Spring 2020

Biophysical and Structural Characterization of Non-Canonical DNA Structures

Linda Yingqi Lin , '20

Biophysical and Structural Characterization of Non-Canonical DNA Structures

Presented as a Senior Honors Thesis in Biochemistry

Linda Yingqi Lin

April 13, 2020

Swarthmore College

Advisor: Liliya A. Yatsunyk

Acknowledgements

The work presented in my thesis would not have been possible without all the people who helped me along the way. First, I would like to thank my research advisor, Liliya Yatsunyk, for helping me develop my skills as a scientist and researcher for the past three and a half years. I am also extremely grateful to all the lab members who welcomed me into the lab and introduced me to the many different techniques we use: Jessica Chen '17, Irene Xiang '18, Barrett Powell '18, Sayed Malawi '18, Deondre Jordan '19, and Allan Gao '19. Additionally, I would like to thank the other amazing people who have been a part of the lab during my time here: Samantha Nyovanie, Yanti Manurung '20, Ariana Yett '21, Hyun Kyung Lee '21, Dana Beseiso '21, Joanne Miao '22, and Sawyer McCarthy. All these lab members really supported me along the way as both fellow researchers and friends.

I am also grateful to all the collaborators I had the chance to work with and the knowledge and ideas they shared with me: William Dean and Brad Chaires (University of Louisville) for the AUC part of the T1 project; Aurore Guédin, Samir Armane, Laurent Lacroix, Jean-Louis Mergny, and Stéphane Thore (University of Bordeaux) for the 19wt project; David Lee and Louis Ghanem (University of Pennsylvania) for an RNA GQ project; and Dawn Carone (Department of Biology) for a cell staining project. Furthermore, I would like to thank Adam Olia (University of Pennsylvania), Daniela Fera (Department of Chemistry & Biochemistry), Jack Nicoludis '12 (University of California San Francisco), Jack Rubien '20 (Department of Biology), Christina Rabeler (Department of Biology), and the Northeastern Collaborative Access Team at the Advanced Photon Source at Argonne National Laboratory for their generous advice and guidance.

Finally, my research work would not have been possible without the numerous professors and mentors who inspired and sustained my interest in science. I would especially like to thank all the faculty and staff in the Departments of Chemistry & Biochemistry, Biology, and Computer Science for helping me grow as a student and scientist throughout my undergraduate career. A very special thank you also goes to the wonderful people who first believed in me as a scientist and inspired my love for scientific research: Hongwei Xu, Flaminia Marucci, Yulong Kuang, and Tianning Diao (New York University), along with Nadia Makar and Susan Fahrenholtz (American Chemical Society Project SEED). I would not be where I am today without them. Last but certainly not least, thank you to my friends and family for being by my side throughout this journey.

Abstract

Non-canonical DNA structures known as G-quadruplexes (GQs) and i-motifs can form from G-rich and C-rich sequences within the genome, respectively, which are prominent within telomeres and oncogene promoters. These non-canonical quadruplexes likely play roles in regulating gene expression and in DNA replication and repair. However, much remains to be understood about their structural features and diversity.

In this work, we investigate the interaction of a GQ-forming sequence with a water-soluble porphyrin, *N*-methyl mesoporphyrin IX (NMM). Biophysical studies revealed an impressively tight, thermodynamically favorable binding interaction. We then solved the GQ-NMM crystal structure at 2.39 Å, which showed that the DNA forms a dimer of parallel GQs with NMM bound at both ends via end-stacking interactions.

Furthermore, we demonstrate that an unusually large, four-tetrad GQ adopts the same structure in solution as it does in crystalline form, thereby validating the crystal structure. We characterize its loop mutants to show that loop interactions fine tune GQ stability, but do not affect GQ folding. Finally, we present ongoing work towards solving the crystal structures of a different GQ-ligand complex and of a monomolecular i-motif.

The work in this thesis advances our understanding of quadruplex structural diversity and ligand binding interactions by contributing to the limited number of solved GQ-ligand crystal structures, as well as by characterizing an atypically large GQ. Non-canonical quadruplexes likely function *in vivo* as regulatory elements. Accordingly, our improved understanding of their structural features informs the design and *in silico* screening of novel drugs, which can selectively recognize these features and subsequently modulate quadruplex stability for therapeutic purposes, particularly against cancer.

Table of contents

List of figures

List of tables

Chapter 1: Introduction to non-canonical DNA structures

Biological significance of non-canonical DNA structures

Genomic context

The most common DNA conformation *in vivo* is the double helix. This double-stranded secondary structure forms from canonical Watson-Crick base pairing between complementary, antiparallel strands of nucleotides. In addition to the hydrogen bonds within base pairs, aromatic stacking interactions between adjacent bases also contribute to the stability of the DNA double helix.

However, a number of biological processes require DNA to exist in the single-stranded state, as depicted in **Figure 1**. During DNA replication and transcription, the two strands of the double helix must be separated to allow the requisite proteins to bind to the appropriate strand and synthesize new DNA or RNA in a complementary fashion. Accordingly, these processes – vital for cellular function and division – necessarily involve single-stranded DNA. At the same time, due to the end-replication problem, single-stranded telomeric overhangs are always present at the ends of eukaryotic chromosomes.

While in its less stable, single-stranded state, DNA can adopt non-canonical conformations.[1] These include quadruplexes, which are tetra-stranded secondary structures that form from non-canonical base pairing interactions and are much bulkier than the canonical double helix. Similarly to the double helix, quadruplexes are stabilized by hydrogen bonding between bases and aromatic base stacking, but can be further stabilized by interactions with specific cations or in particular pH environments. Since any cellular

process that involves DNA in its single-stranded form provides an opportunity for quadruplex formation, quadruplexes may serve a variety of biological functions, particularly in regulation (**Figure 1**).[2]

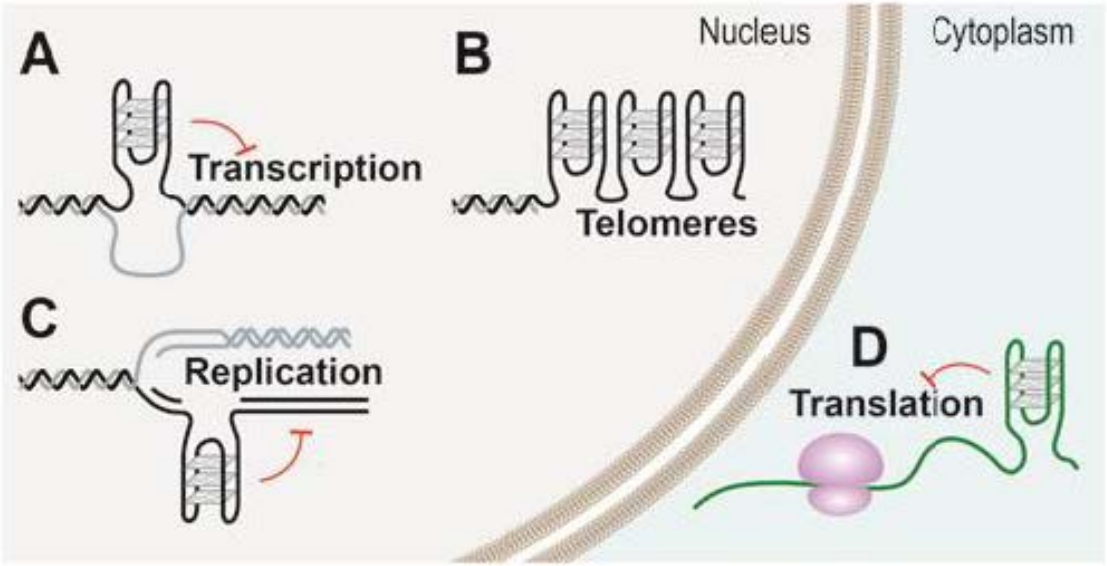


Figure 1. Sites of potential quadruplex formation. Quadruplex structures have the potential to form in DNA (black) that is single-stranded, even if only transiently so. Accordingly, quadruplexes may form during transcription, replication, and translation (in RNA, green), as well as in telomeres. Figure adapted from [2].

The two types of non-canonical quadruplex structures that are the focus of this thesis are G-quadruplexes (GQ), which form from guanine-rich DNA, and i-motifs, which form from cytosine-rich DNA (**Figure 2**). Notably, these structures can form in complementary sequences, so although GQs have been more extensively studied, the identification of a GQ-forming sequence implies that a potential i-motif sequence may exist on the complementary DNA strand.

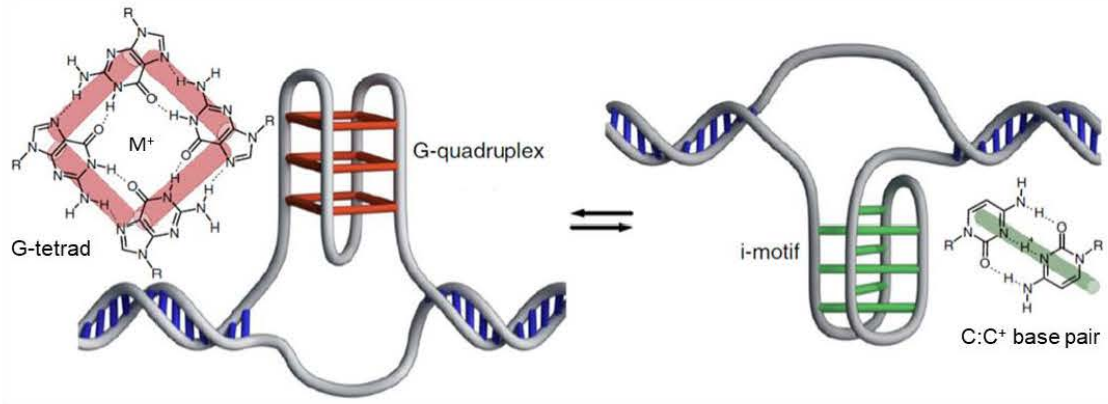


Figure 2. G-quadruplex and i-motif structures. They are shown alongside their structural units, the G-tetrad and the C:C$^+$ base pair, respectively. Figure adapted from [3].

Prevalence of potential quadruplex-forming sequences in genomes

High-throughput sequencing studies have identified over 700,000 sequences in the human genome that can form GQs.[4] These sequences are particularly enriched within regions that are involved in regulation and implicated in disease, such as telomeres and oncogene promoters.[4] Notably, 50% of human promoters and 90% of human DNA replication origins contain at least one potential G-quadruplex-forming motif.[2] Meanwhile, the human telomeric repeat, $(TTAGGG)_n$, is G-rich and contains a repeating G-tract that allows for GQ formation.

Furthermore, certain sequences with quadruplex-forming potential are highly conserved among species and localize to functional genomic regions,[5] which indicates an evolutionary pressure to conserve GQ structures. Therefore, not only are GQ sequences prominent within important regulatory regions of the genome, but their conservation also suggests that they have functional significance.

Potential roles of quadruplexes

As depicted in **Figure 1**, quadruplex structures may inhibit transcription and DNA replication,[2] which are both vital to cellular function. More specifically, they may hinder RNA and DNA polymerase function by preventing its binding to or procession along a DNA strand,[2] likely due to their size and stability. These inhibitory roles can be both beneficial and deleterious. Given that sequences with quadruplex-forming potential are frequently found in oncogene promoters,[6] quadruplex-based transcriptional inhibition can be a viable method to downregulate oncogene expression in cancers.[7] However, the inhibitory role of quadruplexes in replication can cause replication fork stalling, which can lead to genomic instability and potentially cancer.[8] Consequently, quadruplex-targeting anticancer therapies must be selective among different types of quadruplexes to ensure that deleterious off-target effects will not come along with any intended therapeutic effects.

Also shown in **Figure 1** are telomeric quadruplexes. Telomerase, which is upregulated in cancer and catalyzes telomere extension to enable cancer cell immortality, binds single-stranded telomeric overhangs as its substrate. The formation of bulky GQs at the telomere may therefore inhibit telomerase binding and activity.[9] Additionally, telomeric GQs have been shown to serve as a rudimentary telomere cap when natural protein capping is compromised.[10] Notably, sequences from both oncogene promoters[11–14] and telomeres[11] – often complementary to those that form GQs – have been shown to form i-motifs.

The presence of GQ structures *in vivo* was first demonstrated with antibodies against telomeric GQ DNA that bound to the telomeres of *Stylonychia lemnae* macronuclei.[15] Recently, a GQ-binding protein was discovered in *Tetrahymena*

thermophila that binds parallel telomeric GQs,[16] thereby providing compelling evidence for the biological relevance of GQ structures *in vivo*. Similarly, two i-motif binding heterogeneous nuclear ribonucleoproteins were recently identified, suggesting that i-motifs may serve as molecular switches that regulate gene expression.[12–14] Furthermore, both GQs[17] and i-motifs[3] have been visualized in human cells using selective antibodies. These examples are only a few from a growing body of evidence for the relevance of quadruplex DNA structures *in vivo*.

Both GQ and i-motif structures have been shown to be important cis-acting regulatory elements.[18] In humans, genomic defects involving GQ-forming sequences are implicated in diseases such as cancer, fragile X syndrome, Bloom syndrome, Werner syndrome, and Fanconi anemia J.[19] Studies of the connections between GQs and human disease have firmly established GQ DNA as a viable therapeutic target for cancer as well as a variety of other human diseases.[8] However, further understanding of the molecular details of these quadruplex structures and their regulation is necessary to develop quadruplex-targeting therapeutics.

Targeting an i-motif in the MAPK signaling pathway

A gene of great interest in cancer research is the human Ras (HRAS) oncogene, which encodes a GTPase that serves as the first secondary messenger within the mitogen-activated protein kinase (MAPK) signaling pathway.[20] This pathway results in cell proliferation and is constitutively activated in many cancers.[20] The HRAS promoter contains G-rich motifs that can fold into GQs,[21] while the complementary C-rich sequence (iHRAS) forms an i-motif.[14] Although HRAS has been a prominent therapeutic target for

several decades, it has proven difficult to develop a small molecule inhibitor of the HRAS protein because it lacks deep pockets for inhibitor binding.[22] Consequently, targeting HRAS expression itself – such as via understanding the iHRAS i-motif – has become an attractive alternate strategy.[23]

Quadruplex DNA structures

Guanine quadruplexes

GQs can form in guanine-rich regions of the genome. In particular, four guanines can arrange into a square planar conformation, connected by cyclic Hoogsteen hydrogen bonding (**Figure 3**), to form a G-quartet. Also referred to as a G-tetrad, this is the structural unit of the GQ. Four nearby guanine tracts allow for multiple G-quartets to form and assemble through π-π stacking interactions, resulting in a 3D GQ structure (**Figure 2**).[24] GQs are further stabilized by a monovalent cation, notably K^+, that binds in the center between each pair of G-quartets.[25] GQs readily form *in vitro*, with their thermodynamic stability in physiological buffers often rivaling that of duplex DNA.[26]

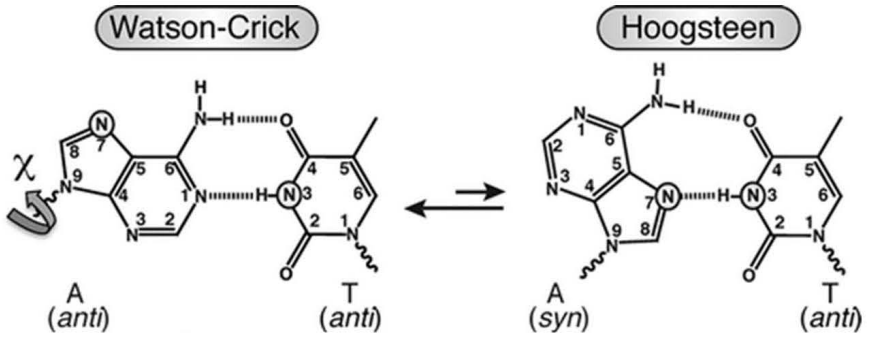


Figure 3. Watson-Crick vs. Hoogsteen base pair. Chemical structure of this classic vs. alternative base pairing geometry. Note that these involve different hydrogen bonding patterns, and that Hoogsteen hydrogen bonding can occur with other nucleotide pairings too. Adapted from [27].

The number of G-tetrads within a biological GQ typically varies between two and four, with three being the most common and four being rare. The guanines can be arranged in either *syn* or *anti* conformations with respect to the deoxyribose sugar (**Figure 4**). These glycosidic conformations, along with the relative directionality of the four constituent guanine tracts, determine the overall topology that a GQ adopts (**Figure 4**). In particular, parallel GQs have all four G-tracts going in the same direction, whereas antiparallel GQs consist of two G-tracts going in one direction and the other two G-tracts going in the other direction in a [2+2] manner. Meanwhile, hybrid GQs have a [3+1] combination of G-tract directionalities.

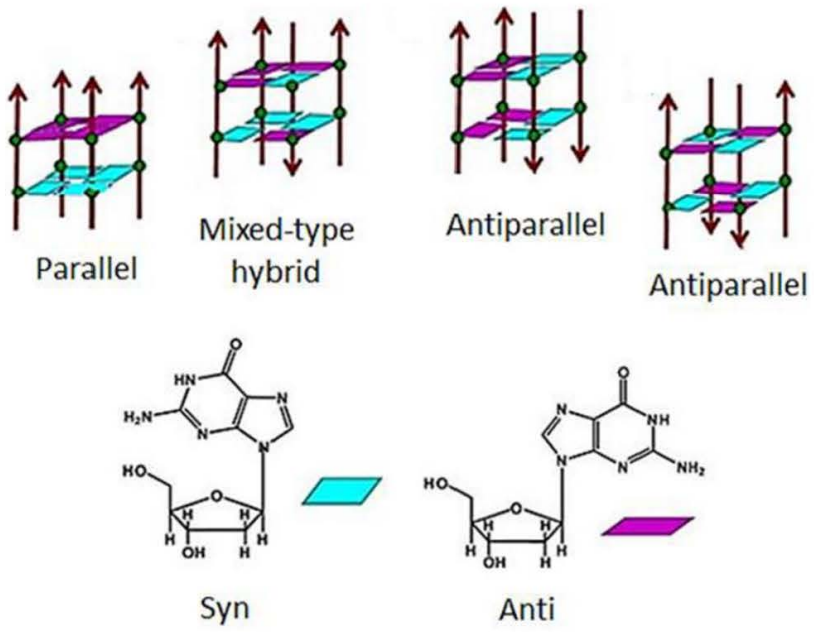


Figure 4. Example G-quadruplex topologies and guanosine glycosidic conformations. Arrows indicate DNA directionality. Possible topologies include parallel, hybrid, and antiparallel, while the possible guanosine conformations are *syn* and *anti*. Figure adapted from [28].

In monomolecular GQs, which are the focus of this thesis, loops connect the G-tracts to form a continuous DNA strand. Loops can connect the G-tracts in different ways, be comprised of any nucleotide(s), vary in length, and may span different regions of the

guanine core. Accordingly, a wide variety of factors can contribute to GQ structural diversity. In addition, quadruplex structures can form with other types of tetrads (such as GCGC and ATAT tetrads),[29] contain bulges,[30] and be connected to duplex DNA.[31] Small molecule binders, which will be discussed shortly, recognize different structural elements within these diverse quadruplexes.

i-motif DNA

Cytosine-rich sequences, which are complementary to guanine-rich sequences, have the potential to form a non-canonical DNA structure known as the intercalated motif, i-motif. I-motifs consist of a stack of intercalated, hemiprotonated C-C^+ base pairs that together form a tetra-stranded structure (**Figure 2**).[32,33] I-motif formation is pH-dependent due to the requirement for cytosine protonation. Under physiological temperature and ionic strength, i-motifs are stable *in vitro* at pH <7.[32] They were previously thought to be unstable under both neutral and basic conditions, and consequently not frequently studied due to their unclear role in physiological conditions.[32] However, it has recently been shown that some i-motifs are stable at physiological pH,[34–36] and there is evidence that i-motifs may serve as molecular switches to regulate gene expression.[12–14] Although GQs and i-motifs can form from complementary DNA strands, the simultaneous formation of these secondary structures may be mutually exclusive due to steric hindrance.[37]

Small molecule ligands

Small molecule ligands that bind GQs can modulate GQ folding and stability[38] and thereby alter or enhance the effects of GQs. Stabilization of a telomeric GQ, for example,

may render it even more difficult for telomerase to extend the telomere. A ligand may also disrupt telomere architecture upon binding, such as by displacing telomere-binding proteins, and thereby enable cancer cell death.[2,7,9,39,40] However, in general, the mechanism underlying GQ ligand selectivity and function is poorly characterized.

Many GQ ligands have been developed[41] with varying affinity and selectivity for GQ DNA. Small molecule ligands can bind GQs in a variety of ways by recognizing different quadruplex structural elements, as depicted in **Figure 5**. The majority of these ligands have affinity in the micromolar range or worse ($K_a < \sim 1\ \mu M^{-1}$, $K_d > \sim 1\ \mu M$).[42] In particular, many planar, aromatic, and positively charged ligands were synthesized with the idea that these properties are most complementary to the aromatic, negatively charged GQ DNA. However, since all nucleic acids are negatively charged, the majority of GQ ligands that can bind GQ DNA tightly lack selectivity for GQ DNA over other DNA structures.

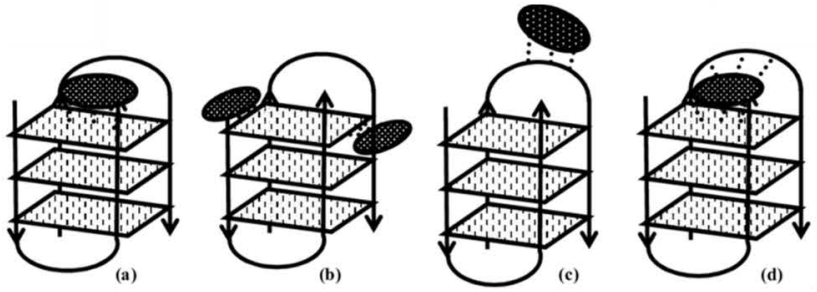


Figure 5. GQ ligand binding modes. (**A**) End stacking, (**B**) edge recognition, (**C**) loop interaction, and (**D**) simultaneous end stacking and loop groove interaction. The ligand is depicted as a solid black oval while the interaction is depicted as dashed lines. Figure adapted from [28].

A much more selective GQ ligand is *N*-methyl mesoporphyrin IX (NMM, **Figure 6**), a water-soluble porphyrin with a distinctive central methyl group. Our lab has previously shown that NMM is not only highly selective for GQ DNA over other DNA structures, but selective for the parallel GQ conformation specifically.[43] The lab has also

demonstrated that NMM can serve as a fluorescent probe for GQ DNA *in vitro*.[44] Its impressive selectivity, along with its fluorescence properties, renders NMM a very promising GQ ligand with a wide array of potential applications (which we recently reviewed[45]). Another GQ ligand studied in this thesis is 3,11-difluoro-6,8,13-trimethyl-8H-quino[4,3,2-kl]acridinium methosulfate (RHPS4, **Figure 6**), which has shown promise in targeting tumor cells.[46,47] I-motif ligands, however are currently scarce.[23]

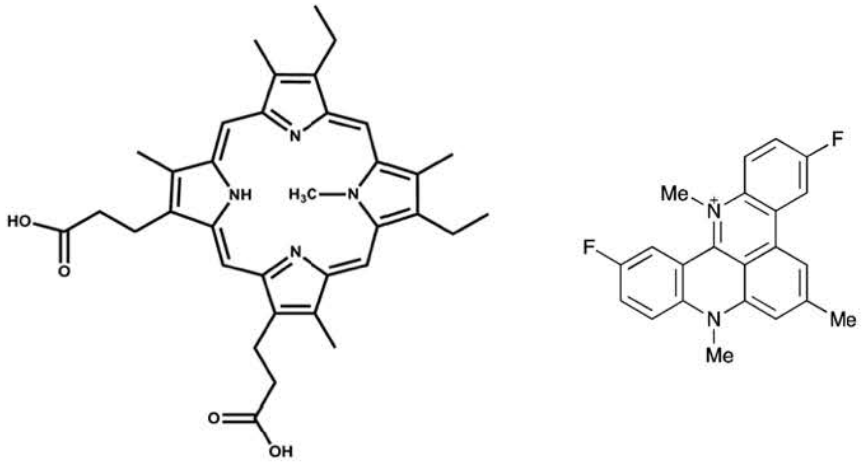


Figure 6. G-quadruplex ligands. (Left) NMM and (Right) RHPS4.

Goals

A large portion of this thesis is dedicated toward elucidating the structures of non-canonical DNA and characterizing their behavior *in vitro*. Specifically, we study the binding of a G-rich telomeric sequence (T1) from the ciliate *Tetrahymena thermophila* to NMM, as well as a longer G-rich sequence (19wt) from the social amoeba *Dictyostelium discoideum*. Both of these organisms are simple eukaryotes with genomes substantially smaller than the human's (104 and 34 Mb, respectively,[48,49] compared to 2.85 Gb in humans),[50] which renders the task of understanding the biological functions of GQs more tractable. We also aim to determine the structure of T1 in complex with RHPS4, as well as the structure of an i-motif which could form near the human RAS promoter (iHRAS).

The work in this thesis mainly involves two approaches. First, we characterize GQ folding and stability, as well as the stoichiometry, strength, and thermodynamic properties of the ligand binding interaction if applicable, via biophysical methods. Second, we aim to elucidate the atomic details of these non-canonical DNA structures via X-ray crystallography. Through doing so, we hope to further our understanding of the structural features and diversity of these DNA structures. Furthermore, this work provides complementary biophysical data and atomic coordinates to inform GQ and i-motif targeting anticancer drug design.

Chapter 2: Materials and methods

DNA and buffers

Lyophilized oligonucleotides were purchased from Integrated DNA Technologies (IDT; Coralville, IA) with standard desalting purification. DNA was hydrated in doubly-distilled water to 1-2 mM and stored at -80 °C. Extinction coefficients for all sequences were obtained using IDT's OligoAnalyzer 3.1 and DNA concentration was determined from UV-vis spectra collected at 95 °C. The full list of DNA sequences used in this work, along with their extinction coefficients and molecular weights, can be found in **Table 1**. To induce GQ formation, DNA was diluted into the desired buffer, heated at 90-95 °C for 5-10 minutes, cooled slowly to room temperature over 4 hours, and equilibrated at 4 °C overnight. NMM stock was prepared in doubly-distilled water and its concentration was determined using an extinction coefficient of 1.45×10^5 $M^{-1}cm^{-1}$ at 379 nm.[51]

All biophysical experiments were performed in 5K buffer consisting of 10 mM lithium cacodylate pH 7.2, 5 mM KCl, and 95 mM LiCl unless otherwise specified. The Na^+ buffers consist of either 5 or 50 mM NaCl and 10 mM lithium cacodylate pH 7.2. Crystallization trials were performed in 20K buffer consisting of 10 mM lithium cacodylate pH 7.2 and 20 mM KCl.

Table 1. DNA sequences studied, extinction coefficients, and molecular weights. Nucleotides added or mutated from the corresponding base sequence (T1 or 19wt) are shown in bold. The 19wt family sequences are spaced for easier identification of loop nucleotides.

Name	DNA sequence	ε, $mM^{-1}cm^{-1}$	MW, g/mol
T1	GGGTTGGGTTGGGTTGGG	173.0	5713.7
T2	GGGTTGGGTTGGGTTGGG**GT**	191.6	6347.1
T3	**G**GGGTTGGGTTGGGTTGGG**GT**	201.7	6676.3
T4	**GG**GGGTTGGGTTGGGTTGGG**GT**	211.8	7005.5
T5	**TGG**GGGTTGGGTTGGGTTGGG**GT**	219.3	7309.7
T6	**TGG**GGGTTGGGTTGGGTTGGG**GTT**	227.4	7613.9
T7	**T**GGGTTGGGTTGGGTTGGG**T**	189.0	6322.1
T8	**C**GGGTTGGGTTGGGTTGGG**G**	189.6	6322.1
T9	**G**GGGTTGGGTTGGGTTGGG**C**	189.2	6322.1
19wt	GGGG GA GGGG TACA GGGG TACA GGGG	270.9	8287.4
M1	GGGG **T**A GGGG TACA GGGG TACA GGGG	270.3	8262.4
M2	GGGG **TT** GGGG TACA GGGG TACA GGGG	264.4	8253.4
M3	GGGG **TT** GGGG TACA GGGG TAC**T** GGGG	259.1	8244.3
M4	GGGG GA GGGG TACA GGGG TA**A**A GGGG	275.3	8311.4
M5	GGGG **T**A GGGG TACA GGGG TA**A**A GGGG	274.7	8286.4
M6	GGGG **A**A GGGG TACA GGGG TA**A**A GGGG	277.2	8295.4
M7	GGGG **T**A GGGG TACA GGGG TA**T**A GGGG	272.8	8277.4
M8	GGGG GA GGGG TA**A**A GGGG TACA GGGG	275.3	8311.4
iHRAS	CGCCCGTGCCCTGCGCCCGCAACCCGA	N/A	N/A

UV-vis spectroscopy

All UV-vis experiments were performed on a Varian Cary 300 UV-vis spectrophotometer equipped with a Cary temperature controller (± 0.3 °C error). Spectra were collected from 220-349 nm for DNA and 352-480 nm for NMM at 0.5 nm intervals

with a 0.1 s averaging time, 300 nm/min scan rate, 2 nm spectral bandwidth, and automatic baseline correction.

UV-vis titration

DNA samples were prepared at 90-170 μM to target a final [DNA]/[ligand] ratio of at least 1.5. NMM samples were prepared in 1 cm methyl methacrylate cuvettes to target ~0.5 absorbance(3-4 μM NMM at 1000 μL). To maintain constant NMM concentration throughout the titration, an equivalent concentration of NMM was added to the annealed DNA sample. In the titration, DNA was added to NMM in increasing increments, and the resulting sample was equilibrated for 2 min before the UV-vis spectrum was collected. The titration continued until no further changes were observed in at least three UV-vis spectra. The volume of DNA added (30-100 μL total), λ_{max}, and absorbance at λ_{max} were monitored throughout the titration. Data was processed using singular value decomposition followed by a direct fit as described below. Reported K_a values represent the average of three trials.

Job plot

Job plot is a continuous variation analysis method that allows for model-independent determination of binding stoichiometry.[52] The method requires two titrations involving equal concentrations of DNA and ligand (~2.50 μM). In the first titration, cuvettes containing 1 mL of NMM were placed into the sample and reference cells in the UV-vis. DNA was titrated into the sample cell while an identical volume of buffer was titrated into the reference cell in increasing increments of 20 – 200 μL. In the second titration, a cuvette containing 1 mL of DNA was placed in the sample cell while a cuvette

containing 1 mL of buffer was placed in the reference cell, and equivalent volumes of NMM were titrated into both cuvettes in increasing increments of 50 – 200 μL. Absorbance differences measured at the wavelengths of minimum and maximum absorbance were plotted against the mole fraction of NMM. The peak or trough in the Job plot indicates the mole fraction of NMM bound to DNA and thereby reports on the stoichiometry of the DNA-ligand complex. Three consistent trials were completed.

Thermal difference spectra (TDS)

TDS were obtained by subtracting UV-vis scans taken at 4 °C after 5 minutes of equilibration from scans taken at 95 °C after 10 minutes of equilibration. In principle, the low and high temperature limits are defined by the temperatures at which the DNA is (mostly) folded and unfolded, respectively. GQs have a characteristic trough in their TDS spectra around 295 nm.[53] A script for processing TDS data can be found in the **Appendix**.

Fluorescence (FL) titration

Fluorescence experiments were conducted on a Photon Technology International QuantaMaster 40 fluorometer at 20.0 °C. Data was collected at the emission range of 560 - 720 nm with 2 nm slit widths, 0.5 nm step size, and 0.5 s integration time. The isosbestic point determined through UV-vis titrations (391 nm for the T1-NMM complex) was used as the excitation wavelength. NMM in a methyl methacrylate fluorescence cuvette (~1.0 μM at 1500 μL) was titrated with small increments of 130-190 μM DNA to target a final [DNA]/[NMM] ratio of at least 1.5. The total volume of DNA added (40-60 μL), λ_{max}, and

intensity at λ_{max} were monitored throughout the titration. Data were analyzed in the same way as UV-vis titration data, and reported K_a values represent the average of five trials.

Quantitative analysis of binding curves

Singular value decomposition (SVD)

SVD is a matrix-based method that allows for global processing of titration data at all wavelengths, thereby providing a strong advantage when compared to traditional fitting of data from only one (or several) wavelengths. First, we imported the data into MATLAB as a matrix $\boldsymbol{M}$ consisting of the signal at each wavelength (columns) for every addition (rows). This matrix was then decomposed into three matrices $\boldsymbol{U}$, $\boldsymbol{S}$, and V such that $\boldsymbol{M} = \boldsymbol{USV}^T$. Matrix $\boldsymbol{U}$ consists of basis columns that represent each component. Meanwhile, the diagonal matrix $\boldsymbol{S}$ contains singular values that serve as weighing factors. Finally, V^T contains amplitude column vectors that indicate how much of each component is present at every addition. DNA concentration was plotted against a vector in V to generate a binding curve. While each column in V represents the same binding event, we used the second column, ***v2***, to fit UV-vis titration data and the first column, ***v1***, to fit fluorescence titration data because they generated curves that looked most representative of a binding event.

Direct fit

Direct fit assumes a simple *DNA + ligand* → *DNA-ligand complex* binding model[54] with 1:1 binding stoichiometry. This method can also be used for higher stoichiometries by assuming equivalent, independent binding sites. However, it is important to keep in

mind that such data treatment is likely an oversimplification. In such cases, the concentration of DNA binding sites was set equal to the concentration of DNA multiplied by the appropriate stoichiometric ratio. Direct fit was applied in GraphPad Prism 4 with the concentration of binding sites at each addition as the independent variable and the appropriate column vector from V as the dependent variable. NMM concentration, as experimentally determined via UV-vis, was either kept constant or allowed to float if it provided a higher quality fit. A refined NMM concentration was accepted only if it deviated less than 20% from the measured value. Direct fit of the data yields the most probable binding stoichiometry and the binding constant, K_a. A variety of binding models (e.g. 1:1, 2:1, and 1:2) were tested. The best fit was the model with the lowest stoichiometry, < 20% error in K_a, random residuals, and a fit curve closely matching the data upon visual inspection.

Isothermal titration calorimetry (ITC)

ITC experiments were conducted on a Malvern MicroCal PEAQ-ITC. The titrations were run in two directions: NMM into DNA (forward) and DNA into NMM (reverse). Forward titrations were performed at 25 °C with 6 µcal/s reference power, high feedback, 750 rpm stir speed, 60 s initial delay, and one 0.4 µL injection followed by 18 injections of 2 µL with 250 s spacing between each injection. The cell contained 11.4-25.0 µM DNA, while the syringe contained 67-180 µM NMM, both prepared in degassed 5K buffer. The reverse experiments were run with the same parameters but with the reference power was set to 10 µcal/s and low feedback. The cell contained 18-24 µM NMM while the syringe contained 350-476 µM DNA. Heat release was monitored during the

experiment, and the observed peaks were baseline corrected, integrated, and plotted against the molar ratio of NMM/DNA to generate a binding curve. Fitting of the data using the instrument software yielded the number of binding sites and dissociation constant, as well as the enthalpy (ΔH), entropy (ΔS), and free energy (ΔG_{298}) associated with the binding event. At least three consistent trials were completed for each direction of the ITC titration.

Circular dichroism (CD) spectroscopy

All CD experiments were conducted on an Aviv 435 circular dichroism spectrophotometer equipped with a Peltier thermal controller (± 0.3 °C error) in 1 cm quartz cuvettes. DNA samples were annealed at ~5 μM alone and with 2 eq. of NMM if applicable in 5K buffer. CD scans were taken at 20 °C from 220 - 330 nm with a 1 s averaging time, 2 nm bandwidth, and 1 nm step. Three scans were collected first for buffer and then for each sample in the corresponding cuvettes. CD data were processed as described in our earlier work,[54] and scripts for data processing can be found in the **Appendix**. CD scans allow for determination of GQ conformation (e.g. parallel, mixed-hybrid, or antiparallel).[55]

CD melting experiments were conducted from 25-95 °C with a 1 °C step, 1 °C/min temperature rate, 15 s averaging time, and 5 s equilibration time. CD signal at the wavelength corresponding to maximum signal in CD scans, was monitored as a function of temperature. A macro for running CD scan-melt-scan experiments can be found in the **Appendix**. Melting temperatures, T_m, were determined via two methods. The first method involves taking the first derivative of the smoothed CD signal (using a 13-point Savitzky-Golay quadratic function) and finding the temperature at the trough through visual inspection (associated with ± 0.5°C error). The second method assumes a two-state model

for DNA unfolding with constant ΔH and should be used for fully reversible melting transitions, i.e. when melting and cooling curves are (nearly) superimposable.[56] Hysteresis was determined as the difference between T_m from the melting and cooling curves. Since the hysteresis never exceeded 3.3 °C, all systems were considered (nearly) reversible and the reported thermodynamic data were obtained using the two-state model. Reported results represent the average of 2-3 trials.

Native polyacrylamide gel electrophoresis (PAGE)

PAGE samples contained 40-50 μM DNA in 5K buffer and were weighted down with 7% w/v sucrose prior to loading. Twenty percent native polyacrylamide gels were made with 5 mM KCl and 1×Tris-Borate-EDTA. Gels were pre-migrated at 150 V for 30 min, loaded with 6-10 μL sample, and run for 120-150 min at 150 V at room temperature. A tracking dye was used to monitor gel progress and an oligothymidylate ladder consisting of dT_{15}, dT_{24}, dT_{30}, and dT_{57} was used as a length marker. DNA bands were visualized using Stains-All and the resulting gel was captured with a scanner.

X-ray crystallography

Crystallization via hanging drop vapor diffusion

Crystallization was achieved at room temperature using the hanging-drop vapor diffusion method. The T1-NMM sample was prepared by annealing HPLC-purified DNA with 1 eq. of NMM at 0.65 mM in 20K buffer. Drops were set manually at 2 μL DNA sample and 1 μL crystallization condition. The original crystal grew in condition 1-31 from the HELIX screen (Molecular Dimensions), which contains 1.0 M sodium formate, 20%

PEG 20000, and 0.05 M Bis-Tris pH 7.0. This condition was then optimized to 0.85 M sodium formate, 17.5% PEG 20000, and 0.05 M Bis-Tris pH 7.0, yielding large hexagonal crystals that grew within 3 weeks to 300 μm in the largest dimension. Crystals were harvested from the growth condition and flash frozen in liquid nitrogen.

The T7-NMM sample was prepared by annealing DNA with 1 eq. of NMM at 0.65 mM in 20K buffer. Drops were set by the TTP Labtech Mosquito Crystal equipped with a humidity chamber at 0.1 μL DNA sample and 0.1 μL of the crystallization condition. Small hexagonal crystals grew within three weeks to 80 μm in the largest dimension from condition C5 of the Natrix screen (Hampton Research), which contains 4.0 M LiCl, 0.01 M $MgCl_2$, and 0.05 M HEPES sodium pH 7.0. Crystals were cryoprotected in the base condition supplemented with 15% ethylene glycol before being flash frozen in liquid nitrogen.

Data collection and structure solution

Datasets were collected at the Advanced Photon Source 24 ID-E synchrotron facility to a maximum resolution of 2.39 Å for T1-NMM and 2.34 Å for T7-NMM. Raw diffraction data was processed using XDS.[57] The structures were solved by molecular replacement (MR) using PHENIX.[58] Three types of 3-quartet parallel GQ models were tested in MR: entire GQs; GQs with only thymine(s) in the loops (with other loop nucleotides removed); and GQ cores consisting of only G-tetrads, with loops and overhangs removed. All models had K^+ ions included. In addition, we tried MR with a single G-tetrad.

The T7-NMM structure was solved via MR using the GQ core of human telomeric DNA in complex with NMM (Protein Data Bank, PDB ID: 4FXM).[59] The initial MR solution was improved using PHENIX AutoBuild, and NMM was placed into the structure with PHENIX LigandFit. Extensive manual model building cycles were performed in Coot[60] followed by PHENIX Refine.

The T1-NMM structure was solved via MR using the individual DNA chains and associated K^+ ions from the T7-NMM structure, followed by AutoBuild, LigandFit, and manual model building, along with continuous refinement cycles. Data collection and refinement statistics for both structures are presented in **Table 2**.

Table 2. Crystallographic statistics for the T1-NMM and T7-NMM complexes.

	T1-NMM	**T7-NMM**
Resolution range (Å)	64.78 - 2.388	51.34 - 2.339
Space group	R 3 2	P 6_3
Unit cell dimensions		
a, b, c (Å)	60.93, 60.93, 194.342	59.28, 59.28, 63.33
α, β, γ (°)	90, 90, 120	90, 90, 120
Unique reflections	5748 (492)	5377 (498)
Redundancy	19.0 (19.0)	9.5 (6.1)
Completeness (%)	97.48 (86.90)	98.84 (92.39)
I/σ	21.4	14.9
R-merge (%)	6.7	7.8
R_{work} / R_{free} (%)	23.52 / 24.91	20.01 / 22.43
Number of atoms	822	862
DNA	701	778
Ligand	115	75
Solvent	0	4
Potassium	6	5
Copies in asymmetric unit	2	2
Overall *B*-factor (Å^2)	115.44	97.15
RMS deviations		
Bond lengths (Å)	0.008	0.01
Bond angles (°)	0.83	1.2
PDB ID	**6PNK**	**6P45**

Model building

The asymmetric units of both the T1-NMM and T7-NMM crystals contain two DNA chains (A and B), each of which is bound to one NMM molecule. The DNA chains form a dimer in the case of T7, whereas they form dimers with symmetry-related molecules in the case of T1. Accordingly, for T7-NMM the dimers are of A-B nature, while for T1-NMM the dimers are of A-A and B-B nature. Both structures include two K^+ ions within each GQ monomer, as well as a K^+ ion at the dimer interface. In the T1-NMM structure, the latter K^+ is at a special position, so it was modeled at 0.5 occupancy for each monomer. Due to a lack of clear electron density, the propionate groups of NMM were not built.

In the T1-NMM structure, the base for T_5 and T_{14} in both chains and for T_4 and T_{15} in chain B was not built due to disorder in the loop regions. Similarly, the sugar for T_5 in chain B was not built. We modeled NMM in chain B in two different orientations each at 0.5 occupancy with the *N*-Me group residing on different pyrrole nitrogens. We also built a sodium formate molecule near G_{18} of chain B (the crystallization condition contained 0.85 M sodium formate).

In the T7-NMM structure, the base for T_{15} in chain A and T_{16} in chain B was not built due to loop disorder. In addition, disorder at the 5' and 3' overhangs of both chains resulted in poor electron density for the phosphate and base for T_1 and T_{20}, which were not built. Structure figures were prepared in PyMOL[61] and Coot.

Structure analysis

To calculate the helical twist within a GQ structure, I wrote a script (**Appendix**) following methodology described by Chung et al.[62] that extracts the coordinates of all

atoms from the PDB file and represents every guanine base as a vector from C8 to the midpoint between N1 and C2 (**Figure 7A**). The angle between the vectors corresponding to every pair of stacked guanines was then calculated. We verified the script using a structure for which helical twist is reported and was calculated using the same theoretical strategy.[62]

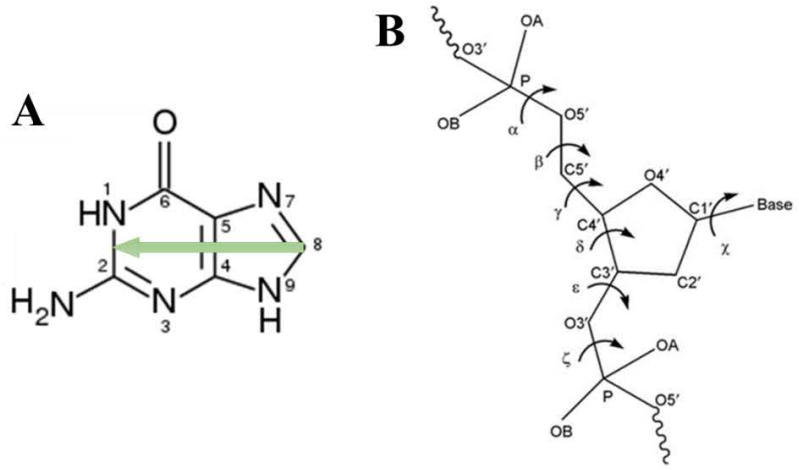


Figure 7. Guanine vector for helical twist calculations and DNA torsional angles. (**A**) Representation of guanine as a vector (green) from C8 to the midpoint of N1 and C2, which was then used to calculate helical twist between stacked guanines. (**B**) Schematic of DNA torsional angles. The backbone dihedral angles α, β, γ, δ, ε, and ξ describe the torsion of the DNA backbone, while the glycosidic torsion angle χ refers to the *syn* or *anti* conformation of the nitrogenous base.

We used principal component analysis to find the centroid of each G-quartet and of NMM (using the atomic coordinates of the 24 central atoms comprising the porphyrin ring, excluding the methyl group). Using this information, we calculated G-quartet (and triad) planarity, as well distances between adjacent G-quartets and between the outermost G-quartet and NMM.[59] We also calculated DNA torsional angles (**Figure 7B**) using 3DNA[63] and visualized their distribution as a wheel plot. The torsional angle plot and the planarity

and distance calculations were all performed using modified versions of MATLAB scripts by Barrett Powell '18.

Finally, root mean square deviation (RMSD) was calculated by aligning all the atoms in each pairing of DNA chains from the T1-NMM and T7-NMM structures in PyMOL (with no outlier rejection). Groove widths were measured in PyMOL as P-P distances, while B-factors were calculated using the Average_b PyMOL script (PyMOL Wiki).

Chapter 3: Binding of *N*-methyl mesoporphyrin IX to the $(GGGTT)_3GGG$ G-quadruplex

In this project, we set out to determine crystal structure of the 5'-$(GGGTT)_3GGG$-3' sequence (T1) alone and in complex with NMM. To increase our chances for successful crystallization, we designed a variety of constructs based on its genomic context. We used BLAST[64] to align T1 with a number of genomes. The T1 sequence is found at position 168,273 on the *Tetrahymena thermophila* genome in the telomeric region. Note that it does not represent the consensus telomeric motif, $(GGGGTT)_n$ (studies are underway in the lab).

The T1 sequence also occurs 11 times in the human genome, in seven distinct regions across six chromosomes (**Table S1**). In particular, it may be involved in regulating β-secretase 2 (BACE2), a gene implicated in Alzheimer's disease. Furthermore, the sequence is found in the bacteria *Neisseria meningitidis*, *Neisseria gonorrhoeae*, the bird pathogen *Escherichia coli* strain APEC O78, and repetitively in *Paenibacillus*.[65,66]

The native constructs (designated T2 – T6) contain the full T1 sequence expanded by 1-3 nucleotides (nt) in the 5', 3', or both directions. In addition, we designed a construct (T7) to prevent the dimerization observed for T1 by adding 5' and 3' thymine tails. Finally, we designed two constructs (T8 and T9) to promote efficient crystal packing via the formation of intermolecular Watson-Crick base pairs by adding cytosine and guanine at the 5' and 3' ends. We choose the C-G base pair for its increased hydrogen bonding in comparison to an A-T base pair. All construct sequences are listed in **Table 1**. We successfully crystallized both T1 and T7 in complex with NMM.

Biophysical characterization of constructs

Prior to crystallization, we determined the TDS and CD signatures of the T1 – T9 constructs to assess their similarity to the target T1 sequence. All constructs have TDS spectra with peaks at 240 and 275 nm and the characteristic trough at 296 nm (**Figure 8A**), indicating that they all fold into GQs. Furthermore, a peak at 262 nm and trough at 241 nm is observed in all CD scans (**Figure 8B**), suggesting a parallel GQ topology. The similarity of the observed signatures (TDS and CD) for the sequences indicates that they fold into similar secondary structures.

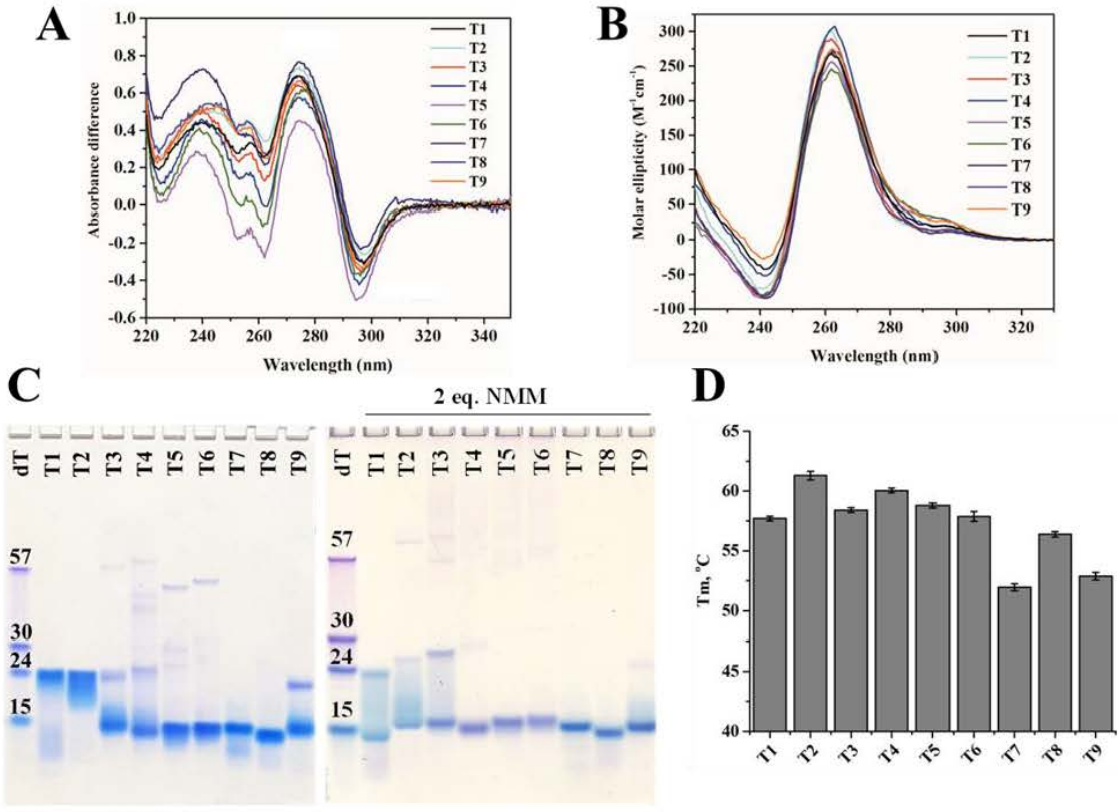


Figure 8. Biophysical characterization of the T1 – T9 constructs. (**A**) TDS and (**B**) CD scans collected at 25 °C. (**C**) Twenty percent native PAGE gels for DNA alone (left) and with 2 eq. of NMM (right). (**D**) T_m for the constructs. All DNA samples were prepared in 5K buffer at ~5 μM except for the gel samples, which were prepared at 40-50 μM.

We then investigated the homogeneity and oligomerization state of all the constructs via PAGE (**Figure 8C**). PAGE reveals that both T1 and T2 predominantly form dimers, although the streakiness of their bands likely suggests the presence of multiple species. The rest of the constructs form monomers, with T3, T4, and T9 also having some amount of dimer. Finally, T3 – T6 also contain a small amount of higher order oligomeric species. Dimer formation in T1 – T4 and T9 can be explained by the presence of a 5' G, which often leads to dimerization.[67–69] In contrast, when a nucleotide other than guanine is found at the 5' end, no dimer formation is observed, as is the case for T5 – T8.

Next, we assessed the stability of all the constructs via CD melting studies (**Figure 8D**). In all cases, the melting transitions were nearly reversible (the hysteresis did not exceed 3.3 °C), thereby allowing us to extract thermodynamic parameters (**Table 3**). T1 melts at 57.7 ± 0.3 °C in the presence of 5 mM K^+. Extension of the T1 sequence only has a mild effect on its stability, with the largest changes in T_m being +3.6 and -5.8 °C. Notably, the T2 – T6 constructs, which are all extended native sequences, have slightly enhanced thermal stability relative to T1 (**Table 3**), likely due to non-essential, stabilizing capping interactions. Meanwhile, addition of non-native nucleotides (in T7 – T9) destabilizes the structure, likely either due to an increase in entropy (if the added nucleotides do not participate in stabilizing interactions) or a lack of dimer formation (**Figure 8C**).

Table 3. Stability and oligomerization states of constructs. Melting temperature, enthalpy, and hysteresis were determined via CD melting experiments, while oligomerization states were determined via PAGE (M – monomer, D – dimer). *relative to T1.

DNA	T_m, °C	ΔT_m, °C*	ΔH, kcal/mol	Hysteresis, °C	Oligomerization
T1	57.7 ± 0.3	-	77 ± 2	3.3	D
T2	61.3 ± 0.4	3.6	83 ± 3	3.2	D, smeary
T3	58.4 ± 0.3	0.7	51.9 ± 0.9	2.5	M+D
T4	60.1 ± 0.3	2.3	50.6 ± 0.8	2.2	M+D
T5	58.8 ± 0.3	1.1	50.8 ± 0.8	2.8	M
T6	57.9 ± 0.4	0.2	53.0 ± 0.4	2.9	M
T7	52.0 ± 0.3	-5.8	48.6 ± 0.1	2.2	M
T8	56.4 ± 0.3	-1.3	46 ± 3	2.9	M
T9	52.9 ± 0.3	-5.0	38.1 ± 0.5	3.1	M+D

It is interesting to compare T8 and T9, which have identical compositions but swapped C/G overhangs. T8 (with a 5' C) appears to be mostly monomeric with a T_m of 56.4 ± 0.3 °C, while T9 (with a 5' G) has a small amount of dimer and a lower T_m of 52.9 ± 0.3 °C. Thus, even the addition of seemingly unimportant overhang nucleotides affects both the oligomerization state and the stability of the resulting GQ, indicating that every nucleotide warrants consideration during construct design. The stability of T1 and T7 under similar conditions was investigated by Largy et al. and yielded similar T_m values.[70]

Dimeric T1 and T2 display significantly higher enthalpy values (~80 kcal/mol) than other constructs (ΔH of ~50 kcal/mol) with the exception being T9, which has a low ΔH of 38 kJ/mol). The higher enthalpies for dimeric GQs indicate that the dimers are maintained by a multitude of bonding interactions, which is consistent with our crystallographic results (presented below).

Biophysical characterization of NMM binding to T1/T7

Effect of NMM via PAGE

Previous work in the lab demonstrated that the selective GQ ligand NMM stabilizes T1 by an impressive 23 ± 2 °C at 2 eq.[43] However, neither the binding strength nor the

binding mode were previously established. To assess the effect of NMM on T1 – T9, we performed PAGE on DNA annealed in the presence of 2 eq. of NMM (**Figure 8C**). As seen through the disappearance of higher order bands when compared to the gel of DNA alone, NMM increases the amount of monomeric GQ at the expense of dimers and higher ordered species.

As T7 and T8 form fully homogeneous complexes with NMM (**Figure 8C**), they were good candidates for crystallographic studies, so we proceeded to characterize these complexes via TDS, CD scans, and CD melting studies. Their TDS signatures, while not similar to that of classical GQ DNA, still retains a small trough at 293 nm, indicative of a GQ fold, and displays a strong trough at 263 nm (**Figure S1A**). This somewhat unusual TDS signature may result from interference by NMM's absorbance. Meanwhile, CD scans demonstrate that NMM causes a small change in CD signal intensity (an increase for T7 and decrease for T1 and T8) and a small red shift of ~2 nm (**Figure S1B-D**). The general CD signature, which is indicative of a parallel GQ, is maintained in the presence of NMM. Finally, CD melting studies demonstrate that NMM greatly stabilizes T1 by 20. ± 1 °C (consistent with earlier work[43]), T7 by 19.9 ± 0.4 °C, and T8 by 17.0 ± 0.5 °C at 2 eq. (**Table S2**). The degree of stabilization is similar among the constructs and is the highest reported in the literature for GQ-NMM complexes. NMM had little effect on ΔH.

To further examine the effect of NMM on the constructs, we used UV-vis and fluorescence spectroscopy, along with ITC, to determine the binding constants and thermodynamic parameters of the binding interaction of NMM with T1 and T7. We also performed Job plots to confirm their binding stoichiometries.

Binding studies via UV-vis and fluorescence titrations

The UV-vis spectrum of NMM displays a 17.9 ± 0.4 nm red shift, low hypochromicity of 1 ± 4 %, and an isosbestic point at 391 nm upon addition of T1 (Figure 9**A**). This data is consistent with values reported for UV-vis titrations of NMM with a variety of parallel GQs.[43] Titration data are best fit to the 1:1 binding model and yield a K_a of 30 ± 20 μM^{-1} for T1 (Figure 9**B**), signifying an impressively tight binding interaction.

NMM displays the 'light-switch' effect in the presence of GQ DNA,[44,71] as its fluorescence becomes dramatically enhanced. We made use of this property to corroborate the K_a determined via UV-vis titrations. Fluorescence titration data yield a 1:1 binding stoichiometry and a K_a of 50 ± 20 μM^{-1} for T1-NMM (Figure 9**C-D**). NMM fluorescence enhancement upon binding T1 was found to be 49 ± 2, which is consistent with the fluorescence enhancement of 40-70 reported for a variety of parallel GQs (**Figure S2**).[44]

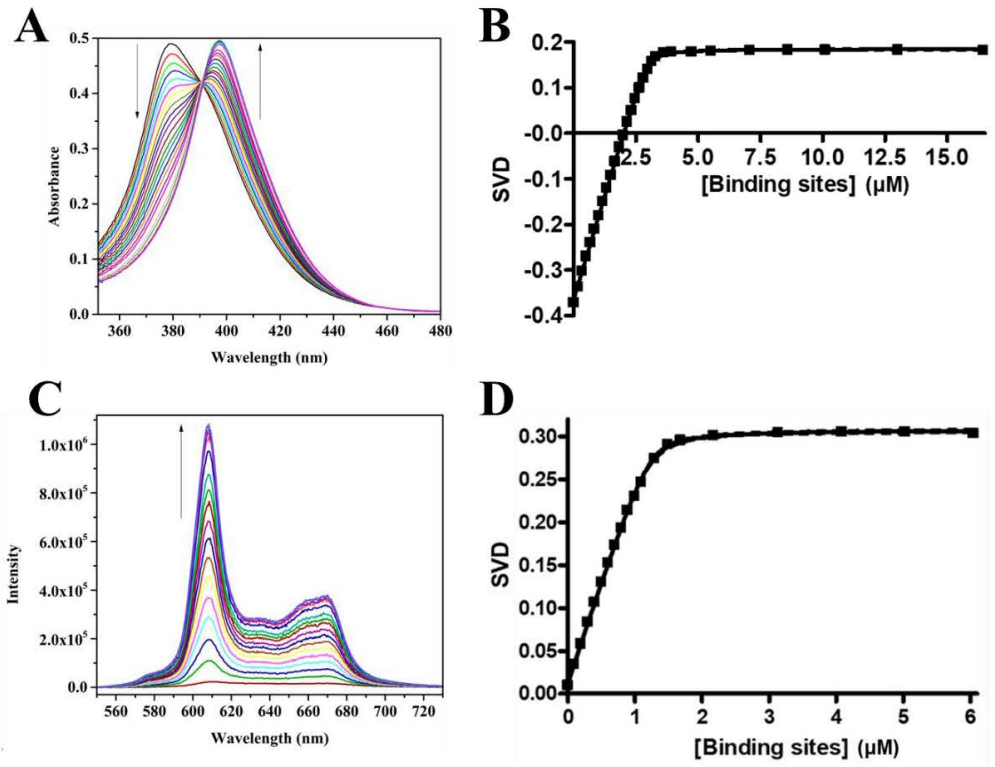


Figure 9. Determination of K_a for T1-NMM via UV-vis and fluorescence titrations. (**A**) Representative UV-vis titration of 3.4 μM NMM with 170 μM T1 to final [T1]/[NMM] of 4.9. (**B** and **D**) SVD fit of titration data (solid squares) to the 1:1 binding model with floating [NMM]. The 95% confidence interval is shown as dashed lines. (**C**) Representative FL titration of 1.0 μM NMM with 150 μM T1 to final [T1]/[NMM] of 6.1 at 20 °C.

Binding studies via ITC

We performed ITC studies to verify the results from UV-vis and FL titrations and to complement the binding data with thermodynamic parameters. ITC data (**Figure 10**) suggest that one NMM molecule binds two T1 GQs (i.e. one NMM binds to one T1 GQ dimer), which is not in line with the 1:1 binding model obtained via UV-vis and FL titrations. The K_a was determined to be 70 ± 20 μM^{-1}, which is somewhat higher, but consistent overall with values obtained from UV-vis and FL titrations. Binding of T1 to NMM is a spontaneous, enthalpically-driven process with the following thermodynamic parameters: ΔG of -10 ± 3 kcal/mol, ΔH of -17 ± 3 kcal/mol, and -TΔS of 7 ± 2 kcal/mol (**Figure 10C**).

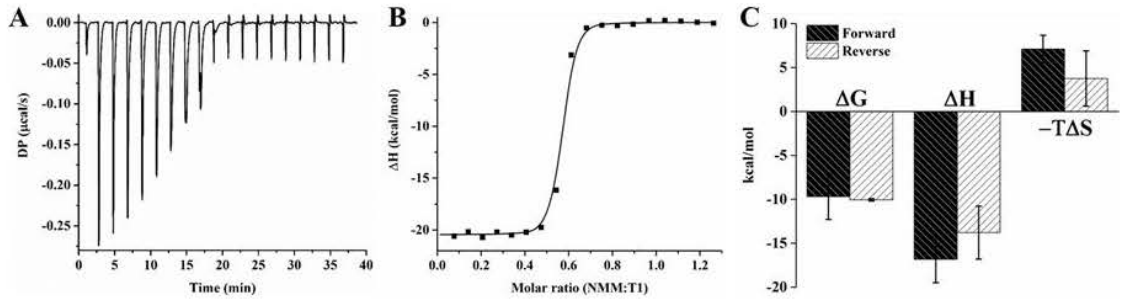


Figure 10. Thermodynamic characterization of T1-NMM binding via ITC. (**A**) Representative raw ITC data for titration of 90.5 μM NMM into 12.4 μM of prefolded T1 in 5K buffer at 25 °C with baseline subtraction. (**B**) Integrated raw heats with a fit generated using the one-site binding model. (**C**) Gibbs free energy deconvoluted into enthalpic (ΔH) and entropic (ΔS) contributions for both the forward and reverse titrations.

We hypothesized that the inconsistency in binding stoichiometry between the three methods (1:1 in UV-vis and FL titrations vs. 2:1 in ITC) could be due to the experimental set up. In both UV-vis and FL titrations, NMM samples were titrated with DNA. This setup was reversed for ITC, in which the macromolecule is conventionally placed in the sample cell and titrated with ligand. It is possible that at the beginning of the ITC titration, under conditions of excess DNA, two T1 GQs (oneT1 dimer) bind one NMM, resulting in the observed 2:1 stoichiometry. Upon further increase in NMM concentration, the stoichiometry could change to 1:1. To test this hypothesis, we performed the reverse ITC experiment: we placed NMM in the sample cell and loaded T1 into a syringe. However, contrary to our hypothesis, the results did not change and once again yielded a 2:1 T1/NMM stoichiometry, along with a K_a of 23 ± 6 μM^{-1}, ΔG of -10.1 ± 0.1 kcal/mol, ΔH of -14 ± 3 kcal/mol, and $-T\Delta S$ of 4 ± 3 kcal/mol (**Figure 10** and **Figure S3**).

Finally, we also assessed the binding of T7 to NMM via ITC (**Table 4** and **Figure S4**). Compared to T1-NMM, the binding is weaker. The thermodynamic parameters for T7-NMM binding are: K_a of 2.5 ± 0.6 μM^{-1}, ΔG of -8.7 ± 0.4 kcal/mol, ΔH of -8.7 ± 0.1 kcal/mol, and $-T\Delta S$ of 0.0 ± 0.6 kcal/mol. The binding stoichiometry is nearly 1:1, as the number of binding sites (n) is 0.86 ± 0.06. A summary of results from all titration experiments is presented in **Table 4**.

1:1 stoichiometry via Job plot

In order to resolve the discrepancy in T1-NMM binding stoichiometry, we used the model-independent Job plot method. Job plot data (**Figure 5**) clearly and unambiguously suggest 1:1 binding, consistent with UV-vis and FL titration data.

Table 4. Summary of binding data. All experiments were performed at 20-25 °C.

Experiment	Model, DNA:NMM	K_a, μM^{-1}	Thermodynamic parameters, kcal/mol
T1-NMM			
UV-vis	1:1	30 ± 20	-
Fluorescence	1:1	50 ± 20	-
ITC Forward*	n = 0.56 ± 0.06	70 ± 20	ΔH = -17 ± 3 $-T\Delta S$ = 7 ± 2 ΔG_{298} = -10 ± 3
ITC Reverse*	n = 0.5 ± 0.1	23 ± 6	ΔH = -14 ± 3 $-T\Delta S$ = 4 ± 3 ΔG_{298} = -10.1 ± 0.1
T7-NMM			
ITC Forward*	n = 0.86 ± 0.06	2.5 ± 0.6	ΔH = -8.7 ± 0.4 $-T\Delta S$ = 0.0 ± 0.6 ΔG_{298} = -8.7 ± 0.1

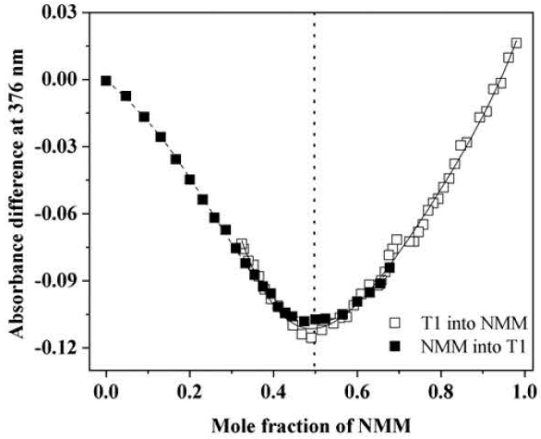


Figure 5. Determination of stoichiometry for the T1-NMM complex via Job plot. Representative Job plot at 20 °C. T1 and NMM were both prepared at 4.0 μM in 5K buffer, and the Job plot was built by plotting absorbance difference at λ_{max}. Data at other wavelengths are consistent with the data presented here.

In summary, the extensive binding studies for T1 with NMM demonstrate that NMM binds T1 with 1:1 stoichiometry and an extremely tight binding constant, K_a, of 50 ± 20 μM^{-1} ($K_d \approx 20$ nM). This binding interaction is thermodynamically favorable and enthalpically-driven. T7, which has two additional terminal thymines, binds more weakly to NMM, also in a 1:1 ratio.

Characterization of the GQ-NMM complexes via X-ray crystallography

Biophysical characterization of T1 – T9 indicated that all the constructs form GQ structures with similar fold and stability to T1. However, as only T5 – T8 produced homogeneous samples alone and/or in complex with NMM (**Figure 8C**), we conducted crystallographic studies on T1 and T5 – T8. We succeeded in producing diffraction quality crystals of T1, T7, and T8 all with NMM, and solved the structure of the T1-NMM and T7-NMM complexes.

Buffer selection for crystallization

Buffer choice is critical for crystallization. The buffer should provide optimal stability for the DNA or DNA-ligand complex, and should have simple composition with low ionic concentration to avoid interfering with the crystallization process and overshadowing the components of the crystallization mixture. We investigated the fold and stability of T1 in 10 mM lithium cacodylate pH 7.2 supplemented with 5-100 mM KCl. An increase in K^+ concentration led to a minor increase in CD signal intensity at 262 nm, higher thermal stability, and higher enthalpy of unfolding (**Figure S5**). Melting of T1 is accompanied by a small hysteresis, which decreases from 3.3 °C at 5 mM KCl to 1.2 °C at 100 mM KCl. Guided by this data, we chose 10 mM lithium cacodylate pH 7.2 and 20 mM KCl as the buffer for crystallographic studies, as it strikes the balance between assuring high thermal stability for T1 (T_m of 65.5 ± 0.9 °C and ΔH of 60 ± 3 kcal/mol) and maintaining a low overall ion concentration.

To determine the effect of ionic strength on T1 stability, we measured the T_m of T1 in 5 mM K^+ buffer in the presence of 0 and 95 mM LiCl. T1 stability is not affected by the presence of LiCl (T_m = 56.2 and 57.7 °C in 0 and 95 mM LiCl, respectively), but the enthalpy of T1 unfolding increases significantly from 47 ± 1 to 77 ± 2 kcal/mol upon addition of LiCl. The data suggest that increased ionic strength facilitates stronger bonding within each monomer and between the monomers that comprise the dimer, which otherwise repel each other due to the negatively charged DNA backbone. Additionally, the high negative charge density in GQ structures can facilitate interactions with Li^+ that contribute to the increased enthalpy of unfolding.

Atomic details of the T1-NMM and T7-NMM crystal structures

Both T1-NMM and T7-NMM produced large hexagonal crystals (**Figure 11A**). The T1-NMM crystal structure was solved in the R 3 2 space group to 2.39 Å, while the T7-NMM structure was solved in the P 6_3 space group to 2.34 Å and has overall higher quality. In both cases, the asymmetric unit contains two DNA chains, each of which binds one NMM molecule. The DNA form a parallel GQ structure, for which the overall architecture is shown in **Figure 11C**.

The RMSD between the two chains is 0.919 Å for T1-NMM and 1.212 Å for T7-NMM (**Table S3**), suggesting that the monomers within the asymmetric units are nearly identical to each other. Furthermore, the GQ structures in T1-NMM and T7-NMM are nearly identical, with an average RMSD among each pairing of DNA chains of 1.3 ± 0.4 Å. Meanwhile, RMSD for the GQ core only (consisting of 12 guanines and excluding any thymines) is 0.64 ± 0.2 Å, suggesting that the TT loops (particularly loops 1 and 3, which

are quite flexible in the crystal structures) are the main source of differences between the structures.

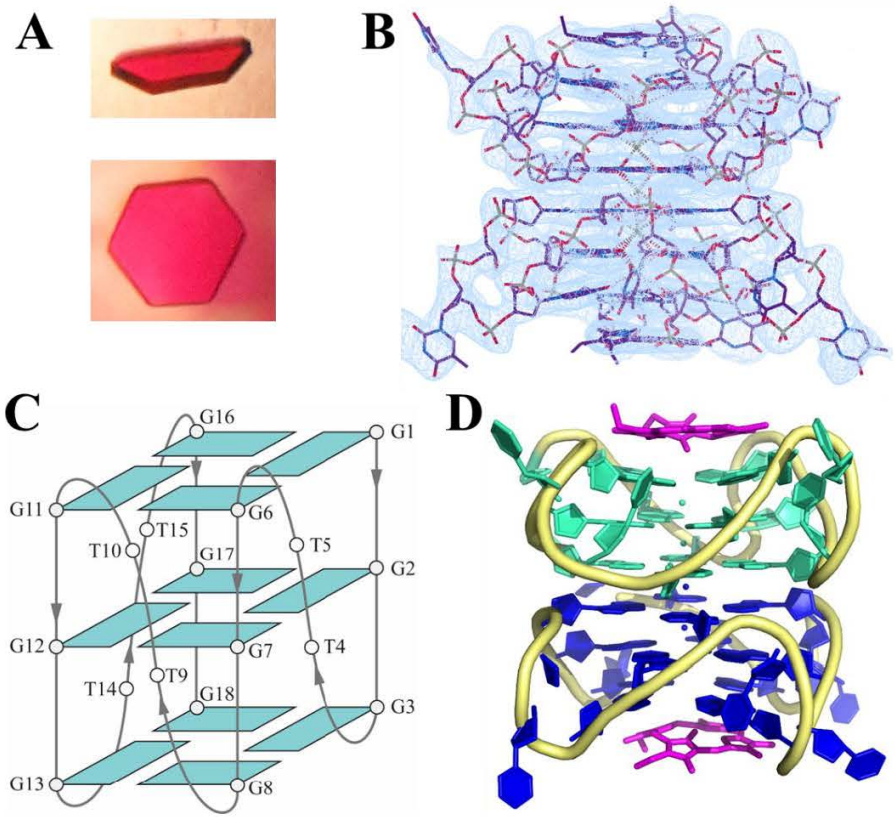


Figure 11. Crystal structure of the T7-NMM complex. (**A**) Representative crystal morphologies: half-hexagonal and hexagonal. (**B**) Electron density shown at 1.0 I/σ around the T7-NMM dimer. Note, electron density for the *N*-methyl group of NMM is not clear at this resolution. (**C**) Schematic representation of the structure. (**D**) Graphical representation of the asymmetric unit: a 5'-5' dimer capped at the 3' ends by NMM. Chain A is colored in teal, chain B is blue, the sugar-phosphate backbone is yellow, and NMM is pink. Potassium ions are depicted as spheres.

The B-factors for these structures are unusually high at 115.44 and 97.15 Å^2 for T1-NMM and T7-NMM, respectively (**Table S4**). Loop flexibility and disorder drives up these B-factors, as the B-factor for loop regions is 148.13 Å^2 for T1 and 128.85 Å^2 for T7. In addition, the 5' and 3' thymine overhangs in T7 (these nucleotides are absent in T1) are highly disordered, with a B-factor of 151.63 Å^2. The flexibility of these terminal thymines

can explain the observed destabilization of T7 with respect to T1 by -5.8 °C, the greatest destabilization among all the constructs (**Table 3**).

In the crystalline state, T1 and T7 form parallel GQs consistent with solution CD data (**Figure 8D**). The GQs dimerize via their 5' G-tetrads, and this dimer is capped on both 3' ends by NMM. This arrangement is consistent with the 1:1 stoichiometry observed in UV-vis, FL, and Job plot experiments. A similar overall arrangement of GQs and NMM is observed in the crystal structure of human telomeric DNA, Tel22, with NMM (PDB ID: 4FXM).[59] The GQ dimers are stabilized by five K^+ ions, all in square antiprismatic coordination of the carbonyl oxygen from the eight guanines in the two adjacent G-tetrads. Four K^+ ions are found between G-tetrads within GQ monomers, while one K^+ is found at the dimer interface.

The dimer in T7-NMM is further stabilized by base stacking between T_5 and T'_{15} of chain B (where the prime (′) notation signifies a symmetry-related molecule). There may be interactions between T_5 of chain A and either T_{20} of chain B or T'_{15} of chain A. However, the density for these two nucleotides is poor and neither nucleotide was fully built. Meanwhile, in T1-NMM, there may be stacking between T_4 in chain A and T_{14} in chain B, but the electron density for T_{14} is poor so its base was not built.

All guanines in the structures adopt the *anti* glycosidic bond conformation, as expected for the parallel GQ topology. The G-quartets are spaced by 3.4 Å, allowing for optimal π–π base stacking interactions (**Table S5**). The average intramolecular helical twist between G-quartets within each GQ monomer is 29 ± 4° for T1-NMM and 28 ± 1° for T7-NMM (**Table S6A**), consistent with the helical twist reported for GQs with a parallel topology[72,73] and for the Tel22-NMM structure (29 ± 1°).[59] The two monomers are twisted

with respect to each other by 116 ± 1°, 117 ± 2° (**Table S6B**), and 120 ± 3°, for T1-NMM, T7-NMM and Tel22-NMM, respectively. The DNA backbone torsional angles are plotted in **Figure S6** and are consistent with those reported for other parallel GQ structures.[73]

Loop arrangement in the T1-NMM and T7-NMM structures

The T1 and T7 sequences have three propeller TT loops. In any loops with clear enough electron density to fully build both thymines, one thymine is tucked into the groove, while the other points out toward the solvent. The structures are type VIII GQs, which have a parallel conformation with four medium grooves.[74] Specifically, in the T7 structure, these grooves have widths of 16.2 ± 0.5, 16.0 ± 0.4, 16.2 ± 0.6, and 16.0 ± 0.6 Å (**Table S7**). Interestingly, the 5' G-quartet – located at the dimer interface – has narrower grooves than the 3' and middle G-quartets (15.5 ± 0.2 Å as opposed to 16.2 ± 0.2 and 16.6 ± 0.3 Å, respectively).

Another structure in the PDB with a TT loop is that of $(TTGGGG)_4$ in Na^+ (PDB ID: 186D).[75] As in our case, the propeller TT loop spans the medium groove (11.8 Å width). One of the thymines is tucked into the groove, interacting with two guanines, while the other is oriented into the solvent. One other parallel GQ structure has a TT loop (PDB ID: 5DWW).[76] In this structure, both thymines in the loop point into the solvent, and one forms intermolecular contacts with one of the T loops.

Atomic details of the interaction between T7 and NMM

For both structures, the central *N*-Me group of NMM is not well resolved, although there is electron density visible for it. The propionic acid side chains are not visible and

were not modeled. However, the macrocycle of NMM is well defined. NMM binds to the 3' terminal G-quartet at a distance of 3.6 Å in both the T1- and T7-NMM structures (**Table S5**), as well as in the Tel22-NMM complex.[59] This distance is only slightly longer than the optimal π-π stacking distance of 3.4 Å. This longer calculated distance may be partially due to the non-planarity of the bound NMM, as distortions in the porphyrin ring may skew its centroid, which was used to calculate distances. Nearly planar GQ ligands are typically located 3.2-3.6 Å from the terminal G-tetrad. Examples include (PDB IDs in parentheses) berberine (3R6R), acridines (3NZ7[77], 1L1H[78], and 3EM2[79]), daunomycin (3TVB[80]), salphen metal complexes (3QSC and 3QSF[81]), a ruthenium polypyridyl complex (5LS8[82]), a dicarbene gold complex (5CCW[83]), and naphthalene diimides (3SC8 and 3T5E[84]).

The G-quartets become increasingly non-planar when moving away from the 5'-5' dimer interface. In this order, the out-of-plane deviations for the G-quartets in T7 are 0.39 ± 0.06, 0.97 ± 0.02, and 2.06 ± 0.06 Å, which are similar to those observed in the Tel22-NMM structure (**Table 5**). NMM itself is also non-planar in these structures, which may allow it to bind more tightly and selectively to these GQs than a nonplanar GQ ligand.

Table 5. Out-of-plane deviations (Å) for G-quartets. Calculated for the T1-NMM, T7-NMM, and Tel22-NMM structures.

Quartet	T1-NMM	T7-NMM	Tel22-NMM[59]
5'	0.40 ± 0.04	0.39 ± 0.06	0.49
Middle	0.5 ± 0.2	0.97 ± 0.02	1.08
3'	1.72 ± 0.09	2.06 ± 0.06	1.85

On its other face, NMM interacts with T_{10} from an adjacent DNA chain (**Figure 12A-B**). This interaction leads to an interesting intermolecular assembly of T7-NMM complexes (**Figure 12C**) that can explain the resulting P 6_3 space group and the hexagonal shape of the T7-NMM crystals (**Figure 11A**). However, it is likely that this NMM-thymine

interaction, observed for both T1-NMM and T7-NMM, is a result of crystal packing forces. The intermolecular assembly for T1-NMM is similar and can be found, along with its asymmetric unit, in **Figure S7**.

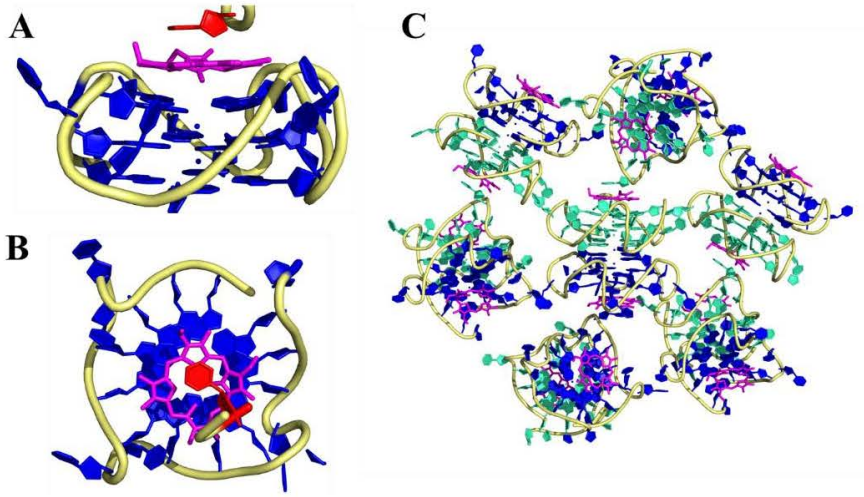


Figure 12. Intermolecular interactions in the T7-NMM crystal structure. (**A**) NMM (magenta) binds a GQ monomer (blue) via π-π stacking on one side and stacking with T_{10} from another GQ monomer (red) on the other side. (**B**) Same as in (A) but rotated by 90°. (**C**) Intermolecular interactions among T7 and NMM molecules. T7 chains are depicted in teal and blue.

Comparison of samples used for crystallization with those prepared for biophysical studies

To verify that the crystal structures are biologically relevant and that the high sample concentrations used for crystallization did not affect DNA-ligand complex folding, we compared the CD signatures of samples prepared for crystallographic studies with those prepared for biophysical studies. The similarity of these CD signatures (**Figure S8A**) indicates that the crystal structures are good representations of the GQ-NMM complexes found in solution. According to PAGE (**Figure S8B**), T1 exists as a mix of monomer and dimer even in the crystallization sample, consistent with that observed at lower concentrations. Meanwhile, T7 is a monomer alone as well as in complex with NMM at

both high and low concentrations. This data suggests that the T1 dimer observed in the crystal structure could exist under physiological conditions. However, the T7 crystallographic dimer is likely an artifact of crystal packing, wherein the thymine overhangs – normally freely moving in solution and preventing dimerization – are tucked into the grooves, thereby opening the 5' G-tetrad to dimerization.

Discussion

We conducted biophysical and crystallographic studies on the interaction of two G-rich DNA sequences, $(GGGTT)_3GGG$ (T1) and $T(GGGTT)_3GGGT$ (T7), with a highly selective porphyrin ligand, *N*-methyl mesoporphyrin IX (NMM). T1 and T7 fold into parallel GQs in solution both alone and in complex with NMM, as well as when bound to NMM in the crystal structures. This observed fold is consistent with a previous study that examined the fold of a variety of DNA with the general sequence $G_3T_nG_3T_nG_3T_nG_3$, where n = 1-5.[85] Short loops with 1-2 Ts, as in our case, lead to the parallel GQ conformation. Meanwhile, loops with 3 Ts result in a hybrid GQ and those with 4-5 Ts lead to an antiparallel GQ fold.

In solution, T1 exists mostly as a dimer of parallel GQs. Addition of thymines at the 5' and 3' ends of T1 (as in T7) fully breaks the dimer, but maintains the parallel GQ fold. Meanwhile, addition of 2 eq. of NMM partially breaks the dimer according to PAGE (**Figure 8C**). We also examined other extended native constructs of T1, T2 – T6. The consistency of their CD and TDS signatures, as well as stabilities, to that of T1 suggest that our findings for the T1-NMM and T7-NMM complexes are applicable to longer DNA sequences with greater biological significance. Meanwhile, extending T1 with non-native

nucleotides (as in T7 – T9) leads to some decrease in stability but preserves the overall GQ fold. The addition of 5' and 3' overhangs fine-tunes GQ stability without significantly affecting the structure. We arrived at a similar conclusion when examining the crystal structure of the 19wt GQ from *Dictyostelium discoideum.*[86]

NMM binding to both T1 and T7 is an enthalpically-driven, thermodynamically favorable process with ΔH of -17 and -8.7 kcal/mol and ΔG of -10. and -8.7 kcal/mol, respectively. The difference in enthalpy between the two GQ-NMM complexes may be compensated by the entropic contribution from the flexible thymine tails of T7 to yield similar ΔG values. Their binding affinity is high, and the K_a of ~50 μM^{-1} (K_d of ~20 nM) for T1-NMM is the highest reported for any GQ-NMM complex.[45] The weaker binding of NMM to T7 as compared to T1 can be explained by the presence of terminal thymines that obstruct the binding surface at the 3' terminal G-tetrad. The lab has determined the binding constants NMM to a variety of GQs using UV-vis, FL, and CD titrations,[43,44,87] and we recently reviewed others' work with GQ-NMM binding as well.[45] Compared to all known GQ-NMM interactions thus far, the binding of NMM to T1 is unprecedentedly tight.

This impressively tight binding is especially intriguing given that NMM does not display the characteristic features of a typical GQ ligand – planarity and a cationic nature. Instead, NMM is non-planar, with an observed out-of-plane deviation of ~0.73 Å[59] and is likely negatively charged under physiological pH. The non-planarity of NMM may be a key factor for its excellent binding affinity to parallel GQs, as it can match the non-planarity of the terminal G-tetrad. Our lab previously showed (through a combination of the Tel22-NMM crystal structure and the energy-minimized structure of free NMM) that the degree of NMM distortion can vary based on its binding partner or the lack thereof.[59]

Furthermore, the tight binding of NMM to T1 is coupled with an impressive stabilizing ability (ΔT_m > 17 °C at 2 eq. of NMM for T1, T7, and T8). Our work demonstrates that NMM, in spite of its non-planarity and negative charge, is an excellent GQ binder when presented with the right GQ, as in the case of T1. Although the field is in search of planar, cationic GQ binders, it is possible that ligands with some degree of non-planarity, or even better, the ability to alter their shape (non-planarity) for an induced fit to their binding partner may allow for both tight binding to particular GQs and selectivity against dsDNA.

The T1 and T7 sequences are not only found in the *T. thermophila* telomere, but also in different parts of the human genome (including a gene involved in Alzheimer's disease),[88] as well as in the genome of at least 34 bacteria (including pathogenic species).[65,66] Our work identifies an unprecedentedly tight binding interaction between T1 and the small molecule ligand NMM and presents the crystal structure of this GQ-ligand complex solved at 2.39 Å, revealing the end-stacking binding mode of NMM. This work expands the small, but growing library of GQ-ligand crystal structures and provides atomic level information about GQ and ligand structural features that promote strong binding. Our work also demonstrates how such binding can be detected and characterized thoroughly via biophysical methods. The atomic coordinates reported here can be used to computationally search for even better drug candidates using available ligand libraries, with the hope that GQ binders may one day serve as anticancer therapeutics.

Acknowledgements: *Barrett Powell '18 assisted with initial model building for T7-NMM. Yanti Manurung '20 ran a few UV-vis/FL titration and Job plot trials. Sawyer McCarthy ran a few CD/TDS trials with NMM. Irene Xiang '18 assisted with the buffer CD study.*

Chapter 4: Biophysical characterization of a four-quartet G-quadruplex from the *Dictyostelium Discoideum* genome

Our lab solved the crystal structure of a four-quartet GQ formed by a G-rich sequence, 19wt, from the *Dictyostelium Discoideum* genome (**Figure 13A-B**). It was the fourth four-quartet monomolecular GQ crystal structure ever published (PDB ID: 6FTU).[86] Since this structure is of an unusually large GQ, understanding its structural details helps us better understand the structural diversity of GQs. However, as GQ conformation may be impacted in during crystallization,[89] it was important to determine if this crystal structure is representative of DNA conformation in solution. Towards this goal, I analyzed the crystal structure to both compare it to and place it in context with other published GQ structures. I then characterized 19wt in solution to verify that it folds into a GQ with an antiparallel conformation as observed in the crystal structure. Finally, I examined the fold and stability of various mutants designed based on loop interactions observed in the crystal structure to assess the role of loop nucleotides in the overall GQ structure.

The work presented in this chapter was published in *Nucleic Acids Research* in 2018.[86] The paper, which can be found in the **Appendix**, includes further discussion of the crystal structure and NMR data on the loop mutants.

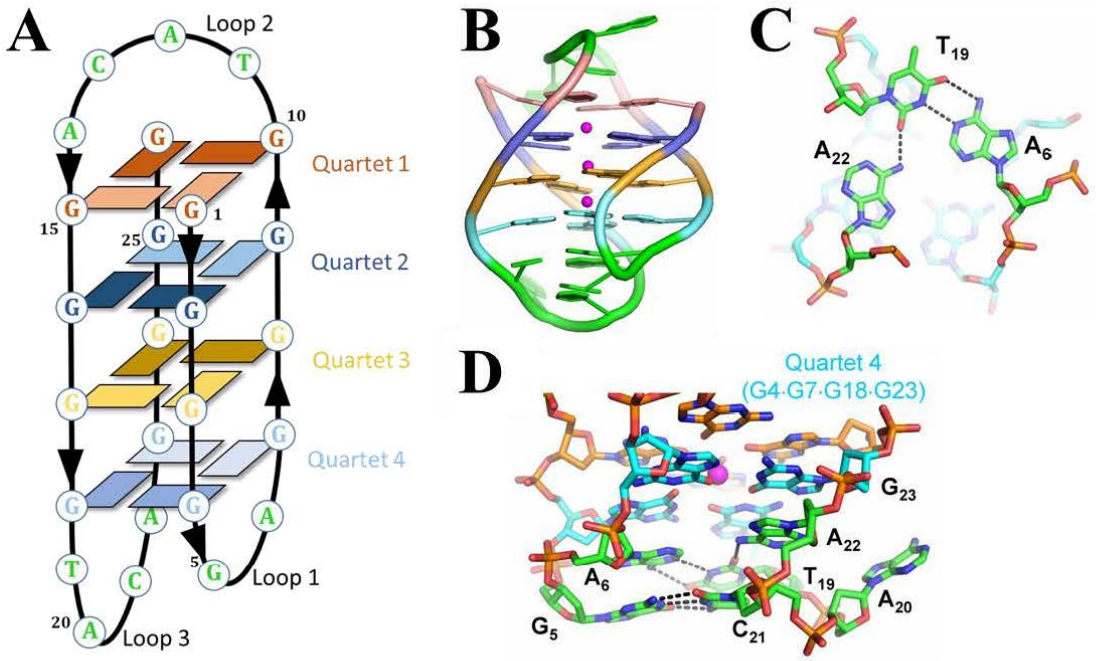


Figure 13. 19wt crystal structure and observed loop interactions. (**A**) Schematic representation of the GQ with each quartet in a different color and loop nucleotides in green. The *anti* and *syn* guanine conformations are indicated by dark and light rectangles, respectively. (**B**) Graphical representation with the same coloring as in (A), with K^+ depicted as magenta spheres. (**C**) The A_6–T_{19}–A_{22} triad. (**D**) Stacking between the G_5–C_{21} Watson-Crick base pair, A_6–T_{19}–A_{22} triad, and Quartet 4 (cyan). Hydrogen bonds are shown as black dashed lines. Adapted from the paper.[86]

Analysis of the crystal structure

We started by analyzing the crystal structure itself so we could compare it to other published GQ crystal structures. First, we examined the distribution of DNA torsional angles (**Figure 7B**) in the 19wt crystal structure (**Figure S9**). The distribution was similar to that reported for other GQ structures,[73] suggesting that although the 19wt GQ is very large, the torsion of the underlying DNA backbone and glycosidic bonds is similar to that observed for other GQs. Moreover, given that GQs have single-stranded loop regions (loop 2 in particular in 19wt, **Figure 13A**) that result in local flexibility, whereas in double-

stranded DNA all nucleotides are engaged in defined base pairs, it is reasonable that we observe a wider distribution of torsional angles for 19wt than for canonical duplex DNA.

Next, we calculated the helical twist between each pair of adjacent G-quartets in the 19wt GQ. The helical twist between Quartets 1–2, 2–3 and 3–4 is 91 ± 6, 145 ± 3 and 89 ± 4°, respectively (**Table S8**). These values are similar to those we calculated for the *Oxytricha nova* telomeric GQ (PDB ID: 1JPQ), which also adopts an antiparallel fold: 95 ± 1, 149 ± 1 and 94 ± 1° for Quartets 1–2, 2–3 and 3–4, respectively. Like canonical B-DNA, all known GQ structures have a right-handed helical twist (with one exception),[62] and the 19wt GQ follows that rule.

These analyses demonstrate that the large, four-tetrad 19wt crystal structure is consistent with other (antiparallel) GQ structures in terms of these summary metrics. Furthermore, although other antiparallel GQ structures have been published in the past, none have been as thoroughly characterized in terms of both torsional angles and helical twist to our knowledge. Accordingly, our analyses inform our understanding of the types of torsional angles and helical twists that antiparallel GQs and/or four-tetrad GQs, as opposed to other GQ conformations, may adopt. These analyses thereby contribute to our knowledge of GQ structural diversity, which is important when designing drugs to selectively target specific GQs.

Fold and stability in different buffers

To determine if the crystal structure of 19wt is similar to its solution state conformation, we studied 19wt in buffers with either Na^+ or K^+ at varying concentrations to examine how its fold and stability would be affected. We studied 19wt with different

cations because K^+ is better at stabilizing GQ structures than Na^+, as its ionic radius allows for better coordination to G-quartets.[90] Consequently, we expected 19wt, if it formed a GQ, to be more stable in K^+ than in an equivalent (or even higher) concentration of Na^+.

First, we assessed the thermodynamic stability of 19wt to see if it is consistent with our expectations for a four-tetrad GQ. Given that a four-tetrad GQ has more π-π stacking and hydrogen bonding interactions than smaller GQs, we expected 19wt to be substantially more stable than a typical GQ under the same conditions. CD melting experiments demonstrated that 19wt has exceptionally high thermodynamic stability, consistent with our expectations. In particular, its melting temperature is greater than 90°C at near physiological ionic conditions (100 mM K^+) and remains above 90°C even when K^+ concentration is lowered to 10 mM. Accordingly, to assess differences in behavior in the presence of different cations, we studied 19wt in 5 mM K^+, as well as in 5 or 50 mM Na^+.

Next, we sought to confirm that 19wt indeed forms a GQ in solution, and that it specifically folds into an antiparallel GQ. The TDS and CD signatures of 19wt in K^+ and Na^+ (**Figure 14A-B**) were quite similar, indicating that 19wt adopts a similar fold in the presence of either cation. TDS displayed the 295 nm trough characteristic of GQs (**Figure 14A**), thereby confirming that 19wt folds into a GQ structure in solution. Meanwhile, CD scans included a peak at 295 nm and trough at 260 nm (**Figure 14B**), which suggest an antiparallel GQ conformation in agreement with the crystal structure.

The main differences for 19wt behavior in different buffers were observed in CD melting experiments (**Figure 14C**). Consistent with expectations, the melting temperature of 19wt is highest in 5 mM K^+ (76.9 ± 0.5°C), lower in 50 mM Na^+ (64.1 ± 0.3°C), and even lower in 5 mM Na^+ (43.8 ± 0.4°C). The dramatic 33.1°C decrease in stability upon

changing the cation from K^+ to Na^+ while keeping the concentration constant at 5 mM further supports 19wt folding into a GQ in solution. Finally, the low hysteresis values (2.0 to 6.4 °C) observed reveal that 19wt undergoes mostly reversible melting transitions in both K^+ and Na^+ buffers.

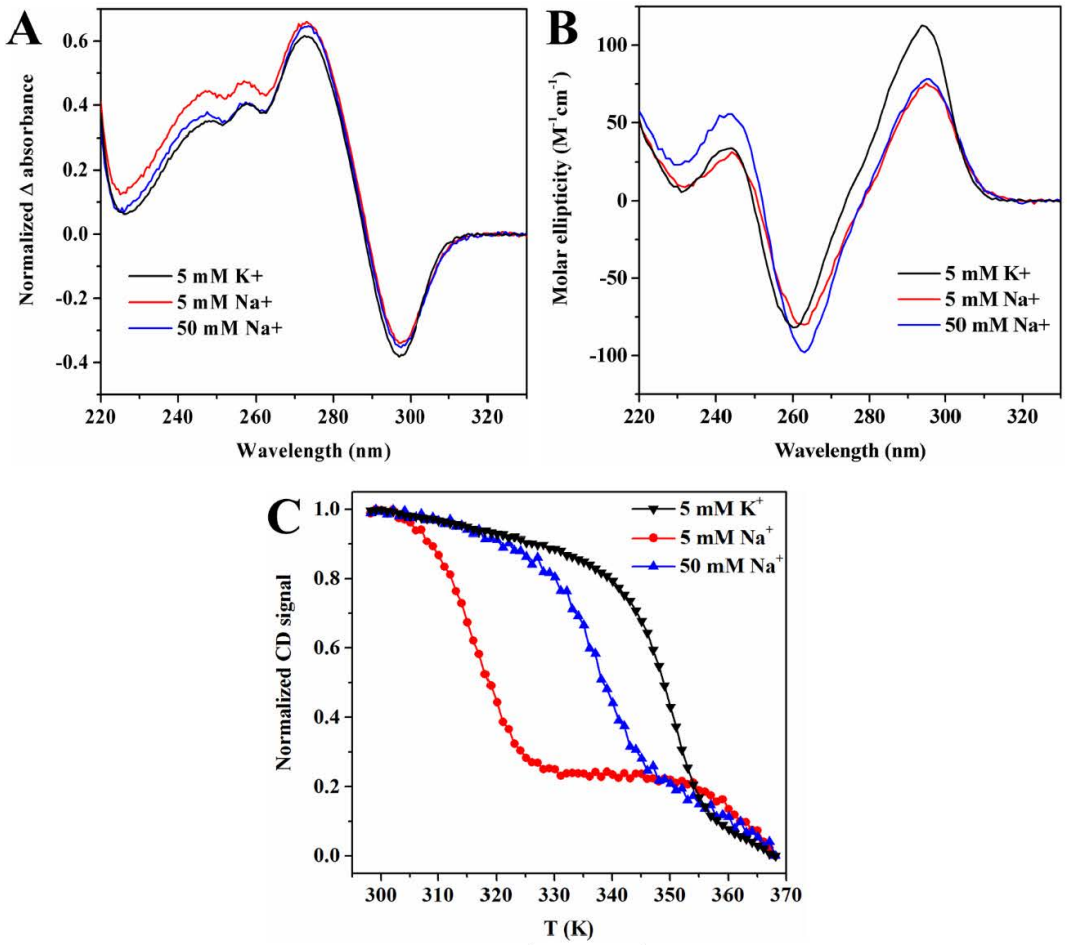


Figure 14. 19wt folding and stability in K^+ and Na^+ buffers. (**A**) TDS, (**B**) CD spectra, and (**C**) CD melting studies monitoring signal at 295 nm.

Characterization of loop mutants

Loop mutant design and initial characterization

In the 19wt crystal structure, in addition to the central four-tetrad GQ core, there is a G_5–C_{21} Watson-Crick base pair, A_6–T_{19}–A_{22} triad, and π-π stacking between C_{13} and

Quartet 1 (**Figure 13**). These loop nucleotide interactions may help with 19wt folding and/or further stabilize the GQ structure. After careful analysis of the 19wt crystal structure, we designed constructs (M1-M8, **Table 1**) with mutations at nucleotides involved in loop interactions within the structure. The selected mutations, along with the loop interactions they are intended to potentially disrupt, are detailed in **Table 6.** By examining the fold and stability of these mutants, we could test whether the targeted loop interactions are essential for GQ formation and thereby potentially biologically relevant, or if they merely result from crystal packing. Notably, if the solution state structure of 19wt is similar to its crystal structure, then these loop mutants should have a measurable effect on the GQ.

Table 6. Interactions tested in loop mutants and thermodynamic parameters. T_m was measured via CD melting experiments. *Note the possibility of forming a new T_5-A_{21} Watson-Crick base pair in place of the G_5-C_{21} base pair.

DNA	Mutation(s)	Potentially disrupts	T_m (°C)	ΔT_m (°C)	Hysteresis (°C)
19wt	None	—	76.7 ± 0.6	—	6.4
M1	G5T	G_5-C_{21}	72.3 ± 0.4	-4.4	4.2
M2	G5T–A6T	G_5-C_{21}, A_6-T_{19}	78.5 ± 1.0	1.8	6.1
M3	G5T–A6T–A22T	G_5-C_{21}, A_6-T_{19}-A_{22}	71.3 ± 0.4	-5.4	4.7
M4	C21A	G_5-C_{21}	76.1 ± 0.5	-0.6	7.3
M5	G5T–C21A	G_5-C_{21} (both nt)*	73.0 ± 0.4	-3.7	4.7
M6	G5A–C21A	G_5-C_{21} (both nt)	69.6 ± 1.6	-7.1	6.7
M7	G5T–C21T	G_5-C_{21} (both nt)	74.1 ± 0.9	-2.6	4.5
M8	C13A	π-π stacking of C_{13} with Quartet 1	76.4 ± 0.3	-0.3	6.2

We first assessed the loop mutants via PAGE (**Figure 15A**) to determine if any of the mutations affect the overall fold of 19wt. The 19wt DNA runs as a single band that is faster than the dT_{15} marker, indicating that it is conformationally homogenous and folds into a single compact, monomolecular structure. All of the mutants have similar mobility

to 19wt, which suggests that none of the introduced mutations affect the overall ability of 19wt to fold into a compact secondary structure.

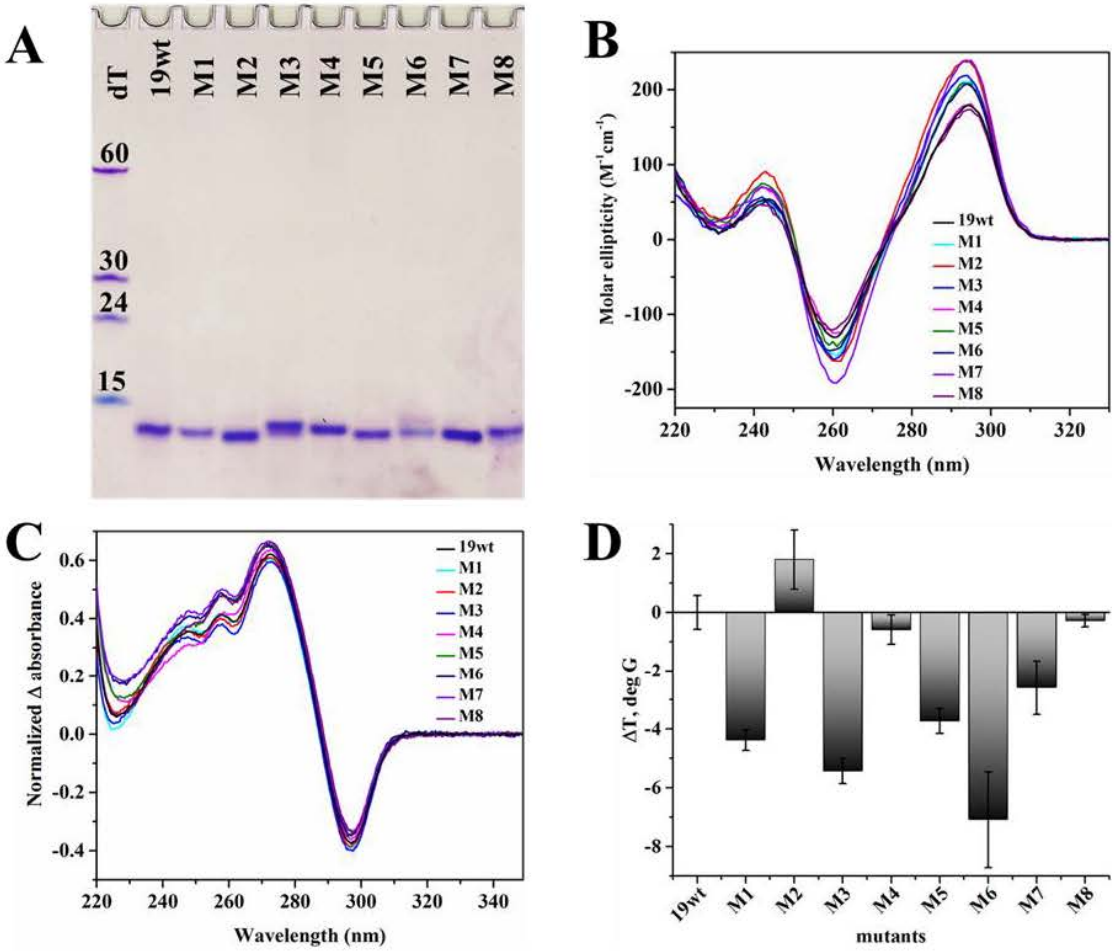


Figure 15. Biophysical characterization of 19wt loop mutants. (**A**) Non-denaturing 15% PAGE, (**B**) CD spectra, (**C**) TDS, and (**D**) ΔT_m relative to 19wt determined via CD melting studies.

CD scans and TDS (**Figure 15B-C**) further support this conclusion, as the signatures for 19wt and its loop mutants are very similar. Taken together, they demonstrate that all the mutants fold into antiparallel GQs like 19wt. Therefore, the loop interactions disrupted by the mutations do not substantially contribute to the antiparallel GQ fold of 19wt. This finding matched our expectations because guanines within the GQ core are the

major contributors to GQ formation, as the Hoogsteen hydrogen bonding between guanines within each G-quartet and the π-π stacking interactions between adjacent G-quartets are greatly stabilizing interactions that facilitate GQ formation.[73]

Loop mutant stability

We then examined the stability of the loop mutants via CD melting studies (**Figure 15D** and **Table 6**). All of the mutants were designed to disrupt at least one loop interaction, as outlined in **Table 6** and thereby potentially destabilize the GQ if the solution state structure of 19wt is similar to its crystalline form. However, given that a mutation may introduce a new stabilizing interaction while abrogating an old one, it was possible that a mutant may have an unintended effect.

All but one mutant was found to be less stable than 19wt (**Figure 15D** and **Table 6**), which was mostly in agreement with our expectations. The sole exception was M2, which was more stable by 1.8 °C. Its unexpected stabilization may be explained by a new, compensatory T_6–A_{22} base pair that can form in place of the disrupted A_6–T_{19} base pair. Similarly, the single mutant M4, which was only slightly less stable than 19wt (by 0.6 °C), may form a new, compensatory G_5–A_{21} non-canonical base pair to maintain the stability of the GQ. In the latter case, this compensatory base pair explanation is supported by ^{1}H NMR data,[86] which reveals an imino peak at ~12.2 ppm for M4 despite disruption of the G_5–C_{21} base pair by the C21A mutation. The only other slightly less stable mutant (by 0.3 °C) was the single mutant M8, which disturbs the π-π stacking interaction between C_{13} and Quartet 1. The lack of destabilization observed with this mutation indicates that this π-π stacking interaction is not a large contributor to the stability of the 19wt GQ.

All other loop mutants (M1, M3, and M5-M7) destabilized the 19wt GQ by 2.6 to 7.1 °C. This is a substantial amount given that one canonical Watson-Crick A–T base pair contributes ~2 °C to the stability of a short duplex while a G–C base pair contributes ~4 °C, as a benchmark for comparison. However, these mutations are not drastically destabilizing either, suggesting that the GQ core retains most of its stability via π-π stacking interactions while these loop nucleotides serve to fine tune GQ stability. All of these destabilizing mutants disrupt the G_5–C_{21} Watson-Crick base pair, thereby demonstrating that this base pair is an important contributor to 19wt structural stability. The triple mutant M3, which disrupts both the G_5–C_{21} base pair and the A_6–T_{19}–A_{22} triad and also prevents new base pairing interactions between loops 1 and 3 from forming, was one of the least stable mutants (ΔT_m = -5.4°C) as expected. Given that M3 is the only mutant that potentially disrupts the A_6–T_{19}–A_{22} triad but it may affect the G_5–C_{21} base pair as well, its substantial destabilization suggests that the triad may be important for 19wt GQ stability, but we lack the data to conclusively say so. Finally, both 19wt and its loop mutants display a low hysteresis (between 4.7 and 7.3 °C, **Table 6**), which reveals that they undergo mostly reversible melting transitions.

In summary, our biophysical studies on the 19wt loop mutants suggest that loop interactions positively contribute to and fine tune GQ stability, although they are not essential for GQ folding. The G_5–C_{21} base pair is particularly important for 19wt stability, while the A_6–T_{19}–A_{22} triad may be as well. Meanwhile, the π-π stacking interaction between C_{13} and Quartet 1 does not greatly contribute to structural stability.

Through characterizing these 19wt loop mutants, we assessed the effects of disrupting loop interactions predicted by the 19wt crystal structure. As we mainly observed

destabilization of loop mutants relative to 19wt, in agreement with our expectations, we can conclude that the 19wt crystal structure represents the major DNA conformation found in solution, thereby rendering it biologically relevant.

Additionally, when probing interactions between specific loop nucleotides in our mutant study, we discovered that alteration of such interactions may lead to the formation of (canonical or non-canonical) compensatory base pairs. Through doing so, we furthered our understanding of loop nucleotide interactions within the GQ, which is important for GQ-targeting drug design because the loop regions of GQs are potential sites of specific ligand recognition Figure **5C/D**).

Chapter 5: Efforts toward solving other non-canonical DNA crystal structures

After successfully solving the T1-NMM and T7-NMM crystal structures and working with the 19wt crystal structure to demonstrate that it is similar to 19wt conformation in solution, we focused our efforts on elucidating other non-canonical DNA structures. Namely, we aim to solve the crystal structure of T1 in complex with a different promising GQ ligand, RHPS4 (**Figure 6**), to determine how its binding mode may differ from that of NMM. Additionally, we hope to solve the crystal structure of iHRAS as part of a venture into another realm of exciting non-canonical DNA structures – i-motifs – which we know much less about than GQs. Although both of these projects are still works in progress, we ultimately hope to further contribute to our understanding of non-canonical DNA structures at the atomic level through them.

Structure of T1 in complex with RHPS4

The T1-RHPS4 project was initiated by Irene Xiang '18, who ran a PAGE gel of T1 with various GQ ligands and saw that T1 forms a clean, homogenous band in complex with RHPS4 (more defined than the T1-NMM band on the same gel). Subsequently, she proceeded to screen T1-RHPS4 for crystallization using the commercial screen HELIX (Molecular Dimensions) and identified the 1-31 condition (1.0 M sodium formate, 20% PEG 20000, and 0.05 M Bis-Tris pH 7.0) as the most promising crystallization hit (**Figure 16A**). Notably, this is the same screen condition that yielded the initial T1-NMM crystals

as well, indicating that the HELIX 1-31 condition may be particularly conducive to crystallization of T1-ligand complexes. I picked up this project after she graduated.

Figure 16. T1-RHPS4 crystals. (**A**) Initial crystallization hit by Irene Xiang '18 in HELIX condition 1-31. (**B**) Optimized crystals that yielded one of the best datasets. (**C**) Looped crystal being shot with an X-ray beam at the synchrotron. Crystals are shown at approximate relative sizes.

I performed extensive optimization of the crystallization conditions to obtain much larger, more geometric, and more reproducible crystals. This optimization greatly improved crystal quality, as initial hits and crystals from early on in the optimization process barely diffracted (or even not at all) when shot with an X-ray beam at the synchrotron. After the extensive optimization process, we were able to reproducibly obtain large, prismatic, diffraction-quality crystals, an example of which is shown in **Figure 16B**. The optimized crystallization conditions for T1-RHPS4 are 0.8-1.1 M sodium formate, 11-16% PEG 20000, and 0.05 M Bis-Tris pH 7.0. Crystallization samples were prepared by annealing 1.0 mM T1 with 1 eq. of RHPS4 in 50 mM KCl and 10 mM lithium cacodylate pH 7.2. Drops were set manually at 1 µL sample and 1 µL crystallization condition in 24-well crystal trays with 400 µL of the crystallization condition in the reservoir.

Notably, however, these crystals take a long time to grow, ranging from approximately one month to over a year at room temperature. Efforts to speed up crystallization by increasing sample, salt, and/or precipitant concentration consistently yielded lower-quality crystals or no crystallization at all. Often, small round crystals or

medium-sized rod crystals will grow quickly on the order of days or weeks (small, round crystallization nuclei are seen surrounding the large crystals in **Figure 16B**), but these crystals are usually not of diffraction-quality. Over time, large, diffraction-quality crystals may grow in the same well. Since the base HELIX 1-31 condition is a cryoprotectant, T1-RHPS4 crystals often do not need to be cryoprotected and can be directly looped and flash-frozen in liquid nitrogen (**Figure 16C**). However, when needed, they are generally tolerant of cryoprotectants consisting of increased PEG, glycerol, and/or ethylene glycol.

We have collected very nice datasets of T1-RHPS4, the best of which are summarized in **Table S9**. The highest resolution dataset is at 2.19 Å, although the best overall dataset has a resolution of 2.32 Å. A frequent problem with T1-RHPS4 crystals is that collected datasets often have high R_{merge} values, indicating low agreement among measurements of the same reflection. It appears that older, larger, and more geometric crystals are more likely to avoid this issue. All datasets presented in **Table S9** have acceptable R_{merge} values of less than or equal to 10%.

We are trying to solve the phase problem for the T1-RHPS4 crystal structure via molecular replacement because T1 from the T1-NMM crystal structure should theoretically be a great search model. Since the T1-NMM asymmetric unit chains two DNA chains, we have tried to run MR (via PHENIX Phaser) using an individual chain or a combination of the two separated chains. However, T1-RHPS4 crystals have an extremely large unit cell (~$35 \times 61 \times 256$ Å^3 or ~$71 \times 223 \times 256$ Å^3 depending on the space group), resulting in a large asymmetric unit consisting of 10 distinct DNA chains (via PHENIX Xtriage). Consequently, many MR attempts yield models that visually seem to somewhat fit the electron density, but various portions of the DNA are misaligned and may clash into each

other. So far, we have not observed apparent density for RHPS4 in any MR solutions, although we are certain that the crystals contain RHPS4 due to its distinctive orange color, which is reflected in the crystals themselves (**Figure 16**). The best MR solution obtained so far came from the merged dataset from June 2019 (**Table S9**), and has all the DNA chains lying in continuous electron density with no large clashes between chains (**Figure 17**). However, it refined to a R_{free} of 55.2%, which is rather close to random (~63%).

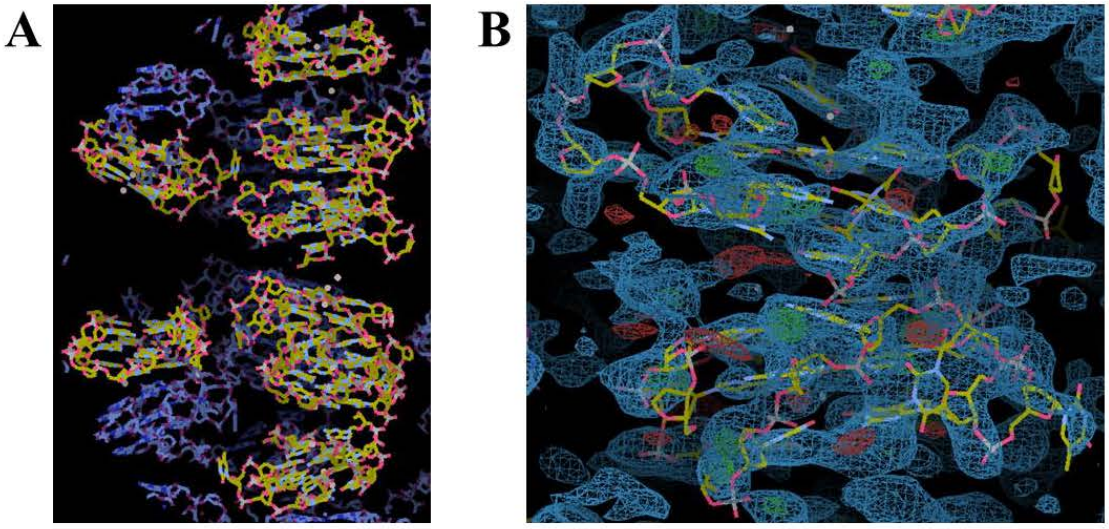


Figure 17. Current best molecular replacement solution for T1-RHPS4. This MR solution has a R_{free} of 56.8%. DNA models are shown as sticks, with symmetry molecules depicted in blue sticks. The electron density map is shown as blue mesh at 1 I/σ, while the FO-FC map is shown as green/red mesh. **(A)** The asymmetric unit, which consists of 10 DNA molecules. (**B**) A close-up view in electron density.

More work needs to be done to find a better MR solution, which will involve trying other datasets (especially the newer ones), models, and search strategies. Afterwards, extensive building and refinement of the 180 nucleotides will need to occur to hopefully arrive at a structure solution. The T1-RHPS4 crystal structure would reveal the binding mode of a different ligand to T1, which will further our understanding of GQ-ligand interactions at the atomic level and can guide GQ-targeting drug design.

The iHRAS i-motif structure

The iHRAS project was started by Deondre Jordan '19, who identified and optimized crystallization hits. Dana Beseiso '21 and Joanne Miao '22 helped to further optimize hits and collect data on the crystals. After Deondre graduated, I worked on solving the iHRAS structure given the datasets collected by the lab.

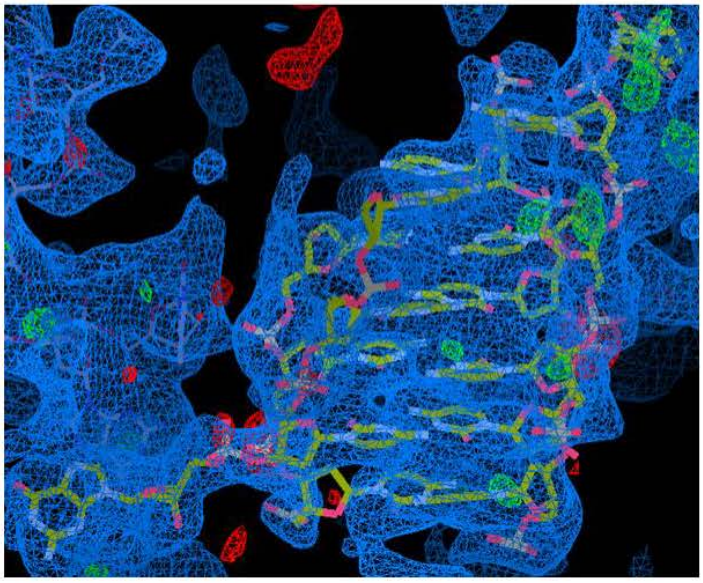
The majority of iHRAS datasets with good statistics are found to have severe anisotropy, as well as occasional severe twinning, when analyzed using PHENIX Xtriage. These are both data quality issues that can render structure solution using these datasets difficult. However, there are two datasets that exhibit only moderate anisotropy and are otherwise sound, which are detailed in **Table S10**. The asymmetric unit for iHRAS contains two DNA chains (via PHENIX Xtriage).

A collaborator performed molecular replacement using one of Deondre's datasets, and I worked with this potential MR solution. Since it only consists of eight separate, but intercalated, CCC segments (four per i-motif), we had to first number the nucleotides in the sequence. We experimented with different possible arrangements given the iHRAS sequence (**Table 1**) and the known intercalated fold. We then proceeded with the sequence numbering that yielded the lowest R_{free} after refinement (50.1%). Afterwards, through manual model building and refinement cycles, I have successfully built seven additional nucleotides (**Figure 18B**), bringing the R_{free} down to 44.1%. The working iHRAS model, which has 31 out of 54 total nucleotides (57%) built, is shown in **Figure 18A**.

I have also tried to do molecular replacement using the i-motif structures with PDB IDs 1CN0,[91] 1BQJ,[91] and 241D[92] with various search procedures using the best dataset (2.02 Å, **Table S10**) to attempt to arrive at a better starting point for model building. I have

not been successful at finding a molecular replacement solution and subsequent sequence numbering that is better than the one discussed above so far. However, there are more numbering arrangements that can be tested, and we can also try doing MR with the 2.27 Å dataset. Additionally, we can continue to build upon the current best model (**Figure 18**).

A



B A: 5′-CG**CCCGT**G**CCC**TGCG**CCC**GCAA**CCCG**A-3′

B: 5′-CG**CCCGT**G**CCC**TGCG**CCCG**CA**ACCC**GA-3′

Figure 18. Working model of the iHRAS crystal structure. (**A**) Refined model with a R_{free} of 44.1%. The DNA model is shown as sticks, while the electron density map is shown as blue mesh at 1 I/σ. The FO-FC map is shown as green/red mesh. (**B**) Schematic of the 31/54 nucleotides (57%) built in the model, with built nucleotides in bold.

There are few published i-motif crystal structures, and the large majority of these consist of tetramolecular or bimolecular structures, which may be less biologically relevant. Therefore, solving the crystal structure of the monomolecular i-motif formed by iHRAS would greatly contribute to our understanding of i-motif structure at the atomic level. Furthermore, it would provide an exciting starting point for an anticancer strategy that targets human RAS expression itself rather than downstream, potentially less druggable components of the MAPK signaling pathway that is activated in many human cancers.[23]

Chapter 6: Conclusion

My thesis work focused on characterizing non-canonical DNA structures via both biophysical and structural methods to inform our understanding of their structural diversity. G-rich, and complementary C-rich, DNA sequences with the potential to form GQs and i-motifs are frequently found at telomeres and oncogene promoters, which are both implicated in disease. These structures may function *in vivo* as regulatory elements, and small molecule ligands that selectively recognize their structural elements can potentially modulate their stability in disease states.

Chapter 3 focused on characterizing the interaction of the T1 GQ with NMM, revealing an unprecedently tight, thermodynamically favorable binding interaction – the tightest reported for any GQ-NMM complex. We turned to crystallography to understand the atomic details of this binding interaction, which revealed the end-stacking binding mode of NMM and adds to the limited number of solved GQ-ligand crystal structures.

Chapter 4 investigated the four-tetrad 19wt GQ to demonstrate that it adopts the same structure in solution as it does in crystalline form. Further biophysical studies of loop mutants suggest that loop interactions fine tune GQ stability, but do not affect GQ folding.

Finally, Chapter 5 presented ongoing work towards solving the crystal structures of the T1 GQ in complex with RHPS4, as well as of the monomolecular iHRAS i-motif.

The work in this thesis contributes to our understanding of GQ-ligand binding interactions and expands our knowledge of quadruplex structural diversity. The crystal structures presented and validated enable a deeper understanding of GQ structural features and interactions, allowing for the design and *in silico* screening of drugs to selectively target them for therapeutic purposes.

References

(1) 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. https://doi.org/10.1016/j.bbrep.2016.01.013.

(2) Rhodes, D.; Lipps, H. J. G-Quadruplexes and Their Regulatory Roles in Biology. *Nucleic Acids Res.* **2015**, *43* (18), 8627–8637. https://doi.org/10.1093/nar/gkv862.

(3) 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. https://doi.org/10.1038/s41557-018-0046-3.

(4) Chambers, V. S.; Marsico, G.; Boutell, J. M.; Di Antonio, M.; Smith, G. P.; Balasubramanian, S. High-Throughput Sequencing of DNA G-Quadruplex Structures in the Human Genome. *Nat. Biotechnol.* **2015**, *33* (8), 877–881. https://doi.org/10.1038/nbt.3295.

(5) Lipps, H. J.; Rhodes, D. G-Quadruplex Structures: In Vivo Evidence and Function. *Trends Cell Biol.* **2009**, *19* (8), 414–422. https://doi.org/10.1016/j.tcb.2009.05.002.

(6) Huppert, J. L.; Balasubramanian, S. G-Quadruplexes in Promoters throughout the Human Genome. *Nucleic Acids Res.* **2007**, *35* (2), 406–413. https://doi.org/10.1093/nar/gkl1057.

(7) Balasubramanian, S.; Hurley, L. H.; Neidle, S. Targeting G-Quadruplexes in Gene Promoters: A Novel Anticancer Strategy? *Nat. Rev. Drug Discov.* **2011**, *10* (4), 261–275. https://doi.org/10.1038/nrd3428.

(8) Maizels, N. G4-Associated Human Diseases. *EMBO Rep.* **2015**, *16* (8), 910–922. https://doi.org/10.15252/embr.201540607.

(9) Wang, Q.; Liu, J.; Chen, Z.; Zheng, K.; Chen, C.; Hao, Y.; Tan, Z. G-Quadruplex Formation at the 3′ End of Telomere DNA Inhibits Its Extension by Telomerase, Polymerase and Unwinding by Helicase. *Nucleic Acids Res.* **2011**, *39* (14), 6229–6237. https://doi.org/10.1093/nar/gkr164.

(10) Smith, J. S.; Chen, Q.; Yatsunyk, L. A.; Nicoludis, J. M.; Garcia, M. S.; Kranaster, R.; Balasubramanian, S.; Monchaud, D.; Teulade-Fichou, M.-P.; Abramowitz, L.; Schultz, D. C.; Johnson, F. B. Rudimentary G-Quadruplex–Based Telomere Capping in *Saccharomyces Cerevisiae*. *Nat. Struct. Mol. Biol.* **2011**, *18* (4), 478–485. https://doi.org/10.1038/nsmb.2033.

(11) Manzini, G.; Yathindra, N.; Xodo, L. E. Evidence for Intramolecularly Folded I-DNA Structures in Biologically Relevant CCC-Repeat Sequences. *Nucleic Acids Res.* **1994**, *22* (22), 4634–4640. https://doi.org/10.1093/nar/22.22.4634.

(12) 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. https://doi.org/10.1021/ja4109352.

(13) Roy, B.; Talukder, P.; Kang, H.-J.; Tsuen, S. S.; Alam, M. P.; Hurley, L. H.; Hecht, S. M. Interaction of Individual Structural Domains of HnRNP LL with the BCL2 Promoter I-Motif DNA. *J. Am. Chem. Soc.* **2016**, *138* (34), 10950–10962. https://doi.org/10.1021/jacs.6b05036.

(14) Miglietta, G.; Cogoi, S.; Pedersen, E. B.; Xodo, L. E. GC-Elements Controlling *HRAS* Transcription Form *i*-Motif Structures Unfolded by Heterogeneous Ribonucleoprotein Particle A1. *Sci. Rep.* **2015**, *5*, 18097. https://doi.org/10.1038/srep18097.

(15) Schaffitzel, C.; Berger, I.; Postberg, J.; Hanes, J.; Lipps, H. J.; Plückthun, A. In Vitro Generated Antibodies Specific for Telomeric Guanine-Quadruplex DNA React with Stylonychia Lemnae Macronuclei. *Proc. Natl. Acad. Sci.* **2001**, *98* (15), 8572–8577. https://doi.org/10.1073/pnas.141229498.

(16) Carle, C. M.; Zaher, H. S.; Chalker, D. L. A Parallel G Quadruplex-Binding Protein Regulates the Boundaries of DNA Elimination Events of Tetrahymena Thermophila. *PLOS Genet.* **2016**, *12* (3), e1005842. https://doi.org/10.1371/journal.pgen.1005842.

(17) Biffi, G.; Tannahill, D.; McCafferty, J.; Balasubramanian, S. Quantitative Visualization of DNA G-Quadruplex Structures in Human Cells. *Nat. Chem.* **2013**, *5* (3), 182–186. https://doi.org/10.1038/nchem.1548.

(18) Kendrick, S.; Hurley, L. H. The Role of G-Quadruplex/i-Motif Secondary Structures as Cis-Acting Regulatory Elements. *Pure Appl. Chem. Chim. Pure Appl.* **2010**, *82* (8), 1609–1621. https://doi.org/10.1351/PAC-CON-09-09-29.

(19) Sissi, C.; Gatto, B.; Palumbo, M. The Evolving World of Protein-G-Quadruplex Recognition: A Medicinal Chemist's Perspective. *Biochimie* **2011**, *93* (8), 1219–1230. https://doi.org/10.1016/j.biochi.2011.04.018.

(20) Dhillon, A. S.; Hagan, S.; Rath, O.; Kolch, W. MAP Kinase Signalling Pathways in Cancer. *Oncogene* **2007**, *26* (22), 3279–3290. https://doi.org/10.1038/sj.onc.1210421.

(21) Membrino, A.; Cogoi, S.; Pedersen, E. B.; Xodo, L. E. G4-DNA Formation in the HRAS Promoter and Rational Design of Decoy Oligonucleotides for Cancer Therapy. *PLoS ONE* **2011**, *6* (9), e24421. https://doi.org/10.1371/journal.pone.0024421.

(22) O'Bryan, J. P. Pharmacological Targeting of RAS: Recent Success with Direct Inhibitors. *Pharmacol. Res.* **2019**, *139*, 503–511. https://doi.org/10.1016/j.phrs.2018.10.021.

(23) Journey, S. N.; Alden, S. L.; Hewitt, W. M.; Peach, M. L.; Nicklaus, M. C.; Schneekloth Jr, J. S. Probing the Hras-1Y i-Motif with Small Molecules. *MedChemComm* **2018**, *9* (12), 2000–2007. https://doi.org/10.1039/C8MD00311D.

(24) Bochman, M. L.; Paeschke, K.; Zakian, V. A. DNA Secondary Structures: Stability and Function of G-Quadruplex Structures. *Nat. Rev. Genet.* **2012**, *13* (11), 770–780. https://doi.org/10.1038/nrg3296.

(25) Spiegel, J.; Adhikari, S.; Balasubramanian, S. The Structure and Function of DNA G-Quadruplexes. *Trends Chem.* https://doi.org/10.1016/j.trechm.2019.07.002.

(26) 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. https://doi.org/10.1093/nar/gkn517.

(27) Nikolova, E. N.; Gottardo, F. L.; Al-Hashimi, H. M. Probing Transient Hoogsteen Hydrogen Bonds in Canonical Duplex DNA Using NMR Relaxation Dispersion and Single-Atom Substitution. *J. Am. Chem. Soc.* **2012**, *134* (8), 3667–3670. https://doi.org/10.1021/ja2117816.

(28) Kaushik, M.; Singh, A.; Kumar, M.; Chaudhary, S.; Ahmed, S.; Kukreti, S. Structure-Specific Ligand Recognition of Multistranded DNA Structures. *Curr. Top. Med. Chem.* **2016**, *16*. https://doi.org/10.2174/1568026616666160530154239.

(29) Lightfoot, H. L.; Hagen, T.; Tatum, N. J.; Hall, J. The Diverse Structural Landscape of Quadruplexes. *FEBS Lett.* **2019**, *593* (16), 2083–2102. https://doi.org/10.1002/1873-3468.13547.
(30) Mukundan, V. T.; Phan, A. T. Bulges in G-Quadruplexes: Broadening the Definition of G-Quadruplex-Forming Sequences. *J. Am. Chem. Soc.* **2013**, *135* (13), 5017–5028. https://doi.org/10.1021/ja310251r.
(31) Lim, K. W.; Phan, A. T. Structural Basis of DNA Quadruplex–Duplex Junction Formation. *Angew. Chem. Int. Ed.* **2013**, *52* (33), 8566–8569. https://doi.org/10.1002/anie.201302995.
(32) Benabou, S.; Aviñó, A.; Eritja, R.; González, C.; Gargallo, R. Fundamental Aspects of the Nucleic Acid I-Motif Structures. *RSC Adv.* **2014**, *4* (51), 26956–26980. https://doi.org/10.1039/C4RA02129K.
(33) 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. https://doi.org/10.1093/nar/gky735.
(34) Jonchhe, S.; Shrestha, P.; Ascencio, K.; Mao, H. A New Concentration Jump Strategy Reveals the Lifetime of I-Motif at Physiological PH without Force. *Anal. Chem.* **2018**, *90* (5), 3205–3210. https://doi.org/10.1021/acs.analchem.7b04661.
(35) Mir, B.; Serrano, I.; Buitrago, D.; Orozco, M.; Escaja, N.; González, C. Prevalent Sequences in the Human Genome Can Form Mini I-Motif Structures at Physiological PH. *J. Am. Chem. Soc.* **2017**, *139* (40), 13985–13988. https://doi.org/10.1021/jacs.7b07383.
(36) 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. https://doi.org/10.1093/nar/gkx090.
(37) Cui, Y.; Kong, D.; Ghimire, C.; Xu, C.; Mao, H. Mutually Exclusive Formation of G-Quadruplex and i-Motif Is a General Phenomenon Governed by Steric Hindrance in Duplex DNA. *Biochemistry* **2016**, *55* (15), 2291–2299. https://doi.org/10.1021/acs.biochem.6b00016.
(38) Mergny, J.-L.; Lacroix, L.; Teulade-Fichou, M.-P.; Hounsou, C.; Guittat, L.; Hoarau, M.; Arimondo, P. B.; Vigneron, J.-P.; Lehn, J.-M.; Riou, J.-F.; Garestier, T.; Hélène, C. Telomerase Inhibitors Based on Quadruplex Ligands Selected by a Fluorescence Assay. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (6), 3062–3067. https://doi.org/10.1073/pnas.051620698.
(39) De Cian, A.; Lacroix, L.; Douarre, C.; Temime-Smaali, N.; Trentesaux, C.; Riou, J.-F.; Mergny, J.-L. Targeting Telomeres and Telomerase. *Biochimie* **2008**, *90* (1), 131–155. https://doi.org/10.1016/j.biochi.2007.07.011.
(40) Neidle, S. Human Telomeric G-Quadruplex: The Current Status of Telomeric G-Quadruplexes as Therapeutic Targets in Human Cancer. *FEBS J.* **2010**, *277* (5), 1118–1125. https://doi.org/10.1111/j.1742-4658.2009.07463.x.
(41) Monchaud, D.; Teulade-Fichou, M.-P. A Hitchhiker's Guide to G-Quadruplex Ligands. *Org. Biomol. Chem.* **2008**, *6* (4), 627–636. https://doi.org/10.1039/B714772B.
(42) Sun, Z.-Y.; Wang, X.-N.; Cheng, S.-Q.; Su, X.-X.; Ou, T.-M. Developing Novel G-Quadruplex Ligands: From Interaction with Nucleic Acids to Interfering with Nucleic

Acid–Protein Interaction. *Molecules* **2019**, *24* (3). https://doi.org/10.3390/molecules24030396.

(43) Nicoludis, J. M.; Barrett, S. P.; Mergny, J.-L.; Yatsunyk, L. A. Interaction of Human Telomeric DNA with N- Methyl Mesoporphyrin IX. *Nucleic Acids Res.* **2012**, *40* (12), 5432–5447. https://doi.org/10.1093/nar/gks152.

(44) Sabharwal, N. C.; Savikhin, V.; Turek-Herman, J. R.; Nicoludis, J. M.; Szalai, V. A.; Yatsunyk, L. A. N-Methylmesoporphyrin IX Fluorescence as a Reporter of Strand Orientation in Guanine Quadruplexes. *FEBS J.* **2014**, *281* (7), 1726–1737. https://doi.org/10.1111/febs.12734.

(45) Yett, A.; Lin, L. Y.; Beseiso, D.; Miao, J.; Yatsunyk, L. A. N-Methyl Mesoporphyrin IX as a Highly Selective Light-up Probe for G-Quadruplex DNA. *J. Porphyr. Phthalocyanines* **2019**, 1–21. https://doi.org/10.1142/S1088424619300179.

(46) Leonetti, C.; Scarsella, M.; Riggio, G.; Rizzo, A.; Salvati, E.; D'Incalci, M.; Staszewsky, L.; Frapolli, R.; Stevens, M. F.; Stoppacciaro, A.; Mottolese, M.; Antoniani, B.; Gilson, E.; Zupi, G.; Biroccio, A. G-Quadruplex Ligand RHPS4 Potentiates the Antitumor Activity of Camptothecins in Preclinical Models of Solid Tumors. *Clin. Cancer Res.* **2008**, *14* (22), 7284–7291. https://doi.org/10.1158/1078-0432.CCR-08-0941.

(47) Phatak, P.; Cookson, J. C.; Dai, F.; Smith, V.; Gartenhaus, R. B.; Stevens, M. F. G.; Burger, A. M. Telomere Uncapping by the G-Quadruplex Ligand RHPS4 Inhibits Clonogenic Tumour Cell Growth in Vitro and in Vivo Consistent with a Cancer Stem Cell Targeting Mechanism. *Br. J. Cancer* **2007**, *96* (8), 1223–1233. https://doi.org/10.1038/sj.bjc.6603691.

(48) Eisen, J. A.; Coyne, R. S.; Wu, M.; Wu, D.; Thiagarajan, M.; Wortman, J. R.; Badger, J. H.; Ren, Q.; Amedeo, P.; Jones, K. M.; Tallon, L. J.; Delcher, A. L.; Salzberg, S. L.; Silva, J. C.; Haas, B. J.; Majoros, W. H.; Farzad, M.; Carlton, J. M.; Smith, R. K.; Garg, J.; Pearlman, R. E.; Karrer, K. M.; Sun, L.; Manning, G.; Elde, N. C.; Turkewitz, A. P.; Asai, D. J.; Wilkes, D. E.; Wang, Y.; Cai, H.; Collins, K.; Stewart, B. A.; Lee, S. R.; Wilamowska, K.; Weinberg, Z.; Ruzzo, W. L.; Wloga, D.; Gaertig, J.; Frankel, J.; Tsao, C.-C.; Gorovsky, M. A.; Keeling, P. J.; Waller, R. F.; Patron, N. J.; Cherry, J. M.; Stover, N. A.; Krieger, C. J.; del Toro, C.; Ryder, H. F.; Williamson, S. C.; Barbeau, R. A.; Hamilton, E. P.; Orias, E. Macronuclear Genome Sequence of the Ciliate Tetrahymena Thermophila, a Model Eukaryote. *PLoS Biol.* **2006**, *4* (9), e286. https://doi.org/10.1371/journal.pbio.0040286.

(49) Eichinger, L.; Pachebat, J. A.; Glöckner, G.; Rajandream, M.-A.; Sucgang, R.; Berriman, M.; Song, J.; Olsen, R.; Szafranski, K.; Xu, Q.; Tunggal, B.; Kummerfeld, S.; Madera, M.; Konfortov, B. A.; Rivero, F.; Bankier, A. T.; Lehmann, R.; Hamlin, N.; Davies, R.; Gaudet, P.; Fey, P.; Pilcher, K.; Chen, G.; Saunders, D.; Sodergren, E.; Davis, P.; Kerhornou, A.; Nie, X.; Hall, N.; Anjard, C.; Hemphill, L.; Bason, N.; Farbrother, P.; Desany, B.; Just, E.; Morio, T.; Rost, R.; Churcher, C.; Cooper, J.; Haydock, S.; van Driessche, N.; Cronin, A.; Goodhead, I.; Muzny, D.; Mourier, T.; Pain, A.; Lu, M.; Harper, D.; Lindsay, R.; Hauser, H.; James, K.; Quiles, M.; Babu, M. M.; Saito, T.; Buchrieser, C.; Wardroper, A.; Felder, M.; Thangavelu, M.; Johnson, D.; Knights, A.; Loulseged, H.; Mungall, K.; Oliver, K.; Price, C.; Quail, M. A.; Urushihara, H.; Hernandez, J.; Rabbinowitsch, E.; Steffen, D.; Sanders, M.; Ma, J.; Kohara, Y.; Sharp, S.; Simmonds, M.; Spiegler, S.; Tivey, A.; Sugano, S.; White, B.;

Walker, D.; Woodward, J.; Winckler, T.; Tanaka, Y.; Shaulsky, G.; Schleicher, M.; Weinstock, G.; Rosenthal, A.; Cox, E. C.; Chisholm, R. L.; Gibbs, R.; Loomis, W. F.; Platzer, M.; Kay, R. R.; Williams, J.; Dear, P. H.; Noegel, A. A.; Barrell, B.; Kuspa, A. The Genome of the Social Amoeba Dictyostelium Discoideum. *Nature* **2005**, *435* (7038), 43–57. https://doi.org/10.1038/nature03481.

(50) International Human Genome Sequencing Consortium. Finishing the Euchromatic Sequence of the Human Genome. *Nature* **2004**, *431* (7011), 931–945. https://doi.org/10.1038/nature03001.

(51) Ren, J.; Chaires, J. B. Sequence and Structural Selectivity of Nucleic Acid Binding Ligands. *Biochemistry* **1999**, *38* (49), 16067–16075. https://doi.org/10.1021/bi992070s.

(52) Huang, C. Y. Determination of Binding Stoichiometry by the Continuous Variation Method: The Job Plot. In *Methods in Enzymology*; Purich, D. L., Ed.; Academic Press, 1982; Vol. 87, pp 509–525. https://doi.org/10.1016/S0076-6879(82)87029-8.

(53) Mergny, J.-L. Thermal Difference Spectra: A Specific Signature for Nucleic Acid Structures. *Nucleic Acids Res.* **2005**, *33* (16), e138–e138. https://doi.org/10.1093/nar/gni134.

(54) Bhattacharjee, A. J.; Ahluwalia, K.; Taylor, S.; Jin, O.; Nicoludis, J. M.; Buscaglia, R.; Brad Chaires, J.; Kornfilt, D. J. P.; Marquardt, D. G. S.; Yatsunyk, L. A. Induction of G-Quadruplex DNA Structure by Zn(II) 5,10,15,20-Tetrakis(N-Methyl-4-Pyridyl)Porphyrin. *Biochimie* **2011**, *93* (8), 1297–1309. https://doi.org/10.1016/j.biochi.2011.05.038.

(55) Kypr, J.; Kejnovska, I.; Renciuk, D.; Vorlickova, M. Circular Dichroism and Conformational Polymorphism of DNA. *Nucleic Acids Res.* **2009**, *37* (6), 1713–1725. https://doi.org/10.1093/nar/gkp026.

(56) Ramsay, G. D.; Eftink, M. R. Analysis of Multidimensional Spectroscopic Data to Monitor Unfolding of Proteins. *Methods Enzym.* **1994**, *240*, 615–645.

(57) Kabsch, W. XDS. *Acta Crystallogr. D Biol. Crystallogr.* **2010**, *66* (2), 125–132. https://doi.org/10.1107/S0907444909047337.

(58) Liebschner, D.; Afonine, P. V.; Baker, M. L.; Bunkóczi, G.; Chen, V. B.; Croll, T. I.; Hintze, B.; Hung, L.-W.; Jain, S.; McCoy, A. J.; Moriarty, N. W.; Oeffner, R. D.; Poon, B. K.; Prisant, M. G.; Read, R. J.; Richardson, J. S.; Richardson, D. C.; Sammito, M. D.; Sobolev, O. V.; Stockwell, D. H.; Terwilliger, T. C.; Urzhumtsev, A. G.; Videau, L. L.; Williams, C. J.; Adams, P. D. Macromolecular Structure Determination Using X-Rays, Neutrons and Electrons: Recent Developments in Phenix. *Acta Crystallogr. Sect. Struct. Biol.* **2019**, *75* (10), 861–877. https://doi.org/10.1107/S2059798319011471.

(59) Nicoludis, J. M.; Miller, S. T.; Jeffrey, P. D.; Barrett, S. P.; Rablen, P. R.; Lawton, T. J.; Yatsunyk, L. A. Optimized End-Stacking Provides Specificity of N-Methyl Mesoporphyrin IX for Human Telomeric G-Quadruplex DNA. *J Am Chem Soc* **2012**, *134* (50), 20446–20456. https://doi.org/10.1021/ja3088746.

(60) Emsley, P.; Lohkamp, B.; Scott, W. G.; Cowtan, K. Features and Development of It Coot. *Acta Crystallogr. Sect. D* **2010**, *66* (4), 486–501. https://doi.org/10.1107/S0907444910007493.

(61) *The PyMOL Molecular Graphics System,*; Schrodinger, LLC.

(62) Chung, W. J.; Heddi, B.; Schmitt, E.; Lim, K. W.; Mechulam, Y.; Phan, A. T. Structure of a Left-Handed DNA G-Quadruplex. *Proc. Natl. Acad. Sci.* **2015**, *112* (9), 2729–2733. https://doi.org/10.1073/pnas.1418718112.
(63) Lu, X.-J.; Olson, W. K. 3DNA: A Versatile, Integrated Software System for the Analysis, Rebuilding and Visualization of Three-Dimensional Nucleic-Acid Structures. *Nat. Protoc.* **2008**, *3*, 1213.
(64) Madden, T. *The BLAST Sequence Analysis Tool*; National Center for Biotechnology Information (US), 2013.
(65) Edgar, R.; Domrachev, M.; Lash, A. E. Gene Expression Omnibus: NCBI Gene Expression and Hybridization Array Data Repository. *Nucleic Acids Res.* **2002**, *30* (1), 207–210. https://doi.org/10.1093/nar/30.1.207.
(66) Barrett, T.; Wilhite, S. E.; Ledoux, P.; Evangelista, C.; Kim, I. F.; Tomashevsky, M.; Marshall, K. A.; Phillippy, K. H.; Sherman, P. M.; Holko, M.; Yefanov, A.; Lee, H.; Zhang, N.; Robertson, C. L.; Serova, N.; Davis, S.; Soboleva, A. NCBI GEO: Archive for Functional Genomics Data Sets—Update. *Nucleic Acids Res.* **2013**, *41* (D1), D991–D995. https://doi.org/10.1093/nar/gks1193.
(67) Lu, M.; Guo, Q.; Kallenbach, N. R. Structure and Stability of Sodium and Potassium Complexes of DT4G4 and DT4G4T. *Biochemistry* **1992**, *31* (9), 2455–2459. https://doi.org/10.1021/bi00124a003.
(68) Wang, Y.; Patel, D. J. Solution Structure of the Human Telomeric Repeat d[AG3(T2AG3)3] G-Tetraplex. *Structure* **1993**, *1* (4), 263–282.
(69) Uddin, M. K.; Kato, Y.; Takagi, Y.; Mikuma, T.; Taira, K. Phosphorylation at 5′ End of Guanosine Stretches Inhibits Dimerization of G-Quadruplexes and Formation of a G-Quadruplex Interferes with the Enzymatic Activities of DNA Enzymes. *Nucleic Acids Res.* **2004**, *32* (15), 4618–4629. https://doi.org/10.1093/nar/gkh766.
(70) 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. https://doi.org/10.1021/jacs.5b13130.
(71) Paramasivan, S.; Bolton, P. H. Mix and Measure Fluorescence Screening for Selective Quadruplex Binders. *Nucleic Acids Res.* **2008**, *36* (17), e106. https://doi.org/10.1093/nar/gkn487.
(72) Reshetnikov, R. V.; Kopylov, A. M.; Golovin, A. V. Classification of G-Quadruplex DNA on the Basis of the Quadruplex Twist Angle and Planarity of G-Quartets. *Acta Naturae* **2010**, *2* (4), 72–81.
(73) Parkinson, G. N. Fundamentals of Quadruplex Structures. In *Quadruplex Nucleic Acids*; Neidle, S., Balasubramanian, S., Eds.; The Royal Society of Chemistry, 2006; pp 1–30. https://doi.org/10.1039/9781847555298-00001.
(74) Karsisiotis, A. I.; O'Kane, C.; Webba da Silva, M. DNA Quadruplex Folding Formalism – A Tutorial on Quadruplex Topologies. *Methods* **2013**, *64* (1), 28–35. https://doi.org/10.1016/j.ymeth.2013.06.004.
(75) Wang, Y.; Patel, D. J. Solution Structure of TheTetrahymena Telomeric Repeat d(T2G4)4 G-Tetraplex. *Structure* **1994**, *2* (12), 1141–1156. https://doi.org/10.1016/s0969-2126(94)00117-0.
(76) Russo Krauss, I.; Ramaswamy, S.; Neidle, S.; Haider, S.; Parkinson, G. N. Structural Insights into the Quadruplex–Duplex 3′ Interface Formed from a Telomeric Repeat:

A Potential Molecular Target. *J Am Chem Soc* **2016**, *138* (4), 1226–1233. https://doi.org/10.1021/jacs.5b10492.

(77) Campbell, N. H.; Smith, D. L.; Reszka, A. P.; Neidle, S.; O'Hagan, D. Fluorine in Medicinal Chemistry: β-Fluorination of Peripheral Pyrrolidines Attached to Acridine Ligands Affects Their Interactions with G-Quadruplex DNA. *Org. Biomol. Chem.* **2011**, *9* (5), 1328–1331. https://doi.org/10.1039/C0OB00886A.

(78) Haider, S. M.; Parkinson, G. N.; Neidle, S. Structure of a G-Quadruplex–Ligand Complex. *J. Mol. Biol.* **2003**, *326* (1), 117–125. https://doi.org/10.1016/S0022-2836(02)01354-2.

(79) Campbell, N. H.; Patel, M.; Tofa, A. B.; Ghosh, R.; Parkinson, G. N.; Neidle, S. Selectivity in Ligand Recognition of G-Quadruplex Loops. *Biochemistry* **2009**, *48* (8), 1675–1680. https://doi.org/10.1021/bi802233v.

(80) Clark, G. R.; Pytel, P. D.; Squire, C. J. The High-Resolution Crystal Structure of a Parallel Intermolecular DNA G-4 Quadruplex/Drug Complex Employing Syn Glycosyl Linkages. *Nucleic Acids Res.* **2012**, *40* (12), 5731–5738. https://doi.org/10.1093/nar/gks193.

(81) Campbell, N. H.; Karim, N. H. A.; Parkinson, G. N.; Gunaratnam, M.; Petrucci, V.; Todd, A. K.; Vilar, R.; Neidle, S. Molecular Basis of Structure–Activity Relationships between Salphen Metal Complexes and Human Telomeric DNA Quadruplexes. *J Med Chem* **2012**, *55* (1), 209–222. https://doi.org/10.1021/jm201140v.

(82) McQuaid, K.; Abell, H.; Gurung, S. P.; Allan, D. R.; Winter, G.; Sorensen, T.; Cardin, D. J.; Brazier, J. A.; Cardin, C. J.; Hall, J. P. Structural Studies Reveal Enantiospecific Recognition of a DNA G-Quadruplex by a Ruthenium Polypyridyl Complex. *Angew. Chem. Int. Ed.* **2019**, *58* (29), 9881–9885. https://doi.org/10.1002/anie.201814502.

(83) Bazzicalupi, C.; Ferraroni, M.; Papi, F.; Massai, L.; Bertrand, B.; Messori, L.; Gratteri, P.; Casini, A. Determinants for Tight and Selective Binding of a Medicinal Dicarbene Gold(I) Complex to a Telomeric DNA G-Quadruplex: A Joint ESI MS and XRD Investigation. *Angew Chem Int Ed* **2016**, *55* (13), 4256–4259. https://doi.org/10.1002/anie.201511999.

(84) Collie, G. W.; Promontorio, R.; Hampel, S. M.; Micco, M.; Neidle, S.; Parkinson, G. N. Structural Basis for Telomeric G-Quadruplex Targeting by Naphthalene Diimide Ligands. *J Am Chem Soc* **2012**, *134* (5), 2723–2731. https://doi.org/10.1021/ja2102423.

(85) Fujimoto, T.; Nakano, S.; Sugimoto, N.; Miyoshi, D. Thermodynamics-Hydration Relationships within Loops That Affect G-Quadruplexes under Molecular Crowding Conditions. *J. Phys. Chem. B* **2013**, *117* (4), 963–972. https://doi.org/10.1021/jp308402v.

(86) Guédin, A.; Lin, L. Y.; Armane, S.; Lacroix, L.; Mergny, J.-L.; Thore, S.; Yatsunyk, L. A. Quadruplexes in 'Dicty': Crystal Structure of a Four-Quartet G-Quadruplex Formed by G-Rich Motif Found in the Dictyostelium Discoideum Genome. *Nucleic Acids Res.* **2018**, *46* (10), 5297–5307. https://doi.org/10.1093/nar/gky290.

(87) Tippana, R.; Xiao, W.; Myong, S. G-Quadruplex Conformation and Dynamics Are Determined by Loop Length and Sequence. *Nucleic Acids Res.* **2014**, *42* (12), 8106–8114. https://doi.org/10.1093/nar/gku464.

(88) Rochin, L.; Hurbain, I.; Serneels, L.; Fort, C.; Watt, B.; Leblanc, P.; Marks, M. S.; De Strooper, B.; Raposo, G.; van Niel, G. BACE2 Processes PMEL to Form the

Melanosome Amyloid Matrix in Pigment Cells. *Proc. Natl. Acad. Sci.* **2013**, *110* (26), 10658. https://doi.org/10.1073/pnas.1220748110.

(89) Li, J.; Correia, J. J.; Wang, L.; Trent, J. O.; Chaires, J. B. Not so Crystal Clear: The Structure of the Human Telomere G-Quadruplex in Solution Differs from That Present in a Crystal. *Nucleic Acids Res.* **2005**, *33* (14), 4649–4659. https://doi.org/10.1093/nar/gki782.

(90) Largy, E.; Mergny, J.-L.; Gabelica, V. Role of Alkali Metal Ions in G-Quadruplex Nucleic Acid Structure and Stability. In *The Alkali Metal Ions: Their Role for Life*; Sigel, A., Sigel, H., Sigel, R. K. O., Eds.; Metal Ions in Life Sciences; Springer International Publishing: Cham, 2016; pp 203–258. https://doi.org/10.1007/978-3-319-21756-7_7.

(91) Weil, J.; Min, T.; Yang, C.; Wang, S.; Sutherland, C.; Sinha, N.; Kang, C. Stabilization of the I-Motif by Intramolecular Adenine–Adenine–Thymine Base Triple in the -structure of d(ACCCT). *Acta Crystallogr. D Biol. Crystallogr.* **1999**, *55* (2), 422–429. https://doi.org/10.1107/S0907444998012529.

(92) Berger, I.; Kang, C.; Fredian, A.; Ratliff, R.; Moyzis, R.; Rich, A. Extension of the Four-Stranded Intercalated Cytosine Motif by Adenine•adenine Base Pairing in the Crystal Structure of d(CCCAAT). *Nat. Struct. Biol.* **1995**, *2* (5), 416–425. https://doi.org/10.1038/nsb0595-416.

Appendix

Supplementary information

Table S1. Occurrences of the T1 sequence in the human genome (GRCh38.p12 primary assembly) identified via BLAST.

Chromosome	Start position	End position	Significance
6	87229964	87229981	Within zinc finger protein 292 gene
7	132317250	132317267	Complement of plexin A4 gene
12	9924441	9924458	Complement of C-type lectin domain family 2 member A gene
	9924446	9924463	Complement of C-type lectin domain family 2 member A gene
	9924451	9924468	Complement of C-type lectin domain family 2 member A gene
	68918645	68918662	Complement of carboxypeptidase M gene
14	99067501	99067518	Complement of LOC107984696 ncRNA
	99067506	99067523	Complement of LOC107984696 ncRNA
	99067532	99067549	Complement of LOC107984696 ncRNA
17	75933313	75933330	Complement of fas binding factor 1 gene
21	41156399	41156416	—

Table S2. Thermodynamic stability of T1, T7, and T8 in the presence of 2 eq. NMM. CD melting studies were performed in 5K buffer. *relative to DNA alone.

	T_m, °C	ΔT_m, °C*	ΔH, kcal/mol
T1 + NMM	78 ± 1	20. ± 1	57 ± 2
T7 + NMM	71.9 ± 0.3	19.9 ± 0.4	53.0 ± 0.9
T8 + NMM	73.4 ± 0.4	17.0 ± 0.5	54.5 ± 0.6

Table S3. RMSD (Å) for the T1-NMM and T7-NMM structures.

	DNA chain		Overall	GQ core
Within structure	T1_A	T1_B	0.9	0.49
	T7_A	T7_B	1.2	0.23
Between structures	T1_A	T7_A	2.0	0.64
	T1_A	T7_B	1.1	0.61
	T1_B	T7_A	1.2	0.66
	T1_B	T7_B	1.0	0.66
	Average		1.3 ± 0.4	0.64 ± 0.02

Table S4. B-factors (Å^2) for the T1-NMM and T7-NMM structures.

	T1-NMM	T7-NMM
Overall	115.44	97.15
GQ w/ loops	115.36	97.75
NMM	104.77	94.01
GQ core	106.46	81.18
TT loops	148.13	128.85
T overhangs	---	151.63
All Ts	148.13	131.87

Table S5. Distances between G-quartets and between outer G-quartets and NMM in the T1-NMM and T7-NMM structures (Å).

	T1			T7		
Quartets	**Chain A**	**Chain B**	**Average**	**Chain A**	**Chain B**	**Average**
5'-middle	3.320	3.315	**3.32** ± 0.01	3.44	3.39	**3.41** ± 0.03
Middle-3'	3.40	3.31	**3.36** ± 0.06	3.41	3.42	**3.41** ± 0.01
3'-NMM	3.5	3.7	**3.6** ± 0.2	3.62	3.65	**3.64** ± 0.02
5'-5' (between monomers)	---	---	**3.38**	---	---	**3.36**

Table S6A. Intramolecular helical twist (°) between each quartet pair in the T1-NMM and T7-NMM structures.

	T1-NMM			T7-NMM		
Chain →	**A**	**B**	**Both**	**A**	**B**	**Both**
Between 5' and middle quartets	34	31		27.4	27	
	29	35		28.1	29	
	30	32		28.6	28	
	32	34		28.8	26	
Average	31 ± 2	33 ± 2	**32** ± 2	28.2 ± 0.6	27 ± 1	**27.8** ± 0.9
Between middle and 3' quartetsß	27	24		27	28	
	29	24		28	26	
	27	25		28	27	
	28	27		31	29	
Average	28 ± 1	25 ± 1	**26** ± 2	29 ± 2	27 ± 2	**28** ± 2
Overall	**29** ± 4			**28** ± 1		

Table S6B. Intermolecular helical twist (°) at the dimer interface for T1-NMM and T7-NMM. A' signifies a symmetry related chain A molecule.

T1-NMM		T7-NMM	
G pair	**Twist**	**G pair**	**Twist**
A1-A'16	117	**A2-B17**	115
A6-A'11	115	**A7-B12**	118
A11-A'6	115	**A12-B7**	118
A16-A'1	117	**A17-B2**	118
Average	**116** ± 1	**Average**	**117** ± 2

Table S7. Groove widths in the T7-NMM structure (Å).

	Groove ->	**1**	**2**	**3**	**4**	**Average (both chains)**
Chain A quartets	**3'**	16.0	16.2	15.9	16.3	16.2 ± 0.2
	Middle	16.7	16.2	16.8	16.5	16.6 ± 0.3
	5'	15.6	15.5	15.3	15.3	15.5 ± 0.2
Chain B quartets	**3'**	16.1	16.2	16.2	16.4	
	Middle	16.9	16.3	17.0	16.2	
	5'	15.8	15.4	15.8	15.1	
Average		16.2 ± 0.5	16.0 ± 0.4	16.2 ± 0.6	16.0 ± 0.6	

Table S8. Helical twist in 19wt. Helical twist was calculated between each pair of adjacent G-quartets within the asymmetric unit of the crystal structure (in all seven DNA chains).

Chain	All	A	B	C	D	E	F	Z
Quartets 1 & 2		90.4	97.6	92.1	97.1	94.5	87.9	92.8
		89.5	94.8	86.2	102.2	94.6	81.7	86.1
		94.9	81.8	94.2	78.4	94.5	86.1	87.4
		94.6	91.5	93.3	83.6	95.5	96.6	92.4
Mean	**91 ± 6**	92 ± 3	91 ± 7	92 ± 4	90 ± 10	94.8 ± 0.5	88 ± 6	90 ± 3
Quartets 2 & 3		147.1	144.1	145.2	142.3	146.7	141.2	146.3
		145.8	148.5	145.6	144.1	147.2	145.4	138.3
		144.7	141.8	143.6	139.5	142.1	149.2	143.3
		151.8	144.9	144.3	142.5	147.1	150.0	149.6
Mean	**145 ± 3**	147 ± 3	145 ± 3	144.7 ± 0.9	142 ± 2	146 ± 3	146 ± 4	144 ± 5
Quartets 3 & 4		88.3	92.2	85.9	89.7	94.3	86.9	85.0
		90.9	92.2	82.1	77.5	92.8	85.2	90.2
		94.2	88.3	88.5	89.0	90.8	89.5	97.2
		97.3	83.7	92.2	85.5	89.2	90.8	89.1
Mean	**89 ± 4**	93 ± 4	89 ± 4	87 ± 4	85 ± 6	92 ± 2	88 ± 3	90 ± 5

Table S9. Statistics for the best T1-PHPS4 datasets.

Collection date	11/24/18	6/30/19	6/30/19	6/30/19	6/30/19	10/25/19	2/13/20
Data collection name	LL_E14_1-400	LL_B10	LL_B10_run2	LL_B10_run3	LL_B10_run3_1 +LL_B10_run2_1	A_15_T1RHPS4_ run1_1_1-197 +A_15_T1RHPS4 _run1_1_343-528	LL_M3_run2_1_15 8-372 +LL_M3_1_83-372
Space group	P 21 21 21	C 2 2 2	P 21 21 21	P 21 21 21	P 21 21 21	P 21 21 21	C 2 2 21
Unit cell (Å)	34.57 59.60 254.63	70.46 121.73 255.96	35.20 60.82 255.75	35.19 60.80 255.73	35.19 60.81 255.74	36.18 62.66 256.62	71.93 125.12 256.05
Unit cell (°)	90.00 90.00 90.00	90.00 90.00 90.00	90.00 90.00 90.00	90.00 90.00 90.00	90.00 90.00 90.00	90.00 90.00 90.00	90.00 90.00 90.00
Rmerge overall	0.076	0.1	0.052	0.055	0.061	0.092	0.097
Resolution (Å)	2.19	2.37	2.26	2.19	2.32	2.45	2.30
Redundancy							
Overall	4.1	4.9	5.3	5.2	10.5	4.0	6.1
Outer shell	3.7	4.4	5.2	5.1	10.7	4.0	6.6
Completeness							
Overall	98.4	98.0	98.8	98.8	99.6	98.4	95.6
Outer shell	96.2	89.3	94.8	93.6	99.6	99.3	97.6
I/σ							
Overall	7.0	8.4	11.9	10.9	18.1	10.1	7.3
Outer shell	0.2	0.9	0.8	0.5	1.8	1.5	1.5
CC(1/2)							
Overall	0.999	0.996	0.999	0.999	?	?	?
Outer shell	0.602	0.832	0.862	0.619	?	?	?

Table S10. Statistics for the best iHRAS datasets.

Collection date	10/25/19	10/25/19
Data collection name	F_11_iHRAS_run4_1	F_12_iHRAS_run2_1
Space group	C 1 2 1	C 1 2 1
Unit cell (Å)	119.06 28.21 53.41	119.13 29.66 54.33
Unit cell (°)	90.00 93.40 90.00	90.00 93.10 90.00
Resolution (Å)	2.27	2.02
Outer shell CC(1/2)	0.546	0.765
Rmerge overall	0.064	0.046
Completeness		
Overall	98.1	97.8
Outer shell	96.5	87.5
Redundancy		
Overall	4.1	3.9
Outer shell	4.2	3.4
I/σ		
Overall	9.6	10.6
Outer shell	0.7	0.8

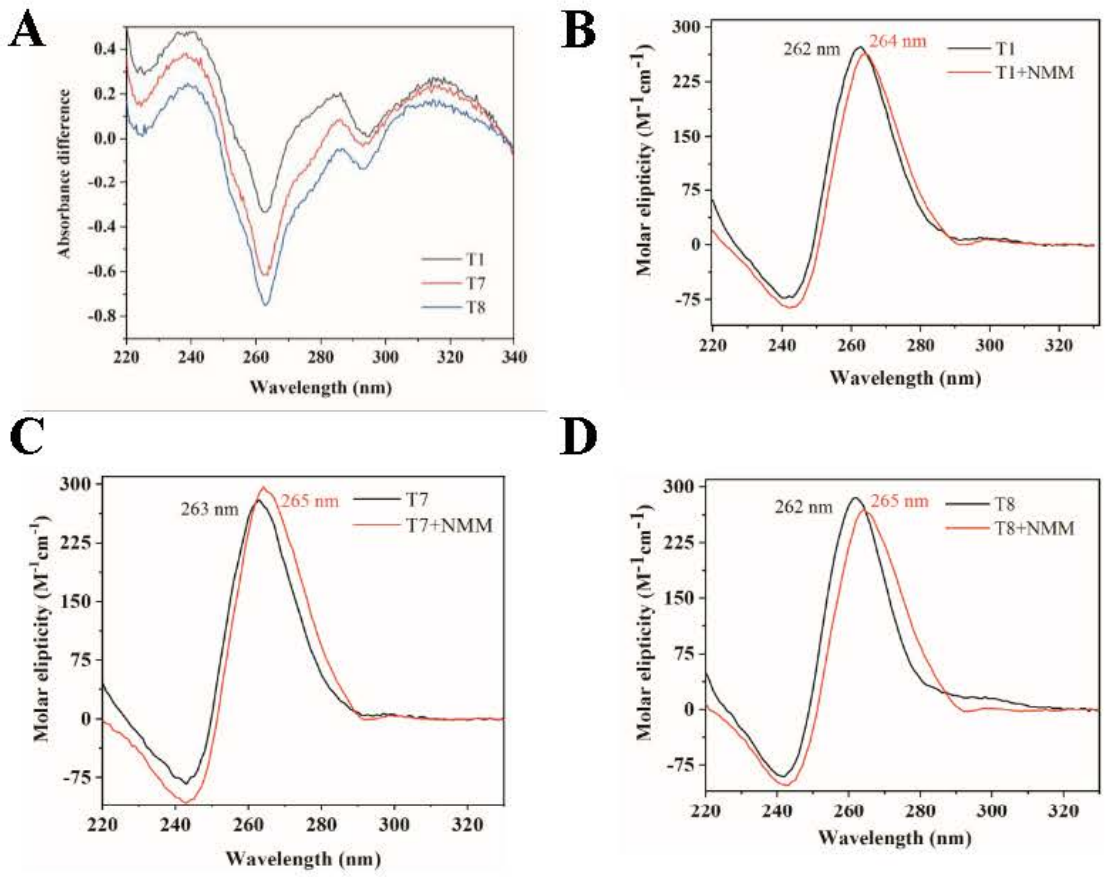


Figure S1. CD and TDS signature of T1, T7, and T8 all in complex with NMM. (**A**) TDS and (**B-D**) CD scans for DNA alone at ~4 μM and with 2 eq. of NMM in 5K buffer at 20 °C.

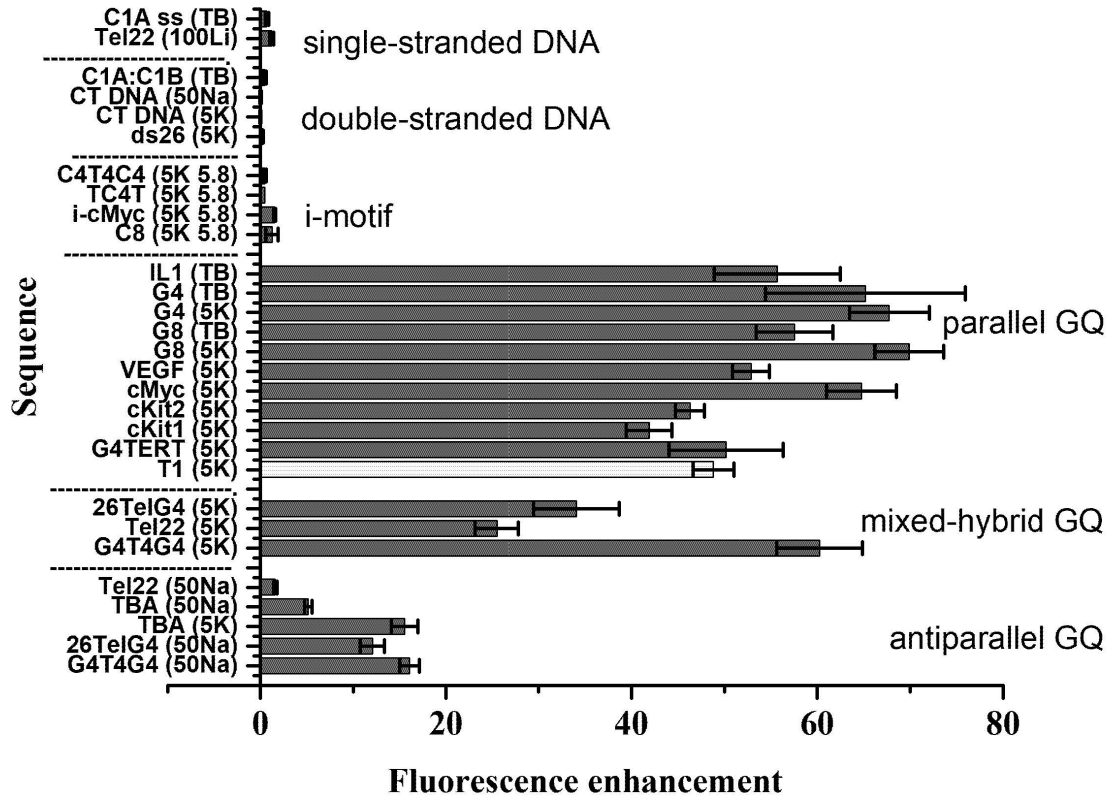


Figure S2. Fluorescence enhancement data for NMM in the presence of 10 eq. of the indicated DNA sequences. T1 data (yellow) was collected with 2-5 eq. of DNA. Figure adapted from [44]. Buffers:

(TB): 50 mM Tris-borate pH 8.3, 10 mM KCl, 1 mM $MgCl_2$

(100Li): 10 mM lithium cacodylate pH 7.2, 100 mM LiCl (5K 5.8): 10 mM lithium cacodylate pH 5.8, 5 mM KCl, 95 mM LiCl

(50Na): 10 mM lithium cacodylate pH 7.2, 50 mM NaCl, 50 mM LiCl.

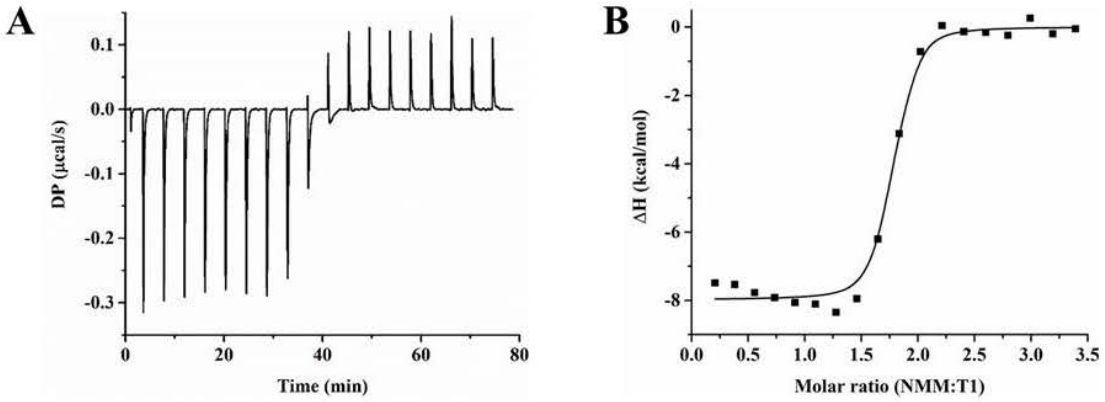


Figure S3. Reverse ITC titration for T1-NMM. (A) Representative raw ITC data for titration of 41.4 µM prefolded T1 into 23.5 µM NMM in 5K buffer at 25 °C with baseline subtraction. **(B)** Integrated raw heats with a fit generated using the one-site binding model.

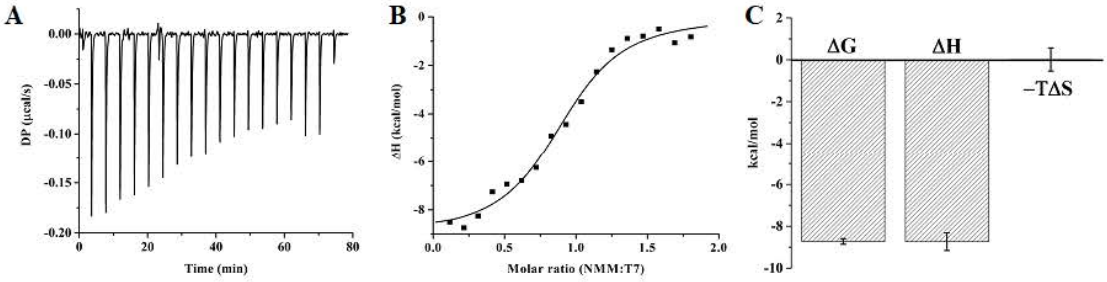
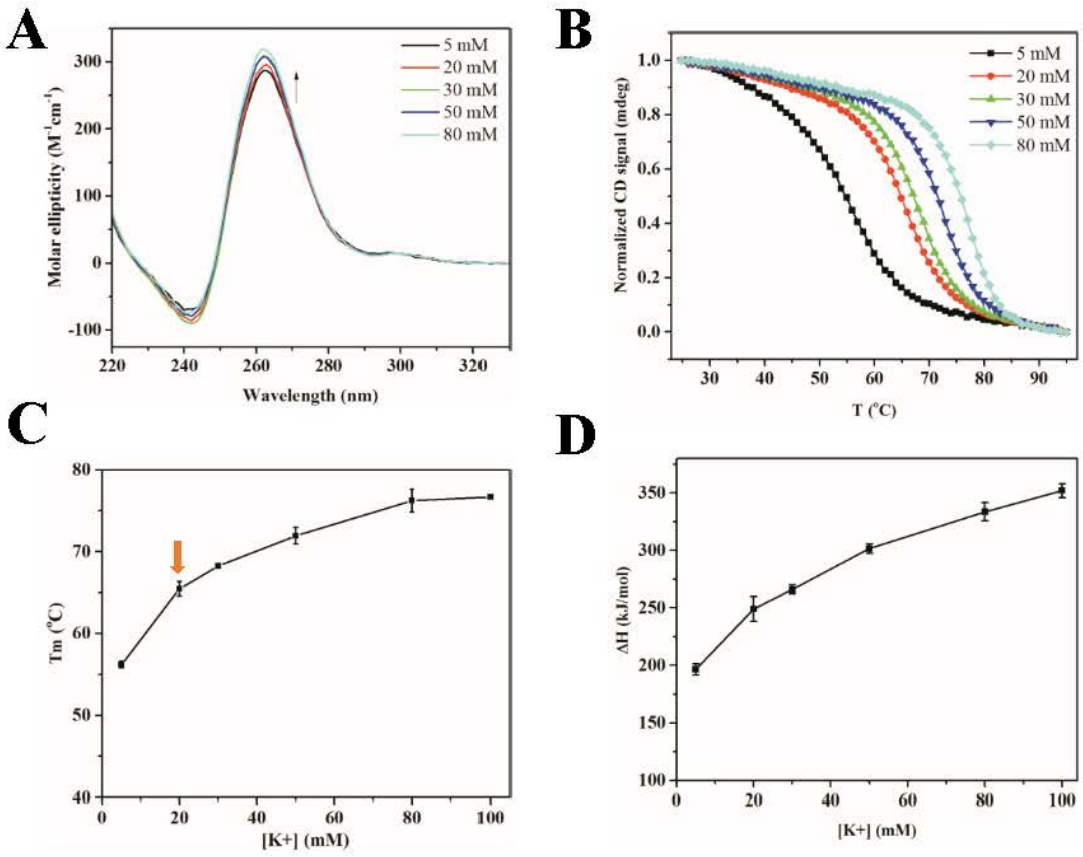


Figure S4. Thermodynamic characterization of T7-NMM binding via ITC. (A) Representative raw ITC data for titration of 89.2 µM NMM into 8.9 µM of prefolded T7 in 5K buffer at 25 °C with baseline subtraction. **(B)** Integrated raw heats with a fit generated using the one-site binding model. **(C)** Gibbs free energy deconvoluted into enthalpic (ΔH) and entropic (ΔS) contributions for both the forward and reverse titrations.



E

[K$^+$], mM	T_m, °C	ΔH, kJ/mol	ΔH, kcal/mol	Hysteresis, °C
5	56.2 ± 0.4	197 ± 5	47 ± 1	3.3
20	**65.5 ± 0.9**	**250 ± 10**	**60. ± 3**	**2.9**
30*	68.3 ± 0.3	266 ± 4	64 ± 1	2.7
50	72 ± 1	302 ± 4	72 ± 1	2.1
80	76.2 ± 1.4	334 ± 8	80. ± 2	1.9
100*	76.7 ± 0.3	352 ± 6	84 ± 1	1.2

Figure S5. Effect of K$^+$ concentration on fold and stability of T1. Experiments were conducted with ~4.5 μM T1 in 10 mM lithium cacodylate pH 7.2 in the presence of 5 - 100 mM KCl. (**A**) CD signatures, (**B**) CD melting curves, (**C**) T_m, (**D**) ΔH, and (**E**) thermodynamic parameters. T_m and ΔH were determined from melting curves. *These numbers were measured once and the errors come from the instrument uncertainty or data fitting.

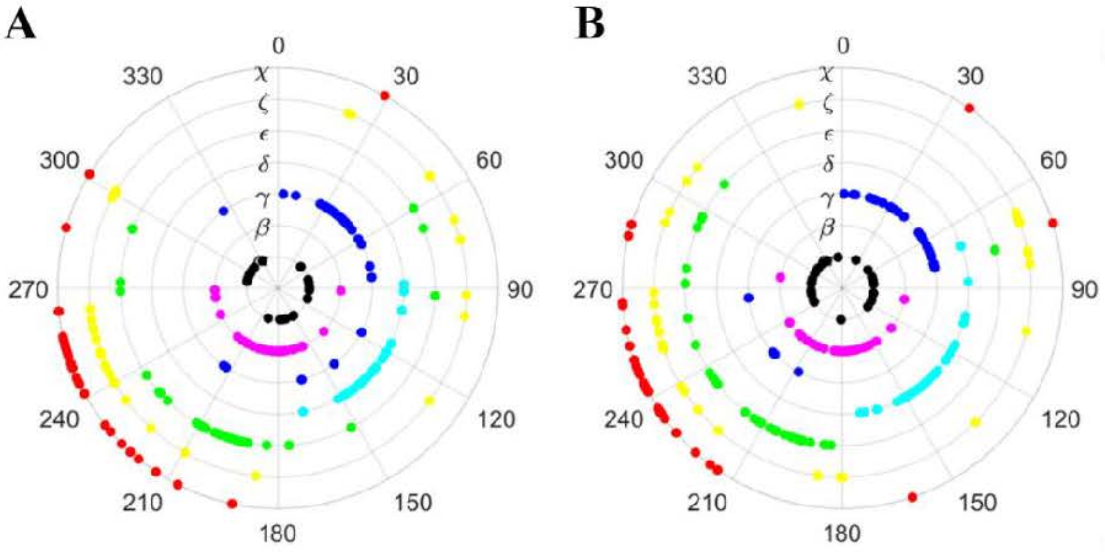


Figure S6. Torsional angle wheel for the (A) T1-NMM and (B) T7-NMM structures. Distribution of DNA torsional angles with each individual angle shown as a dot.

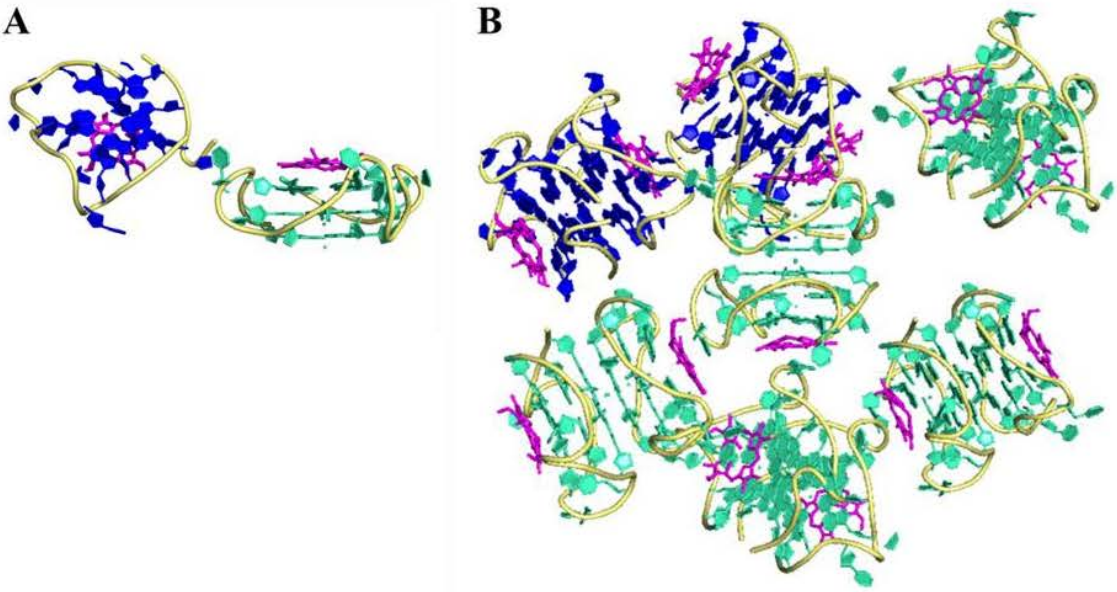


Figure S7. Intermolecular interactions in the T1-NMM crystal structure. (**A**) A graphical representation of the T1-NMM asymmetric unit. NMM stacks on the 3' end of GQ. Chain A is colored in teal, chain B is blue, the sugar-phosphate backbone is yellow, and NMM is magenta. Potassium ions are depicted as spheres. (**B**) Intermolecular interactions among T1 and NMM molecules.

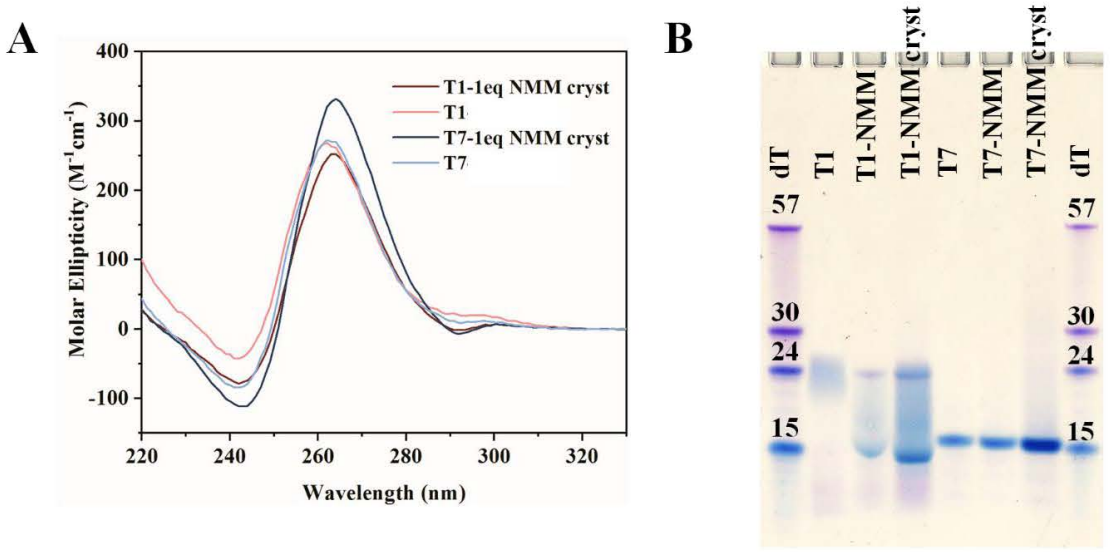


Figure S8. Comparison of samples for crystallization and for biophysical studies. Crystallization samples contained a high concentration of DNA-NMM at 0.65 mM, while PAGE and biophysics samples contained 50 and 5 μM DNA-NMM, respectively. (**A**) CD scans of T1 and T7 alone at 5 μM and in complex with 1 eq. of NMM at 0.65 mM at 25 °C. (**B**) Fifteen percent PAGE of T1 and T7 annealed alone, in complex with 2 eq. of NMM (as in biophysical studies), and at crystallization concentration with 1 eq. of NMM. The crystallization sample was diluted to the same concentration as the others immediately prior to loading the gel. All samples were in 5K buffer except the crystallization samples, which were in 20K buffer.

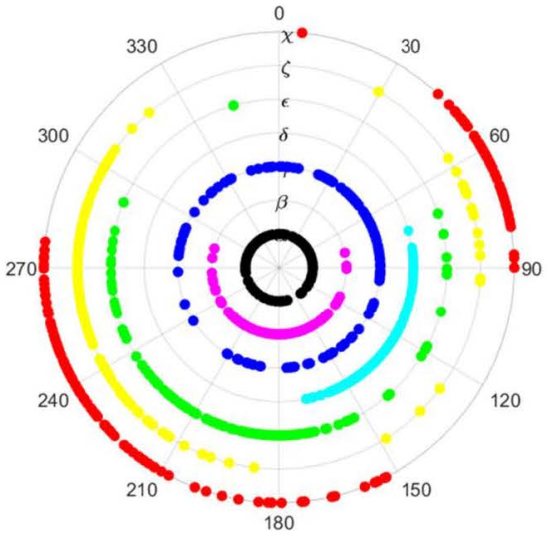


Figure S9. Distribution of DNA torsional angles in the 19wt GQ. Each individual dihedral angle is plotted as a dot.

Scripts

Scripts were written to analyze GQ crystal structures, facilitate running of a standard CD experiment (pre-melt scans, then melt, followed by post-melt scans), and enable more efficient processing of CD and TDS data in Origin. These scripts (along with other variants) can also be accessed online at https://github.swarthmore.edu/yatlab (a private repository shared with Liliya Yatsunyk).

List of scripts:

Helical twist script (Python)

```
"""
Calculates the helical twist given a PDB file of a single GQ.
Note: will likely need to modify for the particular PDB file (see #TODO). This
version works for T1.
"""
import numpy as np
import math
import statistics

def main():
    #dictionary of nucleotide coordinates in GQ chain
    #key = atom #
    #value = [C8, C2, N1] where each atom is a list: [name, x, y, z]
    nucleotide_coordinates = {}
    for i in range(1,19):                          #TODO num nucleotides + 1
        nucleotide_coordinates[i] = []
```

```
#parse pdb file
filename = input("File name: ") + ".pdb"
print()
pdb = open(filename, "r")
for line in pdb:
    info = line.split()
    # print(info)
    atom = info[2]
    nucleotide_number = int(info[5])
    x = float(info[6])
    y = float(info[7])
    z = float(info[8])
    if (atom == "C8") or (atom == "C2") or (atom == "N1"):
        coordinates = [atom, x, y, z]
        nucleotide_coordinates[nucleotide_number].append(coordinates)
    pdb.readline()                                  #TODO remove/include as needed
pdb.close()
                                            #TODO quartet Gs in order; num quartets
quartet1 = [1,6,11,16]
quartet2 = [2,7,12,17]
quartet3 = [3,8,13,18]
quartets = [quartet1, quartet2, quartet3]
Gs = quartet1 + quartet2 + quartet3

#Make vector for each nucleotide
nucleotide_vectors = {}
for G in Gs:
    C8 = nucleotide_coordinates[G][0][1:4]
    N1 = nucleotide_coordinates[G][2]
    C2 = nucleotide_coordinates[G][1]
    midpt = [(N1[1]+C2[1])/2, (N1[2]+C2[2])/2, (N1[3]+C2[3])/2]
    vector = [midpt[0]-C8[0], midpt[1]-C8[1], midpt[2]-C8[2]]
    nucleotide_vectors[G] = vector

for i in range(0,2): #access pairs of stacked quartets  #TODO num quartets
    print("Quartets " + str(i+1) + "+" + str(i+2))
    topQuartet = quartets[i] #top quartet
    bottomQuartet = quartets[i+1] #bottom quartet
    angles = []
    for j in range(0,4): #access each guanine pair
        topG = nucleotide_vectors[topQuartet[j]]
        bottomG = nucleotide_vectors[bottomQuartet[j]]
        twist = angle(topG, bottomG)
```

```
            angles.append(twist)
            print(twist)
        print(statistics.mean(angles))
        print(statistics.stdev(angles))
        # print("Mean: " + str(statistics.mean(angles)))
        # print("SD: " + str(statistics.stdev(angles)))
        print()

def angle(v1, v2):
    """
    Calculates angle between 2 vectors in degrees
    """
    mag1 = np.linalg.norm(v1)
    mag2 = np.linalg.norm(v2)
    angle = np.arccos(np.dot(v1, v2) / (mag1 * mag2))
    angle = math.degrees(angle)
    return angle

if __name__ == "__main__":
    main()
```

Intermolecular twist script (Python)

```
"""
Calculates the intermolecular helical twist (between the tetrads at the
dimer interface) given a PDB file containing GQ dimer
Note: will likely need to modify for the particular PDB file (see #TODO). This
version works for T1 (dimer by symmetry). It is very similar to / a modification of
the helical twist script.
"""
import numpy as np
import math
import statistics

def main():
    #dictionary of nucleotide coordinates in GQ chain
    #key = atom #
    #value = [C8, C2, N1] where each atom is a list: [name, x, y, z]
    nucleotide_coordinates_A = {}
    nucleotide_coordinates_B = {}
    for i in range(1,19):                          #TODO num nucleotides + 1
        nucleotide_coordinates_A[i] = []
        nucleotide_coordinates_B[i] = []
```

```
#parse pdb file
filename = input("File name: ") + ".pdb"
print()
pdb = open(filename, "r")
for line in pdb:
    info = line.split()
    # print(info)
    atom = info[2]
    chain = info[4]
    nucleotide_number = int(info[5])
    x = float(info[6])
    y = float(info[7])
    z = float(info[8])
    if (atom == "C8") or (atom == "C2") or (atom == "N1"):
        coordinates = [atom, x, y, z]
        if chain == 'A':
            nucleotide_coordinates_A[nucleotide_number].append(coordinates)
        elif chain == 'B':
            nucleotide_coordinates_B[nucleotide_number].append(coordinates)
    pdb.readline()                                  #TODO rm if no ANISOU lines
pdb.close()

G_A = [1,6,11,16]                                   #TODO nt number in tetrad
G_B = [16,11,6,1]

#Make vector for each nucleotide
nucleotide_vectors_A = {}
nucleotide_vectors_B = {}
for i in range(0,4):
    C8 = nucleotide_coordinates_A[G_A[i]][0][1:4]
    N1 = nucleotide_coordinates_A[G_A[i]][2]
    C2 = nucleotide_coordinates_A[G_A[i]][1]
    midpt = [(N1[1]+C2[1])/2, (N1[2]+C2[2])/2, (N1[3]+C2[3])/2]
    vector = [midpt[0]-C8[0], midpt[1]-C8[1], midpt[2]-C8[2]]
    nucleotide_vectors_A[i] = vector

    C8 = nucleotide_coordinates_B[G_B[i]][0][1:4]
    N1 = nucleotide_coordinates_B[G_B[i]][2]
    C2 = nucleotide_coordinates_B[G_B[i]][1]
    midpt = [(N1[1]+C2[1])/2, (N1[2]+C2[2])/2, (N1[3]+C2[3])/2]
    vector = [midpt[0]-C8[0], midpt[1]-C8[1], midpt[2]-C8[2]]
    nucleotide_vectors_B[i] = vector
```

```
    angles = []
    for i in range(0,4): #access each guanine pair
        topG = nucleotide_vectors_A[i]
        bottomG = nucleotide_vectors_B[i]
        twist = angle(topG, bottomG)
        angles.append(twist)
        print(twist)
    print(statistics.mean(angles))
    print(statistics.stdev(angles))
    # print("Mean: " + str(statistics.mean(angles)))
    # print("SD: " + str(statistics.stdev(angles)))
    print()

def angle(v1, v2):
    """Calculates angle between 2 vectors in degrees
    """
    mag1 = np.linalg.norm(v1)
    mag2 = np.linalg.norm(v2)
    angle = np.arccos(np.dot(v1, v2) / (mag1 * mag2))
    angle = math.degrees(angle)
    return angle

if __name__ == "__main__":
    main()
```

CD scan-melt-scan macro (Aviv Macro Command)

```
/*
CD macro for scan-melt-scan
Linda Lin & Deondre Jordan 7/17/17

Runs pre-melt scans, melt (+ cool if desired), and post-melt scans for you.

---
Instructions:

-On the CD computer, open Aviv Macro Command Editor
-File -> Read Macro File -> LazyLindaAnnotated
-Save Macro File As -> <WhateverYouWish!>
-Edit lines with comments as needed by double clicking -> save

-Set data save path as normal
->If desired, name cells so data files will include cell name along with cell number
->If cool is also desired, check reverse temperature scan under temperature
configuration as normal
-Configure Experiment -> Macro Experiment -> <MacroFileName> -> Exit/Save Configuration
-Configure Experiment -> Type -> <MacroFileName_MayBeShortened> -> Exit/Save
Configuration
```

```
-Run experiment! Be sure you have enough nitrogen, and then come back when it's done
-Data for each portion of the experiment will be saved and time-stamped when complete.
---
*/


* Exp. Name : 071717Oligo19premelt          //PRE_MELT_SCANS
Set Absolute Temperature : 25.00 degC.      //Temp
Set Bandwidth : 2.00 nm.
Enable Rotor Cells: 0 : 1 : 2 : 3 : 4       //Cells
* Wavelength Start : 330.000 nm
* Wavelength End : 220.000 nm
* Wavelength Step : 1.000 nm
* Wavelength Repeats: 5                     //# scans
Set Data Set Save Options
Save Data to Browser and Hard Drive
Set Averaging Time : 1.00 sec
Run Experiment: Wavelength Scan

* Exp. Name : 071717Oligo19melt             //MELT
* Temperature Start : 25.000 deg C.         //Temp
* Temperature End : 95.000 deg C.
* Temperature Step : 1.000 deg C.
Set Temp. Dead Band : 0.33 deg C.
Set Averaging Time : 15.00 sec
Set Absolute Wavelength : 295.00 nm.
Set Temp. Rate of Change : 1.00 degC/min.
Set Equilibration Time : 0.08 min
Run Experiment: Temperature Experiment

* Exp. Name : 071717Oligo19postmelt         //POST_MELT_SCANS
Set Absolute Temperature : 25.00 degC.      //Temp
* Wavelength Start : 330.000 nm
* Wavelength End : 220.000 nm
* Wavelength Step : 1.000 nm
* Wavelength Repeats: 5                     //# scans
Set Averaging Time : 1.00 sec
Run Experiment: Wavelength Scan
```

TDS script (Origin LabTalk)

```
/*
GENERAL INSTRUCTIONS FOR USING ORIGIN SCRIPTS:

-Have the data file(s) imported as specified in script instructions
-Press Alt + Shift + 3 (Windows) / Option + Shift + 3 (Mac) to open the script window
-Paste the full script
-Edit any lines (commented using "//") and provide any data specified in the
instructions
-Highlight the entire script and press enter

When the comment says:
//Change "Concentrations" to book name
Right click on book -> properties -> name in the "short name" field is the one you want
to use

---
Instructions:
```

```
-Import TDS data with sheet1 named 95C, sheet2 named 4C, wavelength in col(A), abs for
samples in rest of columns, no blank columns. Samples should be in same order in both
sheets.
-Need to edit first & last line with # samples & DNA extinction coefficient
-If normalization does not work, delete the normalized and zeroed portions, normalize
manually, then run the last part of the script only (after normalization)

---
Output:

-TDS, normalized, and zeroed values in a new sheet in original sample order
-Concentrations in a new sheet in original sample order
*/


int samples = 2;                    // Change # samples

//Extract wavelengths
range wavelength = 1[1:259];
int ncol = wks.ncols;

//Create TDS sheet
newsheet cols:=ncol;
wks.name$ = TDS;
wks.col1.width = 8;
col(A)[L]$ = Wavelength;
col(A)[U]$ = nm;
col(A) = wavelength;

//Loop through samples for TDS
for(int i = 2; i <= samples + 1; i++)
{
page.active = 1;
range abs95 = wcol(i);
col(A)[L]$ = Wavelength;
col(A)[U]$ = nm;
page.active = 2;
range abs4 = wcol(i);
col(A)[L]$ = Wavelength;
col(A)[U]$ = nm;
page.active = 3;
wcol(i) = abs95 - abs4;
wcol(i)[L]$ = TDS;
}

//Normalize data
int lastCol = samples + 1;
rnormalize irng:=2:wcol(lastCol) method:=range;


//Zero data
wks.addCol();
for(int i = lastCol + 2; i <= lastCol + 1 + samples ; i++)
{
range zero = wcol(i)[1:59];
int zeroCol = samples + i + 1;
wcol(zeroCol) = wcol(i) - mean(zero);
wcol(zeroCol)[L]$ = Zeroed;
}
```

```
//New sheet for sample concentrations
newsheet;
wks.name$ = Concentrations;
col(A)[L]$ = Sample;
col(B)[L]$ = [DNA];
col(B)[U]$ = uM;

//Loop through samples for concentrations
for(int i = 2; i <= samples + 1; i++)
{
page.active = 1;
double absat260 = cell(179, i)$;
double absat345 = cell(9, i)$;
double baselineCorrected = absat260 - absat345;
page.active = 4;
cell(i-1, 2) = baselineCorrected / 1.73 * 10;      // Replace 1.73 with extinction
coefficient (__*10^5 M^-1cm^-1)
}
```

CD scan averaging script (Origin LabTalk)

```
/*
Script for organizing & averaging CD scan data (7 scans)

---
Instructions:

-Start with worksheet or book with multiple sheets of raw CD scan data
       -If multiple worksheets, script will automatically loop through all sheets for
you
-Wavelength in col(A), CD signal in col(B)
-Data should start on row 1 with no extraneous text above
-Do not change any data/text below the first data row
-Need to edit # of samples in first line
-If using for buffer scans, comment out or delete last 2 segments

---
Output:

-Labeled columns with wavelength, CD signal from each scan, average CD signal
-Blank column for inputting buffer scan info (fill in cuvette # in comments if desired)
-Blank book for inputting concentration info (needed for CD scans - processing script)

---
*/

int samples = 8;      // Change to # samples

loop(i, 1, samples)
{

//Extract wavelength & CD signal data
page.active = i;
range wavelength = 1[1:111];
range scan1 = 2[1:111];
range scan2 = 2[115:225];
range scan3 = 2[229:339];
range scan4 = 2[343:453];
range scan5 = 2[457:567];
range scan6 = 2[571:681];
```

```
range scan7 = 2[685:795];

//Create & name columns;
col(C)[L]$ = Wavelength;
col(C)[U]$ = nm;
wks.col3.type = 4;
col(D)[L]$ = CD signal;
col(D)[C]$ = Scan 1;
col(E)[L]$ = CD signal;
col(E)[C]$ = Scan 2;
col(F)[L]$ = CD signal;
col(F)[C]$ = Scan 3;
col(G)[L]$ = CD signal;
col(G)[C]$ = Scan 4;
col(H)[L]$ = CD signal;
col(H)[C]$ = Scan 5;
col(I)[L]$ = CD signal;
col(I)[C]$ = Scan 6;
col(J)[L]$ = CD signal;
col(J)[C]$ = Scan 7;

//Move scan data;
col(C) = wavelength;
col(D) = scan1;
col(E) = scan2;
col(F) = scan3;
col(G) = scan4;
col(H) = scan5;
col(I) = scan6;
col(J) = scan7;
del Col(A);
del Col(B);
wks.col1.width = 8;

//Average CD scans
col(K)[L]$ = CD signal;
col(K)[C]$ = AVG;
rowstats irng:=2:8 mean:=9 sd:=<none>;

//New column for buffer CD signal info
col(L)[L]$ = CD signal;
col(L)[C]$ = Cuvette;

}


//New book for sample concentrations;
//newbook;
//newbook name:="Concentrations";
//col(A)[L]$ = Sample;
//col(B)[L]$ = [DNA];
//col(B)[U]$ = uM;
```

CD scan processing script (Origin LabTalk)

```
/*
Script for processing CD scan data (7 scans)

---
Instructions:
```

```
-Start with worksheet or book with multiple sheets of CD scan data processed using CD
scans - average script or formatted in the same manner
     -If multiple worksheets, script will automatically loop through all sheets for you
-Manually paste buffer CD signal in the indicated (last) column
-Fill in "Concentrations" book in sheet order (can copy/paste output from TDS script if
used)
-Need to edit 1st line with # samples and middle line with SHORT NAME

---
Output:

-Labeled columns with baseline corrected CD signal, zeroed CD signal, and molar
ellipticity
-Molar ellipticity for samples in sheet order in a new sheet

---
*/


int samples = 10;                           // Change to # samples
samples++;

for(int i = 1; i < samples; i++)
{
page.active = i;

col(M)[L]$ = Baseline corrected;
col(M)[C]$ = Subtract buffer;
col(M) = col(K) - col(L);

col(N)[L]$ = Zeroed;
col(N)[C]$ = Subtract AVG 1st 10 pts;
range zero = col(M)[1:10];
col(N) = col(M) - mean(zero);

col(O)[L]$ = Molar ellipticity;
col(O)[U]$ = M\+(-1)cm\+(-1);
%W = ["Conc"]Concentrations!cell(i, 2)$;    // Change to ["BookName"]SheetName
col(O) = col(N) / (0.03298 * %W);
}

//Create overlay
range wavelength = 1[1:111];
newsheet cols:=samples;
wks.name$ = Overlay;
col(A)[L]$ = Wavelength;
col(A)[U]$ = nm;
col(A) = wavelength;

//Loop through sheets for molar ellipticity
for(int i = 1; i < samples; i++)
{
page.active = i;
range de = 13;
page.active = samples;
int deCol = i+1;
wcol(deCol) = de;
wcol(deCol)[L]$ = Molar ellipticity;
wcol(deCol)[U]$ = M\+(-1)cm\+(-1);
}
```

CD melt script (Origin LabTalk)

```
/*
Script for organizing & initial processing of CD melt data (4-95 C)

---
Instructions:

-Start with book(s) with single worksheet in which raw CD melt data imported
       -If multiple books, script will automatically loop through all books for you
-T in col(A), CD signal in col(B)
-Data should start on row 1 with no extraneous text above
-Script will ignore any extraneous text below data so that is fine

---
Output:

-Labeled sheets & columns
-Column with T in K
-Separate melt & cool sheets with derivative column

---
*/

doc -ef W
{

//Name raw data & insert T (K) column
wks.name$ = raw;
col(A)[L]$ = T;
col(A)[U]$ = C;
col(B)[L]$ = CD signal;
wks.insert(T);
col(T) = col(A)+273.15;
col(T)[L]$ = T;
col(T)[U]$ = K;

//Extract melt & cool data
range meltT = 1[1:92];
range meltCD = 3[1:92];
range coolT= 1[93:184];
range coolCD = 3[93:184];

//Creates sheet for melt data
newsheet;
wks.name$ = melt;
col(A)[L]$ = T;
col(A)[U]$ = K;
col(B)[L]$ = CD signal;
range mT= 1[1:92];
range mCD = 2[1:92];
mT = meltT;
mCD = meltCD;
differentiate iy:=col(B);

//Creates sheet for cool data
newsheet;
wks.name$ = cool;
col(A)[L]$ = T;
col(A)[U]$ = K;
col(B)[L]$ = CD signal;
```

```
range cT= 1[1:92];
range cCD = 2[1:92];
cT = coolT;
cCD = coolCD;
differentiate iy:=col(B);
}
```

Copy of published work

"Quadruplexes in 'Dicty': Crystal Structure of a Four-Quartet G-Quadruplex Formed by G-Rich Motif Found in the *Dictyostelium Discoideum* Genome"

Reference:

Guédin, A.; Lin, L. Y.; Armane, S.; Lacroix, L.; Mergny, J.-L.; Thore, S.; Yatsunyk, L. A. Quadruplexes in 'Dicty': Crystal Structure of a Four-Quartet G-Quadruplex Formed by G-Rich Motif Found in the *Dictyostelium Discoideum* Genome. *Nucleic Acids Research* **2018**, *46* (10), 5297–5307.