

## ABSTRACT

Title of Thesis: DESIGN AND DEVELOPMENT OF THREE-DIMENSIONAL  
DNA CRYSTALS UTILIZING CGAA PARALLEL BASE  
PAIRED MOTIFS

Stephanie Elizabeth Muser, Master of Science, 2013

Thesis directed by: Professor Paul Paukstelis  
Department of Biochemistry

Three-dimensional (3D) DNA crystals hold great potential for various applications such as the development of molecular scaffolds for use in protein structure determination by x-ray crystallography. The programmability and predictability of DNA make it a powerful tool for self-assembly but it is hindered by the linearity of the duplex structure. Predictable noncanonical base pairs and motifs have the potential to connect linear double-helical DNA segments into complex 3D structures. The sequence d(GCGAAAGCT) has been observed to form 3D crystals containing both noncanonical parallel pairs and canonical Watson-Crick pairs. This provided a template structure that we used in expanding the design and development of 3D DNA crystals along with exploring the use of predictable noncanonical motifs. The structures we determined contained all but one or two of the designed secondary structure interactions, depending on pH.

DESIGN AND DEVELOPMENT OF THREE-DIMENSIONAL DNA CRYSTALS  
UTILIZING CGAA PARALLEL BASE PAIRED MOTIFS

By

Stephanie Elizabeth Muser

Thesis 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  
Master of Science  
2013

Advisory Committee:

Professor Paul Paukstelis, Chair  
Professor Dorothy Beckett  
Professor Nicole LaRonde-LeBlanc

© Copyright by

Stephanie Muser

2013

## ACKNOWLEDGEMENTS

Much appreciation and thanks to the Paukstelis lab, and particularly to Dr. Paul Paukstelis for his continued guidance, patience, and advice.

My deepest gratitude goes to my parents, Richard and Debbie Muser, as well, for without their support and love I would not be where I am today.

## TABLE OF CONTENTS

Introduction .....	1
Materials and Methods	
<i>Synthesis and Purification</i> .....	9
<i>Crystallization</i> .....	9
<i>Data Collection of d(GCGAAAGGGCACGTGCCCT) Crystals</i> .....	12
<i>Structure Determination and Refinement of d(GCGAAAGGGCACGTGCCCT)</i> ..	12
Results and Discussion:	
Aim One	
<i>Oligonucleotide d(GCGAAAGGGCACGTGCCCT)</i> .....	13
<i>Canonical anti-parallel duplex region</i> .....	16
<i>Noncanonical parallel base paired region</i> .....	18
<i>Other crystal contacts</i> .....	23
<i>Further Exploration</i> .....	26
Aim Two .....	27
Conclusions .....	31
References .....	33

## INTRODUCTION

Nanotechnology is a rapidly expanding field with increasing applications in various areas such as medicine, energy, and electronics. It is the aim of nanotechnology to control molecules on a nanometer scale. One way to achieve this is through the use of molecular scaffolds. A scaffold presents a way to position molecules for study or to assemble components for future manufacturing. The construction of nanotechnological structures, such as molecular scaffolds, can mainly occur through two ways: ‘top-down’ or ‘bottom-up’ (1). ‘Top-down’ refers to the manipulation of a few atoms or molecules, and ‘Bottom-up’ refers to structure self-assembly through properties of molecular recognition. DNA is a powerful tool for ‘bottom-up’ construction of nanotechnological structures as observed in the formation of two-dimensional (2D) (2) and three-dimensional (3D) (3) crystals.

DNA is predictable and programmable, making it an asset for designing self-assembling 3D lattices. The standard four bases that make up DNA (adenine, guanine, cytosine, and thymine) provide a programming ‘code’ and by simply varying the order, repetition, and length of these four bases, an infinite number of sequences can be constructed. These sequences, when placed in solution with their complement, will come together as a result of adenine preferentially Watson-Crick base-pairing with thymine and guanine preferentially base-pairing with cytosine. These interactions, in a standard solution, will result in a B-form duplex of known rise, pitch, and turn; making DNA a predictable tool for designing 3D crystals.

Duplex DNA is inherently linear however, which presents a predicament when it comes to designing 3D crystal lattices. Constructing complex 3D structures from DNA

requires the use of branching (1). Branching provides the ability to link duplex segments of DNA. Branching of DNA occurs naturally in the form of Holliday junctions during recombination (4, 5), and these Holliday junctions are designed and created with ease in solution (4). However, these types of junctions are too flexible and mobile, making them unsuitable for self-assembling DNA structures. Non-migratory branched junctions have been used in the successful design of DNA structures though (6), and a wide variety of periodic 2D arrays have been designed to self-assemble utilizing the Watson-Crick pairing of sticky ends (7). 3D structures have also been designed and constructed with DNA origami techniques (8). These various designs and structures demonstrate the substantial capability of DNA when it comes to self-assembling nanotechnological structures.

The added ability of DNA to form predictable noncanonical base pairs and motifs offers an alternate route for creating branched structures. While standard Watson-Crick base pairs (observed in duplex DNA) are preferential they are not the only base-pairs possible. Various noncanonical base pairs can readily form, especially in favorable environments or when Watson-Crick pairing is not possible, as the bases are capable of hydrogen bonding through their different faces and sides. Many noncanonical RNA base pairs have been observed, some in conserved three-dimensional motifs (9). Noncanonical DNA motifs have also been observed, such as the G-quadruplex, located in telomeric repeats of chromosomes (10). There is great potential for the rational design of 3D DNA nanostructures utilizing predictable noncanonical motifs for branched junctions. By manipulating the predictability and programmability of DNA along with predictable noncanonical base pairs and motifs, we can design and develop 3D DNA lattices for

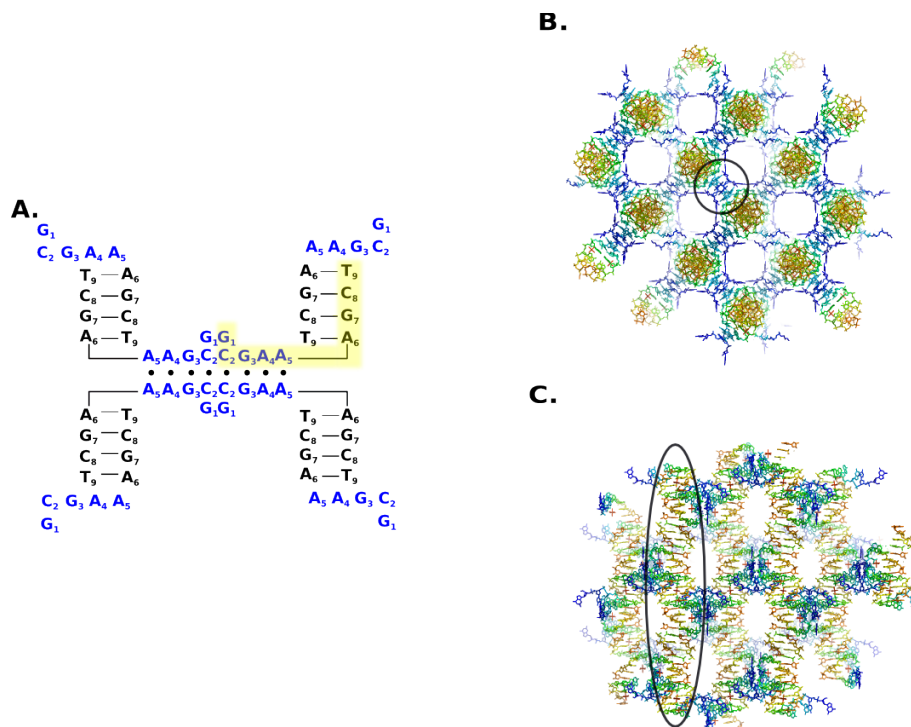
various applications such as optimizing crystallization and structure determination of proteins, and assembling metallic nanoparticles for future nanoelectronic applications.

The first continuously base paired 3D DNA lattice was observed in the crystal structure of the 13-mer d(GGACAGATGGGAG) (*11*). This structure is composed of not only Watson-Crick base pairing, but noncanonical pairings as well. Coaxially stacked anti-parallel Watson-Crick B-form duplexes are held together by noncanonical parallel homopurine base pairs. These parallel homopurine base pairs are formed from the 5'-G<sub>1</sub>G<sub>2</sub>A<sub>3</sub> of one strand parallel base paired with the 5'-G<sub>10</sub>G<sub>11</sub>A<sub>12</sub> of another strand. All of these interactions are formed from the interactions of a single molecule with symmetry related strands. This structure formation resulted in channels of solvent throughout the crystal. Identical 5'-GGA parallel base pairs have been observed in solution as well (*12*), suggesting that these homoparallel base pairs can form predictable noncanonical motifs. These predictable noncanonical motifs can be used to design crystals with larger solvent channels capable of various functions, such as macromolecular sieves (*13*) or scaffolds.

Previous work described the crystal structure of d(G<sub>1</sub>C<sub>2</sub>G<sub>3</sub>A<sub>4</sub>A<sub>5</sub>A<sub>6</sub>G<sub>7</sub>C<sub>8</sub>T<sub>9</sub>), composed of a single molecule in the asymmetric unit, that also contains two different regions of base-pairing, parallel and anti-parallel, resulting in porous crystals with channels of solvent (*14*). (Figure 1) Each strand has a noncanonical parallel base-pairing region, C<sub>2</sub>G<sub>3</sub>A<sub>4</sub>A<sub>5</sub>, followed by a ninety-degree turn and a canonical anti-parallel Watson-Crick base-pairing region, A<sub>6</sub>G<sub>7</sub>C<sub>8</sub>T<sub>9</sub>. The canonical anti-parallel duplexes create B-form helical columns through stacking of the A<sub>6</sub>-T<sub>9</sub> Watson-Crick base pairs. The helical axes of these columns are coincident with the 4<sub>1</sub> screw axis and are connected



by the formation of junctions of the noncanonical parallel base-pairing regions. Figure 1C. These junctions consist of two sets of parallel base-paired duplexes,  $A_5 \bullet A_5$ ,  $A_4 \bullet A_4$ ,  $G_3 \bullet G_3$ , and  $C_2 \bullet C_2$ , stacked  $C_2 \bullet C_2$  to  $C_2 \bullet C_2$ . The 5'-CGA homoparallel motif has been observed in solution previously (15) and is structurally related to the observed 5'-GGA motif. However, instead of the asymmetric G(syn)-G(anti) base-pair at the 5' position seen in the 5'-GGA motif, the 5'-CGA motif contains a symmetric hemiprotonated C•C base pair. Regardless, both of these motifs gain much of their stability from stacking interactions when observed in solution or in crystal formation. Together the homoparallel and anti-parallel base pairing regions of the d(GCGAAAGCT) nonamer form a 3D DNA lattice with an internal network of solvent channels. Figure 1 B and C.



**Figure 1. Secondary structure (A) and lattice views (B and C) of template structure d(GCGAAAGCT) (14)** A. Single strand highlighted in yellow, noncanonical region represented in blue, and canonical anti-parallel Watson-Crick region represented in black. Watson-Crick base pairs and parallel base pairs are represented by black dashes and spheres, respectively. B. Lattice view of the crystal structure looking down the 4-fold screw axis. Single strands are colored from 5' to 3', blue to red. Noncanonical junction region, connecting the canonical columns, is highlighted. Solvent channels that run throughout the crystal are seen. C. Lattice view of the crystal structure looking perpendicular to the 4-fold screw axis. Highlighted are the columns of canonical anti-parallel base pairing regions. PDB 1ixj used for structures

The d(GCGAAAGCT) structure (14) provided inspiration for investigating and expanding the development of self-assembling 3D DNA crystals, because of the porous structure it forms with solvent channels. The noncanonical, stacked parallel base pair motif CGAA formed from C<sub>2</sub> to A<sub>5</sub>, is a model for study as a potential predictable motif, and the columns of stacked Watson-Crick duplex regions also provided a starting point for expansion of the solvent channels.

This research primarily focused on the design and development of predictable noncanonical base pair motifs, not only expanding our knowledge of the versatility of DNA but increasing applications in the design and construction of 3D DNA lattices. The ability to design 3D DNA crystals lattices of varying channel size, especially for use as molecular scaffolds, provides great potential for many applications in nanotechnology, such as the facilitation of protein structure determination and the alignment of nanoparticles for future development of nanowires. The 9-mer structure as well as the first continuously base paired 13-mer provided templates for us to investigate, expand, and rationally design 3D DNA crystal lattices utilizing noncanonical base pairing motifs.

The first aim was to design and crystallize an expanded 3D DNA lattice using predictable CGAA homoparallel base pairs. To do this, we elongated the canonical region of the nonamer structure with the insertion of 10 bases, shown in blue, d(GCGAAAGGGCACGTGCCCT) and successfully crystallized the sequence. No aspects of the original nonamer sequence were altered; maintaining the 5'-CGAA homoparallel motif and extending the Watson-Crick duplex region by approximately one helical turn. The first and last base pairs of the duplex region remain identical to the nonamer structure in order to retain necessary structural interactions. Formation of

columns of Watson-Crick duplexes occurs through the stacking of the A<sub>6</sub>-T<sub>9</sub> base pair, and the backbone transition between the canonical and noncanonical regions is likely stabilized by the presence of a bound hexamine cobalt ion (CoHex), shown in Figure 2. CoHex, bound in the major groove of the nonamer structure at the G<sub>7</sub> position, supports the turn of the backbone at the junction between canonical and noncanonical regions. Magnesium makes water-mediated contacts with the terminal A<sub>6</sub>-T<sub>9</sub> base pairs that stack forming a pseudocontinuous helix. These interactions provide needed support to the structure and were therefore designed in the elongated structure as well.

The second aim was to investigate, design, and crystallize a 3D DNA lattice with expanded noncanonical CGAA homoparallel base pair motifs, using 5'-5' linked sequences. For this expansion, utilizing the strength of Watson-Crick base pairing, we designed sequences with Watson-Crick base insertions between the stacked CGAA homoparallel base pairs.

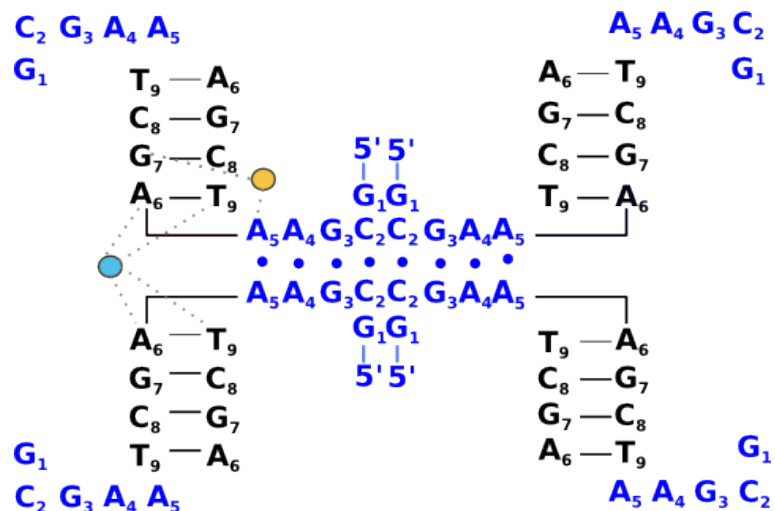


Figure 2. Secondary structure of 3D crystals of template sequence d(GCGAAAGCT). Bound hexamine cobalt, represented by the filled yellow sphere, at the G<sub>7</sub> position stabilizes the orthogonal turn of the backbone from the noncanonical region to the canonical region, shown in blue and black, respectively. Magnesium, represented by the filled blue sphere, makes water-mediated contacts with the terminal A<sub>6</sub>-T<sub>9</sub> base pairs that stack to form helical columns.

## MATERIALS AND METHODS

### *Synthesis and Purification*

Oligonucleotide d(GCGAAAGGGCACGTGCCCT) was synthesized using standard solid-state chemistry (IDT) and purified by 20% 19:1 polyacrylamide 7 M urea gel electrophoresis. The oligonucleotide band was observed and excised by UV shadowing, electro-eluted using Elutrap BT2 membranes by Whatman, and precipitated with 150 mM sodium acetate solution in ethanol. The dried oligonucleotide was dissolved in 5 mM pH 7.0 sodium cacodylate buffer and concentration determined by absorbance at 260 nm.

5'-5' linked oligonucleotides, Table 3, were synthesized using standard solid-state chemistry (including reverse synthesis monomers) on an in-house ABI 8909 Expedite synthesizer and purified using Glen-Pack™ purification cartridges and procedures. Purified oligonucleotides were dialyzed against water and concentrations determined by absorbance at 260 nm.

### *Crystallization*

Crystals of d(GCGAAAGGGCACGTGCCCT) were first grown by mixing 3  $\mu$ L of DNA solution with 1  $\mu$ L of 14 mM  $MgCl_2$ , 56 mM NaCl, 21 mM Cobalt Hexamine, 28 mM pH 6.4 sodium cacodylate, and 17.5% MPD (2-methyl-2,4-propanediol). Drops were incubated at 37°C overnight then cooled down to 24 degrees. Hexagonal shaped crystals grew in 1.5 days. Crystal growth and crystal size were increased through optimization trials, in which the concentrations of  $MgCl_2$ , NaCl, hexamine cobalt, and MPD were varied, along with pH (6.0-7.4), and oligonucleotide concentration (200  $\mu$ M-400  $\mu$ M). Crystal formation was dependent on CoHex and  $Mg^{2+}$ , but independent of pH.

We were able to grow crystals in pHs ranging from pH 3.8 to pH 7.4 using sodium cacodylate and acetate buffers. Crystals used for diffraction studies were grown with 300  $\mu$ M DNA concentration, 3  $\mu$ L DNA mixed with 1  $\mu$ L of 14 mM  $MgCl_2$ , 65 mM NaCl, 21mM hexamine cobalt, 28 mM sodium cacodylate pH 7.0, and 17.5% MPD in a sitting drop and incubated at 37°C overnight before cooling to 24°C. Hexagonal shaped plate-like crystals grew in one day.

A few 5'-5' linked oligonucleotides crystallized under various conditions. Table 1 shows the sequences and the conditions under which they crystallized. All crystals were incubated at 37°C overnight then cooled down to 24 degrees. Crystals grew in ~ 2 days.

**Table 1. 5'-5' linked sequences that crystallized and the varying conditions under which they crystallized. Watson-crick insertion bases shown in red.**

Name	Sequence	# of W-C bases	Parallel Motif	Crystallization Conditions
Seq. 08	3'-AAG <b>CCT AG</b> -5'-5'-CGAA-3'	4	CGAA	2 $\mu$ L 300 $\mu$ M + 2 $\mu$ L 50 mM CaCl <sub>2</sub> , 30 mM sodium cacodylate pH 6, 20 mM KCl, 5% MPD, and 8 mM hexamine cobalt (III) chloride
NCE4-AT-X	3'-AAG <b>CAG CT</b> -5'-5'-CGAA-3'	4	CGAA	2 $\mu$ L 400 $\mu$ M + 2 $\mu$ L 40 mM MgCl <sub>2</sub> , 30 mM sodium cacodylate pH 6-7, 0-5% MPD, and 6-10 mM hexamine cobalt (III) chloride
Seq. 01	3'-AAG <b>CAG TAC T</b> -5'-5'-CGAA-3'	6	CGAA	0.3 $\mu$ L 800 $\mu$ M + 0.3 $\mu$ L 40 mM sodium cacodylate pH 5.5, 40% MPD, and 20 mM hexamine cobalt (III) chloride OR 80 mM NaCl, 20 mM BaCl <sub>2</sub> , 40 mM sodium cacodylate pH 6.0, 45% MPD, 12 mM spermine tetrahydrochloride
Seq. 06	3'-AAG <b>CGG TAC C</b> -5'-5'-CGAA-3'	6	CGAA	2 $\mu$ L 400 $\mu$ M + 2 $\mu$ L 40 mM MgCl <sub>2</sub> , 30 mM sodium cacodylate pH 6.0, and 6 mM hexamine cobalt(III) chloride OR 10% MPD, 40 mM MgCl <sub>2</sub> , 30 mM sodium cacodylate pH 7.0, and 6 mM hexamine cobalt(III) chloride
NCE6G-CG-X	3'-AAG <b>GCG TAC G</b> -5'-5'-GGAA-3'	6	GGAA	0.3 $\mu$ L 400 $\mu$ M + 0.3 $\mu$ L 20 mM MgCl <sub>2</sub> , 50 mM sodium cacodylate pH 7.0, 15% 2-propanol, 1 mM hexamine cobalt (II) chloride, and 1 mM spermine, OR 100 mM KCl, 10 mM MgCl <sub>2</sub> , 50 mM TRIS pH 8.5, and 30% PEG400, OR 200 mM NH <sub>4</sub> Cl, 10 mM CaCl <sub>2</sub> , 50 mM TRIS pH 8.5, and
NCE8G-AT-X	3'-AAG <b>GAG TCG ACT</b> -5'-5'-GGAA-3'	8	GGAA	0.3 $\mu$ L 450 $\mu$ M + 0.3 $\mu$ L 15 mM MgCl <sub>2</sub> , 2 mM BaCl <sub>2</sub> , 50 mM PIPES pH 7.5, 7% 2-propanol, and 0.5 mM Spermine



### *Data Collection of d(GCGAAAGGGCACGTGCCCT) Crystals*

Single crystals were removed from equilibrated drops and flash cooled in the nitrogen cold stream at 100 K or mounted at room temperature with MiTeGen MicroRT loops and capillaries with 45  $\mu$ L of crystallization well solution in the capillary reservoir kept the crystals from drying out. Data was collected on a Brüker Proteum CCD detector with a Brüker Microstart-HF copper rotating-anode X-ray source. Complete data sets were collected on a single crystal in two passes at low and high resolution that were combined, indexed, integrated, and scaled in HKL-2000 (16). Crystals grown at all pHs (3.8 to 7.4) were mounted and tested for diffraction. Those grown at pH 5.5 and below diffracted similarly and indexed to the same space group while those grown at pH 7.0 and above diffracted similarly and indexed to the same space group but had varied  $c$  cell constants. Crystals grown at pHs between 5.5 and 7.0 displayed high mosaicity and could not be indexed. In order to further explore the varying structures of these crystals we exposed them to various conditions. Crystals were soaked in pHs different from which they were grown then transferred to equilibrated drops prior to mounting. MiTeGen mounted crystals were also exposed to varying pHs by carefully exchanging the capillary with a new one containing 0.1 M acetic acid or ammonium hydroxide, ensuring the crystal did not shift. Room temperature collection was performed with the crystals aligned with the  $c^*$  axis in the diffraction plane.

### *Structure Determination and Refinement of d(GCGAAAGGGCACGTGCCCT) Crystals*

The positions of cobalt atoms found by single wavelength anomalous dispersion were used to generate initial phase estimates and Phenix (17) was used for density modification. A-form helices and the nucleobase residues in the noncanonical region

were clearly identified in the experimental electron density maps. Coot (18) was used to build the structure and refinement was performed with REFMAC5 (19). Randomly chosen test set reflections (4.5%) were used for the pH 7.0 data set and the same test set reflections were used for the pH 5.5 structure. Table 2 shows data collection and refinement statistics for both structures. The pH 7.0 structure had a difference in the  $R$  and  $R_{free}$  values (0.07) consistently throughout the refinement process.

## RESULTS AND DISCUSSION

### Aim One

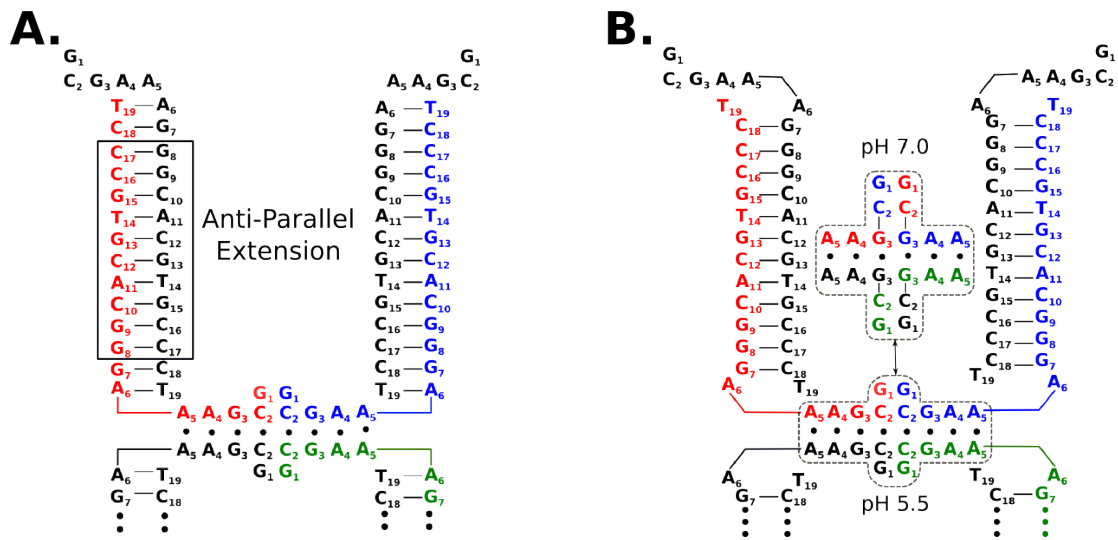
#### *Oligonucleotide d(GCGAAAGGGCACGTGCCCT)*

We solved the crystal structures of d(GCGAAAGGGCACGTGCCCT) grown at pH 5.5 and pH 7.0 by x-ray diffraction. Table 2 shows representative structure determination statistics. At both pHs, crystals diffracted to better than 2.0 Å, were readily phased by single-wavelength anomalous dispersion from bound hexamine cobalt ions, and grew to space group I222 with common  $a$  and  $b$  cell constants. Crystals grown at pH 5.5 or below had a  $c$  cell constant of 85 Å, while crystals grown at pH 7.0 or above had a  $c$  cell constant of 73 Å. These structures contained a single 19-mer in the asymmetric unit forming both the antiparallel Watson-Crick duplex and parallel noncanonical junction with symmetry related strands, as designed. Figure 3 shows the secondary structure of the predicted and solved crystal structures.

**Table 2. Data collection and refinement statistics for pH 7.0 and pH 5.5 structures**

	<b>pH 7.0 Structure</b>	<b>pH 5.5 Structure</b>
<b>Data Collection</b>		
Space group	I222	I222
Unit-cell dimensions (Å, °)	a=31.6 b=44.8 <b>c=73.2</b> , $\alpha=\beta=\gamma=90$	a=31.6 b=44.8 <b>c=85.3</b> , $\alpha=\beta=\gamma=90$
X-ray source	Rotating Copper Anode	Rotating Copper Anode
Wavelength (Å)	1.54	1.54
Resolution (Å)	20.0 - 1.99 (2.0 – 1.93)	20.0 – 1.99 (2.07 – 1.99)
$R_{\text{merge}}$	0.111 (0.323)	0.101 (0.234)
$\langle I/\sigma(I) \rangle$	18.6 (3.1)	17 (4.3)
Completeness (%)	98.2 (86.2)	92.8 (58.7)
Multiplicity*	11.0 (3.5)	13.2 (10.4)
<b>Refinement</b>		
Resolution (Å)	20-1.93	20-2.00
No. reflections	3875	3848
$R_{\text{work}}/R_{\text{free}}$	17.4/24.1	19.4/24.4
R.m.s deviations		
Bond lengths (Å)	0.004	0.005
Bond angles (°)	0.990	1.100
PDBID	4F8G	4F8I

Values in parentheses are for the highest-resolution shell



**Figure 3. Predicted (A) and observed (B) secondary structures of sequence, d(GCGAAAGGGCACGTGCCCT). Identical strands are shown in different colors.**

**A.** Predicted secondary structure that highlights the 10 base insertion in the canonical anti-parallel region **B.** Observed secondary structure with only the noncanonical region varying between pHs. At pH 7.0 and pH 5.5 the noncanonical parallel regions base paired differently. Crystals grown at pH 5.5 formed the same noncanonical regions as the template structure. However, crystals grown at pH 7.0 formed a novel quadruple base pair.

*Canonical anti-parallel duplex region.* The canonical region was designed to be a self-complementary B-form helix with 7 unique base pairs resulting in a 14 base pair duplex. Of the 7 unique pairs expected to form at both pHs, only 6 did, with the A<sub>6</sub>-T<sub>19</sub> base pair missing. (Figure 3) At both pHs, the duplex regions are nearly identical with an RMSD of 0.087 Å for residues 6 through 19.

To our surprise, the Watson-Crick duplex region adopted an A-form conformation instead of the anticipated B-form. While A-form is typically caused by dehydration of the duplex, the duplex is quite hydrated, with many water molecules present, shown in Figure 4. Therefore we speculate that the A-form is adopted due to the presence of hexamine cobalt, and not dehydration. Hexamine cobalt has been previously observed to induce A-form when strings of guanine bases are present (20). Unaware of this phenomena earlier, the sequence was designed with 3 guanine bases in row. Two hexamine cobalt (CoHex) atoms are present in the major groove of the canonical duplex, with CoHex1 making direct hydrogen bonds with the O6 and N7 of G<sub>8</sub> and the O6 of G<sub>9</sub>. CoHex1 also contacts the G<sub>7</sub> phosphate and makes water-mediated contacts with the N7 of the G<sub>7</sub> base. CoHex2 does not make any base-specific contacts but provides charge shielding between the G<sub>7</sub> and G<sub>8</sub> phosphates along with contacting the C<sub>12</sub> phosphate. These interactions of CoHex1 and CoHex2 are believed to be the cause of the collapse of the major groove, resulting in the A-form duplex, and this A-form nature of the canonical region causes the duplex region to be shorter than originally anticipated.

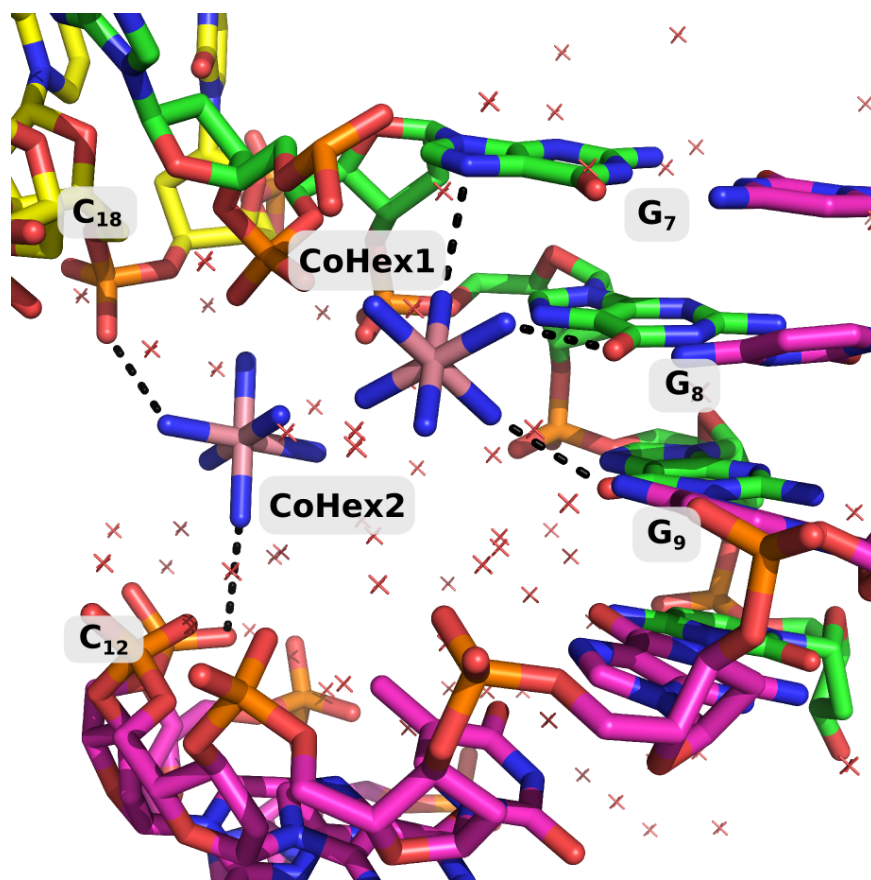
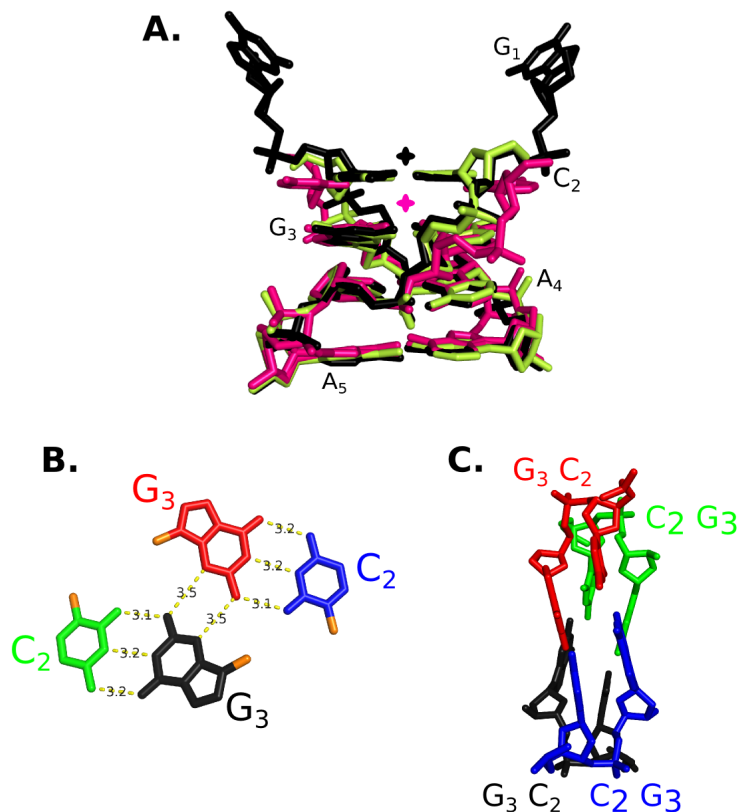


Figure 4. Water and hexamine cobalt (CoHex) binding in the A-DNA duplex major groove of the pH 5.5 structure. Interactions between CoHex ions and DNA are represented by dashed black lines. Three different strands are shown with carbon atoms in three different colors and water molecules are represented by the red crosses. Shown are the base specific contacts CoHex1 makes with the G<sub>7</sub>, G<sub>8</sub>, and G<sub>9</sub> bases, along with the electrostatic interactions CoHex2 makes with C<sub>12</sub> of the duplex strand and C<sub>18</sub> of a symmetry related strand.

*Noncanonical parallel base paired region.* The noncanonical parallel base pairing region was designed to form the same 5'-CGAA homoparallel motif as seen in the nonamer template structure. At both pHs the electron densities of C<sub>2</sub> through A<sub>5</sub> were readily distinguishable, however, density for the G<sub>1</sub> residue was not observed.

At pH 5.5, the anticipated double stranded, stacked CGAA motif was observed, with the A<sub>5</sub>•A<sub>5</sub>, A<sub>4</sub>•A<sub>4</sub>, G<sub>3</sub>•G<sub>3</sub>, and C<sub>2</sub>•C<sub>2</sub> homoparallel base pairs formed. Comparison to the same region as the nonamer structure showed that the regions are highly similar with an RMSD of 0.35 Å for all atoms C<sub>2</sub> through A<sub>5</sub> of both strands. (Figure 5A) At pH 7.0, however, a novel quadruple base pair was observed. Figure 5B and C. The A<sub>5</sub>•A<sub>5</sub>, A<sub>4</sub>•A<sub>4</sub>, and G<sub>3</sub>•G<sub>3</sub> parallel base pairs formed as anticipated, but the hemi-protonated C<sub>2</sub>•C<sub>2</sub> parallel base pair did not. Rather, the C<sub>2</sub> base Watson-crick base pairs with the G<sub>3</sub> of another strand, resulting in bases from four different strands all interacting with each other through hydrogen bonding. Two sets of these quadruple base pairs form and stack on top of each other, completing the pH 7.0 noncanonical junction region. Despite the formation of the quadruple base pairs, the final parallel region structure at pH 7.0 is highly similar to the nonamer template with an RMSD of 1.66 Å. (Figure 5A) Most of this difference comes from the position shifts of the C<sub>2</sub> bases, with an RMSD of 3.23 Å for the two C<sub>2</sub> nucleotides and only 0.76 Å for all other atoms, when compared to the nonamer template structure.



**Figure 5. Noncanonical junction regions.** A. Superposition of the noncanonical base pairing region of the nonamer template (shown in black), pH 5.5 structure (yellow), and pH 7.0 structure (pink). Crossed ovals show intersections of crystallographic two-fold axes. Black ovals are shared by the nonamer template and pH 5.5 structure, and pink ovals are from the pH 7.0 structure. (B and C) Novel quadruple base pair observed in crystal structures at pH 7.0. The four different strands are noted with four different colors. B. Single quadruple base pair shown with ribose bonds shown in orange. Canonical Watson-Crick C-G base pairs are linked through hydrogen bonding between the two guanines at their sugar-edges. C. Two sets of quadruple base pairs stack on top of each other completing the junction region at pH 7.0.

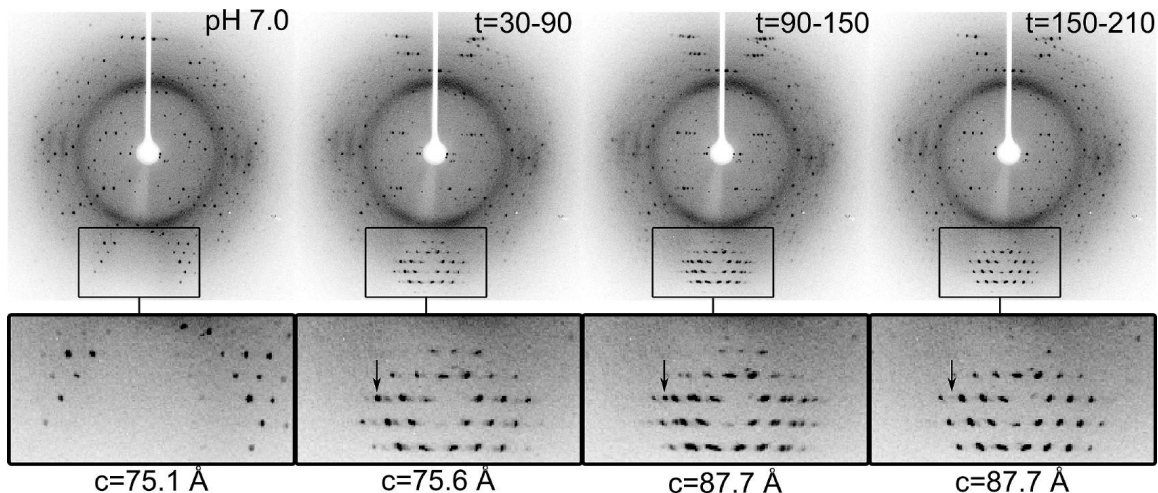


The noncanonical regions, at both pHs, serve as bridges connecting the canonical duplexes, just as in the nonamer template structure. In the nonamer template structure, two sets of noncanonical parallel base pairs stack on top of each other with the crystallographic two-fold axis intersecting between the stacked  $C_2 \bullet C_2$  base pairs. Crystals grown at pH 5.5 have the same such structure. Due to the quadruple base pairs present at pH 7.0, however, the crystallographic two-fold axes intersect between the  $G_3 \bullet G_3$  base pairs, shortening the  $c$  cell constant at pH 7.0 (highlighted in Table 2).

We were able to obtain crystals grown between pH 3.8 and pH 7.4 but unable to determine structures for those crystals grown in pHs ranging from 6.0 to 6.8. Generally, these crystals had a diffraction limit of  $\sim 3.5$  Å but with extreme mosaicity, spot splitting, and/or spot overlaps. This may be due to the presence of a mixed population of  $C_2 \bullet C_2$  and  $C_2 \bullet G_3 \bullet G_3 \bullet C_2$  stacked junctions at these pHs. These localized changes to the noncanonical region throughout the crystal would lead to high mosaicity. In order to investigate this further, we transferred crystals grown at pHs ranging from 6.0 to 6.8 to either low pH, pH 3.8 or 4.2, or high pH, pH 7.4, and soaked them for as little as 5 minutes. Quite remarkably, the crystals diffracted to high resolution ( $\sim 2.0$  Å) after soaking at extreme pHs, and were indexed with  $c$  cell constants consistent with either  $C_2 \bullet C_2$  (pH 3.8 or 4.2) or  $C_2 \bullet G_3 \bullet G_3 \bullet C_2$  (pH 7.4) junction regions. This observation suggests that these junction regions are capable of interconverting within the crystals and maintaining overall lattice packing.

We further explored the capabilities of the junction region to interconvert through some diffraction studies. We mounted crystals, grown at high or low pH, at room temperature and collected diffraction data, then exposed those crystals to the other pH

through vapor exchange (see Materials and Methods). At room temperature, crystals diffracted to lower resolution but were readily indexed with cell constants within 1-2 Å of crystals grown at the same pH. We were able to observe the transitioning of reflections, due to pH change, corresponding to changes in the noncanonical region. Crystals grown at pH 7.0 and induced to interconversion by acetic acid began to show new reflections within 90 s of exposure (capillary exchange) but still had *c* cell constants corresponding to those prior to exposure. Data collected from 90-150 s displayed split reflections, characteristic of multocrystal diffraction while images collected after 150 s displayed reflections corresponding to the larger *c* cell constant of the lower pH structure. Figure 6. Crystals grown at pH 5.5 and induced to interconversion by ammonium hydroxide displayed the corresponding cell shrinkage corresponding to higher pH change. This conversion, however, happened rapidly after capillary exchange and we were not able to follow it by diffraction completely (data not shown).

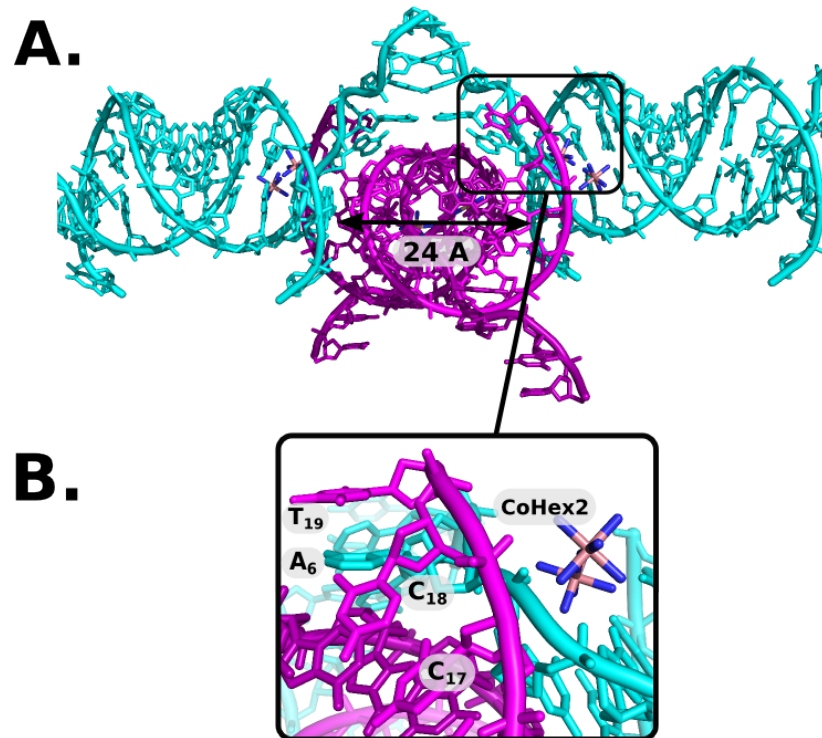


**Figure 6. Structural transition in a single crystal over time, induced by pH. 1 min,  $6^\circ$  oscillation images of a pH 7.0 crystal at room temperature. Magnified boxed regions and indexed  $c$  cell constants using  $5 I/\sigma$  reflections are shown below each full diffraction image. The image on the far left shows diffraction prior to pH perturbation. Subsequent images are of the same crystal in the same orientation following capillary exchange containing a reservoir solution with 0.1 M acetic acid. The time period in seconds following the exchange over which the image was collected is indicated. The readout time for the CCD detector ( $\sim 1$  s) has been ignored. Arrowheads denote the starting position of an example reflection  $(-2,12,-8)$  that appears within 90 s following exchange. This reflection resolves to the right in subsequent frames as the  $c$  cell constant expands throughout the crystal. Similar changes occur throughout the diffraction images demonstrating concerted crystal changes in response to pH perturbations. [This image and caption borrowed from Muser and Paukstelis (21)]**

The rate difference between the conversion of the  $c$  cell constant from high to low pH and low to high pH could most likely be due to the difference in volatility and strength of the acid and base used to induce the pH change. However, intrinsic properties of the junction region may have an influence on the conversion rate as well. Nevertheless, these results demonstrate three significant observations: (i) the noncanonical junction regions of these crystals are capable of interconversion within a single crystal due to pH change, (ii) these interconversions can take place in a relatively short period of time, and (iii) these changes are precise and localized, and do not interrupting lattice packing or diffraction. This observed adaptability of the junction region is a potential new tool for use as DNA crystals as adaptive biomaterials. Furthermore, this feature may also be useful as a method for “locking” guest molecules into a lattice; the induced pH change could shrink the lattice channels after absorption.

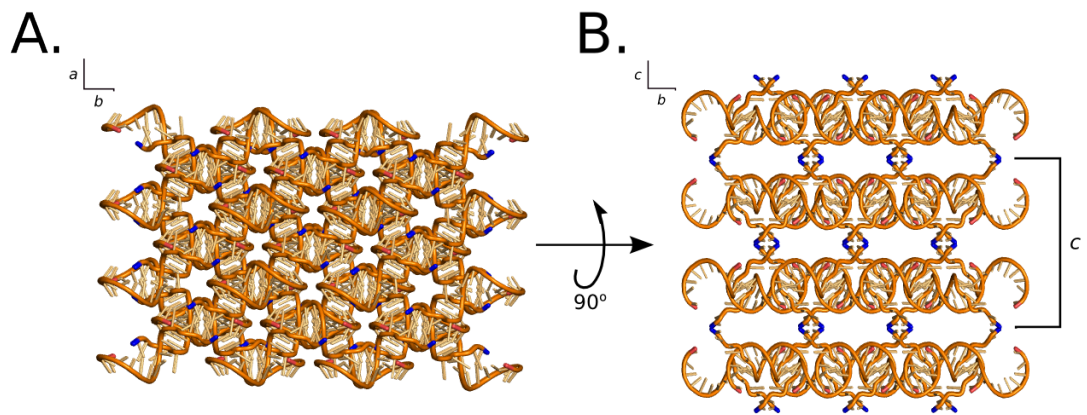
*Other crystal contacts.* While almost all designed base pairing interactions were observed, stacking of the canonical duplexes through the A-T base pair did not occur, as originally predicted from the nonamer template structure. In the nonamer template structure, columns of B-form helices form from the stacking of two A<sub>6</sub>-T<sub>9</sub> base pairs. The equivalent base pair, A<sub>6</sub>-T<sub>19</sub>, did not occur in our designed crystals, however, at either pH. With no A<sub>6</sub>-T<sub>19</sub> base pair formed, the A-form duplexes are unable to stack. The duplexes are still coaxial, however, but the first base pairs (G<sub>7</sub>-C<sub>18</sub>) are separated by ~24 Å. This separation allows a symmetry related duplex, referred to as the central duplex, to fit between these separated coaxial duplexes, referred to as flanking duplexes. (Figure 7A) On each side of the central duplex, the A<sub>6</sub>, from the central duplex, stacks with the T<sub>19</sub> from the flanking duplex, essentially clamping the central duplex in place.

The close contacts of the backbones are stabilized through charge shielding provided by the hexamine cobalt atoms. (Figure 7B)



**Figure 7. Stacking and electrostatic interactions stabilize interwoven A-DNA helices. A. Coaxial canonical anti-parallel helices that are joined through noncanonical regions (cyan) do not stack as predicted but are separated by 24 Å and the insertion of a symmetry-related duplex (violet). B. The tightly packed duplexes are stabilized by stacking of A<sub>6</sub> of the flanking duplexes with T<sub>19</sub> of the central duplex. CoHex2 in the major groove of the flanking duplex provides stabilization through electrostatic interactions.**

These interactions result in the formation of sheets of helices as opposed to the originally designed helical columns. (Figure 8) The variation of the duplex region is also the likely cause of the different space group these structures crystallize to. The nonamer template structure crystallizes to space group  $I4_122$  with helical columns of coaxially stacked duplexes that are concurrent with the 4-fold crystallographic axis. Crystals of our structure, grown at either pH, crystallize to  $I222$ , a lower symmetry space group.



**Figure 8. Lattice assembly of interwoven helical sheets. Cartoon representation of the pH 5.5 lattice structure. 5' ends capped in blue and 3' ends capped in red. A. Helical axes run parallel to the  $a, b$  face diagonals B. Layers of helical sheets are joined through noncanonical parallel regions that extend out of the top and bottom of the sheets. Two sheets make up a unit cell with distances varying by pH. Crystals grown at pH 7.0 have a shorter  $c$  cell constant due to the quadruple base pair formed in the noncanonical region.**

*Further Exploration.* The d(GCGAAAGGGCACGTGCCCT) sequence design and resulting structure has taught us a few things. The 5'-CGAA motif has proved to be quite predictable, when used at low pH, and is adaptable to pH changes. While the 10-base insertion sequence crystal structures resulted in A-form canonical duplexes as opposed to the B-form duplexes the design was predicated upon, the adopted A-form of the duplexes is believed to primarily be due to the presence of hexamine cobalt molecules. As discussed earlier, hexamine cobalt molecules make contact with the G<sub>7</sub>, G<sub>8</sub>, and G<sub>9</sub> bases along with the G<sub>7</sub> phosphate and C<sub>12</sub> phosphate, helping to bridge and collapse the major groove. As the crystals require hexamine cobalt in order to form, we modified the sequence via single base mutations to alter the interactions with the hexamine cobalt atoms and therefore the overall helical form.

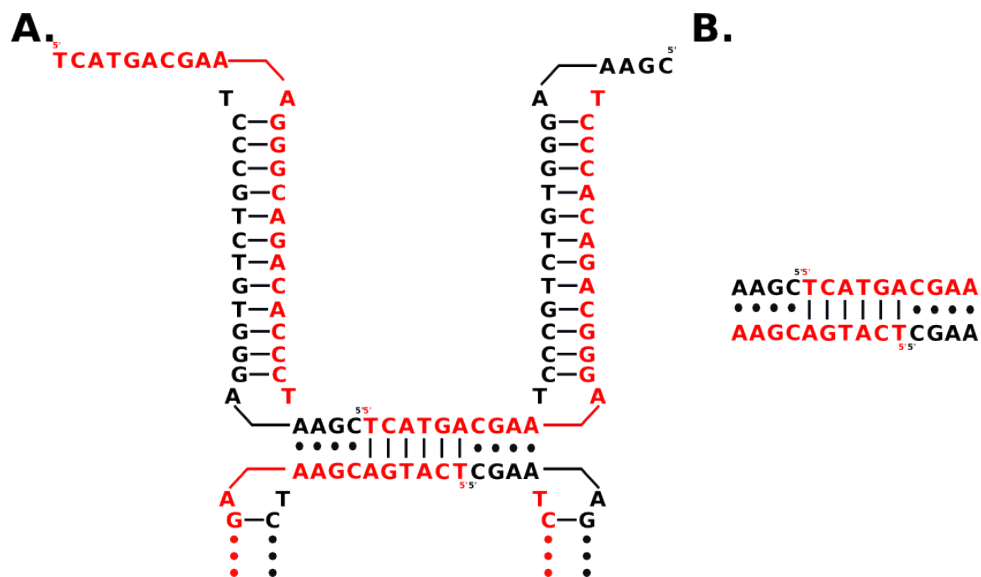
One of the first modifications was the alteration of the A<sub>6</sub>-T<sub>19</sub> base pair to a C<sub>6</sub>-G<sub>19</sub> base pair. The C-G base pair, containing three-hydrogen bonds rather than two as with A-T base pairs, may form more readily. The formation of this base pair would allow for stacking between C-G base pairs, connecting the duplexes. This modification, however, did not result in any crystallization. A second alteration was performed modifying one base at a time of the G<sub>7</sub>G<sub>8</sub>G<sub>9</sub> motif. This modification would disrupt the hexamine cobalt binding sites, possibly allowing the desired B-form helices. Two such sequence changes that would disrupt the hexamine cobalt binding sites are modifications of the G<sub>8</sub>-C<sub>17</sub> base pair to C<sub>8</sub>-G<sub>17</sub> and of the G<sub>9</sub>-C<sub>16</sub> base pair to A<sub>9</sub>-T<sub>16</sub>. Sequence modification of G<sub>8</sub>-C<sub>17</sub> to C<sub>8</sub>-G<sub>17</sub> did not result in crystallization, however, modification of G<sub>9</sub>-C<sub>16</sub> to A<sub>9</sub>-T<sub>16</sub> crystallized with 80 mM sodium chloride, 40 mM sodium cacodylate buffer pH 5.5 to 6.4, 10-20% 2-methyl-2,4-pentanediol, and 12 mM

spermine•tetrahydrochloride. These crystals only diffracted to  $\sim 9$  Å though and most likely have a very large unit cell as the reflections are seen very close to one another. To date, we have been unable to solve this crystal structure.

### **Aim Two**

The second aim of ours was to investigate, design, and crystallize a 3D DNA lattice with expanded noncanonical CGAA homoparallel base-pair motifs. The goal of this aim was to determine and identify sequences and motifs that can be used for elongating the noncanonical region of the template structure and the solved structure described above. We did this through the insertion of Watson-Crick anti-parallel base pairs between the stacked parallel base-pairing regions. Two approaches are: (i) the use of two strands designed to interact, Figure 9A, or (ii) the use of DNA synthesis techniques to create sequences with 5'-5' linkages between the Watson-Crick base-pairing region and the parallel base-pairing region, Figure 9B.





**Figure 9. Predicted secondary structure of extension of the noncanonical junction region. A. Two strands are used, the assembly strand shown in red and the spacer strand shown in black. 5' ends are marked. The assembly strands are self-complementary at the 5' end for the Watson-Crick insertion. The spacer strand is complementary to the assembly strand along the canonical region only. 6 canonical anti-parallel base pairs are inserted between parallel base pairs. B. Construct of the 6 base pair extended noncanonical region only.**

The use of a 5'-5' linkage presented a major advantage as it only requires one strand. We readily synthesized these sequences using an ABS 394 or Expedite 8909 DNA synthesizer with commonly available reagents and phosphoramidites for forward and reverse synthesis with little to no trouble. The 5'-5' linked sequences also allowed us to investigate and optimize the Watson-Crick insertion sequence and noncanonical region alone. By synthesizing only the extended noncanonical regions, we were able to examine the effects of various base modifications within the extension, starting with the terminal

base pairs of the Watson-Crick insertion. These pairs are responsible for the stacking interactions between the Watson-Crick insertion and parallel base pairs, and optimization of these pairs should facilitate crystal formation.

Table 3 shows the sequences we synthesized and tested for crystallization. The crystal structure of the first continuously base-paired 13-mer DNA lattice, highlighted earlier, exhibited a homopurine GGA parallel base-pair motif similar to the CGA parallel base-pair motif used in the template structure described previously. We therefore tested various sequence combinations using a GGAA parallel base-pair motif in place of the CGAA motif. Watson-crick insertions varied from 4 base pairs to 10 base pairs, with all insertions even-numbered. The insertions had to be kept at an even number to continue the use of only one DNA strand. We also explored variations of the terminal Watson-crick base pairs in the insertions. Some sequences tested included the canonical region, specifically an elongation of AGCT, as seen in the template structure.

**Table 3. 5'-5' linkage sequences synthesized and tested for crystallization. Watson-crick regions are highlighted in red.**

Name	Sequence	# of W-C bases	Parallel Motif	Crystallized?	Diffraction?
Seq. 08	3'-AAG <b>CCT AG</b> -5'-5'-CGA A-3'	4	CGAA	✓	
NCE4-AT-X	3'-AAG <b>CAG CT</b> -5'-5'-CGAA-3'	4	CGAA	✓	~ 10 Å, high mosaicity
NCE4G-CG-X	3'-AAG <b>GCT AG</b> -5'-5'-GGA A-3'	4	GGAA		
NCE4G-AT-X	3'-AAG <b>GAG CT</b> -5'-5'-GGAA-3'	4	GGAA		
Seq. 01	3'-AAG <b>CAG TAC T</b> -5'-5'-CGA A-3'	6	CGAA	✓	
Seq. 06	3'-AAG <b>CGG TAC C</b> -5'-5'-CGA A-3'	6	CGAA	✓	
NCE6G-AT-X	3'-AAG <b>GAG TAC T</b> -5'-5'-GGA A-3'	6	GGAA		
NCE6G-GC-X	3'-AAG <b>GGG TAC C</b> -5'-5'-GGA A-3'	6	GGAA		
NCE6G-CG-X	3'-AAG <b>GCG TAC G</b> -5'-5'-GGA A-3'	6	GGAA	✓	
NCE6G-TA-X	3'-AAG <b>GTA GCT A</b> -5'-5'-GGA A-3'	6	GGAA		
Seq. 04	3'-TCG AAA <b>GCG GTA CC</b> -5'-5'-CGA AAG CT-3'	6	CGAA		
Seq. 03	3'-TCG AAA <b>GCA GTA CT</b> -5'-5'-CGA AAG CT-3'	6	CGAA		
Seq. 05	3'-TCG AAA <b>GCC CTA GG</b> -5'-5'-CGA AAG CT-3'	6	CGAA		
Seq. 11	3'-AAG <b>CAG TCG ACT</b> -5'-5'-CGA A-3'	8	CGAA		
Seq. 12	3'-AAG <b>CCG TAT ACG</b> -5'-5'-CGA A-3'	8	CGAA		
NCE8G-AT-X	3'-AAG <b>GAG TCG ACT</b> -5'-5'-GGA A-3'	8	GGAA	✓	~ 2 Å, multi-crystal
NCE10G-AT-X	3'-AAG <b>GAG TCT AGA CT</b> -5'-5'-GGA A-3'	10	GGAA		

We readily observed crystallization of 6 different sequences (Table 1) but only crystals of two sequences diffracted, NCE4-AT-X and NCE8G-AT-X. Unfortunately, these crystals did not give us sufficient data and we were unable to roughly determine space group. We did observe, however, that sequences with insertions of 4 or 6 Watson-Crick base pairs crystallization occurred predominantly with the CGAA parallel motif. This trend though is not conclusive as there may be other sequences we have yet to try. With the GGAA parallel motif we observed crystallization of the 8 base Watson-Crick insertion while the CGAA motif did not. Both of these sequences that crystallized and diffracted have A-T base pairs at the terminal end of the Watson-Crick pairing, stacking on the parallel base pairing region. Terminal end A-T base pairs may be best for stacking interactions facilitating crystallization, though this would need to be explored further. These results, though relatively limited, support continued trials with the CGAA and GGAA parallel motifs, and further investigation elongating the noncanonical region with the insertion of Watson-Crick base pairing.

## **CONCLUSION**

Despite the surprises and pitfalls, our research designing and developing three-dimensional DNA crystals utilizing parallel base-paired motifs was successful. Our continued use of the CGAA parallel motif has furthered our understanding of its predictability and is a tool for the continued construction and development of DNA nanotechnology. Some of the surprises and pitfalls turned out to be quite insightful as well. The two CGAA parallel motifs observed at pH 7 and pH 5.5, and their ability to interconvert, provide great potential for DNA crystals as adaptive materials. While the

elongation of the noncanonical base pairing region has yet to lead to conclusive results, there is still great potential in this area of research and with continued effort we will be able to not only successfully design 3D DNA crystals using CGAA parallel motifs but develop tools for facilitating the continued construction of various 3D DNA crystals for use in DNA nanotechnology and applications.

## References

1. Seeman, N. C. (2003) DNA in a material world, *Nature* 421, 427-431.
2. Winfree, E., Liu, F., Wenzler, L. A., and Seeman, N. C. (1998) Design and self-assembly of two-dimensional DNA crystals, *Nature* 394, 539-544.
3. Zheng, J. P., Birktoft, J. J., Chen, Y., Wang, T., Sha, R. J., Constantinou, P. E., Ginell, S. L., Mao, C. D., and Seeman, N. C. (2009) From molecular to macroscopic via the rational design of a self-assembled 3D DNA crystal, *Nature* 461, 74-77.
4. Seeman, N. C., and Kallenbach, N. R. (1994) DNA Branched Junctions, *Annual Review of Biophysics and Biomolecular Structure* 23, 53-86.
5. Holliday, R. (1964) A mechanism for gene conversion in fungi, *Genetics Research* 5, 282-304.
6. Seeman, N. C. (1998) DNA NANOTECHNOLOGY: Novel DNA Constructions, *Annual Review of Biophysics and Biomolecular Structure* 27, 225-248.
7. Lin, C., Liu, Y., Rinker, S., and Yan, H. (2006) DNA Tile Based Self-Assembly: Building Complex Nanoarchitectures, *ChemPhysChem* 7, 1641-1647.
8. Han, D., Pal, S., Nangreave, J., Deng, Z., Liu, Y., and Yan, H. (2011) DNA Origami with Complex Curvatures in Three-Dimensional Space, *Science* 332, 342-346.
9. Leontis, N. B., Stombaugh, J., and Westhof, E. (2002) The non - Watson-Crick base pairs and their associated isostericity matrices, *Nucleic Acids Research* 30, 3497-3531.
10. Sundquist, W. I., and Klug, A. (1989) Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops, *Nature* 342, 825-829.
11. Paukstelis, P. J., Nowakowski, J., Birktoft, J. J., and Seeman, N. C. (2005) Crystal structure of a continuous three-dimensional DNA lattice *Chemistry & Biology* 12, 497-497.
12. Kettani, A., Bouaziz, S., Skripkin, E., Majumdar, A., Wang, W., Jones, R. A., and Patel, D. J. (1999) Interlocked mismatch-aligned arrowhead DNA motifs, *Structure* 7, 803-804.
13. Paukstelis, P. J. (2006) Three-dimensional DNA crystals as molecular sieves, *Journal of the American Chemical Society* 128, 6794-6795.

14. Sunami, T., Kondo, J., Kobuna, T., Hirao, I., Watanabe, K., Miura, K., and Takenaka, A. (2002) 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* 30, 5253-5260.
15. Robinson, H., Vandermaarel, G. A., Vanboom, J. H., and Wang, A. H. J. (1992) Unusual DNA conformation at low pH revealed by NMR - Parallel-stranded DNA duplex with homo base-pairs, *Biochemistry* 31, 10510-10517.
16. Otwinowski, Z., and Minor, W. (1997) Processing of X-ray diffraction data collected in oscillation mode, In *Methods in Enzymology* (Charles W. Carter, Jr., Ed.), pp 307-326, Academic Press.
17. Adams, P. D., Afonine, P. V., Bunkoczi, G., Chen, V. B., Echols, N., Headd, J. J., Hung, L.-W., Jain, S., Kapral, G. J., Grosse Kunstleve, R. W., McCoy, A. J., Moriarty, N. W., Oeffner, R. D., Read, R. J., Richardson, D. C., Richardson, J. S., Terwilliger, T. C., and Zwart, P. H. (2011) The Phenix software for automated determination of macromolecular structures, *Methods* 55, 94-106.
18. Emsley, P., Lohkamp, B., Scott, W. G., and Cowtan, K. (2010) Features and development of Coot, *Acta Crystallographica Section D* 66, 486-501.
19. Murshdov, G. N. V., Alexei A. and Dodson, Eleanor J. (1997) Refinement of Macromolecular Structures by the Maximum-Likelihood Method, *Acta Crystallographica Section D* 53, 240-255.
20. Robinson, H., and Wang, A. H.-J. (1996) Neomycin, Spermine and Hexaammincobalt(III) Share Common Structural Motifs in Converting B- to A-DNA, *Nucleic Acids Research* 24, 676-682.
21. Muser, S. E., and Paukstelis, P. J. (2012) Three-Dimensional DNA Crystals with pH-Responsive Noncanonical Junctions, *Journal of the American Chemical Society* 134, 12557-12564.