number. For $(CGA)_6$ and $(CGA)_7$, the initial fast phase was followed by a slow phase that did not completely plateau within the 40-minute period. At the end of the 40-minute time course, CD spectra of $(CGA)_6$ and $(CGA)_7$ did not superimpose with the same samples that had equilibrated to the ps-duplex form overnight (Figure 5.8B). The intensity of

the CD signal at 265 nm for $(CGA)_6$ and $(CGA)_7$ was lower than in the equilibrated samples, indicating that the time course sample contained less ps-duplex character than the fully equilibrated samples. This is likely a result of the formation of frameshifted ps-duplexes that have not completely equilibrated to the perfectly registered 18-bp duplex within the 40-minute period.

Chapter 5.2.4. Reversibility

We probed whether d(CGA)-based repeat sequences could undergo reversible pH-dependent structural switching. To do so, we monitored the CD signal at 265 nm where the signal is positive for the ps-duplex form and near-zero in the anti-parallel hairpin form. The experimental time points were selected based on the relative time scales obtained for each structural transition presented in Chapters 5.2.2. and 5.2.3. It was evident that $(CGA)_4$ can reversibility switch between the ps-duplex and anti-parallel hairpin form over several cycles of pH fluctuation (Figure 5.9).

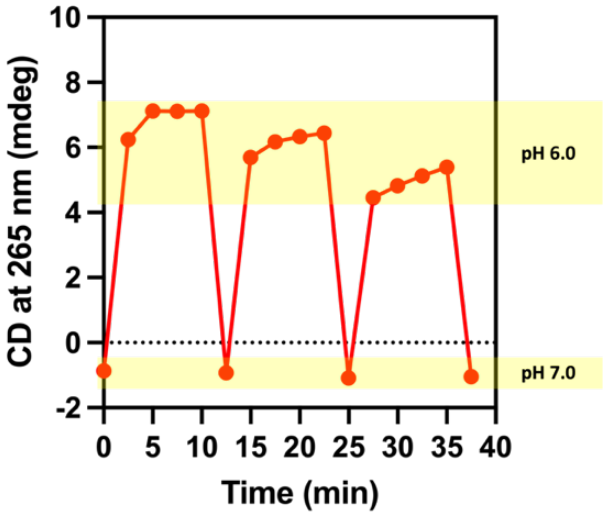

Figure 5.9. $(CGA)_4$ undergoes reversible structural changes driven by pH. $(CGA)_4$ was subjected to several cycles of pH adjustment. Differences in CD signal at 265 nm at each pH demonstrate structural reversibility. The decrease in signal intensity over

time is due dilution effects following each pH adjustment. This data represents one of three reproducible trials.

Chapter 5.3. Summary and conclusions

Here, I have described our CD spectroscopy-based approach to characterize the pH-dependent structural transition of d(CGA)-based oligonucleotides. Our results revealed that $d(CGA)_n$ repeat number (n) can be adjusted to fine-tune the pH at which the structural transition occurs. Additionally, d(TGA) triplets can be incorporated within d(CGA)-based sequences to adjust pH-sensitivity. Kinetic analysis demonstrated that the time scale of the transition is dependent upon the direction of the structural change. Specifically, the formation of the anti-parallel hairpin occurred within seconds, while the formation of perfectly registered ps-duplexes occurred on a longer time scale likely due to triplet frameshifting. CD spectroscopy also confirmed that the structural transition is reversible over several cycles of pH fluctuation. Together, the analysis of the pH range, kinetics, and reversibility of the transition between ps-duplex and anti-parallel hairpin forms provides additional constraints for the incorporation of the d(CGA) triplet repeat motif into the rational design of pH-dependent nanodevices. In the next chapter, I will describe our efforts to optimize a 3D DNA crystal, containing parallel-stranded d(GGA) interactions, as a drug delivery vehicle.

Chapter 5.4. Experimental procedures

Chapter 5.4.1. Oligonucleotide synthesis and purification

DNA oligonucleotides were synthesized on a 1 μmol scale using standard phosphoramidite chemistry on an Expedite 8909 Nucleic Acid Synthesizer (PerSeptive

Biosystems, Framingham, MA) with reagents from Glen Research (Sterling, VA). Oligonucleotides were purified by denaturing 20% (19:1) acrylamide/bis-acrylamide and 7 M urea gel electrophoresis. DNA bands were identified and excised using UV shadowing. DNA was electroeluted from gel slices, ethanol precipitated, and dialyzed against deionized water.

Chapter 5.4.2. Circular dichroism (CD) spectroscopy – equilibrated scans

Circular dichroism (CD) spectra were obtained using a Jasco J-810 spectropolarimeter fitted with a thermostatted cell holder (Jasco, Easton, MD). Samples were prepared using 10 μM DNA in 20 mM sodium cacodylate (pH 5.0 to 7.4) or 20 mM sodium phosphate (pH 8.0), with 100 mM sodium chloride. Samples were incubated at 4°C overnight before data collection. Data were collected at room temperature using a 1 mm pathlength cuvette at wavelengths from 220 to 300 nm and represented as the average of three scans. CD intensity at 265 nm was plotted versus pH for each sample and the data was fit to the following sigmoidal equation using Prism to obtain the transition point pH. "HillSlope" denotes the steepness of the curve between the top and bottom plateaus. In all samples, "TransitionpH" refers to the midpoint of the of the transition curve at which 50% of the sample population adopted the ps-duplex form in reference to maximum CD signal for a given sample.

$$Y = bottom\ plateau + \frac{top\ plateau - bottom\ plateau}{1 + 10^{(\log(transitionpH) - x) * HillSlope}}$$

Chapter 5.4.3. Circular dichroism (CD) spectroscopy – kinetics

Circular dichroism (CD) spectra were obtained using a Jasco J-810 spectropolarimeter fitted with a thermostatted cell holder (Jasco, Easton, MD). DNA

was diluted to 1 μM in buffer containing 100 mM sodium chloride and 20 mM sodium cacodylate, pH 7.0 or 6.0, to a final volume of 400 μL. Samples were equilibrated overnight at 4ºC. A full spectra scan from 220 to 300 nm was collected prior to pH adjustment. To begin the time course, 2 μL of 2.0 M HCl or 2.0 M NaOH was injected into the cuvette. The solution was manually mixed via pipettor for 15 seconds prior to starting data acquisition. Data were collected at room temperature using a 1 cm pathlength cuvette at 265 nm in 1 second increments for up to 40 minutes. Final pH values were confirmed after each pH adjustment and reproducibility was ensured using three individually prepared and tested samples.

Chapter 5.4.4. Circular dichroism (CD) spectroscopy – reversibility

Circular dichroism (CD) spectra were obtained using a Jasco J-810 spectropolarimeter fitted with a thermostatted cell holder (Jasco, Easton, MD). $(CGA)_4$ was diluted to 1 μM in buffer containing 100 mM sodium chloride and 20 mM sodium cacodylate, pH 7.0 to a final volume of 400 μL. The sample equilibrated for 18-20 hours at 4ºC. Data were collected at room temperature using a 1 cm pathlength cuvette at wavelengths from 220 to 300 nm. To adjust the pH, 2 μL of 2.0 M HCl or 2.0 M NaOH was added to the DNA sample. The solution was mixed by pipettor for 15 seconds following each pH adjustment. Scans were collected in increments of 2.5 mins. Final pH values were confirmed after each pH adjustment and reproducibility was ensured using three individually prepared and tested samples.

Chapter 6: Self-assembled non-canonical DNA crystals as drug delivery vehicles

Chapter 6.1. Introduction

Chapter 6.1.1. Overview

Well-established knowledge of the structural programmability, predictability, and stability of Watson-Crick interactions has led to their reliable use in the design of DNA nanoarchitectures[149-151]. A similarly thorough fundamental understanding of non-canonical interactions can be used to fuel the confident incorporation non-Watson-Crick motifs into the creation of well-structured DNA architectures. Dr. Paukstelis was the first to describe a self-assembled 3D DNA crystal stabilized by both Watson-Crick base pairs and non-canonical d(GGA)-based parallel-stranded interactions[42]. The interactions that stabilize the d(GGA) ps-duplex region within these crystals are strikingly similar to the ps-duplex form of the d(CGA) motif. The robust formation of the d(GGA) parallel-stranded interaction-based 3D DNA crystals combined with the biophysical and structural data presented within this dissertation provides strong evidence for the reliable use of d(NGA)-based parallel-stranded motifs in the rational design of self-assembled DNA nanoarchitectures. This inspired us to further evaluate the potential of non-canonical interaction-based structures to function in practical applications.

In this chapter, I describe our characterization of the self-assembled d(GGA)-based 3D DNA crystals as a new platform for the storage and delivery of therapeutic molecules. Our results show that non-canonical interaction-based 3D DNA crystals can load and retain a high concentration of doxorubicin cargo without significant

morphological changes or crystal size reductions. We can tune the physical properties of the loaded crystal depending on the concentrations of doxorubicin and magnesium present during loading. Additionally, we confirmed that the doxorubicin cargo can be released from the crystal. The extent and time scale of cargo release depends upon the amount of doxorubicin loaded within the crystal, divalent and monovalent salt concentrations, and incubation temperature. This study provides the first evaluation of non-canonical interaction-based, self-assembled 3D DNA crystals as vehicles for storage and release of doxorubicin cargo. The initial structural characterization of these crystals by Dr. Paukstelis[42] combined with preliminary studies by Dr. Diana Zhang provided the foundation for this work. Although Dr. Zhang initiated these specific studies, the results presented here are my own.

Chapter 6.1.2. Self-assembled 3D DNA crystals

In the presence of divalent salt, d(GGACAGCTGGGAG) self assembles to form porous 3D DNA crystals (Figure 6.1A).[42] Lattice interactions result in a vast network of solvent channel space that runs throughout the crystal (Figure 6.1B). Per unit cell, solvent channels running perpendicular to the six-fold symmetry axis have a volume of ~14,600 $Å^3$, while solvent channels parallel to the six-fold symmetry axis have a volume of ~17,500 $Å^3$ with a 2 nm diameter.[42]

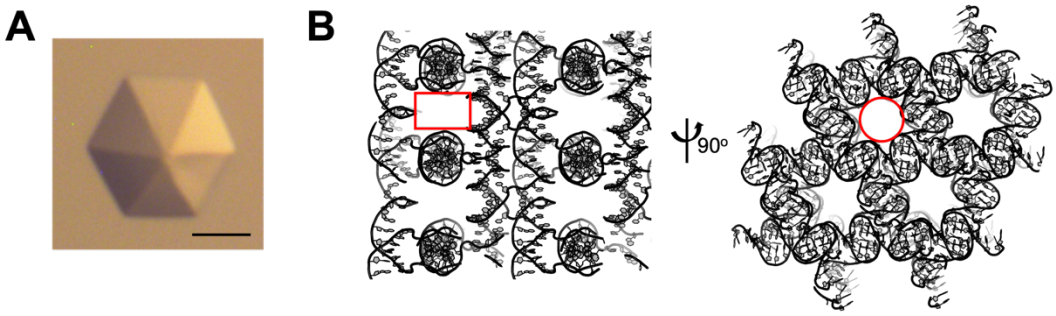

Figure 6.1. Self-assembled 3D DNA crystal. (A) Light microscope image of the hexagonal self-assembled 3D DNA crystal. Scale bars are 70 μm. **(B)** Crystal lattice interactions highlighting the solvent channels that run throughout the crystal. Solvent channels running perpendicular to the six-fold symmetry axis (left) have a volume ~14,600 $Å^3$ per unit cell, while solvent channels parallel to the six-fold axis (right) have a ~17,500 $Å^3$ volume per unit cell and 2 nm diameter.

The self-assembled DNA crystals are stabilized by Watson-Crick and non-canonical interactions (Figure 6.2A).[42] The non-canonical region contains two d(GGA) ps-duplexes stabilized by homo-base pairs that exhibit striking structural similarity to d(CGA)-based ps-duplexes (Figure 1.7). Specifically, the 5′-terminal G-G homo-base pair of the d(GGA) ps-duplex is formed via two hydrogen bonds between the Watson-crick face of G_{10} and the Hoogsteen face of G_1. G_1 adopts the *syn*-orientation to spatially accommodate this pair[42]. Similarly to the G-G and A-A homo-base pairs in d(CGA)-based ps-duplexes, the d(GGA) G_{11}-G_2 homo base pair is formed between two sugar face hydrogen bond interactions, while A_3-A_{12} is formed between two Hoogsteen face hydrogen bond interactions (Figure 6.2B). Again, as observed in d(CGA)-based ps-duplexes, the d(GGA) ps-duplex contains inter-strand G/A base stacking interactions thought to provide significant structural stabilization. Though the base pairing and stacking interactions observed in the G/A step of the d(CGA) and d(GGA) triplets are

highly compatible, d(GGA) cannot form long parallel-stranded duplexes likely due to the spatial requirement of G_1 to adopt the *syn*-confirmation.[30]

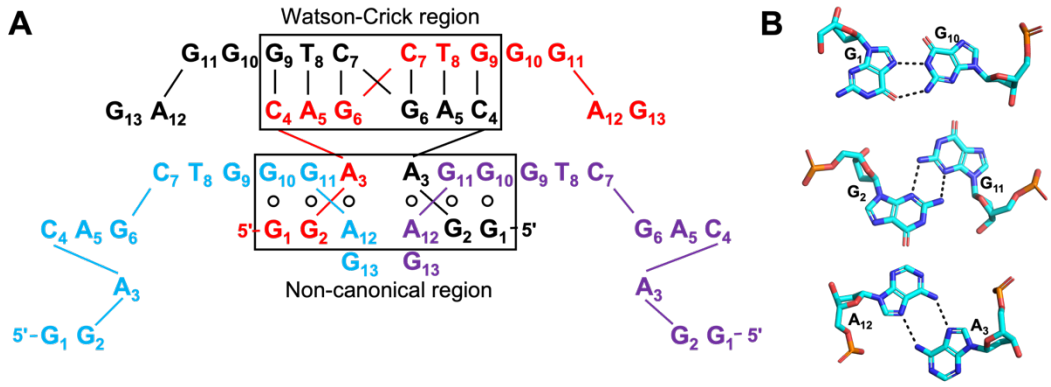

Figure 6.2. Interactions stabilizing the self-assembled DNA crystal. (A) Secondary structure of the 3D DNA crystal highlighting the Watson-Crick and non-canonical interaction regions. Each strand is colored differently to show specific interactions between strands. **(B)** G-G and A-A homo-base pairs stabilizing the d(GGA) ps-duplex within the non-canonical region. Dashed lines represent hydrogen bonds.

The self-assembled DNA crystals have several characteristics that make them viable candidates for use in drug delivery applications. First, the DNA crystals have three unique modes for cargo storage including occupation of solvent channel space, intercalation between stacked base pairs, and programmed attachment via covalent linkage. The distinct storage modes are advantageous for expanding the diversity and capacity of cargo molecules loaded within the crystals. In addition, the high DNA density intrinsic to the DNA crystals provide increased cargo loading capacity and prolonged cellular lifetime by resisting nuclease degradation.[30, 152, 153] Beyond favorable inherent properties, these crystals can be chemically modified to further improve their stability and functionality. Our lab has optimized chemical crosslinking techniques for 3D DNA crystals and have shown that the procedure results in decreased dependence on high cation concentration, enhanced thermal stability, and resistance

nuclease degradation.[154] Finally, layer-by-layer (LbL) crystal assembly can be used to modify and functionalize the crystal exterior.[155] This technique would allow for precise tethering of targeting or tracking domains on the crystal surface which could be used to improve cellular uptake and target specificity.

Chapter 6.1.3. Doxorubicin

For this study, we have selected doxorubicin (DOX) as our cargo of interest. DOX is an effective chemotherapy medication used to treat several cancers including Hodgkin's lymphoma, stomach, lung, ovarian, thyroid, and breast cancer[156]. The classically understood mechanism of action for DOX function is via the inhibition of Topoisomerase II (Topo II). Topo II is an enzyme responsible for resolving topological problems that arise in double-stranded DNA by purposely introducing and repairing double-stranded DNA breaks.[157] Following the formation of a Topo II-induced DNA break, DOX intercalation inhibits subsequent Topo II controlled DNA re-ligation.[158] The resulting double-stranded break causes permanent DNA damage and ultimately induces cell death.[158] Although DOX is an effective anti-cancer agent, severe adverse side effects such as cardiotoxicity remain problematic in its clinical usage.[159] Various drug delivery systems have been developed to encapsulate and target DOX to the cancerous site to reduce off target toxicity while retaining potent anti-cancer efficacy, but these approaches are commonly limited by slow cargo release, rapid clearance from the bloodstream, and lack of biocompatibility.[160-162]

DNA nanostructures have been investigated as a promising alternative approach for DOX delivery due to inherent biocompatibility and biodegradability, the ability to precisely program chemical modifications within the structures, and ease of

cargo loading.[163] To date, the loading capacity of these DNA-based nanostructures are commonly limited to DOX intercalation between DNA base pairs. We plan to explore 3D DNA crystals as a new vehicle for DOX encapsulation. In this approach, we hypothesize that dual loading via diffusion into solvent channels, as well as base pair intercalation will increase DOX loading capacity. Additionally, the unique structural characteristics of this specific crystal (high DNA density, non-canonical interactions) will favor stability. We will use the inherent spectroscopic properties of DOX (480 nm excitation, 590 nm emission, visual red color) to visualize and quantify the amount of DOX loaded within the DNA crystals (Figure 6.3).

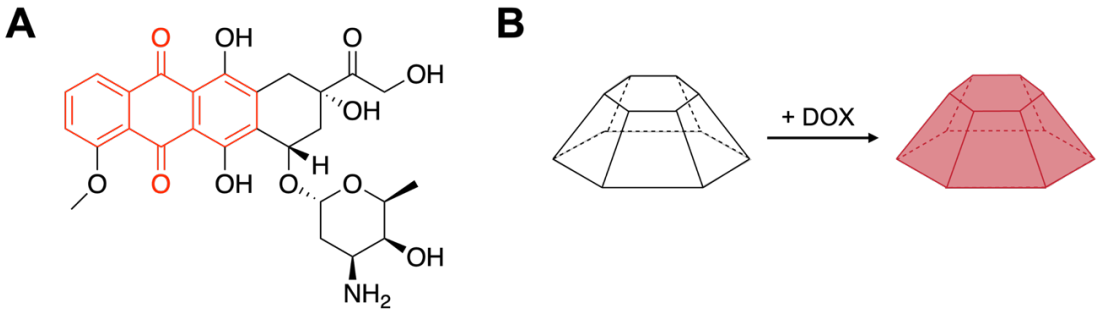

Figure 6.3. Doxorubicin. (A) Chemical structure of DOX where the central anthraquinone region is highlighted in red. This chromophore gives DOX its characteristic red color. **(B)** Schematic diagram of crystals before and after DOX incubation. Prior to incubation, crystals are hexagonal bipyramidal and colorless. Following incubation, DOX loaded crystals will retain hexagonal bipyramidal morphology, but acquire red coloration.

Chapter 6.2. Results and discussion

Chapter 6.2.1. DNA crystals can load DOX cargo

To evaluate the extent to which 3D DNA crystals can encapsulate guest cargo molecules, we grew crystals and incubated them in the presence of DOX. We evaluated the success of the encapsulation by visualization using a light microscope specifically

monitoring crystal morphology, size, and color. The localization of red coloration within the crystal was indicative of successful DOX diffusion from the surrounding buffer solution to within the crystal. Encapsulation was considered successful if crystals did not exhibit morphological changes (i.e. crystal edge rounding or size reduction) following the localization of DOX within the crystal.

Untreated (UT) DNA crystals were incubated with various concentrations of DOX (0 to 500 μM) at 22ºC as described in Chapter 6.4.4 and monitored for 24 hr (Figure 6.4A). All DOX exposed crystals showed significant edge rounding within 2 hrs and were completely degraded within the 24 hr DOX exposure period. Degradation was accelerated for crystals incubated with high concentrations of DOX (250 μM and 500 μM), indicating that DOX could disrupt interactions critical for structural stability. Although these crystals experienced significant edge rounding and were rapidly degraded, we observed localization of red coloration within DOX exposed crystals.

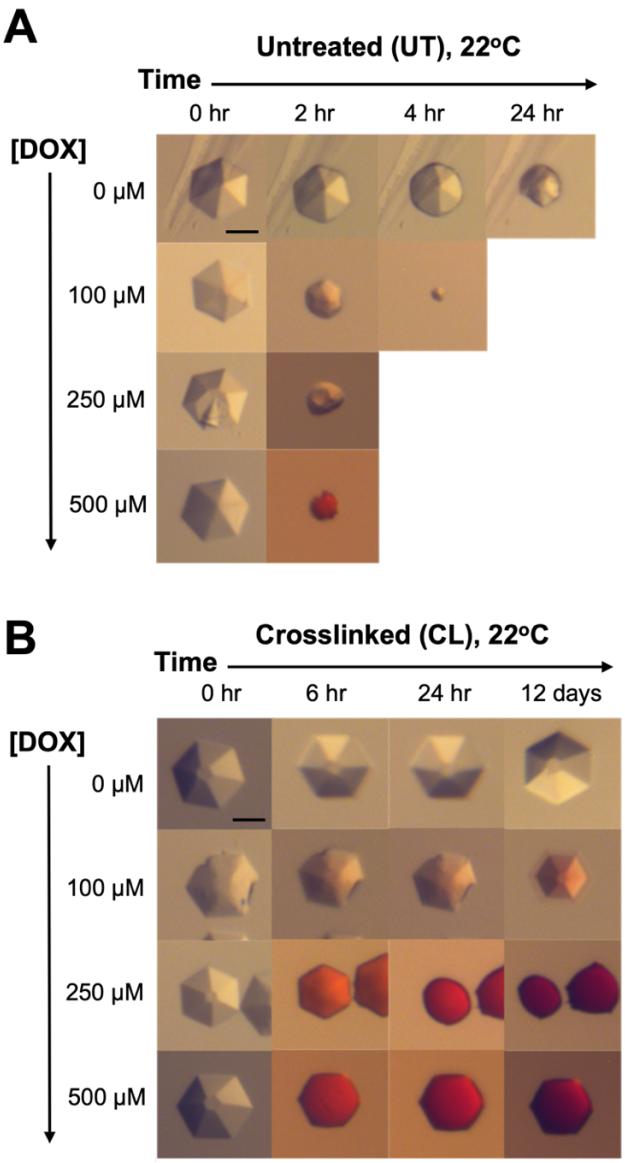

Figure 6.4. Self-assembled DNA crystals incubated with DOX at 22°C. (A) Untreated DNA crystals incubated with 0 to 500 μM DOX and monitored for 24 hours. Crystal rounding was apparent for all samples within 2 hours of incubation. All crystals incubated with DOX acquired red coloration but were completely degraded within 24 hrs. **(B)** Crosslinked crystals incubated with 0 to 500 μM DOX and monitored over 12 days. Crystals incubated with 100 μM DOX had faint red coloration while retaining crystal structural integrity. Crystals incubated with higher DOX concentrations (250 and 500 μM) had intense red coloration and edge rounding within 6 hrs. All scale bars are 70 μm.

We observed a significant improvement in DOX uptake at 22°C when the experiment was repeated with crosslinked (CL) crystals (Figure 6.4B). The crosslinking procedure was previously optimized to improve the stability of these DNA crystals[154] and is described in Chapter 6.4.3. CL crystals incubated with 100 μM DOX showed faint accumulation of red coloration within 24 hours without any accompanying morphological changes. These crystals exhibited a 23% size reduction but retained red coloration and edge sharpness over 12 days. Localization of intense red coloration was observed in CL crystals incubated with higher concentrations of DOX (250 μM and 500 μM) within 6 hrs, but this was accompanied by significant edge rounding. As opposed to the UT crystals, the CL crystals did not undergo complete degradation within the 12-day incubation period. This indicates that the additional rigidity provided by the crosslinking treatment mitigates any detrimental effects associated with diffusion or intercalation of DOX within the crystal.

Size reduction and edge rounding morphologies were reduced when the same experiment was repeated for UT and CL crystals at 4°C (Figure 6.5). At this temperature, crystals incubated with each DOX concentration (100 μM, 250 μM, and 500 μM) retained structural integrity without edge rounding or significant size reduction for the 7-day incubation period. Faint accumulation of red coloration was observed in UT and CL crystals incubated with 100 μM and 250 μM DOX within 6 hrs. The intensity of red coloration within the 250 μM DOX loaded crystals slightly increased with time. Crystals incubated with 500 μM DOX had the most intense localization of red color observed within 6 hrs and reached an apparent maximum loaded state within 4 days. From this, we concluded that incubating UT or CL crystals

with DOX at 4ºC was the most favorable condition for efficient DOX uptake and retention of crystal structural integrity.

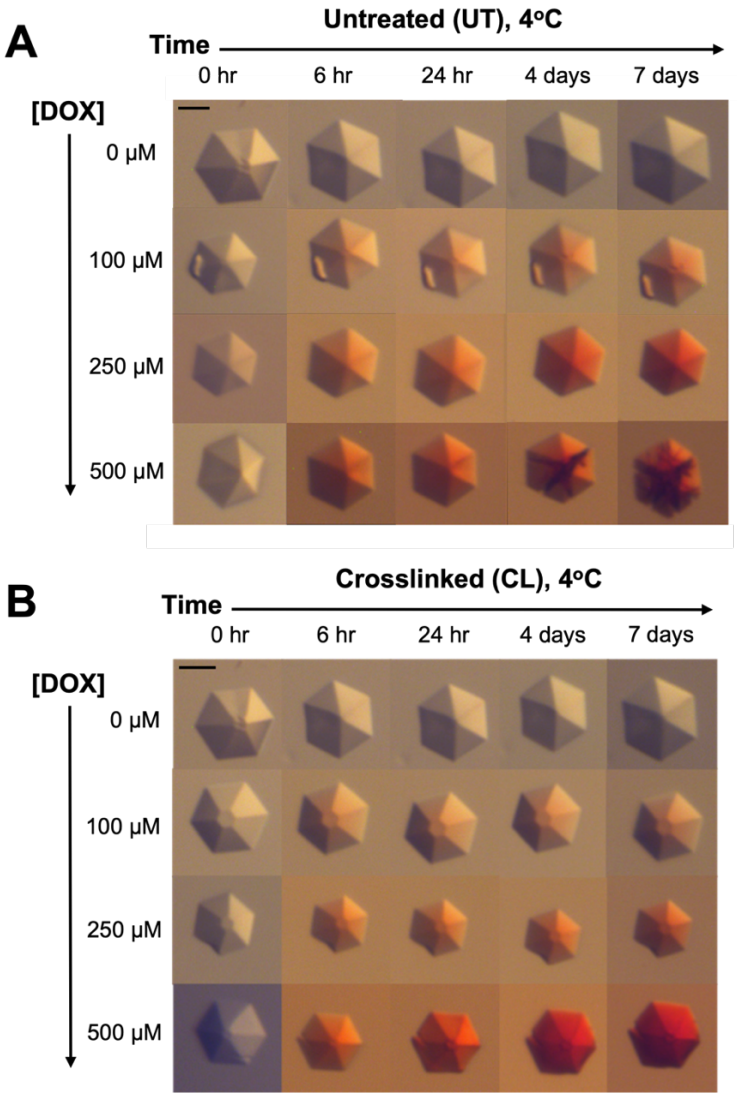

Figure 6.5. Self-assembled DNA crystals incubated with DOX at 4ºC. (A) Untreated DNA crystals incubated with 0 to 500 µM DOX at 4ºC and monitored for 7 days. All crystals incubated with DOX acquired red coloration within 6 hrs. The intensity of red coloration increased over time for samples incubated with 250 µM and 500 µM DOX. Crystals incubated with 500 µM DOX appeared cracked after 4 days of incubation. **(B)** Crosslinked crystals incubated with 0 to 500 µM DOX at 4ºC and monitored for 7 days. Crystals incubated with 100 µM DOX obtained faint red coloration while retaining structural integrity. Crystals incubated with higher DOX concentrations (500 µM) had intense red coloration within 6 hrs. The intensity of red coloration increased over time while retaining crystal integrity. All scale bars are 70 µm.

Chapter 6.2.2. DOX loading conditions can tune crystal properties

We noticed that crystals exhibited two distinct modes of red color localization dependent upon the concentrations of DOX available in the solution. When incubated with low DOX concentration (100 μM), the crystals exhibited a faint red coloration that was localized to the center core of each crystal (Figure 6.6A). The red coloration completely disappeared along the outer edges of the crystal. This produced an apparent gradient where the most intense red color was in the crystal core and transitioned to completely clear along the outer edges. Similarly to other DNA crystals, this type of crystal shattered like glass when perturbed with a nylon loop (Figure 6.6A). As a result, we refer to this type of DOX loaded crystal as "glass-like."

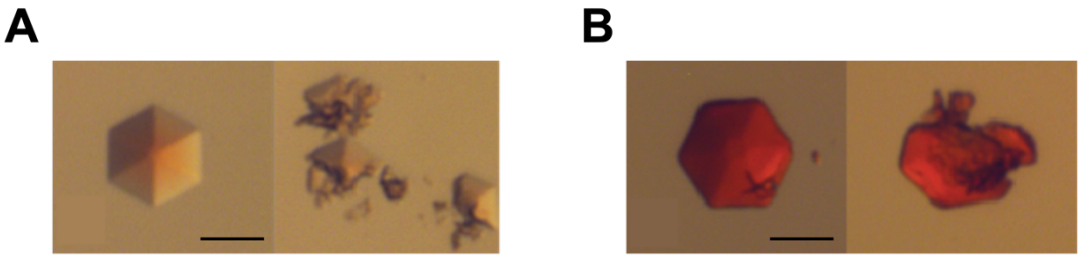

Figure 6.6. DNA crystals incubated with DOX to the glass- and gel-like states. (A) Intact glass-like crystal (left) shatters like glass when crushed by a nylon loop (right). Glass-like crystals have localization of red color concentrated in the core. The intensity of the color fades to completely colorless along the edges. **(B)** Intact gel-like crystal (left) smears like a gel and does not shatter like glass when crushed by a nylon loop (right). The intensity of the red coloration is uniform and intense throughout the entire crystal. All scale bars are 70 μm.

This contrasts with crystals that were incubated with high DOX concentration (500 μM). In this case, the red coloration was intense and uniform throughout the entire crystal (Figure 6.6B). We refer to this type of crystal as "gel-like" because when perturbed with a nylon loop, these crystals smeared like a gel rather than shatter like

glass (Figure 6.6B). We hypothesized that the unique crystal smearing could be a result of lattice destabilization caused by high concentrations of DOX intercalation and packaging within the solvent channels.

Because crystals in the gel-like state resulted in the most intense localization of red coloration, an indication of significant DOX localization, we tuned the incubation buffer to favor the formation of this state. Positively charged Mg^{2+} ions stabilize the crystal by compensating for the dense localization of negative charge along the DNA backbone. The resulting charge neutralization allows for DNA to be tightly packed into a lattice, therefore, we hypothesized that altering the concentration of Mg^{2+} could tune the packing of the lattice and extent of DOX encapsulation.

Consistent with this, we observed an inverse relationship between gel-like crystal formation and magnesium concentration (Figure 6.7). At high magnesium concentration (200 mM) and high DOX concentration (500 μM), crystals were only loaded to the glass-like state over the 5-day incubation period. Alternatively, at low magnesium concentration (50 mM) and high DOX concentration (500 μM), crystals were loaded to the gel-like state within 1 day. These results indicated that the formation of the gel-like state can be tuned using magnesium concentration. This likely occurred due to the decrease in charge neutralization in conditions containing low concentrations of magnesium. The reduced amount of charge neutralization likely induced a higher amount of backbone repulsion and lattice swelling, in turn allowing increased diffusion of DOX cargo within the crystal.

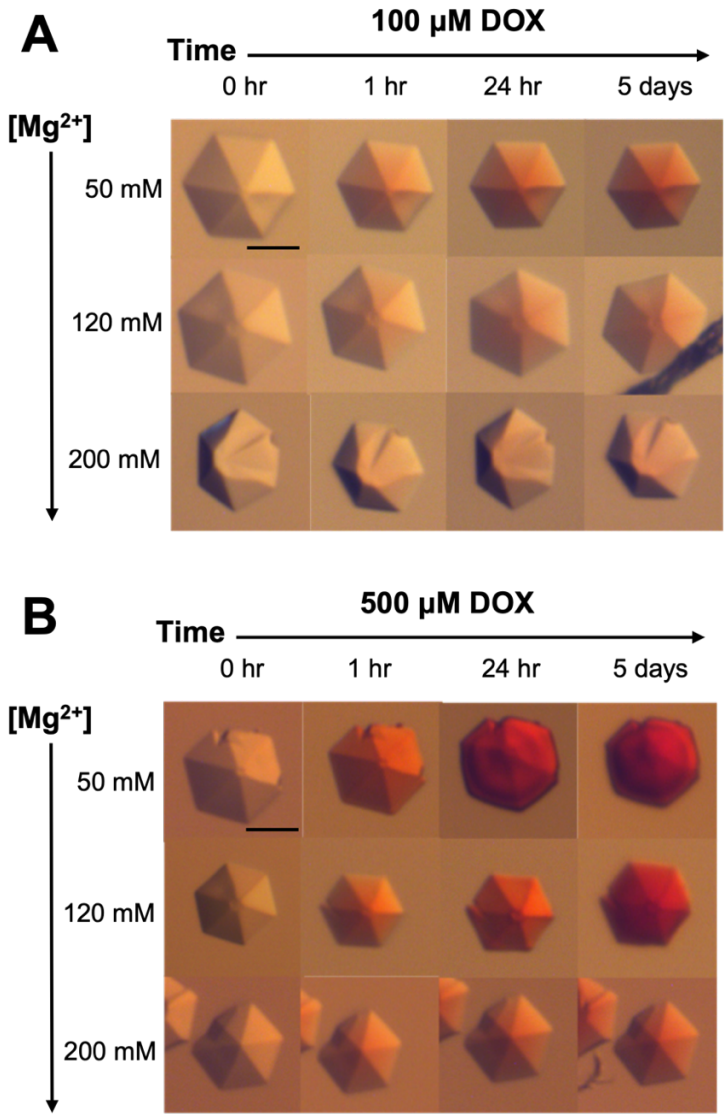

Figure 6.7. Low magnesium concentration favors the formation of the gel-like state. (A) Crosslinked crystals incubated with 100 μM DOX and 50 to 200 mM Mg^{2+} at 4°C. When Mg^{2+} is low (50 mM), the glass-like state is observed within 1 hr. There is minimal red coloration within crystals incubated with high concentration of Mg^{2+}, even up to 5 days. **(B)** Crosslinked crystals incubated with 500 μM DOX show localization of red coloration within 1 hr for all samples. The gel-like state is obtained within 1 day when [Mg^{2+}] is low and [DOX] is high. This suggest that the physical properties of the crystal can be altered by manipulating the incubation solution conditions.

Chapter 6.2.3. DNA crystals have a high DOX loading capacity

Based on visual intensity of concentrated red color within the crystals, we hypothesized that the gel-like crystals could encapsulate the highest concentration of DOX. Therefore, we used crystals loaded to the gel-like state to quantify the maximum loading capacity of DOX-loaded DNA crystals. To quantify this, we generated a standard method to estimate the volume of each DOX-loaded crystal. Specifically, confocal microscopy was used to obtain z-stack images of crystals grown with fluorescently labeled DNA as described in Chapter 6.4.6. The resulting images were compiled to generate 3D reconstructions of crystals spanning various dimensions using ImageJ software (Figure 6.8).

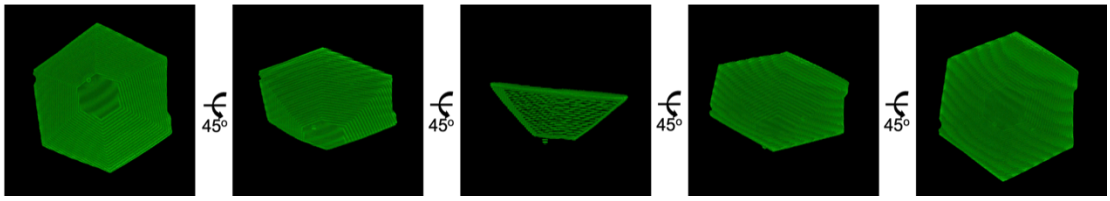

Figure 6.8. 3D reconstruction of a fluorescently labeled DNA crystal. Z-stack images obtained from confocal microscopy were combined to create 3D reconstructions of fluorescently labeled DNA crystals using ImageJ software. Here, images of one 3D reconstruction are rotated 45° to show the full crystal.

The volume of each crystal was estimated using top edge, base edge, and crystal height measurements taken from the 3D crystal reconstructions and Eq. 3 provided in Chapter 6.4.6 (Figure 6.9A,B). Volume estimates were obtained for 45 unique crystals of various sizes (20 to 120 μm base edge length). The estimated crystal volume was subsequently correlated to the length of one edge of the crystal base, a measurement easily obtained using routine light microscopy. The resulting data was compiled to

generate a standard relationship between estimated crystal volume and the length of one base edge (Figure 6.9C, Chapter 6.4.6 Eq. 4). This relationship was used to estimate the volume of new crystals by simply measuring the length of one edge of the crystal base via light microscopy.

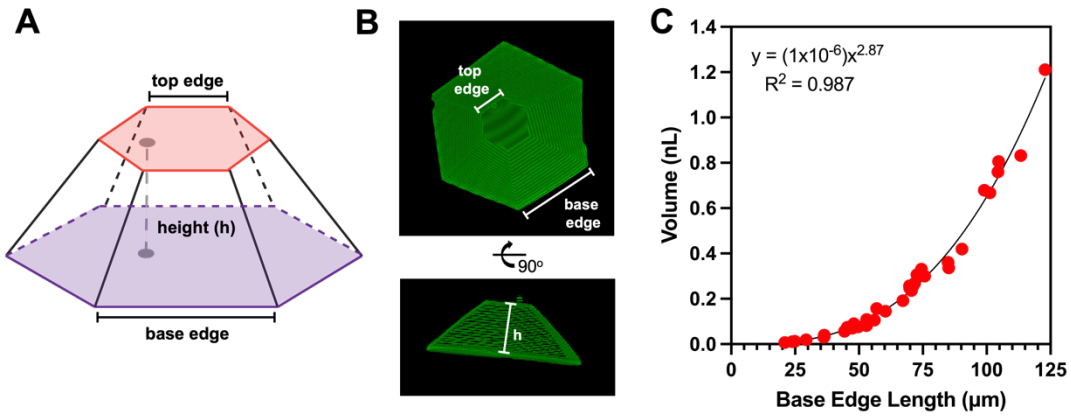

Figure 6.9. Crystal volume estimation. (A) Illustration of a 3D crystal where the dimensions of interest are highlighted. The top hexagon is colored in red and the bottom hexagon is colored in purple. Top edges are measured from the red (top) hexagon, while base edges are measured from the purple (bottom) hexagon. Crystal height (h) is the distance from the purple hexagon to the red hexagon. **(B)** 3D reconstruction generated from z-stack confocal images, rotated 90°. The labeled dimensions were measured six times and the average value was used in Eqs. 1-3 (Chapter 6.4.6) to estimate crystal volume. This procedure was performed for 45 unique crystals of various sizes. **(C)** Crystal base edge length plotted against estimated crystal volume for 45 unique crystal 3D reconstructions. Fitting equation (Eq. 4 in Chapter 6.4.6) and R^2 value are shown. The non-integer exponent indicates that the top edge length is closely, but not perfectly correlated to base edge length, but can still be used to estimate crystal volume. The volume of crystals loaded with DOX was estimated using this relationship (Eq. 4 in Chapter 6.4.6) and the length of a base edge measured from light microscopy.

Using the method described in Chapter 6.4.7, we determined that 8.2 ± 0.2 ng of DOX was loaded per 0.13 nL ± 0.01 nL crystal. This corresponds to an average concentration of 60.9 ± 4.5 mg/mL (112 mM) DOX loaded per individual gel-like crystal. This concentration exceeds the solubility of free DOX in aqueous solution (50

mg/mL) likely due to stabilizing effects related to packing within the crystal. Preliminary calculations based on crystal dimensions, symmetry, and volume indicate that the number of DOX molecules incorporated within the crystal (~9 trillion) is within the number of positions available for intercalation (~14 trillion). The vast network of solvent channels likely facilitates the efficient diffusion of DOX throughout the crystal to maximize access to intercalation positions. Along with intercalation-based loading, we hypothesize that DOX can also be stored in the crystal throughout solvent channels. Further experimentation is needed to confirm solvent channel and intercalation-based dual loading. Specifically, the loading capacity of intercalation-based and non-intercalation-based molecules would be tested individually and in tandem to observe corresponding changes in loading capacity.

Chapter 6.2.4. DOX release from DNA crystals

We assessed the ability of DNA crystals to release DOX cargo by transferring the DOX-loaded crystals to fresh buffer solution without DOX (Figure 6.10A). When gel-like crystals (loaded with 500 μM DOX) were transferred to fresh buffer and stored at 4°C, the intensity of red coloration localized within the crystal did not significantly decrease within 1 day (Figure 6.10B). A clear boarder appeared along the outside of the crystal after 20 mins in the new buffer solution, but the center of the crystal retained the intense red coloration. Contrastingly, when crystals loaded with DOX (100 μM) to the glass-like state were transferred and stored at 4°C, an apparent decrease in the intensity of red coloration was observed within 20 minutes. The glass-like crystals lost all red coloration within 1 day of being transferred to fresh buffer solution (Figure 6.10B). In each case, DOX initially diffuses from the exterior of the crystal as evident

by the formation of a clear border along the crystal edge over time. The diffusion of DOX from the glass-like crystals is faster than gel-like crystals, indicating that cargo release can be tuned depending on the extent of DOX initially loaded to the DNA crystals.

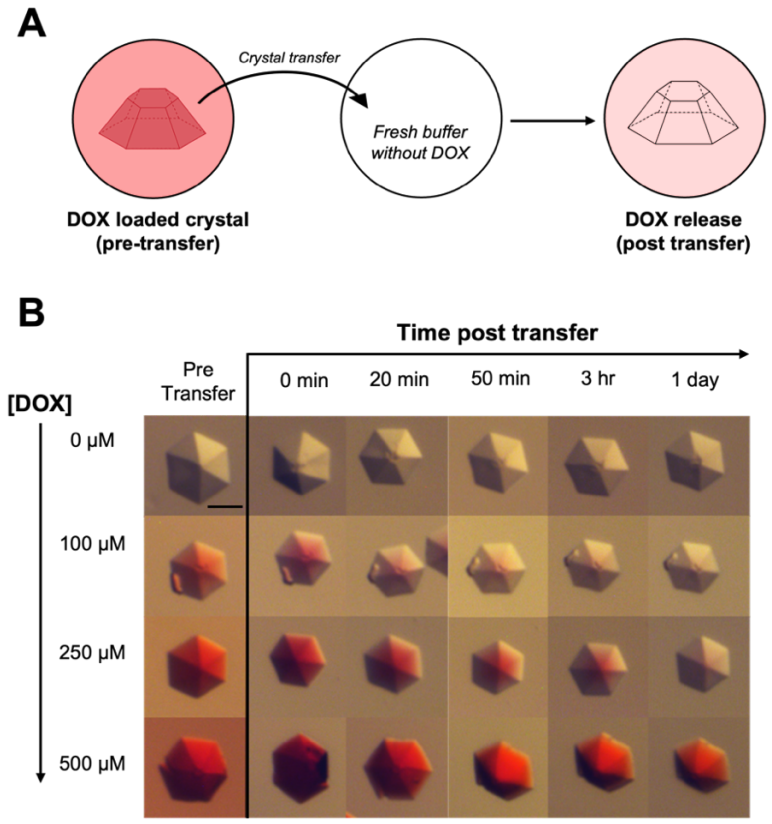

Figure 6.10. DOX release from DNA crystals. (A) Schematic representation of experimental design used to test DOX release from DNA crystals. Crystals were loaded with various concentrations of DOX (0 to 500 µM) and then transferred to a new drop containing buffer solution that was not supplemented with DOX. DOX release was monitored by changes in the intensity of the red color localized within the crystal. A loss of red coloration within the crystal corresponds to release of DOX. **(B)** Time dependent release of DOX cargo for crystals loaded to the glass- or gel-like state. Crystals incubated with 100 µM DOX were loaded to the glass-like state, while crystals incubated with 500 µM DOX were loaded to the gel-like state. DOX cargo is completely released from crystals in the glass-like state within 1 day, as evident by the loss of red coloration within the crystal. Contrastingly, the intensity of red coloration localized within gel-like crystals does not significantly decrease within 1 day. Scale bars are 70 µm.

The difference in cargo retention between the glass- and gel-like states could correspond to a lattice destabilization induced by loading high concentrations of DOX. The normal order of the crystal lattice could be disrupted to accommodate intercalation and repulsive forces caused by loading high concentrations of DOX in the gel-like state. This corresponds to the observed changes in physical properties of crystals loaded to the gel-like state, where when perturbed, these crystals smear like a gel rather than shattering like a normal DNA crystal (Figure 6.6). The structural changes within the lattice could result in blockage of the solvent channel space, therefore, limiting free diffusion from the crystal interior. We observed a rapid diffusion of DOX from crystals loaded to the glass-like state, likely because the concentration of loaded DOX did not exceed the threshold required to induce structural changes in the lattice that would block the solvent channel space.

Preliminary experimental conditions including low incubation temperature and buffers solutions containing high salt concentrations were originally selected to promote crystal stability. While these conditions promote crystal stability, they do not provide insight to DOX release in biological settings. To test this, gel-like crystals were prepared as originally described (crosslinked and incubated with DOX at 4^{o}C), but then transferred to buffer containing physiologically relevant divalent (5 to 120 mM Mg^{2+}) and monovalent (100 mM Na^{+}) salt and incubated at 37^{o}C (Figure 6.11).

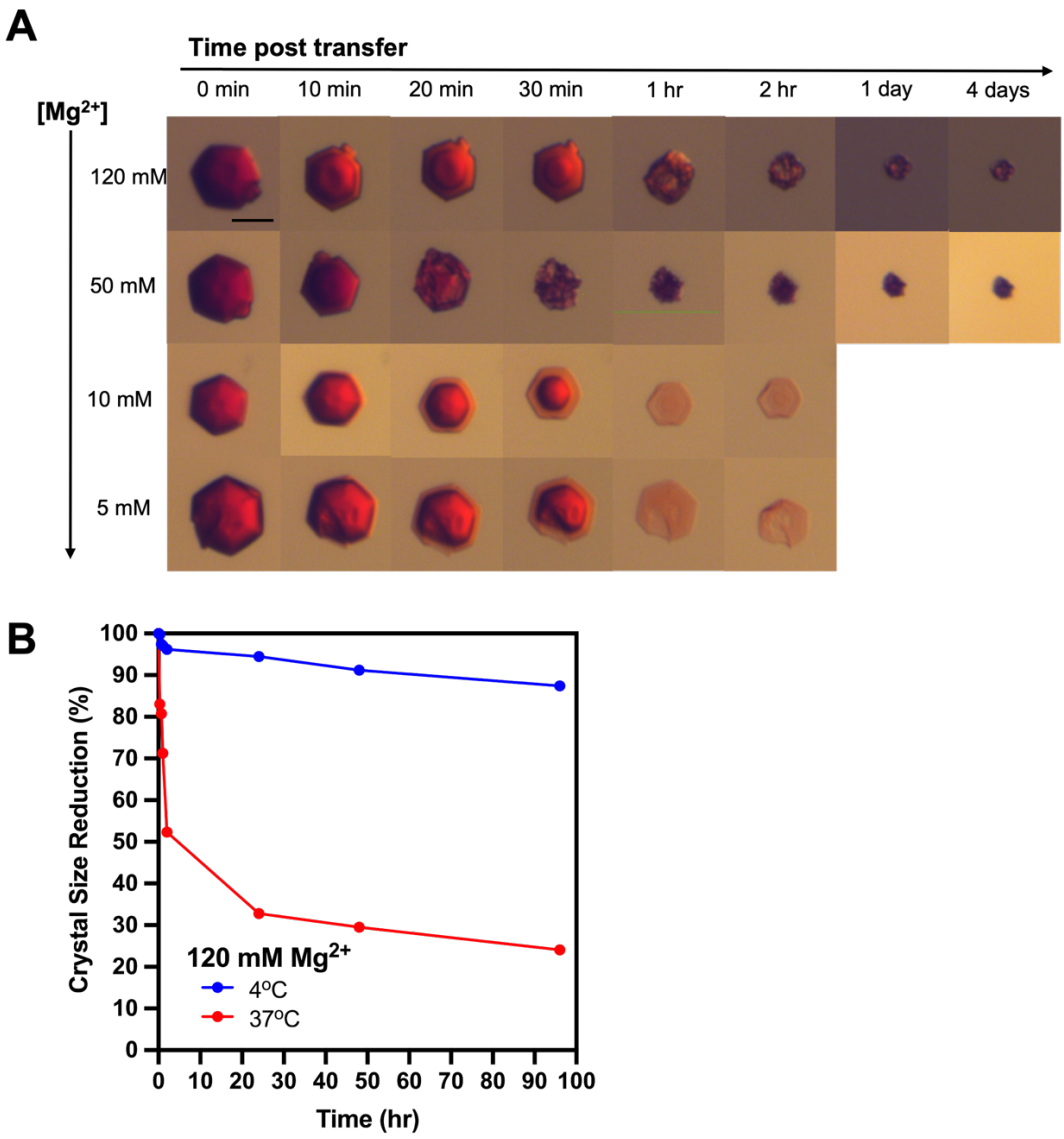

Figure 6.11. DOX release from DNA crystals in physiologically relevant salt concentration and temperature. (A) Crystals loaded with DOX to the gel-like state were transferred to buffer solution containing 120 mM, 50 mM, 10 mM, or 5 mM Mg^{2+} supplemented with 100 mM NaCl without DOX. Crystals were incubated at 37°C and monitored for 4 days. Crystals incubated with 120 mM or 50 mM Mg^{2+} and 100 mM NaCl did not completely degrade within 4 days but exhibited morphological changes and size reduction. Crystals incubated with 10 mM or 5 mM Mg^{2+} and 100 mM NaCl were completely degraded within 1 day. **(B)** Crystal size reduction monitored over time for gel-like crystals transferred to buffer containing 120 mM Mg^{2+} and incubated at 4°C (blue) or 37°C (red). Crystal size reduction was accelerated in samples incubated at high temperature.

As anticipated, increased incubation temperature accelerated crystal degradation compared to low temperature incubation (Figure 6.11B). Crystals transferred to buffer solution containing 120 mM or 50 mM Mg^{2+} supplemented with 100 mM NaCl and incubated at 37°C were not completely degraded within 4 days. These crystals experienced morphological changes (Figure 6.11A) and size reductions (Figure 6.12A,B) that were similar to samples that were not supplemented with NaCl. This indicates that 100 mM NaCl does not significantly affect crystal behavior in the presence of high concentrations of Mg^{2+}.

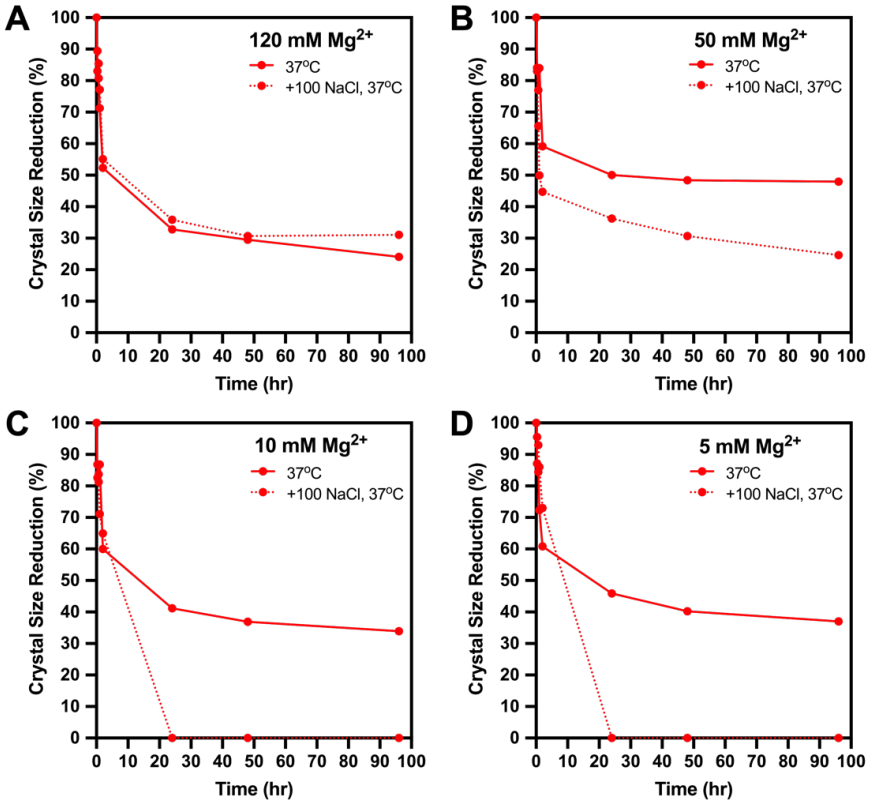

Figure 6.12. Gel-like crystal size reduction in physiologically relevant conditions. Crystal degradation corresponds to DOX release. Crystal size was monitored over time for samples loaded to the gel-like state and transferred to buffer containing **(A)** 120 mM Mg^{2+}, **(B)** 50 mM Mg^{2+}, **(C)** 10 mM Mg^{2+}, or **(D)** 5 mM Mg^{2+}. The buffer either contained only Mg^{2+} (solid red lines) or Mg^{2+} supplemented with 100 mM NaCl (dashed red lines). All samples were incubated at 37°C after crystal transfer.

Crystals transferred to buffer solution containing low concentrations of Mg^{2+} (5 and 10 mM) supplemented with 100 mM NaCl were completely degraded within 1 day (Figure 6.12C,D). DOX cargo was released due to complete crystal degradation where the solvent-exposed exterior of the crystal appeared to shed over time (Figure 6.11A). The concentration of divalent salt is an important factor that influenced the extent of crystal degradation and subsequent DOX release over time, but monovalent salt did not. These results indicate that environmental fluctuations including temperature and salt concentration can be tuned to alter the extent of cargo release from DNA crystals.

Chapter 6.3. Summary and conclusions

Here, I have described our work towards evaluating non-canonical interaction-based 3D DNA crystals as a new platform for storage and release of doxorubicin cargo. Visual identification of red coloration confirmed that DNA crystals can encapsulate DOX cargo. The physical properties of the crystal can be tuned to form the glass- or gel-like state depending on the concentration of DOX and magnesium present during loading. We also determined that gel-like crystals have a high loading capacity for DOX. The extent and time scale of DOX cargo release can be tuned depending on environmental factors including salt concentration and incubation temperature. The results presented here indicate that self-assembled DNA crystals stabilized by the non-canonical d(GGA) ps-duplex could make a suitable vessel for the encapsulation and release of high concentrations of DOX cargo *in vitro*. Further experimentation is required to optimize and assess the behavior of DOX-loaded crystals in cellular environments.

Chapter 6.4. Experimental procedures

Chapter 6.4.1. Oligonucleotide synthesis and purification

Unmodified d(GGACAGCTGGGAG) and 3′-fluorescein labeled (3′-F) d(GGACAGCTGGGAG-Fluorescein) oligonucleotides were synthesized on a 1 μmol scale using standard phosphoramidite chemistry on an Expedite 8909 Nucleic Acid Synthesizer (PerSeptive Biosystems, Framingham, MA) with reagents from Glen Research (Sterling, VA). Oligonucleotides were purified by denaturing 20% (19:1) acrylamide/bis-acrylamide and 7 M urea gel electrophoresis. DNA bands were identified and excised using UV shadowing. DNA was electroeluted from gel slices, ethanol precipitated, and dialyzed against deionized water.

Chapter 6.4.2. Oligonucleotide crystallization

Oligonucleotides were crystallized by mixing 2 μL of 400 μM DNA solution with 1 μL of crystallization solution containing 120 mM magnesium formate, 50 mM lithium chloride, and 10% 2-methyl-2,4-pentanediol. Crystallization was performed in sitting drops, equilibrated against 300 μL of 30% 2-methyl-2,4-pentanediol in the well reservoir, and incubated at 22°C. Crystals were observed within 24 hrs of plating.

Chapter 6.4.3. Chemical crosslinking

Crystals were transferred via pipette from the crystallization drop to a glass crystal tray. The transferred crystals were washed by decanting the crystallization buffer from the drop and replacing with fresh crystallization buffer. The wash procedure was repeated 5 times. Following the wash, the buffer was decanted and the

crystals were soaked in 120 mM magnesium formate for 5 min. 300 mM nornitrogen mustard was added to the crystal drop. The well was sealed and incubated at 4°C overnight. Supernatant was decanted and the crosslinking step was repeated by adding 120 mM magnesium formate and 300 mM nornitrogen mustard solution. The well was sealed and incubated at 4°C overnight. Crosslinking was confirmed by denaturing 20% (19:1) acrylamide/bis-acrylamide and 7 M urea gel electrophoresis.

Chapter 6.4.4. Doxorubicin encapsulation

Crystals were grown as described in Chapter 6.4.2. and used without further treatment (untreated, UT) or used following the crosslinking procedure (crosslinked, CL). Crystals were transferred to a glass crystal tray and washed with soaking buffer containing 10 mM Tris HCl, 120 mM magnesium formate, pH 7.5 for 5 minutes at room temperature. Then crystals were transferred by nylon loop to drops containing 10 mM Tris HCl pH 7.5, 50 to 200 mM magnesium formate, 0 to 100 mM NaCl, and 0 to 500 μM doxorubicin. The drops were sealed and stored at 4°C or 22°C. Crystals were viewed and imaged via Lecia S8Ap0 light microscope with attached CCD camera.

Chapter 6.4.5. Doxorubicin leakage

Crystals were loaded with doxorubicin (0 to 500 μM) as described in Chapter 6.4.4. for 11 days at 4°C. Loaded crystals were transferred via nylon loop to a drop containing 120 mM magnesium formate and 50 mM Tris HCl, pH 7.5. The drops were sealed and stored at 4°C. Crystals were viewed and imaged via Lecia S8Ap0 light microscope with attached CCD camera.

Chapter 6.4.6. Crystal volume estimation

Fluorescent DNA crystals were grown as described in Chapter 6.4.2. using a DNA solution mixture comprised of 99:1 UT:3′-F DNA. Crystals were washed with 100 μL of crystallization buffer, transferred via nylon loop to a 35 mm glass bottom culture dish with 14 mm microwell, and sealed with a cover slip. Crystals were imaged via Leica SP5X confocal microscope with laser excitation at 488 nm and emission detected from 505 to 565 nm. Z-axis scans were collected in 1 um slices and used to create 3D reconstructions of each crystal using ImageJ software. Crystal base edge length, top edge length, and crystal height (h) were measured from the 3D reconstructions. Crystal volumes were estimated from Eqs. 1-3 using the average of six individual measurements of each dimension. Top hexagon refers to the smaller hexagon colored in red in Figure 6.9 and bottom hexagon refers to the larger hexagon colored purple in Figure 6.9. Top edge lengths were used to calculate the area of the top (red) hexagon, while base edge lengths were used to calculate the area of the bottom (purple) hexagon.

$$A_{top\ hexagon} = (3\sqrt{3}(top\ edge\ length)^2)^2 \quad \text{(Eq. 1)}$$

$$A_{bottom\ hexagon} = (3\sqrt{3}(base\ edge\ length)^2)^2 \quad \text{(Eq. 2)}$$

$$V = \frac{h}{3}(A_{top\ hexagon} + A_{bottom\ hexagon} + \sqrt{A_{top\ hexagon} + A_{bottom\ hexagon}}) \quad \text{(Eq. 3)}$$

This procedure was used to estimate the volume of 45 unique crystals spanning various sizes. These values were used to generate the following mathematical relationship (Eq.4) between estimated crystal volume and measured crystal base edge length.

$$Estimated\ crystal\ volume\ in\ nL = 10^{-6} * (base\ edge\ length\ in\ \mu m)^{2.87} \quad \text{(Eq. 4)}$$

Chapter 6.4.7. Loading capacity quantification

Crystals were grown and crosslinked as described Chapter 6.4.2. and 6.4.3. 75 crosslinked crystals were transferred via nylon loop to a glass crystal tray. Buffer was decanted and crystals were washed 5 times with loading buffer containing 70 mM magnesium formate and 10 mM MES, pH 6.5. The crystals soaked in loading buffer for 5 minutes at room temperature. 500 μM doxorubicin was added, the drop was sealed, and crystals were incubated at 4^{o}C for 5 days. Lecia S8Ap0 light microscope was used to measure the base edge of each loaded crystal. The base edge length was used to estimate the volume of each crystal using the relationship in described in 6.4.6 (Eq. 4). The crystals were washed and transferred to a drop with fresh loading buffer to remove unincorporated doxorubicin. The crystals were incubated at 95^{o}C for 1 hr, cooled at room temperature for 10 minutes, and diluted with loading buffer. Crystals were visualized via Lecia S8Ap0 light microscope before and after heating to confirm dissolution. UV absorbance at 480 nm was used to measure total doxorubicin concentration. Crystal volume estimates were used to approximate the concentration of doxorubicin within each crystal. This procedure was repeated in triplicate to obtain an average estimate of DOX loading capacity.

Chapter 7: Conclusions and future work

Chapter 7.1. General summary and implications

In this dissertation I presented the biophysical and structural characterization of the non-canonical d(CGA) triplet repeat motif. In Chapter 2, thermodynamic parameters obtained from UV melting experiments detail the relative stabilities of the hairpin and ps-duplex form. Additionally, comparison of thermodynamic parameters across several sequence variants demonstrates that sequence modifications can be used to tune the stability of the ps-duplex form. In Chapter 3, nuclease sensitivity assays show that the ps-duplex form offers increased stability against single- and double-stranded nucleases, indicating that this motif could offer a unique advantage in cellular environments. CD spectroscopy based kinetic analysis presented in Chapter 5 provides insight into the relative time scale of the structural transition between hairpin and ps-duplex forms. Specifically, we note that the transition from ps-duplex to hairpin is faster than the transition from hairpin to ps-duplex in slightly different conditions. We also use the apparent bi-phasic rate profile to speculate upon the mechanism of ps-duplex formation.

This biophysical data is supplemented by the structure determination of four unique d(CGA)-based ps-duplexes across two crystal structures described in Chapter 4. From the analysis of these structures, we highlight the robust structural predictably of d(CGA) triplets in the ps-duplex form and define the specific structural features of the d(CGA)-based ps-duplex. Additionally, we note that each ps-duplex contains two conformationally distinct d(CGA) triplets defined by differences in backbone torsion angles, hydrogen bonding distances, and base stack overlap areas.

In Chapter 6, we explored the potential for a 3D DNA crystal stabilized by the parallel-stranded d(GGA) triplet motif to be used as a vehicle for drug delivery. We demonstrated that the non-canonical interaction-based 3D DNA crystal can efficiently load, store, and subsequently release high concentrations of Doxorubicin cargo *in vitro*. This directly illustrates that parallel-stranded triplet repeat sequences can be incorporated into DNA architectures that have functional applications.

Together, the data presented within this dissertation have provided the foundation for the rational incorporation of non-canonical interactions, specifically the parallel-stranded d(CGA) motif, into DNA-based nanoarchitectures. This work complements analyses on other well-studied non-canonical motifs (i-motif and G-quadruplexes) used in DNA nanotechnology[164] and inspires the search for additional DNA motifs that form new, unique non-canonical structures. Thorough characterization of both well-established and newly discovered non-canonical structure-forming motifs could be combined to create a pipeline for the "plug-and-play" creation of self-assembled DNA nanoarchitectures. The resulting vast database of biophysical and structural information would streamline the design process and open the door to the development of entirely new architectures with increasingly diverse geometries, precise controllability, and broad functionalities. Such advances could have immediate impact in the rational design of DNA-based architectures for applications in chemical sensing, diagnostics, drug delivery, computing, and molecular assembly.

Chapter 7.2. Future work

This dissertation has delineated the biophysical and structural characteristics of the d(CGA) motif which will directly inform the rational design of non-canonical interaction-based DNA nanoarchitectures. Though this work provides a foundation towards understanding the properties of this motif, there are several opportunities for further experimentation which I will briefly outline below.

Chapter 7.2.1. Structural characterization of the anti-parallel hairpin form

The crystal structures presented in Chapter 4 are the longest ps-duplexes to be solved comprised solely of d(CGA)-based triplets and demonstrate the structural predictability of d(CGA) triplets in the ps-duplex form. Unfortunately, the structure of the d(CGA)-based hairpin form is still elusive. To date, we have limited direct evidence for the specific interactions that stabilize the hairpin structure formed at pH 7.0. UV melting experiments (Chapter 2) indicate that the structure formed at pH 7.0 is unimolecular, CD spectra (Chapter 2) indicate that the strands are in the anti-parallel orientation, and nuclease digestion patterns (Chapter 3) are significantly different at pH 7.0 compared to pH 5.5. NMR-based structural studies performed by Dr. Gao and colleagues showed that at high pH, $(GAC)_3$ (a simple sequence permutation of $(CGA)_3$) forms a duplex stabilized by Watson-Crick G-C base pairs and A-A mismatch homo-base pairs[31]. In combination, these results led us to postulate that d(CGA)-based sequences form hairpin structures at neutral pH stabilized by G-C and A-A base pairs.

In an attempt to bias the formation of an anti-parallel hairpin, we designed two d(CGA)-based sequences that contained complementary WC regions at each hairpin terminus (Figure 7.1A). We hypothesized that stabilizing the hairpin structure with a

programmed WC hairpin terminus would increase the stability and homogeneity of the resulting hairpin structures, in turn potentially forming crystals that would lead to successful structure determination. We obtained several large crystals for these sequences in multiple crystallization buffer conditions, but unfortunately there was very weak or no diffraction for all tested samples (Figure 7.1). We hypothesize that the poor diffraction was due to high disorder or flexibility within the crystals. As a result, future work towards structure determination of the hairpin form should be pursued using other techniques, such as NMR, better equipped to handle the structural flexibility. Structure determination of the hairpin form would provide direct insight into the interactions critical for stabilizing the hairpin form and would further inform the incorporation of this motif into nanoarchitectures.

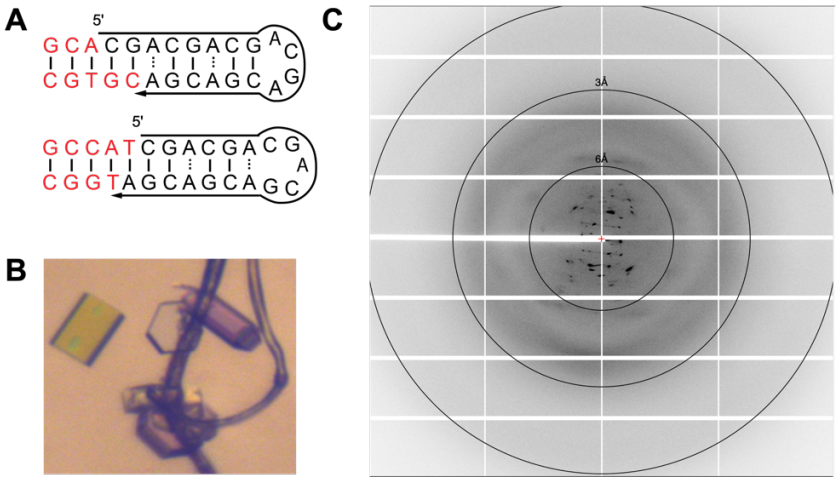

Figure 7.1. Attempted d(CGA)-based hairpin structure determination (A) Two d(CGA)-based sequences were designed to include a complementary WC region installed at each hairpin terminus (added nucleotides are colored red) to bias the formation of an anti-parallel hairpin. (Top) Potential $(CGA)_6$ hairpin containing a 4-nt loop and 5′-end, 2-nt overhang. (Bottom) Potential $(CGA)_6$ hairpin containing a 5-nt loop and 3′-end, 1-nt overhang. **(B)** Light microscopy image of rectangular and hexagonal flat crystals grown from the DNA sequence containing the 5′-end overhang. **(C)** X-ray diffraction image of rectangular crystal from B. Diffraction spots did not extend beyond 6 Å for all crystals tested.

Chapter 7.2.2. Quantitative kinetic analysis

The d(CGA) triplet repeat motif can be used as a building block to construct pH-sensitive DNA nanoarchitectures. Thorough understanding of the kinetics of the transition between structural forms will allow for precise control of rationally designed architectures. The qualitative kinetic analysis described in Chapter 5 provides the foundation for future experimentation by presenting the rate profiles and relative time scales for the transition between structural forms. However, non-automated pH adjustment and manual scan initiation prevented collection of rapid initial datapoints resulting in incomplete profiles. As mentioned in Chapter 5, stopped-flow CD spectroscopy could be used to improve the current data by shortening the collection dead time, automating pH adjustment, and measuring signal on the millisecond time scale. This would allow for the rapid collection of data points immediately following the pH adjustment. Kinetic rates of ps-duplex formation and hairpin formation could be extracted and used to inform the rational incorporation of this motif into nanoarchitectures. Similar experiments could be performed using other sequence variants (including d(NGA) triplets with varying position and frequency within the sequence) to investigate how sequence mutations alter the rate of the transition between structural forms. This data would directly indicate how sequence variants can be used to tune the rate of the structural transitions.

Quantitative rate information would also be valuable in deciphering the mechanism of ps-duplex formation. Our current qualitative data reveals that the transition from hairpin to ps-duplex is more complex than originally anticipated. The multi-phase rate profile suggests that rate of ps-duplex formation is dependent upon the

initial association of two individual strands and subsequent triplet frameshifting to form the perfectly registered ps-duplex (Figure 5.7). Quantitative kinetic analysis of samples with different repeat number (to assess frameshifting) at varied concentrations (to assess molecularity) can be used to identify each phase of the rate profile more confidently. Together, this data will provide a more accurate description towards understanding the mechanism of ps-duplex formation.

The occurrence of triplet frameshifting could be problematic to the confident incorporation and precise tunability of the d(CGA) motif into DNA nanoarchitectures. Using the qualitative data presented in Chapter 5, we hypothesize that the extent of triplet frameshifting is dependent upon sequence length. Sequence length-dependent kinetic analysis (described above) can be used to precisely identify the maximum triplet repeat number that does not correspond to significant triplet frameshifting. If the designed d(CGA)-based architecture requires a sequence length exceeding this triplet repeat threshold, we hypothesize that sequence modifications can be incorporated as an alternative method to force the ps-duplex register and limit the extent of triplet frameshifting. Potential modifications include incorporation of the parallel-stranded polyA motif[165] or a d(CGA) triplet where the 5′-C modified to isocytosine (isoC)[166] (Figure 7.2).

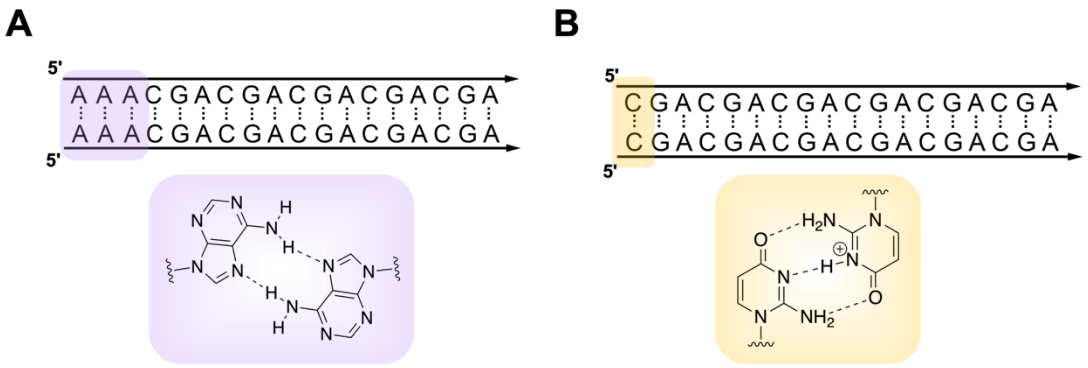

Figure 7.2. Potential sequence modifications to force ps-duplex strand registration. The 5′ terminus of d(CGA)-based sequences can be modified with **(A)** the parallel stranded poly dA motif or **(B)** isocytosine (isoC) to force duplex registration. IsoC contains the same components as natural cytosine, but the hydrogen bond donor/acceptor pattern is altered by switching the functional groups at the 2 and 4 positions. As a result, isoC is expected to only form a homo-base pair with another isoC – forcing strand registration. The specific hydrogen bonding interactions for each base pair are shown.

In each case, the sequence element forcing registration (poly dA or isoC) also adopts the ps-duplex form and is not expected to disrupt the overall pH-dependent structures, but CD spectroscopy should be used to ensure that these modifications do not alter the expected structures formed at either pH. Subsequent quantitative kinetic analysis would identify the extent of successfully forced ps-duplex registration as evident by a corresponding decrease in k_{slow} or the amplitude of the slow phase. This could be coupled with S1 nuclease sensitivity experiments where a lack of S1 digestion would correspond to successful ps-duplex registration.

Chapter 7.2.3. 3D DNA crystals as vehicles for drug delivery

In Chapter 6, we showed that 3D DNA crystals can efficiently load, retain, and release DOX cargo *in vitro*. The next important step towards using this technology as a platform for drug delivery is further optimization for cellular targeting and uptake. Currently, the 3D DNA crystals grow on the micrometer scale (50 to 200 μm). HeLa cells can uptake 3 μm particles via endocytosis, but efficiency of internalization drastically increases with decreasing particle size.[167] To increase efficiency of internalization, the crystal size must be decreased to within the nanoscale. Potential approaches to decrease crystal size include systematic adjustment of crystallization conditions (adjust DNA, precipitant, or salt concentrations) or manually breaking the macroscale crystals to generate nanoscale crystal shards. Transmission electron microscopy can be used to visualize and confirm the size range of nanoscale crystals.

In addition to decreasing crystal size, crystals can be chemically functionalized using layer-by-layer assembly.[155] Functionalizing the crystal exterior with targeting moieties such as cell-penetrating peptides can effectively promote cellular uptake, mitigate off target effects, and improve drug efficacy.[168] Cell-targeting peptide labeled DNA can be generated in solution prior to crystal assembly. Specifically, this can be done using click chemistry where an ethanethiol modified cytosine is covalently linked to a maleimide modified cell targeting peptide. Previously grown unmodified seed crystals can then be incubated with the peptide modified DNA strand to generate a functionalized crystal exterior layer. Once crystal size and exterior functionalization is optimized, the efficacy of DOX loaded 3D DNA crystals can be tested in cell culture models.

Chapter 7.2.4. Development of new d(CGA)-based nanoarchitectures

Using the biophysical and structural parameters presented in Chapters 2-5 as a guide, d(CGA)-based sequences can be rationally incorporated into DNA architectures. The specific pH-dependent structural changes of the d(CGA) motif can be used to program the precise association and dissociation of two separate particles of interest (Figure 7.3). One intriguing opportunity is to construct a pH-sensitive d(CGA)-based cargo delivery nanodevice. In this design, the cargo molecule and delivery vehicle are independently decorated with d(CGA)-based oligonucleotides. Cargo molecules can be subsequently tethered or released from the delivery moiety in a pH-dependent manner via structural changes of the d(CGA) motif. At low pH, the d(CGA)-based sequences will form intermolecular ps-duplexes which will tether the cargo molecule to the delivery vehicle. Following delivery of the complex to the location of interest, a pH increase would trigger the formation of intramolecular hairpin structures, consequently releasing the cargo from the delivery vehicle. This d(CGA) motif-based trigger design is advantageous due to simple architecture, ease of DNA strand synthesis, inexpensive fuel sources (HCl and NaOH), reversibility, and non-hazardous waste products (water and salt). Such d(CGA)-based delivery systems could be specifically developed as pH sensors or drug delivery devices.

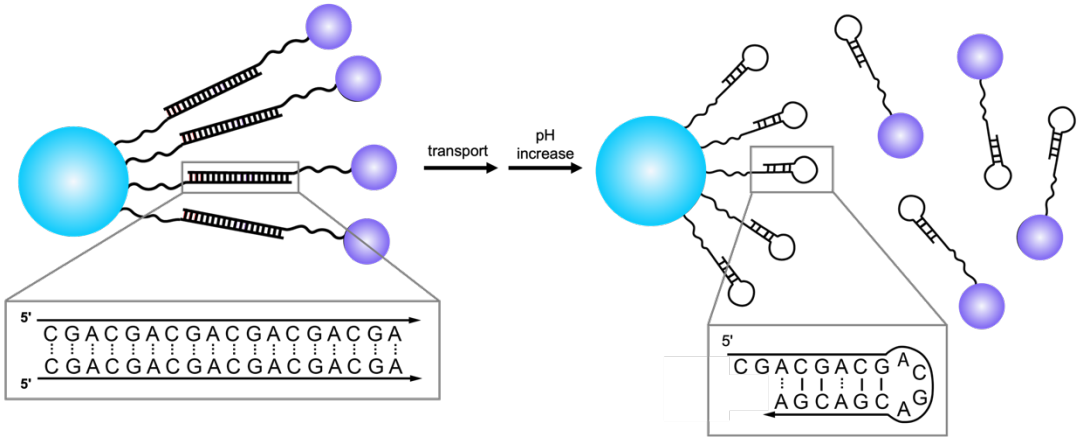

Figure 7.3. d(CGA)-based sequences used to trigger pH-dependent particle localization and dissociation. The blue and purple particles are independently decorated with d(CGA)-based sequences. At pH 5.5 (left), d(CGA) forms intermolecular parallel-stranded duplexes structures, subsequently localizing the blue and purple particles. This complex can be transported to a location of interest and upon a pH increase, the intermolecular ps-duplexes are converted to intramolecular hairpins (right). This will trigger the dissociation of the blue and purple particles. Sequence length and nucleotide composition can be valuable parameters used to tune the stability and length of the d(CGA)-linker sequence.

Chapter 7.3. Prospects for d(CGA)-based ps-duplexes in biology

Though the formation of stable alternative DNA structures may be desirable for the rational design of DNA-based architectures, they may be unfavorable or selected against in biological systems. Repeat sequences that form alternative structures such as triplexes, G-quadruplexes, and hairpins are associated with greater genomic instability due to the challenges they present during replication, transcription, and DNA repair.[169] The resulting aberrant repeat sequence expansions are responsible for numerous degenerative diseases including, but not limited to, Huntington disease, multiple spinocerebellar ataxia, fragile X syndrome, and Friedreich's ataxia.[169]

In addition, there is also the possibility that readily formed, thermodynamically stable alternative DNA structures would be selected against evolutionarily, as they may

have an even greater impact on endogenous replication or repair systems. The mechanism by which ps-duplex structures like those described in this dissertation might form in a biological context is not immediately clear. Intramolecular ps-duplexes would require some form of looping from single-stranded regions to obtain the parallel-stranded orientation, while intermolecular contacts would likely require association from two single-stranded regions.[170] The formation of ps-duplex structures is possible during DNA replication in which long strands of ssDNA are formed in the Okazaki initiation zone during lagging strand synthesis. This is a commonly recognized mechanism for the expansion of trinucleotide repeat sequences associated with hereditary diseases where the formation of alternative structures (i.e. hairpins, G4s, triplexes) results in stalling of the replication fork and template strand misalignment.[83,84] Subsequent resumption of DNA synthesis from the misaligned template results in repeat expansion.

Alternatively, another intriguing possibility could be the formation of RNA/DNA hybrid ps-duplexes formed from transcriptionally active regions, in which nascent RNA triplet product and DNA triplet sense strands are nominally parallel and single-stranded. Such hybrid structures would likely not be identical to the ps-duplex structures presented here, as the 2′-OH of the cytosines would make close contacts with the adjacent guanosine sugars and 2′-OH of the guanosines would make close contacts with adjacent adenosine N_6 (Figure 7.4).

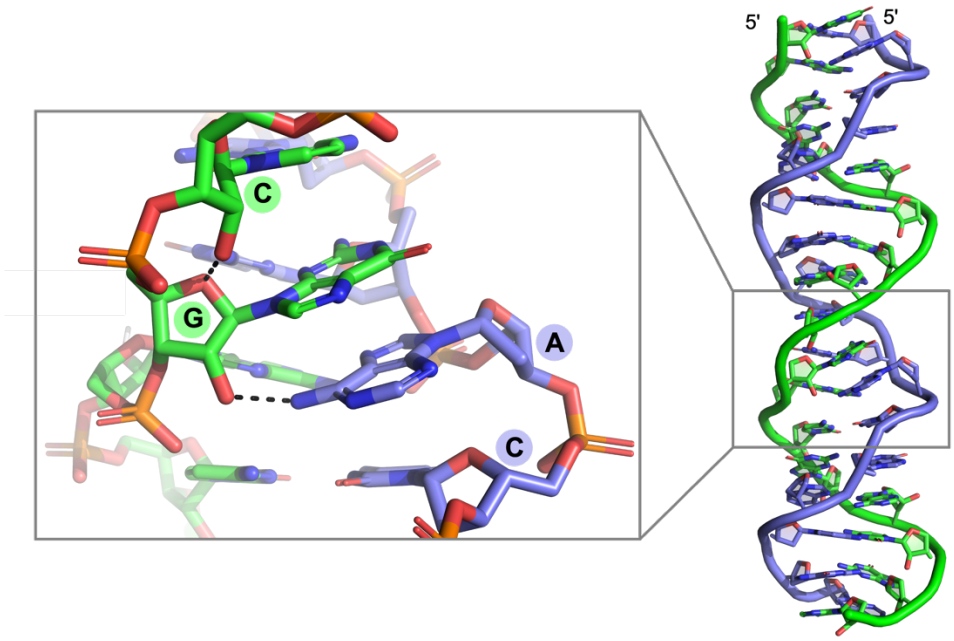

Figure 7.4. RNA/DNA hybrid d(CGA)-based parallel-stranded duplex. 2′-OH were added to one strand of Duplex 1 (from Chapter 4) to model a theoretical RNA/DNA hybrid structure. The RNA strand is colored green, and the complementary DNA strand is colored purple. On the RNA strand, the 2′-OH of each cytosine makes close contact (2.5 Å) to the ribose oxygen of adjacent guanosines. Similarly, the 2′-OH of each guanosine makes close contact (2.7 Å) with the adjacent adenosine N_6 of the opposing DNA strand.

To this point, d(CGA) triplet repeat sequences have not been implicated in human disease and are found least frequently in eukaryotic genomes.[85,87] This raises the interesting possibility that their propensity to form highly stable ps-duplex structures could be a significant factor contributing to the under-representation of the d(CGA) triplet repeat in eukaryotic genomes. Though questions have arisen about the likelihood of C-CH+-dependent structures forming *in vivo*, mounting evidence – including data presented here – suggest that such structures can form at near-neutral pH under crowded conditions.

Appendix

Appendix 1. Matlab code for UV melting data analysis

```
function [output_table] = Luteran(mdca,Epsil85,varargin)
%Luteran Fits sets of melt curves to common parameters
%   mdca = cell array, where each element is an nx2 matrix with columns of T,Abs
%   Since the CT's come from the absorbance at 85 anyway, we will just calculate CT
%   rather than get it from the user.
%   Optional arguments: a vector of pathlengths (default is 1), TLvec and THvec for
%   lo/hi temps.
%   To use for each data set, leave blank or enter vectors with 0 and 100 to use all the
%   data.
%   The data that is used for each melt must include the 85 C point or the program will
%   normalize to the highest T available, with very unpredictable results. Reference
%   temp at the low end is 20 C but data in that range is not needed for fits so far.
%   Third optional argument is the name of the DNA or other string.
%   The demand for a cell array as input is because the melt runs may have different
%   sizes, hence can't put them all into one matrix or 3-D matrix, stuck with cell arrays.
%   The program concatenates them, which allows for future global fitting maybe.
%   To look at hairpin oligos (unimolecular melt), set bimolecular or hairpin variables
%   below by hand.
numsets = size(mdca,1);
%output_ca = cell(numsets,1);
%Abs_ca = cell(numsets,1);
%Tsave = cell(numsets,1);
pathlengths = ones(numsets,1);
TLvec = zeros(numsets,1);
THvec = 100*ones(numsets,1);
Titlestring = 'DNA Melting Curves';
% Set these by hand. One and only one must be true, the other must be false
bimolecular = false(1);
hairpin = true(1);
%
maxnumopts = 4;
numvarargs = length(varargin);
if numvarargs > maxnumopts
   error('Luteran:TooManyInputs', ...
      'Luteran allows at most %d optional inputs',maxnumopts);
end
%
% Now put the new input values from varargin into the optargs cell array, overwriting
% the default values.
%
% But if the user has entered '' then we do not want to replace the default
% There is probably a more elegant way to do this. (This doesn't work yet)
```

```
optargs{1} = pathlengths;
optargs{2} = TLvec;
optargs{3} = THvec;
optargs{4} = Titlestring;
outvarnames = {'DH0','loDH0','hiDHo','DS0','loDS0','hiDS0',...
   'mds','lomds','himds','mss','lomss','himss',...
   'Epsil20','loEpsil20','hiEpsil20','CT','Tm','Normofresiduals','Normofres_CT'};
if length(varargin) > 0
   aretheythere = cellfun(@isempty,varargin);
   for ivarg = 1:length(varargin)
      if ~aretheythere(ivarg)
         optargs(ivarg) = varargin(ivarg);
      end
   end
%optargs(1:numvarargs) = varargin;
   [pathlengths,TLvec,THvec,Titlestring] = optargs{:};
end
meltmat = cell2mat(mdca);
init_temp_index(1) = 1;
for i = 1:numsets
   numtemps(i) = size(mdca{i},1);
   final_temp_index(i) = init_temp_index(i) + numtemps(i) - 1;
   if i < numsets
      init_temp_index(i+1) = final_temp_index(i) + 1;
   end
   meltmat(init_temp_index(i):final_temp_index(i),2) =
meltmat(init_temp_index(i):final_temp_index(i),2)/pathlengths(i);
end
% f1 figure was just for development -- plot plain data
% f1 = figure
% hold on
% for i = 1:numsets
%
plot(meltmat(init_temp_index(i):final_temp_index(i),1),meltmat(init_temp_index(i):f
inal_temp_index(i),2))
% end
f2 = figure('Position',[200,200,1000,1000]);
hold on
f3 = figure('Position',[300,300,1000,1000]);
hold on
for i = 1:numsets
   % At this point truncate the data to fit only the specified range
%    TLvec(i)
   minTindex(i) = find(meltmat(init_temp_index(i):final_temp_index(i),1) >
TLvec(i),1,'first');
```

```
   maxTindex(i) = find(meltmat(init_temp_index(i):final_temp_index(i),1) <
THvec(i),1,'last');
   Tvec = meltmat(init_temp_index(i)+minTindex(i)-
1:init_temp_index(i)+maxTindex(i)-1,1);
   [~,T85] = min(abs(Tvec-85));
%    Uncomment lines below to debug
%    meltmat(init_temp_index(i)+minTindex(i)+T85-2,1)
%    meltmat(init_temp_index(i)+minTindex(i)+T85-2,2)
   CTvec(i) = meltmat(init_temp_index(i)+minTindex(i)+T85-2,2)/Epsil85;
   lb = [-300000,-500,-.005,-0.005,Epsil85/4];
   ub = [-5000,-50,.2,.2,Epsil85];
   TL = 20;
   options = optimset('TolFun',1e-16,'TolX',1e-16,...
         'MaxFunEvals',25000,'MaxIter',5000);
   if (bimolecular)
      initvals = [-75000,-200,.001,.001,0.8*Epsil85];
      [thermvals(i,:),resnorm(i), derres,~,output,~,jacobian] = ...
         lsqnonlin(@Abs_diff_Lut,initvals,lb,ub,options,...
            Tvec,meltmat(init_temp_index(i)+minTindex(i)-
1:init_temp_index(i)+maxTindex(i)-1,2),...
            Epsil85,TL,CTvec(i));
      Tmelting(i) = thermvals(i,1)/(thermvals(i,2) + 1.987*log(CTvec(i))) - 273.15;
   elseif (hairpin)
      initvals = [-40000,-100,.001,.001,0.5*Epsil85];
      [thermvals(i,:),resnorm(i), derres,~,output,~,jacobian] = ...
         lsqnonlin(@Abs_diff_hp,initvals,lb,ub,options,...
            Tvec,meltmat(init_temp_index(i)+minTindex(i)-
1:init_temp_index(i)+maxTindex(i)-1,2),...
            Epsil85,TL,CTvec(i));
      Tmelting(i) = thermvals(i,1)/thermvals(i,2) - 273.15;
   else
      disp('No choice of type of problem?')
      return
   end

%    thermvals(1)
%    thermvals(2)
%    thermvals(3)
%    thermvals(4)
%    thermvals(5)
%    Abs_pred_Lut(thermvals,Tvec,Epsil85,TL,CTvec(i))
%   Tsave{i} = Tvec;
%  Plot scaled results to figure 2
%---------------------
% nlparci provides 95% confidence limits derived from the output of
% lsqnonlin. Returns ci matrix; first column is lower limits second column is upper
```

```
% limits of the fit variables.
% Note the syntax seems to have changed from when I first used it.
   try
      ci = nlparci(thermvals(i,:),derres,'jacobian',jacobian);
      Uplim5(i,:) = ci(:,1)';
            Dnlim5(i,:) = ci(:,2)';
   catch
      disp(['Nlparci failed for iteration ',num2str(i),'. Setting params to best fit.'])
      Uplim5(i,:) = thermvals(i,:);
      Dnlim5(i,:) = thermvals(i,:);
   end
   figure(f2)
   hold on
   box on
   ax = gca;
   ax.ColorOrderIndex = i;
   legf2set(i) =
plot(meltmat(init_temp_index(i):final_temp_index(i),1),meltmat(init_temp_index(i):f
inal_temp_index(i),2)/(CTvec(i)*Epsil85),'s','MarkerSize',10);
   ax = gca;
   ax.ColorOrderIndex = i;
   if (bimolecular)

plot(Tvec,Abs_pred_Lut(thermvals(i,:),Tvec,Epsil85,TL,CTvec(i))/(CTvec(i)*Epsil8
5),'-','LineWidth',2)
   % This line plots a star to indicate the Tm

plot(Tmelting(i),Abs_pred_Lut(thermvals(i,:),Tmelting(i),Epsil85,TL,CTvec(i))/(CTv
ec(i)*Epsil85),'pk','MarkerSize',14)
   elseif (hairpin)

plot(Tvec,Abs_pred_hp(thermvals(i,:),Tvec,Epsil85,TL,CTvec(i))/(CTvec(i)*Epsil85
),'-','LineWidth',2)
   % This line plots a star to indicate the Tm

plot(Tmelting(i),Abs_pred_hp(thermvals(i,:),Tmelting(i),Epsil85,TL,CTvec(i))/(CTv
ec(i)*Epsil85),'pk','MarkerSize',14)
   end
   xlabel('Temperature (∞C)','FontSize',14)
   ylabel('Normalized Absorbance (260 nm)','FontSize',14)
   title(['Normalized ',Titlestring],'FontSize',18)
%   Plot non-scaled results to figure 3
   figure(f3)
   hold on
   box on
   ax = gca;
```

```
   ax.ColorOrderIndex = i;
   legf3set(i)=
plot(meltmat(init_temp_index(i):final_temp_index(i),1),meltmat(init_temp_index(i):f
inal_temp_index(i),2),'o','MarkerSize',10);
   ax = gca;
   ax.ColorOrderIndex = i;
   if (bimolecular)
      plot(Tvec,Abs_pred_Lut(thermvals(i,:),Tvec,Epsil85,TL,CTvec(i)),'-
','LineWidth',2)

plot(Tmelting(i),Abs_pred_Lut(thermvals(i,:),Tmelting(i),Epsil85,TL,CTvec(i)),'pk','
MarkerSize',14)
   elseif (hairpin)
      plot(Tvec,Abs_pred_hp(thermvals(i,:),Tvec,Epsil85,TL,CTvec(i)),'-
','LineWidth',2)

plot(Tmelting(i),Abs_pred_hp(thermvals(i,:),Tmelting(i),Epsil85,TL,CTvec(i)),'pk','
MarkerSize',14)
   end
   xlabel('Temperature (∞C)','FontSize',14)
   ylabel('Absorbance (260 nm)','FontSize',14)
   title(['Absolute ',Titlestring],'FontSize',18);
% Put code in here to rescale to a fixed hyprochromicity and then do a
% global fit to DH0,DS0,mss,mds

end
figure(f2)
legend(legf2set);
figure(f3)
legend(legf3set);
%
output_table =
table(thermvals(:,1),Dnlim5(:,1),Uplim5(:,1),thermvals(:,2),Dnlim5(:,2),Uplim5(:,2),.
..

thermvals(:,3),Dnlim5(:,3),Uplim5(:,3),thermvals(:,4),Dnlim5(:,4),Uplim5(:,4),therm
vals(:,5),...
   Dnlim5(:,5),Uplim5(:,5),CTvec',Tmelting',resnorm',resnorm'./CTvec',...
   'VariableNames',outvarnames);

end

function Absd = Abs_diff_Lut(fitvals,Tvec,Exp_vec,Epsil85,TL,CT)
Absd = Exp_vec - Abs_pred_Lut(fitvals,Tvec,Epsil85,TL,CT);
end
```

```
function Absvec = Abs_pred_Lut(fitvals,Tvec,Epsil85,TL,CT)
R = 1.987;
DH0 = fitvals(1);
DS0 = fitvals(2);
mds = fitvals(3);
mss = fitvals(4);
EpsilTL = fitvals(5);
TK = Tvec + 273.15;
Keq = exp((-DH0./TK + DS0)/R) ;
alpha = 1 + 1./(4*CT*Keq) - sqrt(1./(2*CT*Keq)+1./(16*CT^2*Keq.^2));
Absvec=CT*alpha*EpsilTL.*(1+mds*(Tvec-TL))+CT*(1-
alpha)*Epsil85.*(1+mss*(Tvec-85));
end

function Absd_hp = Abs_diff_hp(fitvals,Tvec,Exp_vec,Epsil85,TL,CT)
Absd_hp = Exp_vec - Abs_pred_hp(fitvals,Tvec,Epsil85,TL,CT);
end

function Absvec_hp = Abs_pred_hp(fitvals,Tvec,Epsil85,TL,CT)
R = 1.987;
DH0 = fitvals(1);
DS0 = fitvals(2);
mds = fitvals(3);
mss = fitvals(4);
EpsilTL = fitvals(5);
TK = Tvec + 273.15;
Keq = exp((-DH0./TK + DS0)/R) ;
alpha = Keq./(1 + Keq);
Absvec_hp = CT*alpha*EpsilTL.*(1+mds*(Tvec-TL)) + CT*(1-
alpha)*Epsil85.*(1+mss*(Tvec-85));
end
```

References

1. Watson, J. D.; Crick, F. H. C., Genetical Implications of the Structure of Deoxyribonucleic Acid. *Nature* **1953,** *171* (4361), 964-967.
2. Watson, J. D.; Crick, F. H. C., Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953,** *171* (4356), 737-738.
3. Franklin, R. E.; Gosling, R. G., Molecular Configuration in Sodium Thymonucleate. *Nature* **1953,** *171* (4356), 740-741.
4. Kaushik, M.; Kaushik, S.; Roy, K.; Singh, A.; Mahendru, S.; Kumar, M.; Chaudhary, S.; Ahmed, S.; Kukreti, S., A bouquet of DNA structures: Emerging diversity. *Biochem. Biophys. Rep.* **2016,** *5*, 388-395.
5. Gellert, M.; Lipsett, M. N.; Davies, D. R., Helix Formation by Guanylic Acid. *Proc. Natl. Acad. Sci. U.S.A.* **1962,** *48* (12), 2013-2018.
6. Ralph, R. K.; Connors, W. J.; Khorana, H. G., Secondary structure and aggregation in deoxyguanosine oligonucleotides. *J. Am. Chem. Soc.* **1962,** *84* (11), 2265-2266.
7. Huppert, J. L., Four-stranded nucleic acids: structure, function and targeting of G-quadruplexes. *Chem. Soc. Rev.* **2008,** *37* (7), 1375-1384.
8. Hardin, C. C.; Watson, T.; Corregan, M.; Bailey, C., Cation-Dependent Transition between the Quadruplex and Watson-Crick Hairpin Forms of d(CGCG3GCG). *Biochemistry* **1992,** *31* (3), 833-841.
9. Włodarczyk, A.; Grzybowski, P.; Patkowski, A.; Dobek, A., Effect of Ions on the Polymorphism, Effective Charge, and Stability of Human Telomeric DNA. Photon Correlation Spectroscopy and Circular Dichroism Studies. *J. Phys. Chem.* **2005,** *109* (8), 3594-3605.
10. Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S., Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.* **2006,** *34* (19), 5402-15.
11. Ma, Y.; Iida, K.; Nagasawa, K., Topologies of G-quadruplex: Biological functions and regulation by ligands. *Biochem. Biophys. Res. Commun.* **2020,** *531* (1), 3-17.
12. Gehring, K.; Leroy, J. L.; Gueron, M., A tetrameric DNA structure with protonated cytosine· cytosine base pairs. *Nature* **1993,** *363*, 561-565.
13. Kang, C.; Berger, I.; Lockshin, C.; Ratliff, R.; Moyzis, R.; Rich, A., Crystal structure of intercalated four-stranded d(C3T) at 1.4 A resolution. *Proc. Natl. Acad. Sci. U.S.A.* **1994,** *91* (24), 11636-11640.
14. Gueron, M.; Leroy, J. L., The i-motif in nucleic acids. *Curr. Opin. Struct. Biol.* **2000,** *10* (3), 326-331.
15. Kanaori, K.; Maeda, A.; Kanehara, H.; Tajima, K.; Makino, K., 1H Nuclear Magnetic Resonance Study on Equilibrium between Two Four-Stranded Solution Conformations of Short d(CnT). *Biochemistry* **1998,** *37* (37), 12979-12986.
16. Leroy, J. L.; Gueron, M., Solution structures of the i-motif tetramers of d(TCC), d(5methylCCT) and d(T5methylCC): novel NOE connections between amino protons and sugar protons. *Structure* **1995,** *3* (1), 101-120.
17. Dai, J.; Carver, M.; Yang, D., Polymorphism of human telomere quadruplex structures. *Biochimie* **2008,** *90* (8), 1172-1183.

18. Fleming, A. M.; Ding, Y.; Rogers, R. A.; Zhu, J.; Zhu, J.; Burton, A. D.; Carlisle, C. B.; Burrows, C. J., 4n–1 Is a "Sweet Spot" in DNA i-Motif Folding of 2′-Deoxycytidine Homopolymers. *J. Am. Chem. Soc.* **2017,** *139* (13), 4682-4689.
19. Wright, E. P.; Huppert, J. L.; Waller, Z. A. E., Identification of multiple genomic DNA sequences which form i-motif structures at neutral pH. *Nucleic Acids Res.* **2017,** *45* (6), 2951-2959.
20. Guédin, A.; Gros, J.; Alberti, P.; Mergney, J.-L., How long is too long? Effects of loop size on G-quadruplex stability. *Nucleic Acids Res.* **2010,** *38* (21), 7858-7868.
21. Abou Assi, H.; Garavís, M.; González, C.; Damha, M. J., i-Motif DNA: structural features and significance to cell biology. *Nucleic Acids Res.* **2018,** *46* (16), 8038-8056.
22. Zhou, J.; Wei, C.; Jia, G.; Wang, X.; Feng, Z.; Li, C., Formation of i-motif structure at neutral and slightly alkaline pH. *Mol. BioSyst.* **2010,** *6* (3), 580-586.
23. Largy, E.; Marchand, A.; Amrane, S.; Gabelica, V.; Mergny, J. L., Quadruplex Turncoats: Cation-Dependent Folding and Stability of Quadruplex-DNA Double Switches. *J. Am. Chem. Soc.* **2016,** *138* (8), 2780-2792.
24. Kankia, B. I.; Marky, L. A., Folding of the Thrombin Aptamer into a G-Quadruplex with Sr2+: Stability, Heat, and Hydration. *J. Am. Chem. Soc.* **2001,** *123* (44), 10799-10804.
25. Day, H. A.; Huguin, C.; Waller, Z. A. E., Silver cations fold i-motif at neutral pH. *Chem. Commun.* **2013,** *49* (70), 7696-7698.
26. Abdelhamid, M. A.; Fabian, L.; MacDonald, C. J.; Cheesman, M. R.; Gates, A. J.; Waller, Z. A., Redox-dependent control of i-Motif DNA structure using copper cations. *Nucleic Acids Res.* **2018,** *46* (12), 5886-5893.
27. Miller, M. C.; Buscaglia, R.; Chaires, J. B.; Lane, A. N.; Trent, J. O., Hydration Is a Major Determinant of the G-Quadruplex Stability and Conformation of the Human Telomere 3′ Sequence of d(AG3(TTAG3)3). *J. Am. Chem. Soc.* **2010,** *132* (48), 17105-17107.
28. Xue, Y.; Kan, Z.; Q., W.; Yao, Y.; Liu, J.; Hao, Y.; Tan, Z., Human Telomeric DNA Forms Parallel-Stranded Intramolecular G-Quadruplex in K+ Solution under Molecular Crowding Condition. *J. Am. Chem. Soc.* **2007,** *129* (36), 11185-11191.
29. Rajendran, A.; Nakano, S.; Sugimoto, N., Molecular crowding of the cosolutes induces an intramolecular i-motif structure of triplet repeat DNA oligomers at neutral pH. *Chem. Commun.* **2010,** *46* (8), 1299-1301.
30. Luteran, E. M.; Kahn, J. D.; Paukstelis, P. J., Stability of the pH-Dependent Parallel-Stranded d(CGA) Motif. *Biophys. J.* **2020,** *119* (8), 1580-1589.
31. Zheng, M.; Huang, X.; Smith, G. K.; Yang, X.; Gao, X., Genetically Unstable CXG Repeats are Structurally Dynamic and Have a High Propensity for Folding. An NMR and UV Spectroscopic Study. *J. Mol. Biol.* **1996,** *264* (2), 323-336.
32. Kejnovska, I.; Tumova, M.; Vorlickova, M., (CGA)4: parallel, anti-parallel, right-handed and left-handed homoduplexes of a trinucleotide repeat DNA. *Biochim. Biophys. Acta.* **2001,** *1527* (1-2), 73-80.

33. Robinson, H.; van der Marel, G. A.; van Boom, J. H.; Wang, A. H., Unusual DNA conformation at low pH revealed by NMR: parallel-stranded DNA duplex with homo base pairs. *Biochemistry* **1992,** *31* (43), 10510-7.
34. Robinson, H.; Wang, A. H., 5'-CGA sequence is a strong motif for homo base-paired parallel-stranded DNA duplex as revealed by NMR analysis. *Proc. Natl. Acad. Sci. U.S.A.* **1993,** *90* (11), 5224-5228.
35. Sunami, T.; Kondo, J.; Kobuna, T.; Hirao, I.; Watanabe, K.; Miura, K.; Takenaka, A., Crystal structure of d(GCGAAAGCT) containing a parallel-stranded duplex with homo base pairs and an anti-parallel duplex with Watson-Crick base pairs. *Nucleic Acids Res.* **2002,** *30* (23), 5253-60.
36. Tripathi, S.; Paukstelis, P. J., Structural Implications of Homopyrimidine Base Pairs in the Parallel-Stranded d(YGA) Motif. *ChemBioChem* **2016,** *17* (12), 1177-1183.
37. Tripathi, S.; Zhang, D.; Paukstelis, P. J., An intercalation-locked parallel-stranded DNA tetraplex. *Nucleic Acids Res.* **2015,** *43* (3), 1937-1944.
38. Wang, Y.; Patel, D. J., Solution structure of the d(T-C-G-A) duplex at acidic pH. A parallel-stranded helix containing C+ .C, G.G and A.A pairs. *J. Mol. Biol.* **1994,** *242* (4), 508-26.
39. Robinson, H.; van Boom, J. H.; Wang, A. H. J., 5'-CGA Motif Induces Other Sequences To Form Homo Base-Paired Parallel-Stranded DNA Duplex: The Structure of (G-A)n Derived from Four DNA Oligomers Containing (G-A)3 Sequence. *J. Am. Chem. Soc.* **1994,** *114* (4), 1565-1566.
40. Kettani, A.; Bouaziz, S.; Skripkin, E.; Majumdar, A.; Wang, W.; Jones, R. A.; Patel, D. J., Interlocked mismatch-aligned arrowhead DNA motifs. *Structure* **1999,** *7* (7), 803-815.
41. Rippe, K.; Fritsch, V.; Westhof, E.; Jovin, T. M., Alternating d(G-A) sequences form a parallel-stranded DNA homoduplex. *EMBO J.* **1992,** *11* (10), 3777-86.
42. Paukstelis, P. J.; Nowakowski, J.; Birktoft, J. J.; Seeman, N. C., Crystal structure of a continuous three-dimensional DNA lattice. *Chem Biol* **2004,** *11* (8), 1119-26.
43. Assi, H. A.; Garavis, M.; Gonzalez, C.; Damha, M. J., I-Motif DNA: structural features and significance to cell biology. *Nucleic Acids Res.* **2018,** *46* (16), 8038-8056.
44. Bacolla, A.; Wells, R. D., Non-B DNA Confirmations, Genomic Rearrangements, and Human Disease. *J Biol Chem* **2019,** *279* (46), 47411-47414.
45. Brooks, T. A.; Kendrick, S.; Hurley, L., Making sense of G-quadruplex and i-motif functions in oncogene promoters. *FEBS* **2010,** *277* (17), 3459-3469.
46. Han, J.; Hsu, C.; Zhu, Z.; Longshore, J. W.; Finley, W. H., Over-representation of the disease associated (CAG) and (CGG) repeats in the human genome. *Nucleic. Acids. Res.* **1994,** *22* (9), 1735-1740.
47. Sundquist, W. I.; Klug, A., Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature* **1989,** *342* (6251), 825-829.
48. Parkinson, G. N.; Lee, M. P. H.; Neidle, S., Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature* **2002,** *417* (6891), 876-880.
49. Lane, A. N.; Chaires, J. B.; Gray, R. D.; Trent, J. O., Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res.* **2008,** *36* (17), 5482-5515.

50. Huppert, J. L.; Balasubramanian, S., Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.* **2005,** *33* (9), 2908-2916.
51. Maizels, N.; Gray, L. T., The G4 Genome. *PLoS Genet.* **2013,** *9* (4), 1-6.
52. Besnard, E.; Babled, A.; Lapasset, L.; Milhavet, O.; Parrinello, H.; Dantec, C.; Marin, J.-M.; Lemaitre, J.-M., Unraveling cell type–specific and reprogrammable human replication origin signatures associated with G-quadruplex consensus motifs. *Nat. Struct. Mol. Biol.* **2012,** *19* (8), 837-844.
53. Schaffitzel, C.; Berger, I.; Postberg, J.; Hanes, J.; Lipps, H. J.; Pluckthun, A., In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. *Proc. Natl. Acad. Sci. U.S.A.* **2001,** *98* (15), 8562-8577.
54. Fernando, H.; Rodriguez, R.; Balasubramanian, S., Selective Recognition of a DNA G-Quadruplex by an Engineered Antibody. **2008,** *47* (36), 9365-9371.
55. Lam, E. Y. N.; Beraldi, D.; Tannahill, D.; Balasubramanian, S., G-quadruplex structures are stable and detectable in human genomic DNA. *Nat. Commun.* **2013,** *4* (1796), 1-8.
56. Paeschke, K.; Simonsson, T.; Postberg, J.; Rhodes, D.; Lipps, H. J., Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nat. Struct. Mol. Biol.* **2005,** *12* (10), 847-854.
57. Verma, A.; Halder, K.; Halder, R.; Yadav, V. K.; Rawal, P.; Thakur, R. K.; Mohd, F.; Sharma, A.; Chowdhury, S., Genome-Wide Computational and Expression Analyses Reveal G-Quadruplex DNA Motifs as Conserved cis-Regulatory Elements in Human and Related Species. *J. Med. Chem.* **2008,** *51* (18), 5641-5649.
58. Hershman, S. G.; Chen, Q.; Lee, J. Y.; Kozak, M. L.; Yue, P.; Wang, L.-S.; Johnson, F. B., Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in *Saccharomyces cerevisiae*. *Nucleic Acids Res.* **2008,** *36* (1), 144-156.
59. Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H., Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc. Natl. Acad. Sci. U.S.A.* **2002,** *99* (18), 11593-11598.
60. Gray, L. T.; Vallur, A. C.; Eddy, J.; Maizels, N., G quadruplexes are genomewide targets of transcriptional helicases XPB and XPD. *Nat. Chem. Biol.* **2014,** *10* (4), 313-318.
61. Sun, D.; Hurley, L. H., The Importance of Negative Superhelicity in Inducing the Formation of G-Quadruplex and i-Motif Structures in the c-Myc Promoter: Implications for Drug Targeting and Control of Gene Expression. *J. Med. Chem.* **2009,** *52* (9), 2863-2874.
62. Xu, Y.; Sugiyama, H., Formation of the G-quadruplex and i-motif structures in retinoblastoma susceptibility genes (Rb). *Nucleic Acids Res.* **2006,** *34* (3), 949-954.
63. Dexheimer, T. S.; Sun, D.; Hurley, L. H., Deconvoluting the Structural and Drug-Recognition Complexity of the G-Quadruplex-Forming Region Upstream of the bcl-2 P1 Promoter. *J. Am. Chem. Soc.* **2006,** *128* (16), 5404-5415.

64. Palumbo, S. L.; Ebbinghaus, S. W.; Hurley, L. H., Formation of a Unique End-to-End Stacked Pair of G-Quadruplexes in the hTERT Core Promoter with Implications for Inhibition of Telomerase by G-Quadruplex-Interactive Ligands. *J. Am. Chem. Soc.* **2009,** *131* (31), 10878-10891.
65. Qin, Y.; Rezler, E. M.; Gokhale, V.; Sun, D.; Hurley, L. H., Characterization of the G-quadruplexes in the duplex nuclease hypersensitive element of the PDGF-A promoter and modulation of PDGF-A promoter activity by TMPyP4. *Nucleic Acids Res.* **2007,** *35* (22), 7698-7713.
66. Brown, S. L.; Kendrick, S., The i-Motif as a Molecular Target: More Than a Complementary DNA Secondary Structure. *Pharmaceuticals* **2021,** *14* (2), 96.
67. Zeraati, M.; Langley, D. B.; Schofield, P.; Moye, A. L.; Rouet, R.; Hughes, W. E.; Bryan, T. M.; Dinger, M. E.; Christ, D., I-motif DNA structures are formed in the nuclei of human cells. *Nat. Chem.* **2018,** *10* (6), 631-637.
68. Kendrick, S.; Akiyama, Y.; Hecht, S. M.; Hurley, L. H., The i-motif in the bcl-2 P1 promoter forms an unexpectedly stable structure with a unique 8:5:7 loop folding pattern. *J. Am. Chem. Soc.* **2009,** *131* (48), 17667-17676.
69. Kendrick, S.; Kang, H. J.; Alam, M. P.; Madathil, M. M.; Agrawal, P.; Gokhale, V.; Yang, D.; Hecht, S. M.; Hurley, L. H., The dynamic character of the BCL2 promoter i-motif provides a mechanism for modulation of gene expression by compounds that bind selectively to the alternative DNA hairpin structure. *J. Am. Chem. Soc.* **2014,** *136* (11), 4161-4171.
70. Kang, H. J.; Kendrick, S.; Hecht, S. M.; Hurley, L. H., The transcriptional complex between the BCL2 i-motif and hnRNP LL is a molecular switch for control of gene expression that can be modulated by small molecules. *J. Am. Chem. Soc.* **2014,** *136* (11), 4172-4185.
71. Brown, R. V.; Wang, T.; Chappeta, V. R.; Wu, G.; Onel, B.; Chawla, R.; Quijada, H.; Camp, S. M.; Chiang, E. T.; Lassiter, Q. R.; Lee, C.; Phanse, S.; Turnidge, M. A.; Zhao, P.; Garcia, J. G. N.; Gokhale, V.; Yang, D.; Hurley, L. H., The Consequences of Overlapping G-Quadruplexes and i-Motifs in the Platelet-Derived Growth Factor Receptor β Core Promoter Nuclease Hypersensitive Element Can Explain the Unexpected Effects of Mutations and Provide Opportunities for Selective Targeting of Both Structures by Small Molecules To Downregulate Gene Expression. *J. Am. Chem. Soc.* **2017,** *139* (22), 7456-7475.
72. Sutherland, C.; Cui, Y.; Mao, H.; Hurley, L. H., A Mechanosensor Mechanism Controls the G-Quadruplex/i-Motif Molecular Switch in the MYC Promoter NHE III1. *J. Am. Chem. Soc.* **2016,** *138* (42), 14138-14151.
73. Shu, B.; Cao, J.; Kuang, G.; Qiu, J.; Zhang, M.; Zhang, Y.; Wang, M.; Li, X.; Kang, S.; Ou, T.-M.; Tan, J.-H.; Huang, Z.-S.; Li, D., Syntheses and evaluation of new acridone derivatives for selective binding of oncogene c-myc promoter i-motifs in gene transcriptional regulation. *Chem. Commun.* **2018,** *54* (16), 2036-2039.
74. Peprah, E., Fragile X syndrome: the FMR1 CGG repeat distribution among world populations. *Ann Hum Genet* **2012,** *76* (2), 178-191.
75. Jin, P.; Warren, S. T., Understanding the molecular basis of fragile X syndrome. *Hum. Mol. Genet.* **2000,** *9* (6), 901-908.

76. Verkerk, A. J.; Pieretti, M.; Sutcliffe, J. S.; Fu, Y. H.; Kuhl, D. P.; Pizzuti, A.; Reiner, O.; Richards, S.; Victoria, M. F.; Zhang, F. P., Identification of a gene (FMR-1) containing a CGG repeat coincident with a breakpoint cluster region exhibiting length variation in fragile X syndrome. *Cell* **1991,** *65* (5), 905-914.
77. The Huntington's Disease Collaborative Research Group, A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes. *Cell* **1993,** *72* (6), 971-983.
78. Budworth, H.; McMurray, C. T., A Brief History of Triplet Repeat Diseases. *Methods Mol. Biol.* **2014,** *1010*, 3-17.
79. Mahadevan, M.; Tsilfidis, C.; Sabourin, L.; Shutler, G.; Amemiya, C.; Jansen, G.; Neville, C.; Narang, M.; Barcelo, J.; O'Hoy, K., Myotonic dystrophy mutation: an unstable CTG repeat in the 3' untranslated region of the gene. *Science* **1992,** *255* (5049), 2153-1255.
80. Campuzano, V.; Montermini, L.; Molto, M. D.; Pianese, L.; Cossee, M.; Cavalcanti, F.; Monros, E.; Duclos, F.; Monticelli, A.; Zara, F.; Canizares, J.; Koutnikova, H.; Bidichandani, S. I.; Gallera, C.; Brice, A.; Trouillas, P.; De Michelle, G.; Filla, A.; De Frutos, R.; Palau, F.; Patel, P. I.; Di Donato, S.; Mandel, J. L. C., S.; Koenig, M.; Pandolfo, M., Friedreich's Ataxia: Autosomal Recessive Disease Caused by an Intronic GAA Triplet Repeat Expansion. *Science* **1996,** *271* (5254), 1423-1427.
81. Gacy, A. M.; Goellner, G. M.; Spiro, C.; Chen, X.; Gupta, G.; Bradbury, E. M.; Dyer, R. B.; Mikesell, M. J.; Yao, J. Z.; Johnson, A. J.; Richter, A.; Melancon, S. B.; McMurray, C. T., GGA Instability in Friedreich's Ataxia Shares a Common, DNA-Directed and Intraallelic Mechanism with Other Trinucleotide Diseases. *Mol. Cell* **1998,** *1* (4), 583-593.
82. Wells, R. D., Non-B DNA conformations, mutagenesis and disease. *Trends Biochem. Sci.* **2007,** *32* (6), 271-278.
83. Mirkin, S. M.; Smirnova, E. V., Positioned to expand. *Nat. Genet.* **2002,** *31* (1), 5-6.
84. Wang, G.; Vasquez, K. M., Effects of Replication and Transcription on DNA Structure-Related Genetic Instability. *Genes* **2017,** *8* (1), 1-17.
85. Kozlowski, P.; de Mezer, M.; Krzyzosiak, W. J., Trinucleotide repeats in human genome and exome. *Nucleic Acids Res.* **2010,** *38* (12), 4027-4039.
86. Andrew, S. E.; Goldberg, Y. P.; Kremer, B.; Telenius, H.; Theilmann, J.; Adam, S.; Starr, E.; Squitieri, F.; Lin, B.; Kalchman, M. A.; Graham, R. K.; Hayden, M. R., The relationship between trinucleotide (CAG) repeat length and clinical features of Huntington's disease. *Nature Genetics* **1993,** *4* (4), 398-403.
87. Astolfi, P.; Bellizzi, D.; Sgaramella, V., Frequency and coverage of trinucleotide repeats in eukaryotes. *Gene* **2003,** *317* (1-2), 117-125.
88. Paiva, A. M.; Sheardy, R. D., Influence of Sequence Context and Length on the Structure and Stability of Triplet Repeat DNA Oligomers. *Biochemistry* **2004,** *43* (44), 14218-14227.
89. Volker, J.; Makube, N.; Plum, G. E.; Klump, H. H.; Breslauer, K. J., Conformational energetics of stable and metastable states formed by DNA triplet repeat oligonucleotides: Implications for triplet expansion diseases. *Proc. Natl. Acad. Sci. U.S.A.* **2002,** *99* (23), 14700-14705.

90. Poggi, L.; Richard, G.-F., Alternative DNA Structures In Vivo: Molecular Evidence and Remaining Questions. *Microbiol. Mol. Biol. Rev.* **2021,** *85* (1), 1-24.
91. Seeman, N. C.; Sleiman, H. F., DNA nanotechnology. *Nat. Rev. Mater.* **2017,** *3*, 1-23.
92. Hagerman, P. J., Flexibility of DNA. *Ann. Rev. Biophys. Biophys. Chem.* **1988,** *17*, 265-286.
93. Winfree, E.; Liu, F. R.; Wenzler, L. A.; Seeman, N. C., Design and self-assembly of two-dimensional DNA crystals. *Nature* **1998,** *394* (6693), 539-544.
94. Li, X.; Yang, X.; Qi, J.; Seeman, N. C., Antiparallel DNA double crossover molecules as components for nanoconstruction. *J. Am. Chem. Soc.* **1996,** *118* (26), 6131-6140.
95. He, Y.; Chen, Y.; Liu, H.; Ribbe, A. E.; Mao, C., Self-assembly of hexagonal DNA two-dimensional (2D) arrays. *J. Am. Chem. Soc.* **2005,** *127* (35), 12202-12203.
96. Rotemund, P. W., Folding DNA to create nanoscale shapes and patterns. *Nature* **2006,** *440* (7082), 297-302.
97. Marchi, A. N.; Saaem, I.; Tian, J. D.; LaBean, T. H., One-pot assembly of a hetero-dimeric DNA origami from chip-derived staples and double-stranded scaffold. *ACS Nano* **2013,** *7* (2), 903-910.
98. Shih, W. M.; Quispe, J. D.; Joyce, G. F., A 1.7-kilobase single-stranded DNA that folds into a nanoscale octahedron. *Nature* **2004,** *427* (6975), 618-621.
99. Aldaye, F. A.; Sleiman, H. F., Modular access to structurally switchable 3D discrete DNA assemblies. *J. Am. Chem. Soc.* **2007,** *129* (44), 13376-13377.
100. Chen, J. H.; Seeman, N. C., Synthesis from DNA of a molecule with the connectivity of a cube. *Nature* **1991,** *350* (6319), 631-633.
101. 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* **2009,** *461* (7260), 74.
102. Shi, J. F.; Bergstrom, D. E., Assembly of novel DNA cycles with rigid tetrahedral linkers. *Angew. Chem. Int. Ed. Engl.* **1997,** *36* (1-2), 111-113.
103. Yang, H.; McLaughlin, C. K.; Aldaye, F. A.; Hamblin, G. D.; Rys, A. Z.; Rouiller, I.; Sleiman, H. F., Metal–nucleic acid cages. *Nat. Chem.* **2009,** *1* (5), 390-396.
104. Chien, M.-P.; Rush, A. M.; Thompson, M. P.; Gianneschi, N. C., Programmable shape-shifting micelles. *Angew. Chem. Int. Ed. Engl.* **2010,** *49* (30), 5076-5080.
105. Engelhard, D. M.; Nowak, J.; Clever, G. H., Copper-Induced Topology Switching and Thrombin Inhibition with Telomeric DNA G-Quadruplexes. *Angew. Chem. Int. Ed.* **2017,** *56* (38), 11640-11644.
106. Liu, D.; Balasubramanian, S., A Proton-Fuelled DNA Nanomachine. *Angew. Chem. Int. Ed.* **2003,** *42* (460), 5734-5736.
107. Muser, S. E.; Paukstelis, P. J., Three-Dimensional DNA Crystals with pH-Responsive Noncanonical Junctions. *J. Am. Chem. Soc.* **2012,** *134* (30), 12557-12564.

108. Modi, S.; G., S. M.; Goswami, D.; Gupta, G. D.; Mayor, S.; Krishnan, Y., A DNA nanomachine that maps spatial and temporal pH changes inside living cells. *Nat. Nanotechnol.* **2009,** *4* (5), 325-330.
109. Sha, R.; Xiang, L.; Liu, C.; Balaeff, A.; Zhang, Y.; Zhang, P.; Li, Y.; Beratan, D. N.; Tao, N.; Seeman, N. C., Charge splitters and charge transport junctions based on guanine quadruplexes. *Nat. Nanotechnol.* **2018,** *13* (4), 316-321.
110. Kotlyar, A. B.; Borovok, N.; Molotsky, T.; Cohen, H.; Shapir, E.; Porath, D., Long, Monomolecular Guanine-Based Nanowires. *Adv. Mater.* **2005,** *17* (15), 1901-1905.
111. Li, Z.; Mirkin, C. A., G-Quartet-Induced Nanoparticle Assembly. *J. Am. Chem. Soc.* **2005,** *127* (33), 11568-11569.
112. Su, Y. B.; Zhao, X.; Chen, L. J.; Qian, H. L.; Yan, X. P., Fabrication of G-quadruplex/porphyrin conjugated gold/persistent luminescence theranostic nanoprobe for imaging-guided photodynamic therapy. *Talanta* **2021,** *233*, 1-9.
113. Nesterova, I. V.; Nesterov, E. E., Rational Design of Highly Responsive pH Sensors Based on DNA i-Motif. *J. Am. Chem. Soc.* **2014,** *136* (25), 8843-8846.
114. Song, L.; Ho, V. H. B.; Chen, C.; Yang, Z.; Liu, D.; Chen, R.; Zhu, D., Efficient, pH-Triggered Drug Delivery Using a pH-Responsive DNA-Conjugated Gold Nanoparticle. *Adv. Healthc. Mater.* **2013,** *2* (2), 275-280.
115. Keum, J. W.; Bermudez, H., DNA-based delivery vehicles: pH-controlled disassembly and cargo release. *Chem. Commun.* **2012,** *48* (99), 12118-12120.
116. Ma, W.; Sun, H.; Chen, B.; Jia, R.; Huang, J.; Cheng, H.; He, X.; Huang, M.; Wang, K., Engineering a Facile Aptamer "Molecule-Doctor" with Hairpin-Contained I-Motif Enables Accurate Imaging and Killing of Cancer Cells. *Anal. Chem.* **2021,** *93* (43), 14552-14559.
117. Chandrasekaran, A. R.; Rusling, D. A., Triplex-forming oligonucleotides: a third strand for DNA nanotechnology. *Nucleic Acids Res* **2018,** *46* (3), 1021-1037.
118. Liao, W. C.; Riutin, M.; Parak, W. J.; Willner, I., Programmed pH-Responsive Microcapsules for the Controlled Release of CdSe/ZnS Quantum Dots. *ACS Nano* **2016,** *10* (9), 8683-8689.
119. Chen, Y.; Mao, C., Reprogramming DNA-Directed Reactions on the Basis of a DNA Conformational Change. *J. Am. Chem. Soc.* **2004,** *126* (41), 13240-13241.
120. Robinson, H.; Wang, A. H., 5'-CGA sequence is a strong motif for homo base-paired parallel-stranded DNA duplex as revealed by NMR analysis. *Proc. Natl. Acad. Sci.* **1993,** *90* (11), 5224-5228.
121. Robinson, H.; van der Marel, G. A.; van Boom, J. H.; Wang, A. H., Unusual DNA conformation at low pH revealed by NMR: parallel-stranded DNA duplex with homo base pairs. *Biochemistry* **1992,** *31* (43), 10510-10517.
122. SantaLucia, J.; Kierzek, R.; Turner, D. H., Functional group substitutions as probes of hydrogen bonding between GA mismatches in RNA internal loops. *J. Am. Chem. Soc.* **1991,** *113* (11), 4313-4322.
123. Xia, T.; SantaLucia, J.; Burkard, M. E.; Kierzek, R.; Schroeder, S. J.; Jiao, X.; Cox, C.; Turner, D., Thermodynamic Parameters for an Expanded Nearest-Neighbor Model for Formation of RNA Duplexes with Watson−Crick Base Pairs. *Biochemistry* **1998,** *37* (42), 14719-14735.

124. Allawi, H. T.; SantaLucia, J., Thermodynamics and NMR of Internal G·T Mismatches in DNA. *Biochemistry* **1997,** *36* (34), 10581-10594.
125. Petersheim, M.; Turner, D. H., Base-stacking and base-pairing contributions to helix stability: thermodynamics of double-helix formation with CCGG, CCGGp, CCGGAp, ACCGGp, CCGGUp, and ACCGGUp. *Biochemistry* **1983,** *22*, 256-263.
126. Siegfried, N. A.; Bevilacqua, P., Thinking Inside the Box: Designing, Implementing, and Interpreting Thermodynamic Cycles to Dissect Cooperativity in RNA and DNA Folding. *Methods Enzymol.* **2009**, *455*, 365-393.
127. Suck, D.; Oefner, C., Structure of DNase I at 2.0 A resolution suggests a mechanism for binding to and cutting DNA. *Nature* **1986,** *321* (5), 620-625.
128. Prince, W. S.; Baker, D. L.; Dodge, A. H.; Ahmed, A. E.; Chestnut, R. W.; Sinicropi, D. V., Pharmacodynamics of recombinant human DNase I in serum. *Clin. Exp. Immunol.* **1998,** *113* (2), 289-296.
129. Sambrook, J.; Russell, D. W., *Molecular Cloning: A Laboratory Manual*. 3rd ed.; Cold Spring Harbor Laboratory Press: Cold Spring Harbor, New York, 2001.
130. Desai, N. A.; Shankar, V., Single-strand-specific nucleases. *FEMS Microbiol. Rev.* **2003,** *26* (5), 457-491.
131. Koval, T.; Oxtergaard, L. H.; Lehmbeck, J.; Norgaard, A.; Lipovova, P.; Duskova, J.; Skalova, T.; Trundova, M.; Kolenko, P.; Fejfarova, K.; Stransky, J.; Svecova, L.; Hasek, J.; Dohnalek, J., Structural and Catalytic Properties of S1 Nuclease from Aspergillus oryzae Responsible for Substrate Recognition, Cleavage, Non-Specificity, and Inhibition. *PLoS One.* **2016,** *11* (12), 1-25.
132. Saoji, M.; Paukstelis, P. J., Sequence-dependent structural changes in a self-assembling DNA oligonucleotide. *Acta Crystallogr D Biol Crystallogr* **2015,** *71* (12), 2471-2478.
133. Muser, S. E.; Paukstelis, P. J., Three-dimensional DNA crystals with pH-responsive noncanonical junctions. *J Am Chem Soc* **2012,** *134* (30), 12557-64.
134. Zheng, M.; Han, X.; Gao, X., (GAC)3 parallel duplex. **2003**.
135. Weiss, M. S.; Hilgenfeld, R., On the use of the merging R factor as a quality indicator for X-ray data. *J. Appl. Cryst.* **1997,** *30*, 203-205.
136. Kobuna, T.; Sunami, T.; Kondo, J.; Takenaka, A., X-ray structure of d(CGAA), parallel-stranded DNA duplex with homo base pairs. *Nucleic Acids Res.* **2002,** *2* (1), 179-180.
137. Cui, J.; Waltman, P.; Le, V. H.; Lewis, E. A., The Effect of Molecular Crowding on the Stability of Human c-MYC Promoter Sequence I-Motif at Neutral pH. *Molecules* **2013,** *18* (10), 12751-12767.
138. Tang, W.; Niu, K.; Yu, G.; Jin, Y.; Zhang, X.; Peng, Y.; Chen, S.; Deng, H.; Li, S.; Wang, J.; Song, Q.; Feng, Q., In vivo visualization of the i-motif DNA secondary structure in the Bombyx mori testis. *Epigenetics & Chromatin* **2020,** *13* (12), 1-12.
139. Dzatko, S.; Krafcikova, M.; Hansel-Hertsch, R.; Fessl, T.; Fiala, R.; Loja, T.; Krafcik, D.; Mergny, J. L.; Foldynova-Trantirkova, S.; Trantirek, L., Evaluation of the Stability of DNA i-Motifs in the Nuclei of Living Mammalian Cells. *Angew. Chem. Int. Ed.* **2018,** *57* (8), 2165-2169.
140. Delano, W. L., **2002**, *PYMOL*. http://www.pymol.org.

141. Lu, X. J.; Olson, W. K., 3DNA: a software package for the analysis, rebuilding and visualization of three-dimensional nucleic acid structures. *Nucleic Acids Res.* **2003,** *31* (17), 5108-5121.
142. Leslie, A. G. W.; Powell, H. R., Processing Diffraction Data with Mosflm. In *Evolving Methods for Macromolecular Crystallography*, Read, R. J.; Sussman, J. L., Eds. Springer: 2007; pp 41-51.
143. Kabsch, W., XDS. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66* (2), 125-132.
144. Evans, P. R.; Murshudov, G. N., How good are my data and what is the resolution? *Acta Crystallogr. D Biol. Crystallogr.* **2013**, *69* (7), 1204-1214.
145. McCoy, A. J.; Grosse-Kunstleve, R. W.; Adams, P. D.; Winn, M. D.; Storoni, L. C.; Read, R. J., Phaser crystallographic software. *J. Appl. Cryst.* **2007,** *40* (4), 658-674.
146. Afonine, P. V.; Grosse-Kunstleve, R. W.; Echols, N.; Headd, J. J.; Moriarty, N. W.; Mustyakimov, M.; Terwilliger, T. C.; Urzhumtsev, A.; Zwart, P. H.; Adams, P. D., Towards automated crystallographic structure refinement with phenix.refine. *Acta Crystallogr. D Biol. Crystallogr.* **2012,** *68* (4), 352-367.
147. Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K., Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66* (4), 486-501.
148. Wallimann, P.; Kennedy, R. J.; Miller, J. S.; Shalongo, W.; Kemp, D. S., Dual Wavelength Parametric Test of Two-State Models for Circular Dichroism Spectra of Helical Polypeptides: Anomalous Dichroic Properties of Alanine-Rich Peptides. *J. Am. Chem. Soc.* **2003,** *125* (5), 1203-1220.
149. Douglas, S. M.; Dietz, H.; Liedl, T.; Hogberg, B.; Graf, F.; Shih, W. M., Self-assembly of DNA into nanoscale three-dimensional shapes. *Nature* **2009,** *459* (7245), 414-8.
150. Wei, B.; Dai, M.; Yin, P., Complex shapes self-assembled from single-stranded DNA tiles. *Nature* **2012,** *485* (7400), 623-626.
151. Ke, Y.; Ong, L. L.; Shih, W. M.; Yin, P., Three-Dimensional Structures Self-Assembled from DNA Bricks. *Science* **2012,** *338* (6111), 1177-1183.
152. Li, H.; Zhang, B.; Lu, X.; Tan, X.; Jia, F.; Xiao, Y.; Cheng, Z.; Li, Y.; Silva, D. O.; Schrekker, H. S.; Zhang, K.; Mirkin, C. A., Molecular spherical nucleic acids. *Proc. Natl. Acad. Sci. U.S.A.* **2018,** *115* (17), 4340-4344.
153. Cao, Z.; Huang, C. C.; Tan, W., Nuclease resistance of telomere-like oligonucleotides monitored in live cells by fluorescence anisotropy imaging. *Anal Chem* **2006,** *78* (5), 1478-84.
154. Zhang, D.; Paukstelis, P., Enhancing DNA Crystal Durability through Chemical Crosslinking. *Chem Bio Chem* **2016,** *17* (12), 1163-1170.
155. McNeil, R., Jr.; Paukstelis, P. J., Core-Shell and Layer-by-Layer Assembly of 3D DNA Crystals. *Adv Mater* **2017,** *29* (28), 1701019-1701025.
156. Rivankar, S., An overview of doxorubicin formulations in cancer therapy. *J. Cancer Res. Ther.* **2014,** *10* (4), 853-858.
157. Nitiss, J. L., DNA topoisomerase II and its growing repertoire of biological functions. *Nat. Rev. Cancer.* **2009,** *9* (5), 327-337.
158. van der Zanden, S. Y.; Qiao, X.; Neefjes, J., New insights into the activities and toxicities of the old anticancer drug doxorubicin. *FEBS* **2021,** *288* (21), 6095–6111.

159. Rawat, P. S.; Jaiswal, A.; Khurana, A.; Bhatti, J. S.; Navik, U., Doxorubicin-induced cardiotoxicity: An update on the molecular mechanism and novel therapeutic strategies for effective management. *Biomed. Pharmacother.* **2021**, *139*, 111708.
160. Deepa, K.; Singha, S.; Panda, T., Doxorubicin Nanoconjugates. *J. Nanosci. Nanotechnol.* **2014,** *14* (1), 892–904.
161. Zhao, N.; Woodle, M. C.; Mixson, A. J. Advances in delivery systems for doxorubicin. *J Nanomed Nanotechnol* **2018**, *9* (5), 519.
162. Ghasemiyeh, P.; Mohammadi-Samani, S. Solid lipid nanoparticles and nanostructured lipid carriers as novel drug delivery systems: applications, advantages, and disadvantages. *Res Pharm Sci* **2018**, *13* (4), 288-303.
163. Kumar, V.; Palazzolo, S.; Bayda, S.; Corona, G.; Toffoli, G.; Rizzolio, F., DNA Nanotechnology for Cancer Therapy. *Theranostics* **2016,** *6* (5), 710-725.
164. Mergny, J.-L.; Sen, D., DNA Quadruple Helices in Nanotechnology. *Chem. Rev.* **2019,** *119* (10), 6290-6325.
165. 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.* **2009,** *37* (9), 2810-7.
166. Roberts, C.; Bandaru, R.; Switzer, C., Theoretical and Experimental Study of Isoguanine and Isocytosine: Base Pairing in an Expanded Genetic System. *J. Am. Chem. Soc.* **1997,** *119* (20), 4640-4649.
167. Gratton, S.; Ropp, P. A.; Pohlhaus, P. D.; Luft, C. J.; Madden, V. J.; Napier, M. E.; DeSimone, J. E., The effect of particle design on cellular internalization pathways. *Proc. Natl. Acad. Sci. U.S.A.* **2008,** *105* (33), 11613-11618.
168. Copolovici, D. M.; Langel, K.; Eriste, E.; Langel, U., Cell-Penetrating Peptides: Design, Synthesis, and Applications. *ACS Nano* **2014,** *8* (3), 1972-1994.
169. Khristich, A. N.; Mirkin, S. M., On the wrong DNA track: Molecular mechanisms of repeat-mediated genome instability. *J. Biol. Chem.* **2020,** *295* (13), 4134-4170.
170. Tchurikov, N. A.; Chernov, B. K.; Golova, Y. B.; Nechipurenko, Y. D., Parallel DNA: generation of a duplex between two Drosophila sequences in vitro. *FEBS Lett.* **1989,** *257* (2), 415-418.