Method Design for Separation of Non-Canonical DNA Structures via Capillary Electrophoresis

Erin V. Chen, ’23

Follow this and additional works at: https://works.swarthmore.edu/theses

Part of the Chemistry Commons, and the Engineering Commons

Recommended Citation

This work is licensed under a Creative Commons Attribution-NonCommercial-Share Alike 4.0 International License. Please note: the theses in this collection are undergraduate senior theses completed by senior undergraduate students who have received a bachelor's degree. This work is brought to you for free by Swarthmore College Libraries' Works. It has been accepted for inclusion in Senior Theses, Projects, and Awards by an authorized administrator of Works. For more information, please contact myworks@swarthmore.edu.
Method Design for Separation of Non-Canonical DNA Structures via Capillary Electrophoresis

Erin Chen

Advisors: Carr Everbach (Engineering), Liliya Yatsunyk (Chemistry)

Spring 2023

Table of Contents

Abstract 3
1. Introduction 4
   1.1 Non-canonical G-quadruplex DNA structures and their diversity 4
   1.2 Electrophoresis methods for separation 8
   1.3 Method development goals 11
2. Methods 13
   2.1 DNA sequence selection and design 13
   2.2 Capillary Electrophoresis (CE) and Laser-Induced Fluorescence (LIF) detection 14
   2.3 UV-Vis spectroscopy: Thermal Difference Spectra (TDS) 15
   2.4 Circular Dichroism (CD) spectroscopy 15
   2.5 Native polyacrylamide gel electrophoresis (Native PAGE) 16
3. Results 18
   3.1 Method design for confirmation of folded G-quadruplex structures in CE 18
      3.1.1 Fluorescent dye testing and analysis 18
      3.1.2 Sample ionic strength effects 21
      3.1.3 Kinetic folding studies of T9 DNA 23
      3.1.4 Voltage, sample volume, and capillary length effects on separation 26
   3.2 Method design for model DNA 28
      3.2.1 Species identification in low concentration environments 28
      3.2.2 Species comparison under various dyes 30
      3.2.3 Buffer ionic strength effects on separation 32
4. Discussion 34
   4.1 Reproducibility and reliability of results 34
   4.2 Effective method designs for future use 36
   4.3 Project progress and future recommendations 37
   4.4 Engineering design 39
Appendices 42
   Appendix A: Acknowledgements 42
Abstract

G-quadruplex DNA structures are highly relevant subjects of biochemical study, implicated in human diseases and cancer. Conventional methods for G-quadruplex DNA structure separation are time-consuming, environmentally unfriendly, and low throughput. Thus, there is a pressing need for alternative methods of separation to be developed for G-quadruplex DNAs. Here we present work in designing a separation method using the capillary electrophoresis instrument. Capillary electrophoresis is a highly automated, reusable, and quantitative technique that separates analytes based on their size and charge. A capillary method for G-quadruplex separation would be a valuable analytical tool to supplement the biophysical methods available by addressing the pitfalls of conventional separation techniques. Through a series of informed experiments, we attempted to design a broadly applicable capillary electrophoresis method for G-quadruplex structural separation. Although a final method was not completed, we have engineered coupled experimental techniques to aid in the overall design process. Additionally, we demonstrated the presence of multiple structures within the capillary via non-symmetric peaks in the electropherogram.
1. Introduction

Here we present work on method design for separation, identification, and quantification of non-canonical G-quadruplex DNA structures via optimization and manipulation of the capillary electrophoresis technique. To understand the importance of studying G-quadruplex DNA and the utility of developing the capillary electrophoresis method for G-quadruplex separation, we first present a background on these topics.

1.1 Non-canonical G-quadruplex DNA structures and their diversity

Deoxyribonucleic acid, or DNA, is one of the building blocks of life and an important part of the central dogma of biology. Canonical Watson-Crick DNA adopts a two-stranded antiparallel duplex structure formed from a polymer of four nucleotide building blocks: adenine (A), thymine (T), cytosine (C), and guanine (G). The strands of DNA have a specific directionality (ends are labeled either 5’ or 3’) that correspond to the orientation of the sugar-phosphate backbone that is linked to the bases. In these canonical DNA, the base pairing pattern is such that adenine pairs with thymine and cytosine pairs with guanine (Fig. 1).

Figure 1. Schematics of DNA bases and backbone directionality. Canonical Watson-crick base pairing is shown in the schematic. The colors are used to highlight the differences between the bases.
However, DNA is conformationally diverse, and there exists a wide range of DNA structures such as triplex, hairpin, i-motif, and G-quadruplex DNA structures, among others. These structures fall under the umbrella of non-canonical DNA, an extremely broad group of structures united by breaking canonical Watson-Crick pairing motif and structural expectations. Shown to play important biological functions in the human body, these non-canonical DNA structures are linked to diseases and are highly relevant subjects of study [1].

G-quadruplexes (GQs) are in the center of a rapidly growing biochemical field, with over 30,000 journal articles published on the topic in 2022. GQs are unique due to their structure of four guanine DNA bases arranged in a square planar tetrad conformation, called a G-tetrad (Fig. 2A). The guanine bases in G-tetrads are held together by Hoogsteen hydrogen bonds, and further stabilized by (1) the stacking of multiple tetrads and (2) metal cations situated in the central channel of the G-tetrads (Fig. 2B).

**Figure 2.** (A) Schematic of square planar arrangement of guanine bases to form a G-tetrad. Red dashed lines indicate Hoogsteen hydrogen bonds. Red numbering indicates conventional carbon atom numbering. (B) An example of G-tetrads stacking to form a GQ [2].

GQ structure is highly studied as sequences with GQ-forming potential are often located in regions of the genome implicated in cancer and other diseases, such as telomeres, centromeres,
and oncogene promoters. Over 700,000 sequences with GQ forming potential have been identified in the human genome [3], and the presence of GQ structures in vivo has been repeatedly verified [4]. The exact biological role of GQs is unclear, but research suggest a diversity of potential functions. For example, GQs can pose a hindrance to the regular function of cells. Unresolved, the presence of GQs in the genome could lead to double stranded breaks during DNA replication that may lead to cell death or issues in replication, which would associate it with genomic diseases [5]. However, the presence of GQs can also be beneficial. The mechanism of stabilizing GQs in cancer cells to induce apoptosis is being studied in the context of developing new cancer therapeutics [6]. Much evidence points to the natural occurrence of GQs in the body where GQs act as regulatory agents [7]. In the same vein, GQs have also been studied within the context of being an aptamer or drug component, rather than the target, such as aptamers to target HIV envelope glycoproteins [8].

GQs are structurally diverse. Given that G-tetrads are the defining characteristic of GQs, there are numerous ways guanine bases can come together to form tetrads. These heteromorphic structures can vary in number of G-tetrads, molecularity (the number of unconnected strands that form a GQ), strand direction (parallel, antiparallel, or hybrid), handedness (left-, right-handed, or both), loop type, and oligomeric states (Fig. 3). In addition, other structural features add diversity to GQ folds – bulges and overhangs.
**Figure 3.** Summary of variable GQ features. Molecularity – different colors represent different strands. Strand orientation – antiparallel with two strands pointing up and two strands pointing down; parallel with all strands pointing in the same direction; and hybrid or 3+1, with three strands down and one strand up. Loop type – three primary loop types in GQs. Monomer vs. dimer – different colors represent different monomeric units. Handedness – colors highlight the groove of the GQ. [2], [9], [10]

The conformation, structural diversity, and equilibrium between GQ structures for a DNA sequence are highly dependent on environmental factors. The presence of ligands can lead to conformational selectivity [11], while the presence of other molecules, such as molecular crowding agents, can also alter the solvent behavior and influence GQ folding [12]. However, the most influential component to GQ formation is the presence of a cationic species of appropriate size. The GQ assembly necessitates a positively charged ion in the central channel to stabilize the structure (**Fig. 2**). The ion positions itself equidistantly between two adjacent tetrads, and thus only a few ions are sized to fill this position. Potassium and sodium are the most common and are also
present in the human body in significant quantities, however other ions such as barium can also lead to GQ formation but are only found in trace amounts in the body. Thus, it is important to prepare DNA samples in ionic conditions that promote stable GQ folding.

1.2 Electrophoresis methods for separation

Electrophoresis methods for GQ analysis typically involve the conventional native polyacrylamide gel electrophoresis method. This is a highly time-consuming and resource-intensive experiment, in which an experimentalist must spend multiple hours to prepare a non-reusable gel matrix (oftentimes requiring hazardous chemicals such as the neurotoxin acrylamide) and run 10-15 samples on the gel over the course of 2-4 h. These experiments yield qualitative information on homogeneity and the number of species by separating them by their size, charge, and shape. Sometimes quantitative analysis can be applied if intensity of the bands varies. In general, this method is time-consuming, low throughput, environmentally unfriendly, hazardous, and costly. Thus, there remains a pressure to develop alternate methods of electrophoretic separation to analyze GQ species in a way that addresses the failings of more conventional electrophoretic techniques.

Capillary electrophoresis (CE) is a separation method based on migration rates of differently charged and sized species under an applied electric field. CE is less preparation- and time-intensive than other forms of electrophoresis and aligns with principles of green chemistry and engineering. The sample volumes for CE are on the order of nanoliters and nanomolar and are largely recoverable. The capillary itself is also reusable, and most of the process is automatic thus allowing for extremely efficient and reproducible experiments to be run.

Gel and other zone electrophoresis techniques were popularized in the late 1940s, early 1950s, however CE was not popularized until the 1990s. Many different CE techniques exist, and
a significant body of research is available on protein separation, analysis, and binding. Past research into GQs with CE has largely covered DNA-compound binding studies where DNA is studied in complex with a protein. For example, numerous studies have examined the binding of the protein thrombin with a DNA GQ aptamer [13]–[15]. These studies have used non-equilibrium capillary electrophoresis and fluorescence to assess binding kinetics, stability, and the effect of monovalent cation concentration. Other studies have examined conformational dynamics of GQ folding in the presence of ligand with kinetic capillary electrophoresis paired with mass spectrometry [16].

The CE instrument contains an automated system for high-throughput experiments. Within the instrument, a narrow fused-silica capillary is filled with a conductive background electrolyte, called the separation buffer. Upon injection of an analyte, a potential difference is applied across the capillary, inducing analyte migration. Near the end of the capillary, a detector records signal over retention time. A schematic of the CE instrument is shown in Fig. 4.

The CE separation, based on size and charge, is linked to two dominant forces within the capillary. The first, electroosmotic flow ($\mu_{eo}$), is the net movement of the ions towards the cathode. It originates from the Coulombic force of the applied voltage on the electrolytes of the solution. It only is significant in small channels, as a layer of mobile ions forms and ‘pushes’ everything forward through the capillary. The second, electrophoretic mobility ($\mu_{ep}$), is from the interaction of charges in the analyte and capillary. The capillary is typically negatively charged, thus positive ions will cluster closer to the walls. Increased charge of an analyte increases the effect of electrophoretic mobility. Additionally, larger particles decrease the effect of electrophoretic mobility due to frictional forces. This electrophoretic force is what enables separation of a mixture. As the electroosmotic flow is greater than the effects of electrophoretic mobility, all components
within the capillary are pushed to the detector. The reaction of different types of molecules and their charges to these forces is illustrated in Fig. 4B.

Figure 4. (A) Schematic of CE device. (B) Depiction of how electroosmotic ($\mu_{eo}$) and electrophoretic ($\mu_{ep}$) flow affect differently charged particles. (C) Illustration of layer of mobile ions (bulk solution) form the electric double layer upon voltage application.

CE can be used in combination with multiple types of spectroscopies to enable detection of the compounds as they travel through the capillary. By changing the source and detector out, different spectroscopies can be utilized with the CE instrument. UV-Vis and fluorescence spectroscopy are popular options. However, these methods are limited by their ability to only scan one or two wavelengths in the duration of the experiment.

Fluorescence-coupled CE requires either a naturally fluorescing analyte or an external dye that binds to an analyte to produce detectable signal. DNA does not naturally fluoresce, and although one can purchase fluorescently tagged DNA, it comes at an expense. Fluorescent dyes, however, are more affordable, but introduce additional complications to a system. Dyes bind to DNA under a certain equilibrium, and the effects of different components (i.e. buffer composition, binding method, temperature, etc.) can alter the efficacy of dye binding. For a fluorescent dye,
binding also varies between DNA structures. Therefore, these are important factors when considering the accuracy or quantitative interpretability of a signal in the electropherogram. Multiple methods exist to fluorescently non-covalently tag DNAs with a dye – either the sample is prepared with the dye (pre-column loading), or the dye is integrated in the separation buffer of the capillary (on-column loading). Most dyes fluoresce when bound to DNA, but do not fluoresce alone in solution, a so-called “light switch” property. This “light switch” property enables the detection of the analyte in the capillary under a laser induced fluorescence source and detector at the appropriate wavelengths.

Designing a separation method requires balancing the need for a meaningful separation of peaks against the natural broadening that occurs in the capillary. Capillary transient isotachophoresis (ctITP)-based nonequilibrium capillary electrophoresis of equilibrium mixtures (NECEEM) is a kinetic CE technique that can be used to improve the resolution of an electropherogram. cTIP-based NECEEM involves using leading and terminal ions, contained in the sample and separation buffers, with electrophoretic mobilities greater than or less than the analyte (DNA) in question. These ions aid in focusing the analyte band and, as the sample progresses through the capillary, the ions de-stack and analytes are separated.

1.3 Method development goals

It is difficult to detect and identify GQ conformations in heterogenous samples (Fig. 3). Circular dichroism spectroscopy (which utilizes chiral light) is frequently used to detect signals specific to conformations, but due to signal overlap, it is impossible to infer whether one or more GQ conformations are present in a solution. Methods like native polyacrylamide gel electrophoresis can help separate GQ conformations, but it is not quantitative, is time-consuming, and utilizes non-reusable resources. DNA and GQs are negatively charged, owing to the sugar-
phosphate backbone of the DNA which faces outwards in GQ structures. Thus, CE presents a viable method for studying mixtures of GQ structures as the changes in conformation, sequence, size, and oligomeric state may alter the charge and size properties of GQ structures and thus influence the distance traveled in the capillary electropherogram.

As far as we know, the separation of GQ conformations has not been studied via capillary electrophoresis. Previous capillary electrophoresis work involving GQs predominantly involves characterization of the binding of a well-known two-tetrad antiparallel GQ that targets the protein thrombin [17]. Our goal is to build off the framework established by this past research to design an optimized method for separation. During method development, we are specifically limited by the need for our method to be broadly applicable. Thus, significant constraints include the need for a low-cost, accessible method with minimal environmental impact. Through a series of informed experiments whose results shape the design process, we sought to develop a novel method of using capillary electrophoresis for separation and quantification of conformations of GQ DNA (Fig. 3). To achieve our goal, we optimized the capillary’s dimension, buffer components, and various instrument parameters.

We have demonstrated the presence of non-symmetric peaks within the electropherograms, with the peak structure indicating the presence of multiple conformations. However, we were unable to design a complete, broadly applicable separation method. We present the results of our optimization, in hope that it informs further research into GQ conformational separation via capillary electrophoresis. We additionally present a method for the separation and identification of folded and unfolded G-quadruplexes via capillary electrophoresis coupled with circular dichroism.
2. Methods

2.1 DNA sequence selection and design

Lyophilized oligonucleotides were purchased from Integrated DNA Technologies (IDT; Coralville, IA) with standard desalting purification. We purchased tris (hydroxymethyl) amino methane (Tris-base) and glycine (Gly) from Sigma (St. Louis, MO). DNA was hydrated in doubly-distilled water to 1-4 mM and stored at -80 °C.

The sequences studied in this work (Table 1) were chosen due to their well characterized nature [18], [19] and small number of conformations, lending the sequences to act as model systems.

**Table 1.** DNA sequences studied in this work and their properties.

<table>
<thead>
<tr>
<th>DNA Name</th>
<th>Sequence</th>
<th>Extinction Coefficient (mM(^{-1}) cm(^{-1}))</th>
<th># of Expected Species - Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>T9</td>
<td>G(GGGT)(_3)GGGC</td>
<td>189.2</td>
<td>1 - parallel monomer (+ parallel dimer at high concentrations)</td>
</tr>
<tr>
<td>T1</td>
<td>(GGGGT)(_3)GGG</td>
<td>173.0</td>
<td>2 - parallel dimer, parallel monomer</td>
</tr>
<tr>
<td>T7</td>
<td>T(GGGGT)(_3)GGGT</td>
<td>189.0</td>
<td>1 - parallel monomer</td>
</tr>
<tr>
<td>TET25</td>
<td>GTT(GGGGT)(_3)GGGG</td>
<td>240.3</td>
<td>3 - parallel monomer, hybrid monomer, and one other species</td>
</tr>
<tr>
<td>TET26</td>
<td>GTT(GGGGT)(_3)GGGGT</td>
<td>248.8</td>
<td>3 - parallel monomer, two other monomeric species</td>
</tr>
<tr>
<td>S3</td>
<td>(ATGGA)(_4)</td>
<td>217.4</td>
<td>Non-canonical hairpin alone. Canonical duplex when combined with S3D.</td>
</tr>
<tr>
<td>S3D</td>
<td>(TCCAT)(_4)</td>
<td>176.6</td>
<td>Unfolded alone. Canonical duplex when combined with S3.</td>
</tr>
</tbody>
</table>
Samples for analysis were prepared at 2 μM in Tris-HCl buffer, composed of 50 mM Tris-base titrated to pH 8.2 with 1 M HCl. When relevant, 0-100 mM KCl was added to the samples. DNA was diluted with buffer to a desired concentration and annealed by heating at 95 °C for 5 minutes followed by slow cooling to room temperature over a period of 4-5 hr. These samples were stored overnight at 4 °C before further use. All biophysical data were processed using Origin (Origin Version 9.1, OriginLab Corporation, Northampton, MA, USA.).

2.2 Capillary Electrophoresis (CE) and Laser-Induced Fluorescence (LIF) detection

Capillary transient isotachophoresis (ctITP)-based nonequilibrium capillary electrophoresis of equilibrium mixtures (NECEEM) was performed on a Beckman Coulter P/ACE MDQ CE System with laser-induced fluorescence (LIF) detection (Ar-ion laser excitation at 488 nm, 520 nm long pass, and 488 nm notch filters). 32 KARAT software was used for CE control and analysis. The leading ion is in the sample buffer (chlorine anion) and the terminal ion is in the separation buffer (glycine anion).

Tris-Gly separation buffer was formulated from 25 mM Tris-base and 192 mM glycine, unless otherwise specified. The pH was recorded (approximately pH 8.5). All buffers were filtered with a 0.20 μm filter prior to use. DNA was labeled “on column” via integration of the fluorescent dye in the separation buffer. Fluorescent dyes SYBR Gold, Green 1, and Green 2 [20]–[22] were purchased from Life Technologies (Carlsbad, CA). SYBR Gold was used at 1:100,000, while SYBR Green 2 was used at ratios 1:10,000 or 1:100,000.

The capillary was uncoated fused-silica with 50 μm internal diameter (Polymicro Technologies, Phoenix, Arizona), unless otherwise specified. Two lengths were used: 60.2 cm total length, 50.0 cm effective length, from inlet to detector; or 30 cm total length, 20 cm effective length, from inlet to detector. The capillary and the inlet and outlet vials were filled with Tris–Gly.
separation buffer. We injected 2 µM DNA samples onto the capillary at 0.4 psi for 5 sec and separated at 5.0 kV, unless otherwise specified. Some samples were diluted to 200 nM in Tris-HCl buffer and injected at 4 psi for 5 sec. A one-hour rinse series (H₂O, 1 M NaOH, and separation buffer) and sample blank were run before experimentation for the day; a half hour flush (H₂O and 1 M NaOH) terminated experimentation for the day; and the capillary was stored in water when unused. Separation buffer was made fresh approximately once a week.

2.3 UV-Vis spectroscopy: Thermal Difference Spectra (TDS)

UV-Vis spectra were collected on an Agilent Cary 3500 UV–Vis spectrophotometer equipped with a Peltier temperature controller (± 0.5 °C error) using 1 cm pathlength quartz cuvettes. Data were collected using 1 nm intervals, 0.02 s averaging time, 2 nm spectral bandwidth, no baseline correction, and 220–350 nm range.

Concentrations of DNA were determined from scans of unfolded DNA at 90 °C via Beer’s law, \( A = \varepsilon l c \), where \( A \) is UV-Vis absorbance, \( \varepsilon \) is the extinction coefficient of the DNA, \( l \) is the cuvette path length, and \( c \) is the concentration of the sample.

Thermal difference spectra experiments involved a UV-Vis scan at 20 °C (corresponding to a state in which the DNA is folded) and at 90 °C (corresponding to a state in which the DNA is unfolded) after 10 minutes of equilibration. Data at 20 °C were subtracted from data at 90 °C and zeroed with data from 341-350 nm, yielding the TDS signatures. These signatures have characteristic peaks and troughs that correspond to different non-canonical DNA structures. The characteristic feature for a GQ is a trough at 295 nm [23].

2.4 Circular Dichroism (CD) spectroscopy

CD spectra were collected on samples ~2 µM in 1-cm quartz cuvettes at 20 °C on an Jasco J-1500 instrument with a Peltier thermocontroller (± 0.5 °C error). Instrument parameters include
a 2 nm bandwidth, 1 s digital integration time, 200 nm/min scanning speed over a 220-330 nm window. Five scans were recorded and averaged. Data were processed via subtraction of a cuvette baseline (solvent: water), zeroing with 321-330 nm data, and calculation of molar ellipticity using the following formula, where \( l \) is the cuvette path length in cm and \( c \) is the concentration of the sample in \( \mu \text{M} \).

\[
\text{Molar Ellipticity} \ (M^{-1} \text{cm}^{-1}) = \frac{\text{CD Signal}}{0.03298 \cdot \ l \cdot \ c}
\]

Kinetic CD experiments were conducted such that a spectrum, as described above, was taken of the initial sample; 1 M KCl stock was added to the final concentration of 10 mM KCl; the samples were inverted and mixed; and the time-dependent program was initiated after the cuvette was reinserted in the instrument. The time-dependent program consisted of single scans, with parameters as described above, recorded every one or five minutes along the full 220-330 nm window for at least two hours. Singular value decomposition through MATLAB was performed on the full CD spectra to obtain the singular values and left- and right- singular vectors. Singular value decomposition is a more robust method of accounting for change over an independent variable, as it considers full spectra rather than the change of a single point of the spectra. The right-singular vectors (v1, v2, \ldots) were plotted against time.

2.5 Native polyacrylamide gel electrophoresis (Native PAGE)

Native polyacrylamide gel was prepared at 15% in Tris-borate-EDTA (TBE) buffer supplemented with 10 mM KCl. TBE with 10 mM KCl was used as a running buffer. The gel was premigrated for 30 min at 150 V at room temperature. Annealed DNA samples were prepared in Tris-HCl buffer to either 10 \( \mu \text{M} \) for low concentration samples or 50 \( \mu \text{M} \) for high concentration samples. Samples were weighted 1:3.33 v/v with 9% sucrose, loaded (25 \( \mu \text{L} \) for low concentration
samples or 10 µL for high concentration samples) onto the gel, and run for ~2 h at 150 V. Oligothymidylate markers 5’ dTn (where n = 15, 24, 30, 57) were used as migration standards.

DNA bands were visualized in two ways: (1) with only StainsAll or (2) first with one of three fluorescent stains, rinsed, and then with StainsAll [24]. Filtered StainsAll was stained for 7-15 minute and destained in ddH₂O. The fluorescent stains were 1:10,000 SYBR Gold, SYBR Green 1, or SYBR Green 2 [20]–[22] prepared in running buffer with a 30 minute staining and no destaining. Gels stained with StainsAll were photographed with a smartphone camera over a light box. Gels stained with the fluorescent SYBR dyes were imaged with a GE Healthcare Amersham Imager 680 imager with excitation with 460 nm and detection at 525 nm. Depending on the gel, exposure time ranged from 0.1 - 2 seconds, chosen to optimize visibility of bands.
3. Results

Here we present a summary of relevant observations made from experimental testing, in search of an optimized separation method for GQ structural separation under a capillary electrophoresis methodology. A full summary of experimental results, chronologically, is detailed in Appendix C-F.

3.1 Method design for confirmation of folded G-quadruplex structures in CE

To develop an appropriate separation method, we first needed to ensure a reproducible, detectable signal appeared under the desired cITP-based NECEEM method with LIF detection upon binding of DNA with a fluorescent dye. Moreover, we wanted to ensure that the detectable signal corresponded to a folded GQ structure throughout the experiment. Our initial conditions were obtained from work on GQ aptamers binding the thrombin protein [25].

Most results in Results 3.1 center the representative model DNA of T9, a homogeneous DNA sequence expected to fold into a single parallel monomeric species in the presence of potassium chloride (Table 1, Fig. 10). In capillary electrophoresis experiments, we expect to observe only one peak for this DNA. If there is no potassium chloride, we expect only one peak that corresponds to an unfolded species.

3.1.1 Fluorescent dye testing and analysis

Three DNA-binding fluorescent dyes were analyzed for use in this project: SYBR Gold, SYBR Green 1, and SYBR Green 2. Each is reported to bind different DNA structures and structural features differently [20]–[22], thus the effects of these different dyes on representative structural samples were tested with native PAGE (Fig. 5). The structural samples included the oligothymidylate markers (expected as unfolded structures), a T9 GQ sequence without KCl (expected as an unfolded structure, but likely folds into a GQ structure due to the presence of 10
mM KCl in the running buffer), a T9 GQ sequence with KCl (predominantly folds into a single monomeric GQ structure, may have minor dimeric properties), and a combined sample of S3 and S3D sequences (folds into representative Watson-Crick duplex).

**Figure 5.** Fluorescent dye binding analysis via native PAGE. Fifteen percent gel was prepared in 1 × TBE supplemented with 10 mM KCl. The gels were first stained with the appropriate fluorescent dye, visualized, rinsed, and then re-stained with Stains-All and visualized once more. Size markers correspond to dTₙ (n = 15, 24, 30, 57). The DNA samples were prepared the following way: lane 1 is 50 µM T9 in Tris-HCl with 0 mM KCl; lane 2 is 50 µM T9 in Tris-HCl with 10 mM KCl; lane 3 is 50 µM S3 and 50 µM S3D in 10 mM Tris-acetate at pH 8.3 with 50 mM potassium acetate and 5 mM magnesium acetate. Exposure times were 0.9 s, 2 s, and 0.1 s.

We observe that SYBR Gold binds all DNA structures strongly, with some preferential binding to unfolded and duplex structures (Fig. 5). SYBR Green 1 binds duplex DNA strongly, with very weak binding to GQ DNA and almost no binding to dT markers. SYBR Green 2 binds duplex DNA strongly, GQ DNA to a moderate extent, and no binding to dT markers. Connecting these data to the performance of the dyes in a capillary electrophoresis setting, we initially decided to use SYBR Gold as the fluorescent dye used in on-column labeling within the capillary. However, we grew uncertain of whether the signal in the capillary corresponded to a folded or
unfolded species. Figure 6 illustrates the electropherograms for T9 under 0 and 10 mM KCl. Both electropherograms look highly similar within a trial, yet do not yield symmetric peaks. In both 0 and 10 mM cases, the Trial 1 data show a shallow peak following the main peak, while the Trial 2 data show an asymmetric mean peak and shallow trailing shoulder. This result potentially suggested that the T9 sequences were experiencing multiple conformations within the capillary, which would contribute to the asymmetry of the electropherogram.

Thus, to test if the T9 sequences were unfolding or experiencing an equilibrium of folded and unfolded species within the capillary, we conducted experiments utilizing the SYBR Green 2 dye, which was shown in Fig. 5 to have no binding to the unfolded oligothymidylate markers. We expected that under SYBR Green 2, the T9 0 mM KCl sample would exhibit no peak and the T9 10 mM KCl sample would exhibit a singular peak. The results of this experiment are shown in Fig. 6. We observe two partially overlapping peaks. In the 0 mM KCl sample, these peaks are of approximately equal intensity. In the 10 mM KCl sample, the earlier peak possesses increased intensity. These results did not agree with our expectations. However, studies of oligothymidylate markers and other unfolded DNA sequences (Appendix C-D) in the capillary indicate that SYBR Green 2 binding to single-stranded DNA is not negligible and does present itself in the CE experiment, although it does not present itself in the native PAGE experiment (Fig. 5). This occurrence attests to the high sensitivity of the CE technique, as well as potentially explains the dual peak observed if one peak is the unfolded state and one peak is the folded parallel monomeric GQ.
Figure 6. Effects of different fluorescent dyes. Representative electropherograms of T9 DNA (A) annealed with 0 mM KCl, with an unfolded structure expected and (B) annealed with 10 mM KCl, with a GQ structure expected. Samples were injected at 0.4 psi for 5 sec and a 5 kV voltage was applied across the 20 cm effective length capillary. Data are from Exp. 5.1 (SYBR Gold Trial 1), 6.1 (SYBR Gold Trial 2), and 8.1 (SYBR Green 2). See Appendix C-D for more details.

3.1.2 Sample ionic strength effects

Even ignoring SYBR Green 2’s potential to bind to unfolded DNA sequences, the dual peak expression is still concerning (Fig. 6). We previously expected one dominant signal for each sample. Thus, we probed the possibility that our 0 mM KCl sample was impure and whether an equilibrium between folded and unfolded states existed within our 10 mM KCl sample.

Thus, the effects of sample ionic strength, predominantly studied through increasing amounts of KCl, were examined in systems of T7 and TET25. Supplementary experiments of circular dichroism and thermal difference spectra were performed to evaluate if there was an impurity in the samples. The presence of a trough at 295 nm in the thermal difference spectra indicates a G-quadruplex structure within the sample, and this is only observed for the >0 mM KCl samples as expected (Fig. 7C). Additionally, circular dichroism spectra (Fig. 7D) involve characteristic peaks and positions that correspond to particular strand orientations of the DNA
(parallel, antiparallel, hybrid). We observe the lack of these peaks/troughs in the 0 mM T7 and TET25 samples. In the 1 mM T7 sample, we observe a strong peak at 260, indicative of a parallel conformation as expected. In the 1 mM TET25 sample, we observe a spectra with peaks at 295 and 260 nm, indicating a hybrid structure or a mix of parallel/antiparallel/hybrid structures within solution, also consistent with our expectations. Thus, this experiment has confirmed the behavior of our samples in Tris-HCl buffer against well-established experimental methods. Similar supplemental experimental data for T9 are given in Appendix C-F.

Figure 7. Effects of sample ionic strength. Electropherogram (A) T7 DNA and (B) TET25 DNA annealed with 0-100 mM KCl. Samples were diluted to 200 nM in buffer before injection at 4 psi
for 5 sec and application of 18 kV voltage across the 75 μm diameter, 50 cm effective length capillary. (C) Thermal difference spectra of select undiluted samples used in CE. (D) Circular dichroism spectra of select undiluted samples used in CE. Data are from Exp. 4.1. See Appendix C-F for more details.

We were additionally able to use these results to understand the influence of ionic strength on the performance of the samples in capillary electrophoresis (Fig. 7A-B). We observe a general trend that increasing the ionic strength increases the retention time of the samples. This result agrees with predictions that a higher ionic strength will lead to a slower migration given its heightened ability to conduct current. However, the T7 10 mM and 100 mM samples and the TET25 1 mM and 10 mM samples are inconsistent with the trend. The non-linearity of the trend may be a result of general instrument error.

With regards to peak shape, we observe an increasing amount of asymmetry with increasing ionic strength. This phenomenon may manifest due to the increased retention time of samples with greater ionic strength: the samples are given more time to interact with the capillary walls, thus improving the separation, but also broadening the signals. These results point to the multiple conformations of TET25 present in solution.

3.1.3 Kinetic folding studies of T9 DNA

Thus, to more definitively examine the dual peak nature of our CE electropherograms (Fig. 6), we designed a kinetic experiment that probes the identity of the peaks. By adding additional potassium chloride to a sample prepared with 0 mM KCl that has exhibited dual peak behavior on the electropherogram, we hope that the GQ-peak will grow in magnitude and the unfolded peak will shrink with magnitude. The results of this method are shown in Fig. 8A – indeed, an increase in the earlier peak and a decrease in the later peak is observed over a period of ~45 minutes, until a homogeneous peak is observed. We can conclude that the earlier peak corresponds to a folded GQ conformation in T9, while the later peak corresponds to an unfolded conformation.
We have coupled such results with the same experimental setup, simply executed with circular dichroism spectroscopy (Fig. 8B). The same results are obtained. With simply the addition of an appropriately sized cation species, GQ DNA folds from unfolded DNA without the need to anneal the DNA. The CD data show that a previously unfolded T9 spectra rapidly converts to the parallel conformation. These data were processed through singular value decomposition to analyze the change over time, and the v1 vector is plotted in Fig. 8C. A qualitative analysis of the kinetic progression of the v1 vector compared to the CE electropherograms is highlighted with the colored lines. At ~45 minutes, the v1 vector is very close to the asymptotic limit of folded GQ structures.

Thermal difference spectra likewise confirm these results, showing high similarity between the T9 sample annealed with 10 mM KCl and the T9 sample with 10 mM KCl added after (Fig. 8D).

Thus, we have demonstrated the efficacy of a kinetic coupled CE-CD method to identify CE peak identity between folded and unfolded states. This method is an important tool that other researchers may be able to use in their GQ-CE experiments. It effectively takes advantage of the changing equilibrium upon KCl addition to accurately identify peaks. This development can likely be pushed further – first, via more accurately aligning timing of CE and CD experiments; and second, in use not only in assessing folded/unfolded states, but also assessing the changing equilibrium between different GQ conformations upon KCl addition.
Figure 8. Kinetic effects of KCl addition to unfolded GQ sample. (A) Electropherograms of T9, 0 mM KCl pre- and post- the addition of 10 mM KCl. Multiple injections into the capillary were taken from the post-KCl sample, approximately once every 14 minutes. Samples were injected at 0.4 psi for 5 sec and a 5 kV voltage was applied across the 20 cm effective length capillary. (B) Circular dichroism data of T9, 0 mM KCl sample pre- and post- the addition of 10 mM KCl. Multiple scans of the cuvette were taken from the post-KCl sample, once every 5 min for 0-170 minutes. Purple coloring corresponds to earlier scans, red coloring corresponds to later scans. (C) Singular value decomposition of CD data. The colored lines correspond to the samples in the CE, beginning with 0 min post KCl addition. (D) Thermal difference spectra of T9 pre-, post- and annealed with 10 mM KCl. Note the black and blue lines are almost perfectly overlaid. Data are from Exp. 9.1. T9 with 10 mM KCl data are from Exp. 8.1. See Appendix C-F for more details.
3.1.4 Voltage, sample volume, and capillary length effects on separation

Here we also present the initial optimization of methodology, executed during analyses of the simple T9 system. Appropriate voltage, sample volume, and capillary length are all vital components of the system. Figure 9A demonstrates the effects of injection volume on the separation and clarity of peaks. A T9 0 mM sample was injected at regular volume and double such volume, and we see the doubled injection volume significantly increases the peak fluorescent intensity, but also reduces the resolution of the separation. The early shoulder of the regular volume injection’s peak is not resolved in the double volume injection. Thus, we can conclude that the balance between peak resolution and meaningful fluorescence intensity is important to method development. (It is of note that the balance between separation and intensity can also be influenced by increased dye concentration and is a potential optimization target.)
Figure 9. Effects of sample injection volume and voltage on separation. Electropherograms of (A) T9, 0 mM KCl DNA in SYBR Green 2 with a doubled sample injection volume, (B) T9, 0 mM KCl DNA in SYBR Green 2 under two separation voltages, and (C) T9, 10 mM KCl DNA in SYBR Gold under three separation voltages. Unless otherwise specified, samples were injected at 0.4 psi for 5 sec and a 5 kV voltage was applied across the 20 cm effective length capillary. Injection data from Exp. 11.1; voltage data from Exp. 11.1 (4, 5 kV) and Exp. 5.1 (5, 8, 10 kV).

With regards to voltage, we observe expected trends as the voltage increases (Fig. 9B-C). Increased voltage leads to shortened retention times, which makes sense as a stronger potential across the capillary will increase the strength of the electroosmotic force. These effects were tested for both SYBR Gold and SYBR Green 2 systems.
The decreased retention time of a higher voltage can be counteracted by lengthening the capillary. However, it is very costly and time consuming to test capillary lengths, as completely new capillaries must be created and equilibrated to test different lengths. Oftentimes, creating new capillaries, especially at specialized lengths, requires independent cartridges and individualized tubing components. Thus, only two effective lengths, 20 cm and 50 cm, were tested. The 50 cm effective length was tested predominantly under 18 kV conditions with SYBR Gold dye. Upon the identification of SYBR Green 2 as a more promising target, the 50 cm capillary was not used in experiments to yield data specific to this GQ system that could be compared against the 20 cm capillary. However, the effects of a longer capillary are well known [26], and we ultimately deemed the 20 cm system to be a better choice for teasing out the nuances of the peaks without broadening the signals to uninterpretable extents.

3.2 Method design for model DNA

Here we detail the results of our optimization experiments in pursuit of a method to separate different strand orientations of GQ DNA. We have thus far confirmed the presence of folded GQ structures within the capillary and turned our sights to more complicated model systems of T1, T7, TET25, and TET26 in hopes of separating specific conformations.

3.2.1 Species identification in low concentration environments

First, it was important to recharacterize our model systems under a lower DNA concentration. The DNA used in our capillary electrophoresis experiments are injected at a 2 µM concentration and may further dilute while traveling through the capillary. Thus, it is imperative to confirm the number of expected species in the capillary. Figure 10 illustrates native PAGE experiments at a 10 µM low DNA concentration sample compared against a 100 µM sample, a concentration closer to previous characterization of these sequences [19]. The native PAGE
technique precludes low concentrations of DNA as low concentrations require increased sample loading volumes, of which there is a maximum limit. From these data, we observe that T9 exists predominantly in the monomer state at low concentrations and a mixture of dimer and monomer state at high concentrations, as evidenced by the presence of one band at ~15, which is shorter than its sequence length of 20 nt (and therefore corresponds to a monomer), vs. one band at ~15 and one band at ~26, which is longer than its sequence length (and therefore corresponds to a dimer). The low concentration species distribution for T1, T7, TET25, and TET26 match previous experimenters’ observations at higher concentrations [18], [19]. T1 exhibits one dimeric species on the gel (which potentially exists as both monomer and dimer in capillary); T7 exists as one monomeric species; and TET25 and TET26 exist as three monomeric species.
Figure 10. Effect of low concentration on equilibrium of different GQ DNA structures. Fifteen percent gel was prepared in 1 × TBE supplemented with 10 mM KCl. The DNA samples were prepared at 10 µM, unless otherwise specified.

3.2.2 Species comparison under various dyes

To first screen the performance of our established base conditions (0.4 psi 5 sec injection, 5 kV across a 20 µm effective length capillary), we tested our model systems under SYBR Gold (when applicable) and SYBR Green 2 dyes (Fig. 11). For T1, we tested under only SYBR Green 2 (Fig. 11A). The native PAGE predicted two T1 species in capillary, and in Trial 1, we observe two distinct peaks and one potential later shoulder. In Trial 2, we observe one peak with a potential early shoulder and a trailing end asymmetry. For T7, we predict very simple results, just one symmetric peak that ideally corresponds to one GQ conformation. However, under both dyes, we see a level of asymmetry in the peaks (Fig. 11B). Under SYBR Gold, there is a trailing end and a small peak occurring after the main peak (~ 10% of the intensity). The two SYBR Green 2 trials yield peaks with varied amounts of asymmetry. For TET25, we predict three conformations and observe in both SYBR Gold and SYBR Green 2 complicated spectra composed of an early shoulder, a main peak, and a later low intensity broad signal (Fig. 11C). The latter broad signal is not smooth, however, and is composed of lumps and asymmetric shapes. A similar prediction of three conformations and similar electropherograms are observed for TET26 under SYBR Green 2 (Fig. 11D). These TET25 and TET26 data are consistent with the data from Fig. 10, where one species is displayed more prominently than the other two much weaker species.

To an extent, the data are relatively consistent with the predictions, provided that there is a generous interpretation of various shoulders. We do observe simpler electropherograms for T7 and more complicated electropherograms for TET25 and TET26, but we are also unable to make definitive assignments to the peaks. However, this is a solid starting point for optimization and
method development, as multiple species are clearly present, but simply require optimization. The conformations are in the sample, but developing a separation method requires teasing the overlapping peaks out to present in well-resolved, disparate, narrow peaks.

Within the SYBR Green 2 data, the results between Trials 1 and 2 are rather drastically different. For two experimental set ups that should yield essentially identical results, it is of rather significant concern that the results are not reproducible.

**Figure 11.** Effect of different fluorescent dyes on complicated model systems of (A) T1, in which two species (parallel monomer and parallel dimer) are expected, (B) T7, in which one species is expected (parallel monomer), (C) TET25, in which three species are expected (monomeric hybrid, parallel, and one other conformation), and (D) TET26 DNA, in which three
species are expected (monomeric parallel and two other conformations). Samples were injected at 0.4 psi for 5 sec and a 5 kV voltage was applied across the 20 cm effective length capillary. SYBR Gold data are from Exp. 7.1 and SYBR Green 2 data are from Exp. 10.1 and Exp. 11.1. See Appendix C-D for more details.

3.2.3 Buffer ionic strength effects on separation

The lack of reproducibility is a concerning result, observed both in Fig. 11 and in the multiple trials run in Fig. 6. Additionally, we observe high variance in the retention time for identical samples, both from different experiments as well as within sequential data (i.e. Fig. 8A). As supplemental experiments consistently proved the reliability of the sample preparation (Appendix E-F), we turned to the other significant component of the capillary experiment: the separation buffer. Thus, to probe the variation between experiment to experiment, we sought to analyze the effects of slight variations in buffer ionic strength.

Here, we find significant variation between three seemingly identical trials (Trial 1, Trial 2, and Trial 3) for T1, T7, TET25, and TET26. We observe similar trends of the location of peaks/shoulders/trails in the spectra, yet these features vary in intensity, separation, and retention time. When the ionic strength of the buffer was reduced, we observe improved separation of the main peak from the shallow, lumpy trailing peak in TET25 and TET26. However, in TET25 and TET26, the main peak is severely attenuated. The lower ionic strength also led to improved peak separation in T1, achieving a separation seen in some, but not all previous trials. In contrast, increasing the ionic strength of the buffer yielded significantly less signal separation for all samples except T9, which yielded the dual peak structure that was seen in Fig. 6. Curiously, some samples were largely attenuated in intensity.
Figure 12. Effects of buffer ionic strength on complicated model systems of (A) T1, (B) T7, (C) TET25, (D) TET26, and (E) T9 GQ DNA, all in 10 mM KCl. Samples were injected at 0.4 psi for 5 sec and a 5 kV voltage was applied across the 20 cm effective length capillary with on-

Tris-Gly is 25 mM Tris, 192 mM Gly.
Tris-Gly2 is 28 mM Tris, 200 mM Gly.
Tris-Gly3 is 23.33 mM Tris, 166.7 mM Gly.
column SYBR Green 2 loading. Data are from Exp. 9.1-12.1. See Appendix C-D for more details.

4. Discussion

An iterative, informed set of experiments were performed to design and optimize an experimental separation method for G-quadruplex DNA. The initial goal included confirming the presence of a folded GQ structure within the capillary – a goal which took the majority of the project time to accomplish. Yet, this initial goal yielded a CE-CD coupled kinetic experimental method of analyzing GQ folding. This method not only confirmed the appropriate behavior of the system, but could also prove useful in terms of peak identification once a final optimized separation procedure is produced.

However, the difficulties in the first stage of the project meant that the separation method design for more complicated model systems did not progress far, inhibited partially by the variability of results and other outside constraints.

4.1 Reproducibility and reliability of results

One of the most significant issues that plagued the development of an effective separation method was the lack of reproducibility between seemingly identical trials. For example, Fig. 6 illustrates this in the SYBR Gold Trial 1 and Trial 2 data. The peaks are similar, however the trailing end of the peak (for both samples) exhibits as a shallow, partially merged peak in Trial 1 but a short, sharp shoulder in Trial 2. The same type of inconsistency is seen in SYBR Green 2 Trial 1 and 2 in Fig. 11. The shape of the shoulders and trailing ends of the peaks varies, although the most notable difference is in Trial 2, where the exceptional separation of Trial 1 is not observed. Three experimental set-ups intended to be identical, Fig. 12A-D, also suffered the same issue. Most concerningly, one rerun of the kinetic CE experiment in Fig. 8A yielded similar results (Exp.
8.1, Exp. 9.1), but the other two reruns (Exp. 10.1, Exp 11.1) yielded no noticeable separation of the peaks (Appendix C-D).

From these data, we can observe that the issues are not random. The variations are clearly associated with the day of experimentation. This is evidenced by the fact that the changed effects typically influenced the entire trial/experimental day’s results, rather than occur randomly for different samples. Peak shapes (Fig. 6A) and retention time (Fig. 11, Fig. 12A-D) were often influenced in the same way. Some days would yield less-resolved peaks and earlier retention times across all samples, while other days would yield the opposite. As supplementary experiments performed consistently (Appendix C, E-F), this suggests the variation occurs within the capillary system. The variation (± 1.5 min in retention time) is not negligible, although acceptable standards for retention time are not absolute.

Another significant insecurity were the results in Fig. 6 of the SYBR Green 2 data. Here, we observe two peaks, which were later assigned as the GQ peak (the earlier retention time) and the unfolded peak (the later retention time) via kinetic coupled CE-CD experiments (Fig. 8). However, there is still a level of suspicion given that we expected near homogeneous unfolded and folded samples. Both electropherograms suggest that a mixture of the two forms is observed. Specifically, that the unfolded sample has approximately half folded and half unfolded sample, while the folded sample has majority folded sample and a small amount of unfolded sample. We believe there are two potential reasons for this discrepancy, although the investigation of such is outside the scope of this project. First, that the equilibrium between folded and unfolded sample is more complex or more observable in the capillary system than in pure sample conditions (i.e. circular dichroism spectroscopy or thermal difference spectra). Second, that the dye’s binding of folded DNA and unfolded DNA is not comparable (Fig. 5) and thus the relative binding of
unfolded DNA is underrepresented. For example, if \( \frac{1}{2} \) of the unfolded molecules are bound by the dye, but \( \frac{3}{4} \) of the folded molecules are bound, an electropherogram of a sample with equivalent ratios of each would appear to have more folded molecules than reality. This suggests that the unfolded sample would have a minor amount of folded DNA, while the folded sample has a significant level of unfolded DNA. The latter point is not inconsistent with other supporting CD and TDS studies – if folded GQ samples do have significant amounts unfolded DNA, they would not be readily observable in either method. Method development progressed regardless of these results, however it was important to keep in mind in case unforeseen complications arose in optimization work.

4.2 Effective method designs for future use

In Results 3.1, we presented the effective design of a coupled CE-CD method for DNA peak identification within CE data. This method is adaptable to most CE separations, capitulating on the fact that DNA conformational equilibrium changes upon introduction of increased amounts of potassium chloride. By monitoring the time-dependent changes to the potassium chloride in CE (via the changing intensity of separated peaks) and CD (via the changing character of the CD spectra), peaks in an electropherogram can be qualitatively identified and assigned.

This method is not only constricted to the dynamics between folded and unfolded DNA in response to the addition of potassium chloride, but also may be used to identify peaks corresponding to different GQ conformations if the equilibrium is potassium-dependent. The utility of this technique relies on the external stimulant to affect the DNA, thus will only be effective if this property of the DNA system is available. Not all GQ sequences spontaneously fold upon the addition of potassium chloride. However, this designed preliminary method is an important tool to inform work done on GQ separation in capillary electrophoresis.
4.3 Project progress and future recommendations

Given the large amount of project time required to establish that folded GQ samples traveled through the capillary system, relatively little time was dedicated to optimizing and designing a separation method that would work on more complicated DNA sequences (T1, T7, TET25, TET26). Plagued by reproducibility issues, the experimental work suggested that the correct number of species was present in the peaks (Fig. 12), and attested to the ability to visualize multiple GQ conformations in the capillary electrophoresis electropherograms. However, teasing these peaks out proved to be difficult. Attempting to optimize separation and minimize peak broadness was challenging while dealing with a large amount of unreliability in the results and a short time left in the project.

If this study is continued, our first recommendation would be to introduce a marker compound for the experiments. This marker will likely help normalize the retention times and reduce the error in the separation, as it will have known behavior and therefore the electropherograms collected under the method can all be normalized against the marker. Additionally, the marker would allow for a more tunable system given that if the marker passes the detector too early, the method could be subtly adjusted that day (i.e. reducing voltage) to ensure separation is consistent.

With regards to the attempts at separating TET25 and TET26, the most complicated model systems, we believe it would be beneficial to introduce simple homogeneous hybrid and antiparallel systems (> 2 tetrad) to the study. These would provide a basis for understanding the peak shape and intensity, as it is not unimaginable that the binding modes of the SYBR dyes to these other conformations are altered or attenuated compared to the parallel strand orientation. This potential discrepancy may lead to shapes such as the low intensity, later broad peak observed
in Fig. 11C-D. Without truly understanding the typical behavior of the dye with strand orientations besides parallel, we do not have the full introductory knowledge to predict how the samples would react in capillary and appear in electropherograms.

Capillary electrophoresis is a nuanced method for separation. The array of parameters able to be optimized is nearly infinite, and only a limited number of these parameters were able to be tested in the duration of this project. Notably, we tested the capillary length and diameter, the instrument voltage, the fluorescent dye used, injection volume and pressure, separation buffer pH and ionic strength, sample concentration and ionic strength, among many other possibilities. The experimental approach to the CE separation was largely based over ctITP-based NECEEM, however, it may be beneficial to explore other capillary electrophoresis methods in this project. For example, Mironov et al. used equilibrium capillary electrophoresis of equilibrium mixtures (ECEEM) coupled with mass spectrometry to separate folded and unfolded DNAs. Additionally, other methods such as gel capillary electrophoresis may bring the system closer to a native PAGE system (via a reusable capillary filled with a gel matrix) and therefore provide an easier reliability of separation [27].

Ultimately, we believe that the continuation of this project under a NECEEM separation would require first establishing more reliable results and reproducibility prior to optimization of the method design. If this is successful, then we believe it is beneficial to introduce purely antiparallel and purely hybrid DNA sequences as a model DNA studied, and then pursue the design of a robust separation method. However, if unsuccessful, we would recommends turning design goals to other forms of capillary electrophoresis (ECEEM or gel) or searching for a new method entirely.
4.4 Engineering design

Here we presented work in designing a robust capillary electrophoresis method for G-quadruplex DNA conformation separation. Like all projects, we found a basis and a launching point off past capillary electrophoresis work done on these complexes (primarily in complex with proteins), but we aspired to manipulate and design a method to suit a specific goal: separating conformations. We presented our attempts to design a method usable by other laboratories that study GQs, create a more sustainable alternative to traditional electrophoresis methods, and fill the gap where the GQ field lacks a quantitative way of assessing conformations in a sample. We designed this set-up to follow the principles of green science, prioritizing reusability, waste management, and sample size concerns to improve the accessibility and utility of such a method. These principles acted as constraints for our method design. Although no significant external restraints exist, we are still constrained by the need to create a method that is inherently useful and usable by other laboratories. Thus, we are constrained by functions such as the cost of an experiment under the method we develop and the accessibility of the instruments and chemicals we are using. We are also constrained by what instrumentation we have available – in particular, the detector and laser-induced fluorescence source are already owned by our institution, but also are significant limiting factors to what fluorescent dyes we can use, as only a certain range of wavelengths are accessible with the instrumentation. The acquisition of a new source and detector would be extremely costly and thus acted as a constraint on how we are able to optimize our method and what we were able to test in our optimization choices.

To regulate our method and adjust for these constraints, we developed requirements for our method at the beginning of the project. With regards to the separation itself, it was required that the separation would yield clear, well-resolved peaks that are able to be matched to specific GQ
conformations. This is a common requirement for separation methods, not only for electrophoresis. By analyzing the electropherograms of fluorescence intensity vs. retention time, we were able to determine if our method met such requirements. With regards to the method’s accessibility, our requirements were such that (1) the cost for our specific method does not exceed what typical use for this instrument entails, (2) the time it takes to run the method is no more than double the typical use for this instrument entails, (3) the number of additional experiments required to support this method is minimized, preferably two or fewer. Although our method is thus far incomplete, it does obey such self-imposed requirements. The cost largely spans the cost of chemicals and DNA, all of which GQ laboratories should have readily available. The method run-time does not differ significantly from general use of the instrument, facilitated by the shorter capillary length. Additionally, the hands-off element to the instrument alleviates some of the experimental burden as well. With regards to the third point, only circular dichroism spectroscopy has been shown to be necessary to support the results from the method, although laboratories studying these sequences likely already use a wider range of biophysical characterization methods. It is of note that the coupled circular dichroism spectroscopy does significantly reduce the accessibility of our method, restricting it to use by institutions that have both instruments. Notably, a CE system typically costs $100,000-$150,000, while a CD system additionally costs approximately $100,000. They are long-lasting systems and widely used throughout the fields of chemistry and biochemistry, however the high-cost of the systems is a significant point with regards to the applicability of this work.

There were no significant professional standards or legal codes that governed the design of this method. Although there are always ethical considerations around what DNA sequences are used, our DNA was synthetically manufactured and the analyzed sequences are common
throughout many species’ genomes and pose no ethical concern. The field of GQ research is largely non-standardized, however all researchers publish their own reports of how their experimentation is conducted, and a general procedure for experiments varies between instruments and laboratories.
Appendices

Appendix A: Acknowledgements

This project would be nowhere near what it is today without the support and contributions from so many different people. First, I would like to thank my advisors Professor Liliya A. Yatsunyk and Professor Carr Everbach for providing the best support and mentorship I could ask for in pursuing my E90 project, even when things looked like they were spiraling. I would also like to give a huge thanks to Professor Kathryn Riley, both for very graciously letting me use the capillary electrophoresis instrument, as well as for her advice and expertise on the subject. I’d also like to thank Professor Dawn Carone and Christina Rabeler in the biology department for letting me use the fluorescence imager. More thanks to Cassy Burnett and Professor Matt Zucker for the organization and administration behind the E90 course both semesters, and to Professor Nelson Macken and Professor Carr Everbach (again!) for their seminars on planning, ethics, and design constraints.

In a similar vein, I’d like to thank Chris Chung for very generously sharing the capillary electrophoresis instrument time with me, even as he was working on his thesis. I’d also like to thank Rory, Sandy, Gwen, Mac, and Maya for taking the time to come to my midsemester presentation and providing all the feedback on it. And huge thanks to everyone in the Yatsunyk and Riley labs – past and present – (especially Kailey Martin!) who have made the day-to-day of working on this project so much fun. It really does take a village, and I’m so grateful to all these individuals for their endless support.
Appendix B: References


Appendix C: Experimental overview

Table C1. Summary of experimental conditions for capillary electrophoresis experiments in Appendix D. Multiple listings for a condition indicates that multiple conditions were tested in that experiment. All DNA samples were prepared at 2 μM in Tris-HCl sample buffer with varying amounts of KCl. All samples were dyed with the on-column method. Notation is such that Exp. N.1 equivalent to Exp. N.

<table>
<thead>
<tr>
<th>Exp. #</th>
<th>Loading Conditions</th>
<th>Separation Buffer</th>
<th>Capillary Diameter / Eff. Length</th>
<th>Dye</th>
<th>Voltage</th>
</tr>
</thead>
<tbody>
<tr>
<td>1.1</td>
<td>Diluted (200 nM DNA. 4 psi 5 sec injection.)</td>
<td>Tris-Gly (25 mM Tris-base and 192 mM Glycine)</td>
<td>50 μm / 50 cm</td>
<td>SYBR Gold 1:100,000</td>
<td>18 kV</td>
</tr>
<tr>
<td>1.2</td>
<td>Diluted</td>
<td>Tris-Gly</td>
<td>50 μm / 50 cm</td>
<td>SYBR Gold 1:100,000</td>
<td>18 kV</td>
</tr>
<tr>
<td>2.1</td>
<td>Diluted</td>
<td>Tris-Gly</td>
<td>50 μm / 50 cm</td>
<td>SYBR Gold 1:100,000</td>
<td>18 kV</td>
</tr>
<tr>
<td>3.1</td>
<td>Diluted</td>
<td>Tris-Gly</td>
<td>50 μm / 50 cm</td>
<td>SYBR Gold 1:100,000</td>
<td>18 kV</td>
</tr>
<tr>
<td>4.1</td>
<td>Diluted</td>
<td>Tris-Gly</td>
<td>75 μm / 50 cm</td>
<td>SYBR Gold 1:100,000</td>
<td>18 kV</td>
</tr>
<tr>
<td>5.1</td>
<td>Diluted and Undiluted (0.4 psi 5 sec injection.)</td>
<td>Tris-Gly</td>
<td>50 μm / 20 cm</td>
<td>SYBR Gold 1:100,000</td>
<td>5 kV, 8 kV, 10 kV</td>
</tr>
<tr>
<td>5.2</td>
<td>Diluted and Undiluted</td>
<td>Tris-Gly</td>
<td>50 μm / 20 cm</td>
<td>SYBR Gold 1:100,000</td>
<td>5 kV</td>
</tr>
<tr>
<td>6.1</td>
<td>Diluted and Undiluted</td>
<td>Tris-Gly</td>
<td>50 μm / 20 cm</td>
<td>SYBR Gold 1:100,000</td>
<td>5 kV</td>
</tr>
<tr>
<td>7.1</td>
<td>Undiluted</td>
<td>Tris-Gly</td>
<td>50 μm / 20 cm</td>
<td>SYBR Gold 1:100,000</td>
<td>5 kV</td>
</tr>
<tr>
<td>7.2</td>
<td>Undiluted</td>
<td>Tris-Gly</td>
<td>50 μm / 20 cm</td>
<td>SYBR Green 2:1:100,000</td>
<td>5 kV</td>
</tr>
<tr>
<td>8.1</td>
<td>Undiluted</td>
<td>Tris-Gly</td>
<td>50 μm / 20 cm</td>
<td>SYBR Green 2:1:10,000</td>
<td>5 kV</td>
</tr>
<tr>
<td>9.1</td>
<td>Undiluted</td>
<td>Tris-Gly</td>
<td>50 μm / 20 cm</td>
<td>SYBR Green 2:1:10,000</td>
<td>5 kV</td>
</tr>
<tr>
<td>10.1</td>
<td>Undiluted</td>
<td>Tris-Gly</td>
<td>50 μm / 20 cm</td>
<td>SYBR Green 2:1:10,000</td>
<td>5 kV</td>
</tr>
<tr>
<td>11.1</td>
<td>Undiluted and 0.4 psi 8 sec injection</td>
<td>Tris-Gly</td>
<td>50 μm / 20 cm</td>
<td>SYBR Green 2:1:10,000</td>
<td>4 kV, 5 kV</td>
</tr>
<tr>
<td>12.1</td>
<td>Undiluted</td>
<td>Tris-Gly2 (28 mM Tris, 200 mM Gly) Tris-Gly3 (23 mM Tris, 167 mM Gly)</td>
<td>50 μm / 20 cm</td>
<td>SYBR Green 2:1:10,000</td>
<td>5 kV</td>
</tr>
</tbody>
</table>
Appendix D: CE experimental data

**Figure D1.** Electropherograms for (A) Exp. 1.1 and (B) Exp. 1.2. Exp. 1.2 was run on the day of Exp. 2.1, but with the samples that were used in Exp. 1.1. Exp. 1.1/1.2 samples were made in 0 mM KCl.

**Figure D2.** Electropherograms for (A) Exp. 2.1 and (B) Exp. 3.1. Exp. 2.1 samples were made in 5 mM KCl. Exp. 3.1 samples were made in 100 mM KCl.
Figure D3. Electropherograms for (A) T7 and (B) TET25 in Exp. 4.1. Samples were made in KCl as indicated by the figure legend.

Figure D4. Electropherograms for Exp. 5.1 (A) T9 samples in KCl as indicated by the figure legend under diluted and undiluted loading conditions, and (B) T9 samples in 10 mM KCl under voltages as indicated by the figure legend.
**Figure D5.** Electropherograms for (A) Exp. 6.1 T9 samples in undiluted loading conditions, and (B) T9 samples from Exp. 5.2 and Exp. 6.1. Exp. 5.2 was run on the day of Exp. 6.1, but with the samples that were used in Exp. 5.1. Sample KCl concentration is indicated in figure legend.

**Figure D6.** Electropherograms for (A) T9 kinetic experiment upon addition of 10 mM KCl to 0 mM KCl sample, and (B) samples in 10 mM KCl. Note that the dTs have no significant fluorescence intensity. Data are from Exp. 7.1.
Figure D7. Exp. 7.2. electropherograms of assorted GQ-forming sequences. Exp. 7.2 was run on a separate day from Exp. 7.1, but with the samples that were used in Exp. 7.1. KCl concentration in samples is indicated in the figure legend. Note the low fluorescence intensity and high noise resulting from a diluted dye.

Figure D8. Electropherograms for (A) T9 kinetic experiment upon addition of 10 mM KCl to 0 mM KCl sample, and (B) samples in 10 mM KCl, unless otherwise specified. Note that the dTs have no significant fluorescence intensity. Data are from Exp. 8.1.
Figure D9. Electropherograms for (A) T9 kinetic experiment upon addition of 10 mM KCl to 0 mM KCl sample, and (B) samples in 10 mM KCl, unless otherwise specified. Data are from Exp. 9.1.
**Figure D10.** Electropherograms for (A) Exp. 9.2 samples in 10 mM KCl, (B) Exp. 10.1 samples in 10 mM KCl, and (C) Exp. 10.1 T9 kinetic experiment upon addition of 10 mM KCl to 0 mM KCl sample. Exp. 9.2 was run on the day of Exp. 10.1, but with the samples that were used in Exp. 9.1.

**Figure D11.** Electropherograms for (A) T9 0 mM KCl sample with varying injection volumes and voltages in Exp. 11.1, and (B) GQ DNA samples in 10 mM KCl in Exp. 11.1 (C) Exp. 11.1 T9 kinetic experiment upon addition of 10 mM KCl to 0 mM KCl sample. (D) T9 0 mM KCl samples in Exp. 11.2. Exp. 11.2 was run on the day of Exp. 12.1, but with the samples that were used in Exp. 11.1.
Figure D12. Electropherograms for DNA in 10 mM KCl (unless otherwise specified in the figure legend) in (A) Tris-Gly2 and (B) Tris-Gly3 buffers. Data are from Exp. 12.
Figure E1. Circular dichroism spectra for select samples from (A) Exp. 1 in 0 mM KCl (B) Exp. 2 in 5 mM KCl (C) Exp. 3 in 100 mM KCl (D) Exp. 4 with KCl concentrations specified in the figure legend.
Figure E2. Circular dichroism spectra for select samples from (A) Exp. 5 (B) Exp. 6 (C) Exp. 7 (D) Exp. 8. dT and S3+S3O were prepared in 10 mM KCl.
Figure E3. Circular dichroism spectra and analysis for T9 0 mM KCl sample in a kinetic experiment. (A) Exp. 9 with 5 minute intervals between post-KCl scans for 170 minutes. (B) Singular value decomposition analysis of (A), with colored lines to guide the eye corresponding to CE data in Fig. D9. (C) Exp. 10 with 1 minute intervals between post-KCl scans for 120 minutes. (D) Singular value decomposition analysis of (C).
Figure E4. Circular dichroism spectra and analysis for T9 0 mM KCl sample in a kinetic experiment in the presence of 1:10,000 SYBR Green 2. (A) Exp. 12 with 1 minute intervals between post-KCl scans for 120 minutes. (B) Singular value decomposition analysis of (A).
Appendix F: TDS experimental data

**Figure F1.** Thermal difference spectra for samples from (A) Exp. 1 in 0 mM KCl (B) Exp. 2 in 5 mM KCl (C) Exp. 3 in 100 mM KCl (D) Exp. 4 with KCl concentrations specified in the figure legend.
Figure F2. Thermal difference spectra for select samples from (A) Exp. 5 (B) Exp. 6 (C) Exp. 7 (D) Exp. 8. dT and S3+S3O were prepared in 10 mM KCl.
**Figure F3.** Thermal difference spectra for select T9 samples from (A) Exp. 9 (Post KCl Addition data) and Exp. 8 (0 and 10 mM KCl data). (B) Exp. 10.

**Figure F4.** Thermal difference spectra for select T9 samples from Exp. 12 (post KCl) and Exp. 11 (in 0 mM KCl).
Appendix G: Unedited gel experimental data

Figure G1. Unedited fluorescent gel photographs from Fig. 5. Exposure times were 0.9 s, 2 s, and 0.1 s for SYBR Gold (left), SYBR Green 1 (center), and SYBR Green 2 (right), respectively.

Figure G2. Unedited Stains-All gel photographs from Fig. 5.
Figure G3. Unedited Stains-All gel photographs from Fig. 10.
Appendix H: DNA-Ligand binding studies

Ligand binding studies were performed as capillary electrophoresis, circular dichroism, and thermal difference spectra experiments (Methods), however the DNA samples were annealed with 1 or 5 equivalents of the chosen ligand. Data in Fig. H1 show the results of these preliminary tests.

**Figure H1.** Characterization of TET25 in complex with NMM or PyDH2. (A) Electropherogram of TET25 in complex with NMM. (B) Electropherogram of TET25 in complex with PyDH2. (C) Circular dichroism spectroscopy data of TET25 samples. (D) Thermal difference spectra data of TET25 samples. CE data were run with parameters from Exp. 4. DNA samples were made in 10 mM KCl.