12-2012

The dynamic equilibrium of human telomeric G-quadruplexes.

Robert Buscaglia
University of Louisville

THE DYNAMIC EQUILIBRIUM OF HUMAN TELOMERIC G-QUADRUPLEXES

By

Robert Buscaglia
B.S., Northern Arizona University, 2007

A Dissertation
Submitted to the Faculty of the
School of Medicine of the University of Louisville
in Partial Fulfillment of the Requirements
for the Degree of

Doctor of Philosophy

Department of Biochemistry and Molecular Biology
University of Louisville
Louisville, KY

December 2012

THE DYNAMIC EQUILIBRIUM OF HUMAN TELOMERIC G-QUADRUPLEXES

By

Robert Buscaglia

B.S., Northern Arizona University, 2007

A Dissertation Approved on

November 19, 2012

by the following Dissertation Committee:

_____

Jonathan B. Chaires, Ph.D., Dissertation Director

_____

Thomas E. Geoghegan, Ph.D.

_____

William L. Dean, Ph.D

_____

John O. Trent, Ph.D.

_____

Robert D. Gray, Ph.D.

Alla mia bella famiglia, per il loro amore e sostegno.

ACKNOWLEDGEMENTS

I am indebted to my mentor, Dr. Jonathan B. Chaires. His guidance has been invaluable. Without him none of this would have been possible. I also want to express gratitude to the many individuals who have given me their time and encouragement. Specifically, Dr. Robert D. Gray, for his mentorship and patience with my endless barrage of questions, as well as each of my committee members. This dissertation is an example of what can be obtained by the synergistic interaction of multiple people. The quality and relevance of this work was achieved by the unparalleled contribution of many. I credit all the brilliant and helpful individuals whom I have had the privilege to work with for making this dissertation successful.

ABSTRACT

THE DYNAMIC EQUILIBRIUM OF HUMAN TELOMERIC G-QUADRUPLEXES

Robert Buscaglia

November 19, 2012

G-quadruplexes are thought to have biological importance, with studies based on small molecule interactions and quadruplex-interactive antibodies demonstrating their potential for formation *in vivo*. One potential biological function of quadruplex structures is the regulation and maintenance of telomeres. Telomeres are nucleoprotein complexes involved in chromosome stability. Human telomeres are composed of the repeated DNA sequence, 5'-d(TTAGGG), that terminates in a 3' single-stranded overhang. DNA sequences with homology to the human telomere are capable of quadruplex formation *in vitro*. Specifically, sequences containing four-guanine stretches (e.g. 5'-d(AGGG(TTAGGG)$_3$)) are capable of forming at least five distinct unimolecular structures. Which structure is favored is believed to be linked to solvent composition and the addition of 3'- and 5'-flanking residues.

This dissertation provides an essential biophysical investigation of the polymorphic equilibrium displayed by human telomeric quadruplexes. Multiple biophysical techniques are utilized to assemble a thermodynamic description of the influences of hydration and molecular crowding on conformational selection and elucidate complex unfolding mechanisms with unique intermediate states. This dissertation provides the first application of phasor diagrams in the study of quadruplexes. Phasor diagrams are shown to be sensitive to alterations in quadruplex

structure (i.e. folding and unfolding) by monitoring changes in the complex lifetime distribution of 2-aminopurine. This dissertation contains the first multi-faceted biophysical investigation of the underlying mechanism of co-solvent driven conformational changes of human telomeric quadruplexes. The thermodynamic study illustrates that quadruplexes are stabilized by dehydration, a behavior opposite that of canonical duplex structures. Additionally, the ability of PEGs to drive the conformational selection of a parallel quadruplex through differential binding is clarified, addressing unsubstantiated claims that the propeller form is the most biologically relevant conformation. Finally, an in-depth thermodynamic investigation of the thermal unfolding of human telomeric quadruplexes is conducted. Multiple spectroscopic techniques are used to evaluate the thermal unfolding process and characterize potential intermediates states. This dissertation work is the first to apply spectroscopic deconvolutions to demonstrate that human telomeric quadruplexes unfold through sequential mechanisms requiring intermediate species. These results are highlighted by the recovery of an intermediate species whose biophysical description is best characterized by an ensemble of triple-helical conformations.

TABLE OF CONTENTS

Page

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

GENERAL INTRODUCTION OF QUADRUPLEXES.

The discovery of the double-helical structure of DNA was a landmark moment in understanding genetic material.[1] The double helix and Watson-Crick base pairing provided the basis for determining the mechanism of DNA replication and understanding protein/DNA interactions.[2] However, it is becoming evident that higher-order nucleic acid structures (e.g. hairpins, triplexes, G-quadruplexes) also have biological relevance.[3] G-quadruplexes are a particular higher-order DNA structure whose field of study has become of increasing interest due to their potential biological functions and use in nanotechnology and as therapeutics.

This dissertation provides an essential thermodynamic investigation of the dynamic equilibrium displayed by human telomeric quadruplexes. The first application of phasor diagrams for monitoring changes in quadruplex conformation is presented. Evaluation of time-resolved fluorescence through phasor diagrams is found to be sensitive to changes in the human telomeric quadruplex conformation at high potassium concentrations, a conformational change not easily captured by other biophysical techniques. The first multi-faceted biophysical investigation of the co-solvent driven conformational change of human telomeric quadruplexes is also conducted. Biophysical studies clarify the influence of PEGs on the conformational state of the human telomeric quadruplex and demonstrate PEGs act through a conformational selection mechanism by differential binding to an energetically unfavorable all-parallel quadruplex form. Finally, this dissertation provides the first in-depth spectroscopic deconvolution of denaturation spectra obtained for the thermal unfolding of human telomeric

quadruplexes. Deconvolution of thermal denaturation spectra demonstrates quadruplex unfolding is complex, requiring a mechanism beyond a simple two-state transition. Complex analytical models are applied to thermal denaturation results and illustrate thermal unfolding of human telomeric quadruplexes occurs by sequential mechanisms including one or two intermediate states. These sequential mechanisms are found to be general for human telomeric quadruplexes independent of starting conformation. Additionally, based on combined spectroscopic methods, an intermediate state is found to have a biophysical description that is best characterized by an ensemble of triple-helical conformations.

Structural Characteristics of Quadruplexes. In 1963, ten years after the discovery of the double-helix, X-ray diffraction of guanonsine monophosphate gels observed the helical association of four guanine residues into planar structures known as G-quartets.[4] It is now known that guanine-rich nucleic acids can associate into structures displaying stacked G-quartets known as G-quadruplexes. Quadruplexes are four-stranded helical structures capable of forming inter- or intramolecularly and can include unimolecular, bimolecular and tetramolecular associations (Figure 1).[5,6] Quadruplexes typically contain two or more stacked G-quartets and are stabilized by the coordination of cations with the electronegative 06 atoms oriented toward the central core. Biological studies primarily use Na^+ and K^+, but quadruplex formation is also possible in the presence of Pb^{2+}, Cu^{2+}, Sr^{2+}, and other cations whose complexes with quadruplexes were recently reviewed.[7]

The directionality of interacting nucleic acid strands and glycosidic torsion angles are important quadruplex characteristics.[8,9] Quadruplexes are said to be parallel when all strand segments have the same 5'→3' orientation. Anti-parallel quadruplexes are characterized by strands having a (2+2) or (3+1) mixture of 5'→3' directionalities. The glycosidic conformation
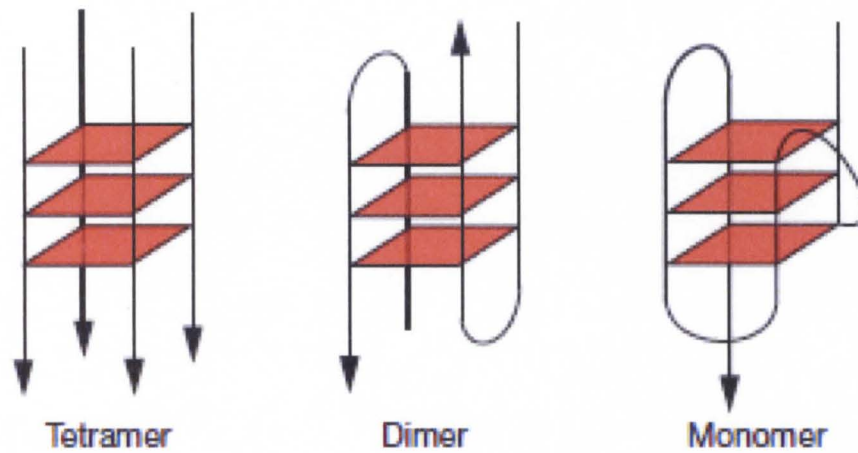
Figure 1. Inter- and intramolecular association of quadruplex structures. Quadruplexes can form from the tetramolecular association of four strands (tetramer), bimolecular association of two strands (dimer), or self-association of a single strand (monomer). Quadruplexes are also characterized by their strand orientations. Tetramer illustrates a parallel configuration, dimer a (2+2) anti-parallel configuration, and monomer a (3+1) anti-parallel configuration. Figure used with permission from Bates et al. 2007.[6]

of quadruplexes can contain syn or anti guanine torsion angles. The stacking arrangement of syn and anti residues effect the chirality of the quadruplex structures, a key aspect involved in circular dichroism (CD) measurements.[10] Quadruplex structures are additionally characterized by the connecting loop segments.[11] Loops are commonly classified into three major types: diagonal, edgewise, and double-chain reversal (side), although other rare loop patterns have been observed.[12] Diagonal and edgewise loops connect non-adjacent and adjacent strand segments, respectively, across the G-quartet face, while side loops connect adjacent strand segments across the side of the quadruplex structure. Side loops are the only loop type capable of maintaining parallel orientation of the strand segments. Unimolecular quadruplexes that contain three loop regions have 27 potential loop segment permutations (Figure 2), although some are sterically impermissible.[13]

Applications in Nanotechnology, Sensors, and Therapeutics. Quadruplex based systems are attractive for use as semi-conductors, nanoelectronics, and nanomaterials.[14] Guanines are of particular interests as conductors due to their low ionization potential and the ability for guanine radicals to migrate significant distances. One example uses quadruplex formation as a DNA nanoswitch.[15] The quadruplex is capable of being repeatedly folded and unfolded through the use of K^+ and crown ether. The folded quadruplex represents an "on" state that increases the conductivity of the DNA sequence 40-fold due to improved π-overlap caused by G-quartet stacking. The self assembly of guanine into supramolecular architectures is also an area of interest.[14] Guanosine derivatives, in particular lipophilic derivatives, are capable of self association into a variety of elongated linear polymers including G-ribbons of unlimited length.[16]
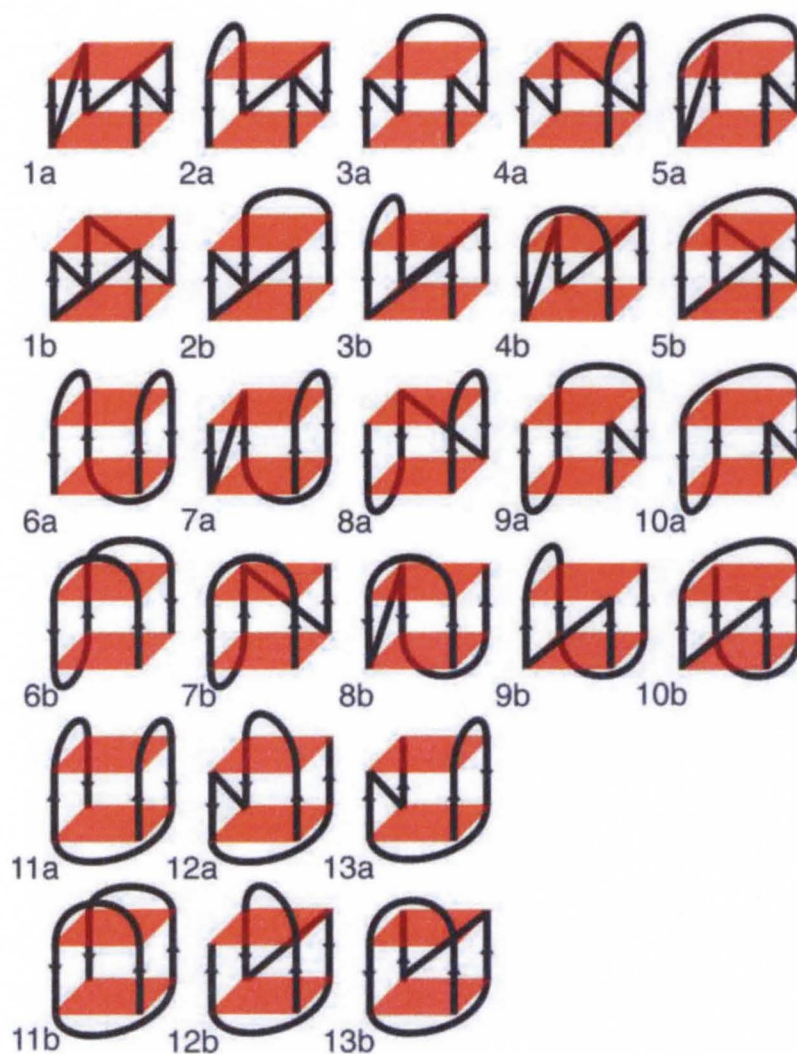
Figure 2. Permutations of loop segments for unimolecular quadruplex structures. The potential for different distinct quadruplex conformations highlights the complexity of unimolecular quadruplex conformations. The potential conformational diversity of quadruplex structures makes them an appealing target for specific targeting by small-molecules for disease intervention. Conformation not pictured is diagonal/edgewise/diagonal due to its unlikely steric configuration. Figure used with permission from Webba da Silva 2007.[13]

Quadruplex systems are also capable of being used as sensors for the detection of Pb^{2+}, Cu^{2+}, K^+, and other metals.[17-19] Quadruplex sensors are designed using bound porphyrin whose fluorescence emission can be monitored. The binding of the metal ions causes allosteric displacement of the porphyrin from its bound state, resulting in a fluorescent emission change that can be monitored as a function of metal concentration. These label-free quadruplex systems are capable of detecting Cu^{2+} as low as 3.0 nM.[19]

Quadruplex structures are also being explored as aptamers for potential disease treatment.[20,21] G-rich oligodeoxynucleotides demonstrate increased uptake by cancerous cells over sequences in which secondary structure is unlikely. AS1411 is an unmodified 26nt oligodeoxynucleotide capable of inhibiting proliferation and inducing cancer cell death by acting as an aptamer of nucleolin.[22] The biological effects of AS1411 correlate with disrupting molecular interactions and localization of nucleolin, with studies now focused on determining models for its cancer-specific antiproliferative activity.[23]

Biological Potential of Quadruplexes. Of most interest is the biological relevance of quadruplex structures. Recent reviews have highlighted quadruplex formation *in vivo* and discussed their potential roles in telomere maintenance and each stage of the central dogma (Figure 3).[8,24] Quadruplex formation *in vivo* is predicted to occur intramolecularly with a generalized G-rich sequence motif of $5'-d(G_{nG}N_{L1}G_{nG}N_{L2}G_{nG}N_{L3}G_{nG})$.[25] The human genome is predicted to have >350,000 quadruplex forming motifs based on the assumption that quadruplex formation *in vivo* would require three or more stacked G-quartets ($nG \geq 3$). Computational studies have additionally demonstrated that putative quadruplex sequences (PQS) are non-randomly distributed throughout the human genome and are evolutionarily conserved amongst yeast.[26]
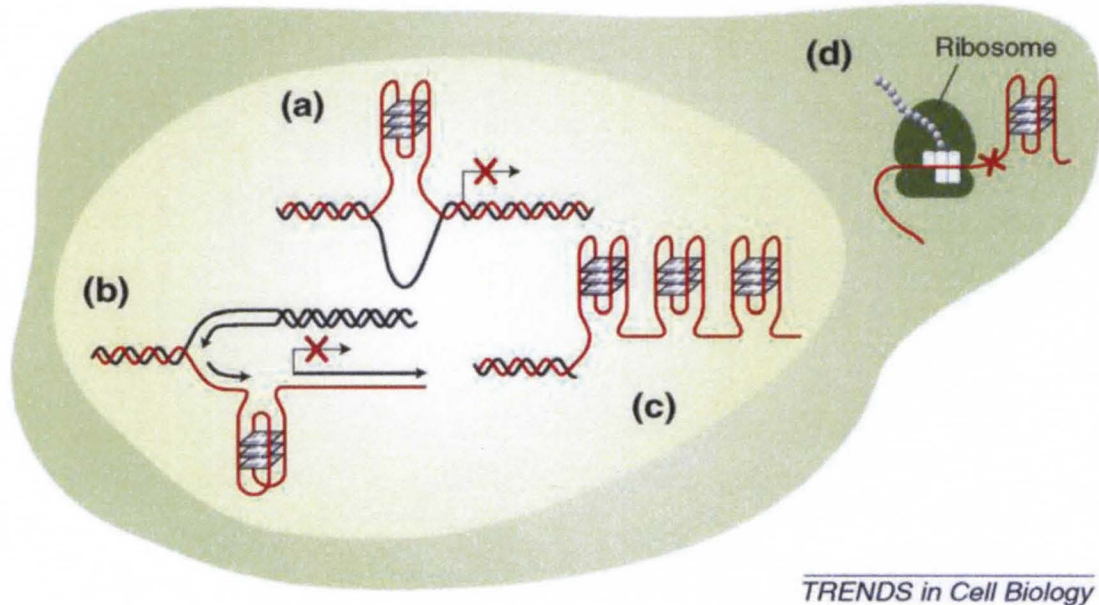
Figure 3. Potential biological functions of quadruplexes. Figure highlights the potential for transient single-stranded states during A) replication and B) translation capable of supporting quadruplex structures. Quadruplex formation may also occur in single-stranded genomic regions such as C) the 3'-terminal overhangs of telomeres or from D) RNA. Figure used with permission from Lipps and Rhodes 2009.[24]

Protein control of quadruplex formation *in vivo* has been demonstrated by studies conducted using *Stylonychia lemnae*.[27] Telomere end-binding proteins (TEBP) were found to control the formation of anti-parallel quadruplexes dependent on cell-cycle phosphorylation of TEBPβ. Antibodies made for specificity to *Stylonychia* telomeric quadruplexes were shown to interact with macronuclei *in vivo*. Antibodies made for specific interactions with parallel quadruplexes had minimal interaction compared to antibodies with affinity for anti-parallel quadruplexes.[28] Furthermore, small molecule ligands designed for preferential interaction with quadruplexes have been shown to localize to chromosomal termini in both human normal and cancer cells.[29] These studies illustrate the potential for both quadruplex formation and quadruplex regulation *in vivo*.

Transcriptional Repression and Stimulation. PQSs were found to correlate with downstream regions of transcription start sites, and have considerable overlap with gene promoters and transcription activation sties.[25] Negative supercoiling during gene transcription can lead to destabilization of duplex DNA to form transiently single-stranded regions, making promoter quadruplex structures a potential target for transcriptional regulation.[30] Quadruplex structures have been proposed to both stimulate and suppress gene transcription.[8] A study of single nucleotide polymorphisms (SNP) demonstrated a correlation between aberrant gene expression and the presence of SNPs in promoter quadruplex forming sequences.[31] One of the best studied promoter quadruplex systems is the nuclease hypersensitivity element III_1 (NHE III_1), which contains sequences that form quadruplexes *in vitro*.[32] NHE III_1 has been shown to control over 80% of MYC expression. Mutations aimed to destabilize quadruplex formation in the NHE III_1 increase MYC expression, while use of small molecules capable of stabilizing quadruplexes reduce expression.[33,34]

Replication and Epigenetic Regulation. During replication, duplex DNA is unwound to single-stranded templates. The discontinuity of lagging strand replication presents a transiently single-stranded state capable of supporting quadruplex formation. Additionally, the potential for formation of quadruplex structures in regions such as gene promoters (discussed above) could present steric impediment of replication machinery. It is therefore likely that replication requires the aid of helicases capable of efficiently unwinding quadruplex structures. Quadruplex helicase activity has been demonstrated by a variety of enzymes, several of which were recently reviewed for their ability to resolve telomeric quadruplexes.[35]

Two helicases, PIF1[36] and FANCJ[37], have important biological functions as quadruplex helicases. PIF1 has been shown to bind quadruplex DNA preferentially over duplex DNA and is capable of efficiently unwinding quadruplex structures.[38] Absence of PIF1 slows replication[39] and increases genomic instability[40] at regions where PIF1 is correlated to quadruplex binding. FANCJ is another quadruplex specific helicase whose absence results in large genomic deletions in regions enriched in quadruplex forming sequences.[37] The influence of quadruplex structures on replication may also have potential effects in maintaining epigenetic markers. REV1 is an enzyme involved in regulating the translesion synthesis of DNA at regions of stalled replication. Cells deficient in REV1 demonstrate a loss in the processive replication through quadruplex DNA, synthesis of DNA that is uncoupled from histone recycling, and increased expression of silenced genes.[41] FANCJ was recently shown to coordinate REV1 and WRN/BLM, which act by two unique mechanisms to maintain epigenetic stability at quadruplex forming regions.[42]

Translational Control through Quadruplex Formation. RNA represents an attractive source for quadruplex formation due to its natural single-stranded conditions. RNA quadruplexes were recently reviewed for their potential influences on translation.[43] RNA

9

quadruplexes typically form parallel structures and have higher stability than their DNA counterpart.[44] RNA quadruplexes have seen an explosion of interest for their potential regulation in untranslated regions (UTR) of mRNA, with 4000 5'-UTRs and 5000 3'-UTRs having been found that contain PQSs.[45] The set of 5'-UTR quadruplexes includes mRNA sequences from NRAS, BCL-2, and VEGF oncogenes. The 5'-UTR quadruplex from NRAS and BCL-2 were shown to modulate translation, with disruption of the quadruplex forming sequence increasing protein production.[46,47] This is believed to represent quadruplex structures blocking the progression of the translational complex after cap-dependent translation initiation. Interestingly, 5'-UTRs have also been found to have quadruplex forming sequences important for internal ribsome entry sites (IRES). The VEGF mRNA was found to contain a PQS within the IRES. Mutation of the quadruplex sequence reduced protein production, indicating the quadruplex structure was essential for cap-independent translational initiation.[48] Together these study show RNA quadruplexes have potential roles in translation inhibition and stimulation.

The Human Telomere. Human chromosomal ends, known as telomeres, are composed of the repeated DNA sequence 5'-(TTAGGG), where the 3'-terminus is an ~200 nt single stranded overhang. The human telomere is a nucleoprotein complex whose biology and function were recently reviewed.[49] Although critical for many cellular functions, the progressive shorting of telomeres is an important regulatory component in cell aging and senescence.[50] Telomerase acts to elongate telomeric DNA and has become a target for cancer therapy due to its over-expression in cancer cells.[51] Human telomere quadruplexes offer targetable structures for small-molecule inhibition of telomerase by stabilization of the single-stranded template into secondary DNA structures.[5]

The telomere is protected by the association of a six-protein complex known as shelterin, which serves to regulate and protect telomeric DNA.[49] Shelterin contains three proteins that directly bind telomeric DNA (TRF1, TRF2, and POT1). The single-stranded DNA binding protein POT1 is known to interact specifically with human telomere repeats to maintain a single-stranded template.[52] POT1 is proposed to disrupt quadruplex structures to aid telomerase elongation.[53]

Although the shelterin complex serves a functional role, the actual structure of human telomeres is of some debate. Human telomeres were originally observed in a "t-loop" formation, where the 3'-overhang folds back to intrude double-stranded telomeric DNA.[54] Although the t-loop has served as a model for telomere protection, without chemical treatment the frequency of observing looped telomere ends is reduced.[55] New models for telomere structure include the D-loop, which suggests both intrusion of double-stranded telomeric DNA and quadruplex formation may occur.[56] Models have also been suggested for the oligomeric formation of quadruplex structures along the 3'-overhang.[57] Experimental and computational approaches have shown structures containing multiple quadruplex units are stable, and that specific interactions between quadruplex loop regions are possible.[58]

Human Telomere Quadruplexes. Many investigations have focused on resolving human telomere quadruplex structures to develop more specific quadruplex-interactive compounds.[59,60] The bimolecular association of 5'-d(TAGGGTTAGGGT) was shown to contain a complex inter-conversion of parallel and anti-parallel conformations,[61] while intramolecular conformations from four telomere repeats have been shown to adopt five distinct topologies. Quadruplex conformation is sensitive to the cation present and the addition of 5'- or 3'-flanking residues. The first intramolecular human telomere quadruplex was resolved by NMR using a 22-

nt human telomere sequence, 5'-d(AGGG(TTAGGG)$_3$) in the presence of Na$^+$.[62] The structure was determined to be a "basket" conformation containing three stacked G-quartets in an anti-parallel topology with one diagonal and two edgewise loops. The same sequence was then used in an X-ray crystallographic study that determined an all-parallel "propeller" conformation containing three side loops in the presence of K$^+$.[57]

Besides the original Na$^+$ and K$^+$ conformations, three additional anti-parallel conformations have been described in the presence of K$^+$.[63-70] Two of the conformations, known as hybrid-1 and hybrid-2, contain a unique (3+1) anti-parallel strand orientation. Hybrid-1 and hybrid-2 conformations have one side and two edgewise loops, although the 5'→3' order of the loops is different. The fifth conformation, hybrid-3, resembles the original conformation determined in Na$^+$ but differs from each of the previous conformations by containing only two-stacked G-quartets. Each of the above structures is now considered to be part of an ensemble of structures formed by human telomere quadruplexes. The dynamic equilibrium of these quadruplexes serves as the system in this dissertation.

Dissertation Objectives.

In vitro biophysical studies offer abundant information on the dynamic equilibrium of human telomere quadruplexes. Prior to the elucidation of the hybrid conformations, biophysical techniques including analytical ultracentrifugation, CD and fluorescence combined with computational approaches were used to demonstrate the propeller conformation did not match the resulting hydrodynamic description.[71] The propeller quadruplex is now believed to be only a minor component in aqueous environments.[72] Biophysical investigation of quadruplexes are critical for understanding the behavior of human telomeres. A detailed biophysical understanding of quadruplex stability and unfolding can aid the design of more specific quadruplex-interactive compounds and improve predictive computations for quadruplex formation. Sensitive biophysical techniques, such as time-resolved fluorescence, may be capable of resolving the complex heterogeneity of conformations displayed by human telomere quadruplexes. This dissertation works toward the central objective of defining the energetic landscape of human telomere quadruplexes through the hypothesis that:

> G-quadruplexes are polymorphic in solution. A dynamic equilibrium exists among an ensemble of structures that is thermodynamically linked to solvent composition and ligand binding. Investigation of G-quadruplexes through biophysical techniques will provide essential thermodynamic information about G-quadruplex polymorphism.

The hypothesis is addressed through three specific aims.

> **Specific Aim I.** To probe G-quadruplex structure by evaluation of loop environment and conformation via the development of time-resolved fluorescence methods.
>
> **Specific Aim II.** To evaluate the effect of dehydration on telomeric G-quadruplex thermodynamics including changes to stability, kinetics and conformational polymorphism.
>
> **Specific Aim III.** To investigate and characterize the thermodynamic profiles of human telomeric G-quadruplex structures.

Specific Aim I. Time-resolved fluorescent techniques confer increased sensitivity to multiple solution species that can be partitioned based on their fluorescent lifetimes and relative contribution to the overall fluorescence.[73,74] The complex mixture of conformations adopted by human telomere quadruplexes represents a system whose study could be aided by the resolution offered by time-resolved fluorescent techniques. A commonly used fluorescent adenine analog, 2-aminopurine (2AP), is useful in the study of local DNA environments but displays a complex multiple component lifetime upon incorporation into DNA oligomers.[75] The complex fluorescent decay of 2AP can be difficult to deconvolute, but monitoring changes in its lifetime can be correlated to structural changes through phasor diagrams.[76]

Phasor diagrams are the transformation of time-resolved fluorescent results into polar coordinates that serves as a model-independent analysis technique of complex fluorescent decays.[77] Phasor diagrams have been previously shown to be an effective technique to evaluate protein lifetimes *in vitro*, capable of capturing changes in protein conformation and protein/protein interactions.[78] Specific Aim I extends the use of phasor diagrams into nucleic acid systems by monitoring 2AP-labeled human telomere quadruplexes. Unlike earlier protein studies that evaluated protein mixtures, 2AP represents a single fluorophore that demonstrates a multi-component fluorescent decay. Phasor diagrams simplify the analysis of 2AP-labeled samples, removing the need to perform non-linear regression analysis to evaluate quadruplex conformational changes. Phasor diagrams are capable of tracking changes in quadruplex conformation upon cation titrations, nuclease digestions, and acrylamide quenching.

Specific Aim II. The ability for PEGs to perturb the conformational equilibrium of human telomere quadruplexes was originally studied to rationalize the differences resulting between biophysical investigations and X-ray crystallographic studies.[71] High concentrations of PEG are

capable of causing a conformational change to the propeller form. Preliminary work (Chapter 4) demonstrated acetonitrile was also capable of driving the conformational change, with suggestions that the hydration state of the quadruplexes was a key aspect in their conversion.[79] Continued work in this field has suggested that PEGs exert an excluded volume effect on the quadruplex system, and that through a molecular crowding mechanism the quadruplexes are driven toward the propeller conformation.[80] Such studies have hypothesized that the biologically relevant quadruplex conformation must therefore be the propeller form. These claims have not been substantiated by rigorous biophysical investigations. Specific Aim II set forth to clarify and evaluate the conformational change induced by the addition of acetonitrile and PEG through a collaborative effort utilizing both experimental and computational approaches. Results suggest a conformational selection model best describes the conversion of human telomere quadruplexes based on the differential binding of PEGs.

Specific Aim III. The final specific aim is addressed by evaluating the thermal stability of human telomere quadruplexes. Such studies are essential for understanding the mechanisms driving quadruplex folding and the inter-conversion of multiple solution species. The thermal denaturation of quadruplexes has normally been treated as a simple two-state unfolding mechanism. A recent review discussed the inconsistencies of published thermodynamic data and suggests complex unfolding mechanisms are required to accurately obtain thermodynamic parameters from thermal denaturation experiments.[81] Specific Aim III attempts to clarify and evaluate the complexity of human telomere quadruplex thermal unfolding.

The first study (Chapter 5) contains a detailed and multi-faceted evaluation a 22-nt human telomere quadruplex forming sequence.[82] The thermal unfolding of the 22-nt was evaluated for significant spectral components using singular value decomposition (SVD), a

model-independent spectral deconvolution technique.[83] Five unique analytical models were evaluated and used to characterize the unfolding mechanism of the 22-nt sequence. CD, Förster resonance energy transfer (FRET) and fluorescent emission results are combined to show quadruplex thermal unfolding is best modeled by a four-state sequential mechanism, with one intermediate state being consistent with an ensemble of triple-helical conformations.

The second study (Chapter 6) expands the use of the analytical models, evaluating the CD melting profiles of eight distinct human telomere quadruplex forming sequences. The eight quadruplex forming sequences correspond with three distinct human telomere topologies: hybrid-1, hybrid-2, and hybrid-3. Thermodynamic characterization of the thermal unfolding was evaluated in aqueous and 50% (v/v) acetonitrile solutions. Results indicate that the human telomere quadruplexes may contain triple-helix intermediates similar to that found for the 22-nt quadruplex regardless of the starting solution conformations. Each of the eight quadruplexes were also found to unfold by a four-state sequential mechanism under aqueous conditions.

Purpose. The dissertation work completed here provides a biophysical description of the dynamic equilibrium of human telomeric quadruplexes. Biophysical studies require techniques that are capable of resolving complex conformations. Application of phasor diagrams as a model-independent analysis technique is sensitive to changes in the state of quadruplex conformations, illustrating sensitivity to unique states not easily captured by other biophysical techniques. Continued development of time-resolved fluorescence for investigation of quadruplexes may have potential to characterize the solution properties of quadruplex conformations and provide resolution of intermediate states.

This dissertation also clarifies the underlying mechanism of co-solvent driven conformational changes. The human telomeric quadruplex is shown to be sensitive to changes

in solvent composition, indicating intracellular conditions may strongly influence the conformation of quadruplexes. These studies also illustrate the ability for quadruplex conformations to be influenced by a linked equilibrium with binding, suggesting proteins may be capable of inducing conformational transitions under biological settings.

Finally, this dissertation aspires to provide new and substantial thermodynamic information regarding the complexity of conformations displayed by human telomere quadruplexes. The thermodynamic description developed in this dissertation shows that the unfolding of quadruplexes is complex and contains multiple intermediate states. These studies define the stability of human telomeric quadruplexes and illustrate energetic barriers that must be overcome during unfolding. The mechanisms developed in this dissertation may also serve as potential models for how enzyme complexes, such as TPP1-POT1[84], unfold and maintain quadruplexes structures in single-stranded form.

CHAPTER 2

QUADRUPLEX STRUCTURE AND STABILITY ILLUMINATED BY 2-AMINOPURINE PHASOR PLOTS.

Introduction.

Fluorescence techniques offer valuable tools for investigating the structure, polymorphism and interactions of nucleic acid systems. The fluorescent adenine analog 2-aminopurine (2AP) has been widely used in the study of nucleic acid systems because of its favorable fluorescent properties, particularly its high quantum yield and its excitation wavelength that is significantly red-shifted from the absorbance region of protein and the natural occurring bases.[85] Because of its advantageous fluorescence properties, 2AP allows for the investigation of protein/DNA systems including DNA and RNA polymerases,[86,87] and has been widely used to characterize RNA and DNA structure and dynamics.[88-92]

Although the steady-state fluorescence of 2AP is relatively simple and can provide information on the nucleic acid system through emission intensity changes, time-resolved measurements have the potential to provide more insight into underlying photophysical processes. The lifetime of 2AP free in aqueous solution is a single exponential decay with a lifetime of 10–11 ns.[75,93] Upon incorporation into a nucleic acid polymer or oligonucleotide, the decay of 2AP fluorescence becomes complex with multiple lifetimes ranging from 8 ns to 50 ps.[93] Recent investigations have attempted to describe the fluorescent lifetime of 2AP within a nucleic acid polymer as a distribution of lifetimes rather than discrete lifetime states.[75,94] The photophysical changes of 2AP upon incorporation into nucleic acid polymers have been extensively studied. Titration of 2AP with nucleosides or nucleoside triphosphates established

the importance of base stacking in the quenching of 2AP fluorescence and the emergence of multiple lifetimes.[95] Computational approaches evaluating the electronic structure of 2AP through time-dependent density functional theory confirmed the static and dynamic quenching of the 2AP fluorescence arising from base stacking. Other computational studies also revealed multiple electronic states upon excitation that were proposed to result in shortened excited state lifetimes and the formation of new non-radiative transitions.[96,97]

Interpretation of time-resolved fluorescence data typically requires model fitting to determine lifetimes and fractional intensities (with the exception of the Maximum Entropy Method[98]). Non-linear regression analysis for lifetime determination depends on the model chosen to best represent the data. Model fitting can become increasingly complicated when solutions contain multiple lifetimes. Samples containing multiple lifetimes, arising either from a mixture of fluorophores or fluorophores with complex lifetime distributions, require increased time for analysis and present difficulty in resolving unique models. This complexity has limited the use of time-resolved methods in nucleic acid studies.

Simplification of time-resolved data analysis may be accomplished through the use of phasor diagrams, which are model-free graphical representations of transformed time-resolved fluorescence results.[77] Phasor plots have antecedents in Cole–Cole plots,[99] and appear under a variety of names in fluorescence studies, such as AB-plots[100] and polar plots[101]. Equations 1 and 2 give the (x,y) transformation of frequency-domain lifetime measurements into a 2D (G,S) phasor plot for graphical representation of time-dependent results.

$$x = G = M \cos \varphi \qquad \text{(Equation 1)}$$

$$y = S = M \sin \varphi \qquad \text{(Equation 2)}$$

Measurement of the modulation ratio (M) and phase delay (φ) are directly acquired from frequency-domain instrumentation, but there exists direct transformations to the (G,S)

space for time-domain measurements.[77,101-103] This simplified analysis removes both the non-linear regression and model-dependent assumptions, facilitating an immediate analysis of time-dependent results. At a fixed modulation frequency, fluorophores that exhibit a single exponential decay will result in a phasor point that falls on the so-called universal circle, described by a semicircle with radius one-half centered at (0.5, 0). The position of the lifetime along the universal circle is dependent on the modulation frequency used for the time-dependent measurement. An immediate conclusion can be drawn that a single lifetime is present in solution if the phasor point lies on the universal circle. If the solution contains a mixture of fluorophores with unique single-exponential lifetimes or a single fluorophore that exhibits multiple exponential decays, the phasor point will fall within the universal circle. Although the phasor analysis will be unable to directly determine the exact distribution of lifetimes in solution without further analysis, any phasor points that fall within the universal circle are known to contain a mixture of fluorescent decays. Excited state reactions cause movement of the phasor points outside the universal circle. Examples include FRET monitored by lifetime changes of the acceptor fluorophore and solvent relaxation.[77,104] The placement of the phasor point of a complex mixture can be described as the linear combination of single exponential phasor points (Figure 4). As stated earlier, two fluorophores having single exponential lifetimes will exhibit phasor points that fall along the universal circle. A mixture of the single-exponential fluorophores will result in a phasor point that lies at a position within the universal circle along the line segment connecting the two single exponential phasor points.[77] Importantly, any movement in the position of the phasor point indicates changes in the lifetimes or fractional intensities. Sensitive time-dependent measurements are simplified to movements of the phasor point that establish interactions and dynamics of the system. Phasors allow for rapid interpretation of time-dependent lifetime results and can be used to determine changes in
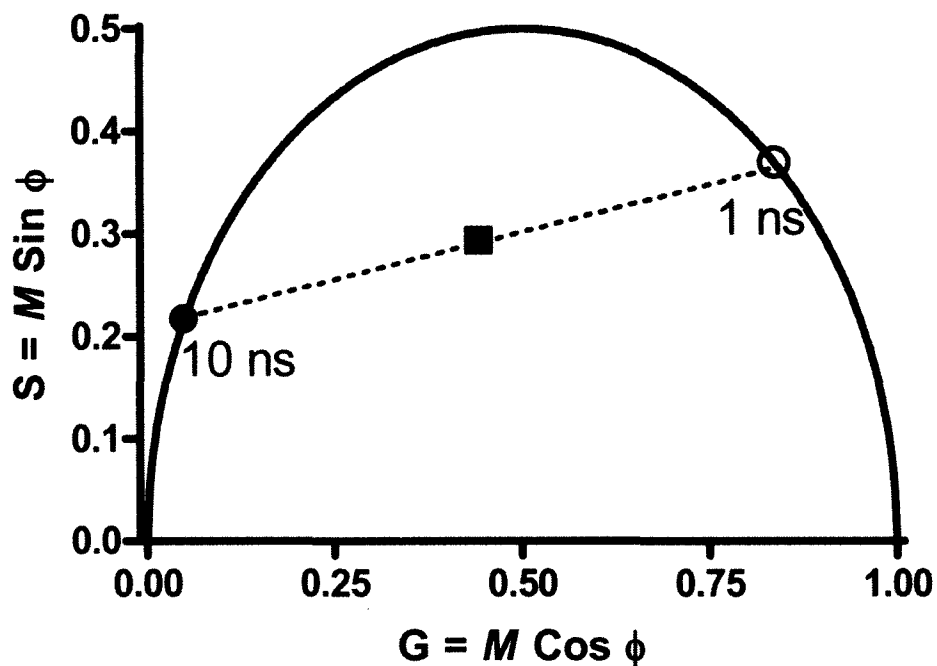
Figure 4. Construction of a phasor point for a mixture of two single-exponential lifetimes in solution. Phase and modulation data were simulated using a frequency of 70 MHz. The phasor points corresponding to single exponential decays of 10 ns (closed circle) and 1 ns (open circle) fall on the universal circle. A mixture of these two fluorophores in solution would result in two exponential decays, resulting in a phasor point falling within the universal circle. The phasor point would fall along the line segment connecting the two unique single exponential decays, with the position of the point dependent on the fractional intensity composition of the mixture. The phasor point shown (closed box) represents 1:1 fractional intensities of 10 and 1 ns species in solution, resulting in a phasor point that lies directly between the two points on the universal circle.

the local environment of the fluorophore being monitored, resulting in simplified analysis of fluorophore mixtures, structural changes and binding interactions.

Phasor diagrams have become an important tool for time-dependent fluorescence microscopy studies due to improved data analysis routines and simplification of lifetime results and FRET efficiency calculations.[101,105-107] Recently, in vitro studies of protein systems and their analysis by phasor diagrams were presented, simplifying complex analysis of time-dependent data and making interpretation of often complex results accessible to non-experts.[77,78] Phasor analysis has been used to enhance fluorescence microscopy and in vitro protein studies, whereas extension of the method into nucleic acid studies has yet to be described. The effectiveness of phasors for analysis of complex nucleic acid systems is described here through the use of the 22-nt 5'-d(AGGG(TTAGGG)$_3$) human telomeric G-quadruplex forming oligodeoxynucleotide (ODN).

G-quadruplexes are structures containing two or more stacked square-planar G-tetrads composed of four Hoogstein hydrogen bonded guanines. G-quadruplexes may form from the tetramolecular association of four strands, the bimolecular association of two strands, or from the unimolecular folding of one strand. In the later two cases, strand segments align in parallel or anti-parallel arrangements joined by a variety of single-strand loops.[108] G-quadruplex structures are stabilized by the coordination of a cation, typically sodium or potassium, with the O6 of the guanine residues within the cavity of the G-tetrads.[109] The human telomeric G-quadruplex sequence folds into a variety of conformations depending on solution composition and sequence modifications.[5] These diverse conformations feature unique strand segment orientations and loop types. G-quadruplex structures adopt conformations containing parallel, anti-parallel or (3+1) mixed strand segment orientations. Loop types include diagonal, edgewise and double-chain reversal. Each of these strand orientations and loop types are found in

telomeric G-quadruplex DNA structures; an all parallel double chain reversal structure,[57] two anti-parallel structures containing one diagonal and two edgewise loops,[62,68,69] as well as two (3+1) mixed structures containing two edgewise and a double-chain reversal loop.[63-66,70] Telomeric G-quadruplex formation in the presence of potassium favors 'hybrid' structures with two edgewise loops and one double-chain reversal loop, whereas only a single predominant anti-parallel 'basket' structure has been found in the presence of sodium.[5]

Because the core quartet stacks of telomeric G-quadruplex structures are similar, studies that focus on strand orientation and loop-type can potentially differentiate telomeric G-quadruplex structures. Techniques sensitive to changes in local environment, particularly fluorescence, are especially important for this purpose. Although steady-state fluorescence techniques have been widely used to investigate G-quadruplex structures, only a relatively few time-resolved fluorescence investigations have been reported.[110,111] Incorporation of 2AP into the loops of the telomeric G-quadruplex sequence has aided thermodynamic and kinetic studies of G-quadruplex folding, providing a signal that is sensitive to a local structural environment that complements the measurement of global folding or other structural transitions.[71,72,112] Solution FRET and single-molecule FRET studies have elucidated conformational heterogeneity of the telomeric structures.[113,114] The use of FRET pairs plays an important role in high-throughput thermal melting assays to explore the stability of G-quadruplex structures and potential ligands.[115,116]

This study evaluates 2AP-incorporated or FRET-labeled telomeric G-quadruplex forming ODNs through the use of time-resolved instrumentation and phasor diagram analysis. Our results show that such time-resolved methods reveal subtleties in a number of G-quadruplex structural transitions that are not evident in steady-state measurements, aiding the elucidation of the underlying reaction mechanisms.

Materials and Methods.

Oligodeoxynucleotides and sample preparation. Unlabeled and 2AP-labeled 5'-d(AGGG(TTAGGG)$_3$) were purchased from Integrated DNA Technologies (Coralville, IA, USA). All ODNs were purchased on a 5-μmol scale with standard desalting. Four unique 2AP-labeled ODNs were purchased with substitutions at positions 1, 7, 13 and 19 corresponding to a single adenine change to 2AP (Diagram 1). Fluorescently tagged 6FAM-d(AGGG(TTAGGG)$_3$)-TAMRA ODNs were purchased from Sigma (St Louis, MO, USA) on a 10-μmol scale with standard desalting and HPLC purification. ODNs are quality checked by commercial sources through the use of either MALDI-TOF or Electrospray Ionization mass spectrometry. All purchased ODNs were used without further purification. Fluorescent tag abbreviations are 6-carboxyfluorescein (6FAM) and tetramethylrhodamine (TAMRA). Prior to use, DNAs were diluted into folding buffer (10 mM tetrabutylammonium phosphate, 1 mM EDTA acid, pH 7.1) to a stock concentration of 1 mM and stored at 4 °C. Folding buffer represents a potassium and sodium-free solvent resulting in minimal G-quadruplex formation prior to the addition of added cation. Folding buffer solutions are used to provide a defined initial state so the effects of added cation can be clearly identified. Unless otherwise stated, all experimental samples were annealed prior to use by placement in boiling water for 10 min followed by quenching on ice for 10 min. All fluorescent standards and reagents were purchased at the highest available grade from Sigma. Fluorescent standards were prepared at stock concentrations of 1 mg ml^{-1} and diluted to experimental concentrations for matched intensities with samples using the following solvents: p-terphenyl (τ = 1.05 ns) in ethanol, fluorescein (τ = 4.1 ns) in 0.1M NaOH and 2AP (τ = 10.5 ns) in folding buffer. Stock solutions of 1M NaCl, 1M KCl and 3M acrylamide were prepared in folding buffer.

A. 1AP - 5`-**X**GGGGTTAGGGTTAGGGTTAGGG

7AP - 5`-AGGGTT**X**GGGTTAGGGTTAGGG

13AP - 5`-AGGGTTAGGGTT**X**GGGTTAGGG

19AP - 5`-AGGGTTAGGGTTAGGGTT**X**GGG

B.

Adenine   2-Aminopurine

C.

Basket (143D)   Hybrid-1 (2HY9)
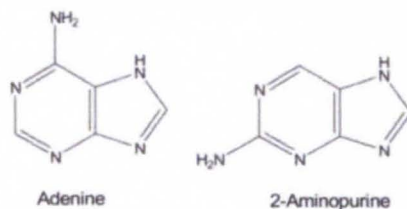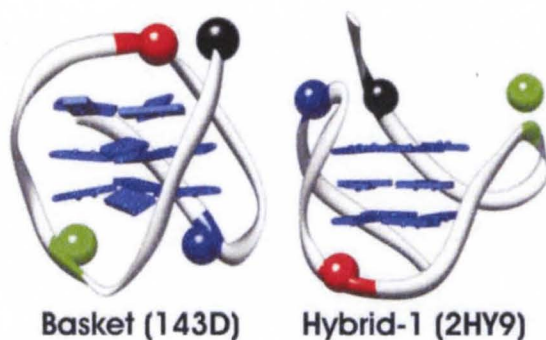
Diagram 1. (A) 2AP substituted G-quadruplex forming sequences derived from the human telomeric repeat. The diagram gives the abbreviations used for the different 2AP-substituted 22-nt deoxyoligonucleotide sequences along with a color-coded format used for all phasor plots. The bold X in each sequence corresponds to the placement of the 2AP in the 22-base sequence. (B) A comparison of the chemical structures of adenine and its fluorescent analog 2AP. (C) A comparison of two published human telomeric G-quadruplex conformations (143D and 2HY9) highlighting where the 2AP substitutions (colored spheres) would be located within these distinct G-quadruplex conformations. Blue planar rectangles represent guanine residues in the formation of G-tetrad cores. The sequentially labeled telomeric forming G-quadruplex sequences will cause the fluorescent 2AP probe to localize differently dependent on the conformation and can be used to locally probe the loop environment.

A concern whenever an oligonucleotide is modified by a base substitution is that the native structure might be perturbed. For 2AP substitutions, circular dichroism and thermal denaturation measurements were used to show that the global G-quadruplex structure was minimally perturbed in modified sequences, as was observed and reported earlier.[71] In addition, steady-state fluorescence properties of 2AP substituted at specific loop sites were found to correlate with solvent accessible surface areas of adenine residues in atomic-resolution models of G-quadruplex structures,[112] suggesting that the modified base accurately reports on the local loop environment.

Steady-state and time-resolved fluorescence spectroscopy. All steady-state scans were collected on a Horiba Scientific (Edison, NJ, USA) Fluoromax-3 equipped with a circulating water bath. Time-resolved data were acquired on an ISS (Champaign, IL, USA) K2 multi-frequency cross-correlation phase and modulation fluorometer equipped with a circulating water bath and either a 300-nm LED for 2AP-labeled samples or a 472-nm LED for 6FAM-labeled samples. All time-resolved scans were completed with polarizers set to magic angle conditions and with emission path length cutoff filters passing wavelengths >380nm for the 300-nm LED and 515nm for the 472-nm LED to negate the collection of stray light.[117] Sample and references were scanned using 1 cm path length Starna (Atascadero, CA, USA) quartz microvolume cuvettes with an optimal volume of 800 ml. Samples were maintained at a constant temperature of 20 °C. Non-linear regression analysis of fluorescence lifetime data was conducted using Globals software.[118]

Phasor preparation. For each set of phase delay (φ) and modulation ratio (M) values collected, a phasor point was constructed using an Excel spreadsheet to calculate the values of G and S (Equations 1 and 2) at the reported modulation frequency. Deviations of S and G were calculated using Equations 3 and 4.

$$\partial_G = \sqrt{\left(S \times \partial_\phi\right)^2 + \left(\cos\phi \times \partial_M\right)^2} \qquad \text{(Equation 3)}$$

$$\partial_S = \sqrt{\left(G \times \partial_\phi\right)^2 + \left(\sin\phi \times \partial_M\right)^2} \qquad \text{(Equation 4)}$$

Equations 3 and 4 represent the SD of the G (δ_G) and S (δ_S) components where δ_φ, the SD of the phase delay, and δ_M, the SD of the modulation ratio, are recorded from time-resolved measurements. Evaluation of the errors associated with S and G can be used to establish significant movements of the phasor point. A universal circle was constructed using Mathematica 9.0 to generate 1000 points of a semicircle centered at (0.5, 0) with a radius of 0.5. All phasor diagrams were plotted using Prism GraphPad software.

Cation titrations. 2AP-labeled ODNs or dual-labeled FRET ODNs were diluted from stock to an approximate strand concentration of 2 µM using folding buffer. Background checks of folding buffer alone and cation into folding buffer gave total intensity counts of <1% of the sample signals, resulting in negligible contributions from background to the total observed intensity. Cation titrations were completed by the addition of aliquots of either 100 mM or 1M NaCl or KCl into the sample cuvette followed by thorough mixing and 5-min equilibration periods. Single frequency phase and modulation scans were collected by setting a fixed modulation frequency (93 MHz) and iterating scans over 360 s with a time step of 30 s, resulting in the acquisition of 13 scans per sample point. Instrumental scan parameter ESE% was set to 0.18 to reduce collected phase and modulation value deviations using an iteration maximum of 250. Cation concentration ranged from 0 to 100 mM over 28 points.

Mung bean nuclease digestions. Digestion of 2AP-labeled ODNs by the single-strand specific mung bean nuclease was monitored using single frequency phase and modulation data acquisition (93 MHz). Samples were prepared by dilution of 2AP-labeled ODNs from stock to 2 mM strand concentration using a digestion buffer (50mM potassium acetate, 100mM potassium

chloride, 1mM zinc sulfate, pH 5.0). Single frequency scans were acquired in the absence of

mung bean nuclease to establish an initial phasor point. Digestion of the labeled ODNs was

monitored over 26010 s with data acquisition every 45 s after the addition and thorough mixing

of mung bean nuclease to 1 U/μl. Digestions were similarly completed and monitored using

steady-state fluorescence emission intensities. Samples were prepared in digestion buffer at

strand concentrations of 2 μM. Data collection was conducted by excitation of the sample at

305 nm with emission monitored at 370 nm with excitation and emission slit widths of 5 nm and

1 s integration times. Data acquisition was trigged after thorough mixing of mung bean

nuclease into the sample at 1 U/μl and monitored over 20000 s with data collection every 5 s.

Digestions were completed at 20 °C.

Acrylamide quenching. Steady-state fluorescence emission intensity and time-resolved

lifetimes of 2AP were monitored the addition of the fluorescence quencher acrylamide.

Separate samples for each acrylamide concentration were prepared by dilution of stock ODNs to

2 μM strand concentration. Samples contained 25 mM KCl or 50 mM NaCl and acrylamide

ranging from 0 to 0.55 M. The samples were diluted to a final volume of 2 ml by folding buffer.

After sample preparation, ODNs were allowed to equilibrate overnight at room temperature in

the dark. Data collection was then acquired on all samples using both steady-state fluorescence

emission scans and multiple-frequency phase and modulation scans. Steady-state emission

scans were collected with excitation at 305 nm and emission scanned from 320 to 440 nm every

1 nm, with 1 s integration and 5-nm excitation and emission slit widths. Lifetime scans were

acquired using multi-frequency phase and modulation scans from 10 to 150 MHz over 10 points

on a log base 2 scale. ESE% was set to 0.18 with a maximum of 250 iterations. The average of

four scans is reported for all samples. This sample preparation and data acquisition protocol

allowed identical samples to be scanned in both steady-state and time-resolved fluorometers

for direct comparison of quenching profiles using fluorescence emission intensity and lifetimes.

Phasor diagrams were prepared using the 49-MHz phase and modulation scans for all

acrylamide concentrations.

Results and Discussion.

Time-resolved evaluation of 2AP-substituted telomeric G-quadruplex DNA. The four

ODNs shown in Diagram 1 were used to probe loop regions of the G-quadruplex structure

formed by the human telomere 22-base sequence 5'-d(AGGG(TTAGGG)$_3$). The structure of 2AP

is shown to highlight the chemical change to the purine necessary for increased quantum yield

and improved fluorescence studies. The 2AP differs from adenine by relocation of the amine

group from position 6 to position 2. Diagram 1 highlights two unique human telomeric G-

quadruplex conformations and their distinctive 2AP environments. The four chosen

substitutions place a 2AP residue in successive loops, beginning with a single 5'-flanking residue

and stepping through each of the three loops formed by the monomolecular telomeric G-

quadruplex. These strategic substitutions confer the ability to explore the localized

environment of the 2AP residue within the G-quadruplex loop, and can be used to draw

conclusions about the loop environment. The same 2AP substitutions have been previously

reported from steady-state investigations and were used to draw conclusions about possible G-

quadruplex conformations present in solution.[71] The resulting steady-state fluorescence

emission curves for the four 2AP-substiuted ODNs were measured under the experimental

conditions used for this study (Figure 5). The normalized fluorescence emission intensities show

the ability of 2AP to probe the local loop environments. Differences in 2AP fluorescence at

specific positions are seen in the presence of sodium or potassium. Such differential 2AP

fluorescence quenching indicates different loop environments, probably due to differences in

base stacking.[95-97] Although measurement of emission intensities provide information about the

local environment of G-quadruplex loops, steady-state methods are of limited resolution and

may not reveal local or global conformational heterogeneity.[5]
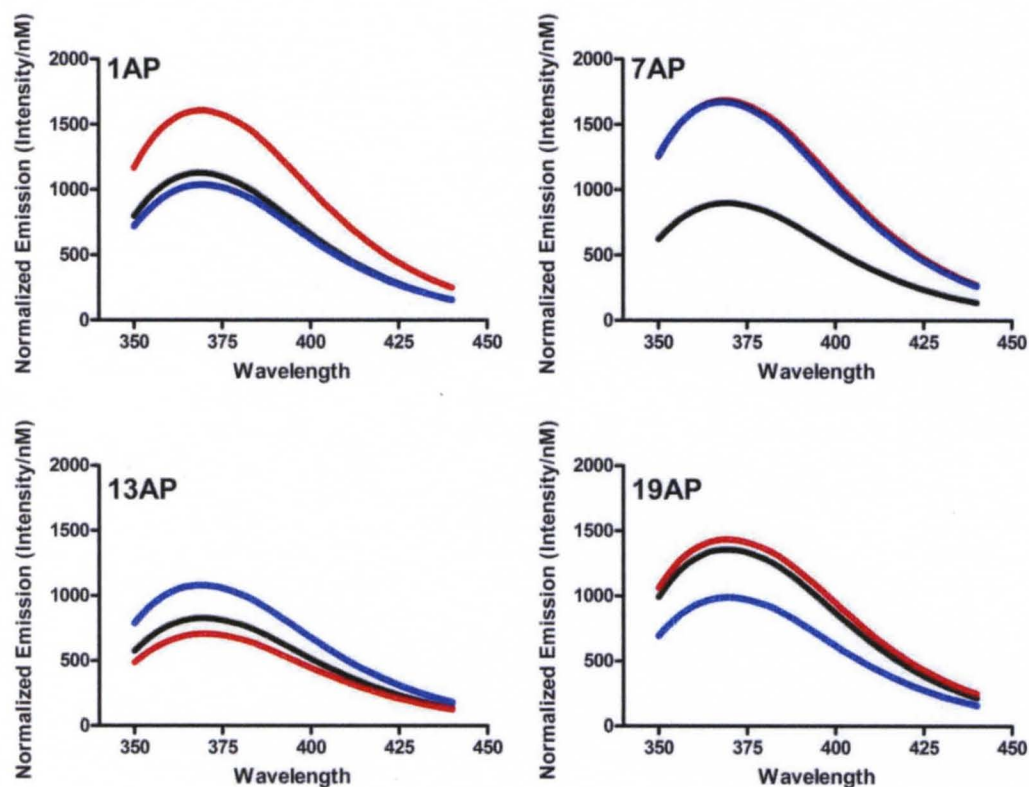
Figure 5. Normalized steady-state fluorescent emission intensities for the four unique 2-aminpurine substituted deoxyoligonucleotides in the presence of no cation (black), potassium (red) or sodium (blue). Resulting emissions scans were collected after the annealing process described, resulting in a comparison of 2AP intensities from different loops of equilibrated human telomeric G-quadruplex conformations. Fluorescent emission intensities allow for a comparison of the 2AP environment to be made based on total intensity at 370 nm, but does not yield information on G-quadruplex solution heterogeneity.

Time-resolved measurements offer the opportunity to evaluate more subtle differences in loop environments, with potential for revealing conformational heterogeneity. Figure 6 illustrates the complication of measuring the lifetime of 2AP upon incorporation into a nucleic acid polymer. Measurement of a 2AP residue free in solution (Figure 6A) results in a single exponential decay with a lifetime of 10.4 ns, in agreement with published results.[75,93] Upon substitution into G-quadruplex structures, however, the data (Figure 6B) cannot be fitted by a single lifetime, and more complex models are required. Table 1 reports both the lifetimes and fractional intensities resulting from non-linear regression analysis of all four individual 2AP-substituted ODNs in the presence of 25 mM KCl. Several different models were tested for each data set, resulting in a reduction of Chi-squared values that reach a minimum when using three discrete exponential decays. Models that included Lorentz distributions were also tested and resulted in almost equally reduced Chi-squared values while requiring fewer fitting parameters. Analysis of the 2AP residue incorporated at position 13 of the ODN (13AP) results in multiple lifetimes best represented by three exponential decays with lifetime estimates of 0.354, 3.1, and 9.07 ns, and fractional intensities of 0.166, 0.345 and 0.489, respectively. The same substitution at the 13 position was previously reported as a single exponential decay of 0.34 ns.[110,111] Although our results are similar for the fastest lifetime component, the rigorous analysis done in this work elucidates a more complex 2AP lifetime profile similar to the reports of multiple fluorescence decays of 2AP upon incorporation into a nucleic acid polymer.[93]

Model	1AP	7AP	13AP	19AP
Single Discrete				
χ^2	1381.76	1512.61	2701.60	1333.61
F_1	1.00	1.00	1.00	1.00
τ_1	3.76	3.68	2.97	3.93
Single Lorentz				
χ^2	62.20	71.13	130.12	59.33
F_1	1.00	1.00	1.00	1.00
τ_1	3.71	3.49	0.38	7.18
W_1	7.14	7.51	9.81	3.97
Two Discrete				
χ^2	20.15	21.97	35.57	17.55
F_1	0.18	0.19	0.25	0.18
τ_1	0.79	0.77	0.62	0.84
F_2	0.82	0.81	0.75	0.82
τ_2	6.17	6.23	6.55	6.39
Two Lorentz				
χ^2	3.96	6.09	17.70	4.28
F_1	0.45	0.32	0.48	0.34
τ_1	0.00	0.00	0.00	0.00
W_1	6.34	4.38	3.69	5.34
F_2	0.55	0.68	0.52	0.66
τ_2	6.26	6.38	7.69	6.48
W_2	0.00	1.82	0.04	1.34
Three Discrete				
χ^2	2.48	4.95	6.36	4.52
F_1	0.10	0.11	0.17	0.09
τ_1	0.37	0.41	0.35	0.42
F_2	0.33	0.33	0.35	0.28
τ_2	2.94	3.03	3.10	2.78
F_3	0.58	0.56	0.49	0.63
τ_3	7.82	7.69	9.07	7.66
Discrete/Lorentz Mixture				
χ^2	2.52	5.10	6.60	4.64
F_1	0.08	0.09	0.14	0.09
τ_1	0.40	0.44	0.35	0.53
F_2	0.93	0.91	0.86	0.91
τ_2	5.30	5.38	5.31	5.66
W_2	3.65	3.55	4.97	3.39

Table 1. Model-dependent analysis of 2AP-incorporated ODNs. Analysis includes single discrete exponential, single Lorentz distribution, two discrete exponentials, two Lorentz distributions,

three discrete exponentials, and discrete/Lorentz mixture. $\chi 2$ is the reduced chi-squared, F is the fractional intensity, τ is the lifetime, and W is the width of the Lorentz distribution. Analysis used fixed errors of 0.20 for phase delay and 0.004 for modulation ratio.

.
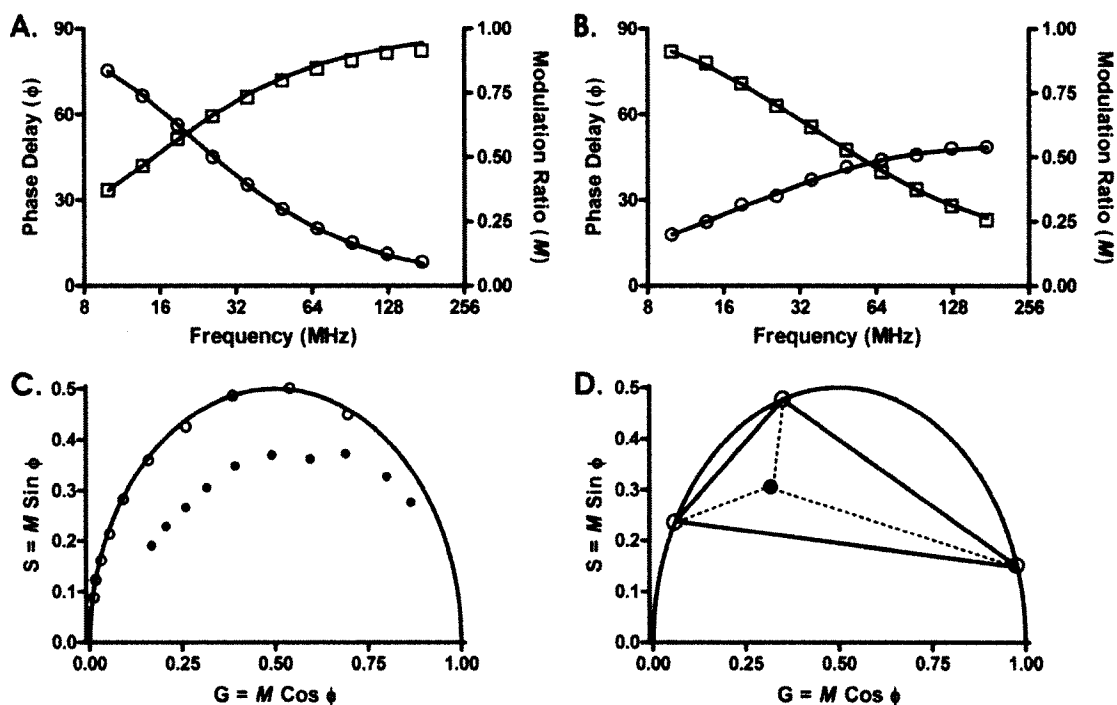
Figure 6. Phase (φ, open circles) and modulation (M, open squares) multi-frequency lifetime

plots for free 2AP and 13AP in the presence of 25mM KCl showing best fit lines determined from

curve fitting analysis using Global software. (A) 2AP free in solution shows a single exponential

decay with a lifetime of 10.4 ns. (B) The substituted 2AP deoxyoligonucleotide, 13AP, is best

represented with three exponential decays with lifetimes of 0.354, 3.1, and 9.07 ns and

corresponding fractional intensities of 0.166, 0.345, and 0.489, respectively. (C) These phase

and modulation plots can be rapidly converted to phasor plot representations and analysis

without the requirement of curve fitting to models. The frequency dependence is highlighted by

conversion of all collected frequency data to unique phasor points for free 2AP (open circles)

and 13AP (closed circles). Free 2AP demonstrates that single lifetime fluorophores will follow

the arch of the universal circle when plotted against frequency. Complex decays, such as that

from 13AP, exhibit similar arching behavior when plotted against frequency while remaining

within the universal circle. (D) The phasor point for 13AP at 70MHz (filled circle) falls within the

35

universal circle and represents a complex mixture of fluorescent decays. Through use of the

three exponential decays found from non-linear regression analysis (fractional intensity and

lifetime values given above), the phasor point construction is illustrated as the mixture of the

three single exponential decays (open circles).

Simplifying time-resolved data analysis using Phasor diagrams. Phasor diagrams offer

a model free evaluation of time-resolved results by transforming the data into phasor plot

coordinates. The phase and modulation results for free 2AP and 13AP in the presence of 25mM

KCl are transformed into phasor points over the entire range of collected modulation

frequencies using Equations 1 and 2 (Figure 6C). This highlights the dependence of phasor points

on the chosen frequency. The phasor points are shown to have arching trajectories dependent

on the fixed frequency when it is changed from 10 to 150 MHz. The free 2AP phasor points

follow the arching 'universal circle' indicating that a single lifetime is present, whereas 13AP

phasor points fall within the 'universal circle' due to the complex distribution of lifetimes. Figure

6D illustrates the meaning of the complex 13AP phasor points more exactly. Three lifetimes

were resolved by non-linear regression for the 13AP substitution (Table 1). These three

lifetimes are indicated as phasor points on the universal circle in Figure 6D. The phasor point for

the 13AP complex lifetime mixture lies within line segments connecting the three unique single

exponential lifetimes. Line segments connecting the phasor points of each single lifetime and

phasor point are shown in Figure 6D. These indicate that the phasor is not equidistant from

each phasor point, reflecting the relative fractional intensity of each lifetime component. The

exact location of the phasor point thus arises from the underlying distribution of fluorescence

lifetimes and fractional intensities, and provides a unique snapshot of that complexity. Factors

that influence either the underlying lifetimes or fractional intensities would result in a change in

the position of the phasor point.

Phasor points were constructed for all four 2AP-labeled ODNs in the presence of 25 mM

potassium or 50 mM sodium (Figure 7A). The four 2AP-substituted ODNs are described by points

on the phasor plot corresponding to unique environments around each of the substitutions. All
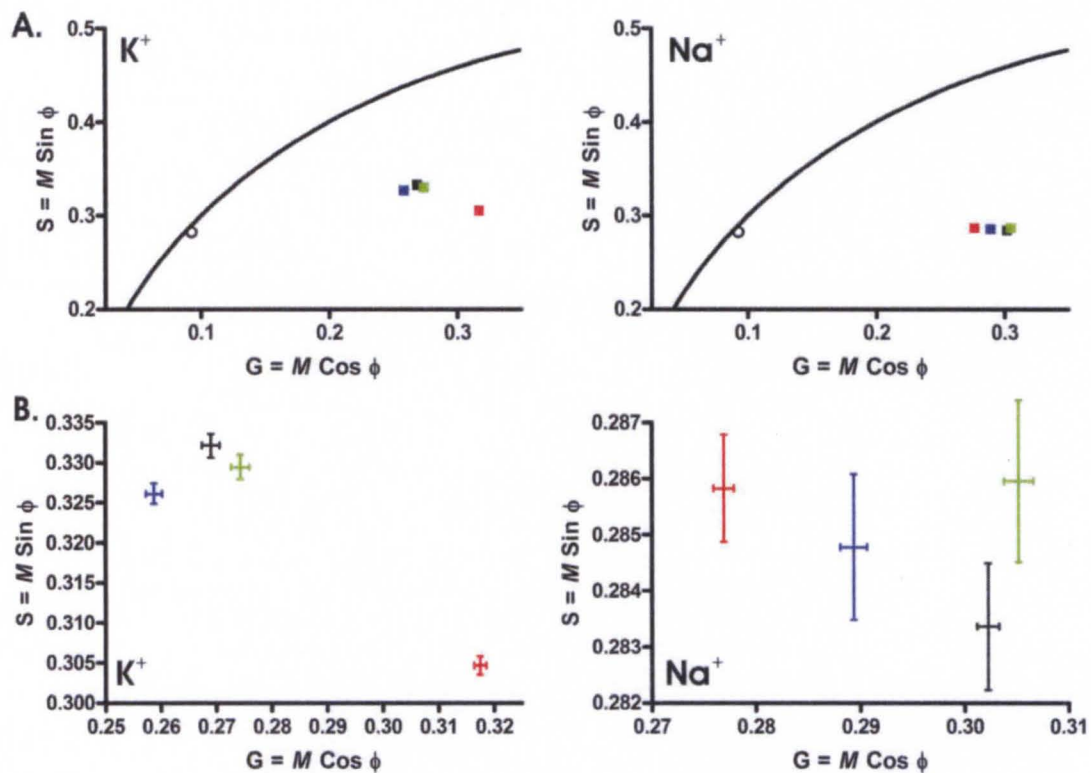
Figure 7. (A) Phasor diagram representations at 70 MHz of the four different 2AP-labeled deoxyoligonucleotides in the presence of either potassium or sodium with colors corresponding to Diagram 1. The 'universal circle' is represented by a solid arching line whereas the open circle phasor point demonstrates the placement of the single exponential decay from free 2AP. (B) The evaluation of errors associated with the phasor points for both potassium and sodium highlights differences in phasor point locations dependent on the G-Quadruplex loop in which the 2AP is placed. Any difference in the phasor point reflects a 2AP environment distinct from the placement at other positions. Both potassium and sodium phasor representations show differences between the environment of all four positions used to probe the G-quadruplex conformation.

four 2AP-labeled ODN phasor points fall within the universal circle indicating a mixture of fluorescent lifetime decays. The phasor points are constructed through model free analysis, requiring no prior knowledge of the lifetime mixture or non-linear regression analysis. The significant advantage of these phasor plots is that they show, at a glance, a specific signature of particular G-quadruplex conformations. The signature for the Na^+ "basket" form is clearly distinctive from that of the K^+ "hybrid" form.

The errors associated with the phasor points of the four different 2AP-labeled ODNs in the presence of either KCl or NaCl were generated using Equations 3 and 4 and are shown in Figure 7B. Although the phasor points in Figure 7A show overlapping regions, this is only because the size of the points used to represent the phasors is larger than the error of the phase and modulation measurements. The enlarged phasor diagrams show that the points are well resolved on the phasor plot. Each of the 2AP-substitutions occupies a unique location on the phasor diagram reflecting the fluorophore's sensitivity to distinct G-quadruplex loop environments. A comparison between the 2AP-labeled ODNs sorted by position further demonstrates that the environment of each 2AP is unique in sodium or potassium solutions (Figure 8). The four 2AP-labeled ODN samples result in unique phasor points between folded and unfolded conditions, as well as differences between sodium and potassium conformations. These distinct phasor points established analysis of 2AP-labeled G-quadruplexes by phasor diagrams can be used to follow changes in solution conformation.

Monitoring G-quadruplex formation through phasor diagrams. The formation of G-quadruplex structure was monitored by addition of sodium or potassium to a solution of the human telomeric G-quadruplex forming ODN. These types of experiments have determined optimal cation concentrations for G-quadruplex formation, the kinetics of G-quadruplex structural formation, and the mechanism for G-quadruplex folding.[72,119,120]
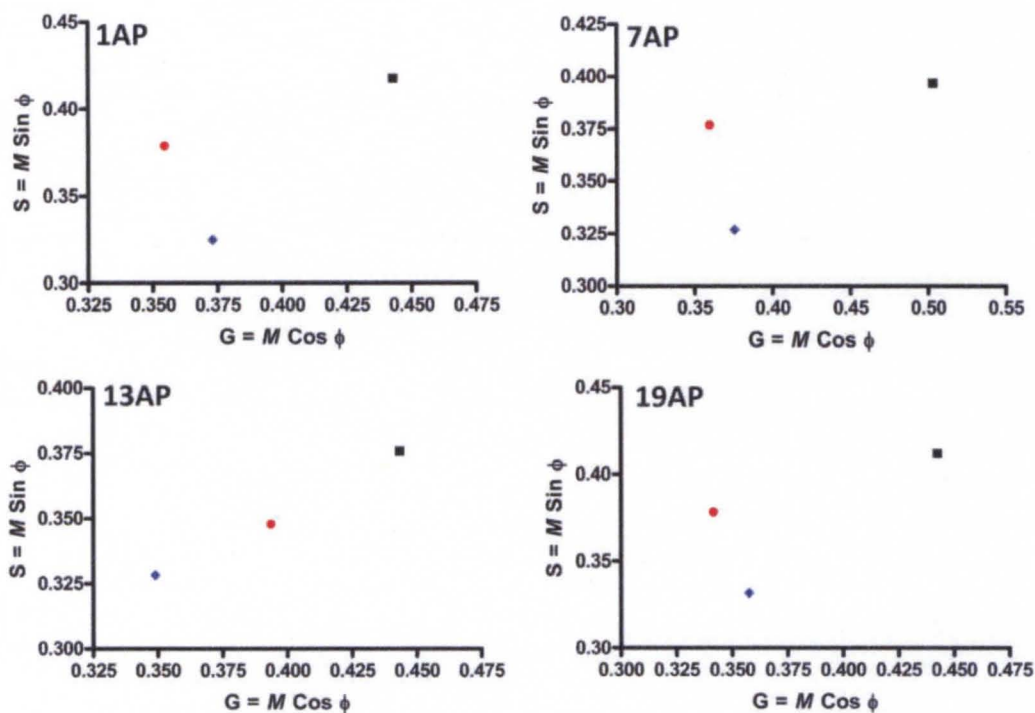
Figure 8. Phasor plot comparison at 70 MHz of 2AP substitution in the presence of no cation (black square), potassium (red circle) and sodium (blue diamond). Unique phasor diagram locations for all 2AP substitutions reveal the local fluorophore environment in the three conditions studied have different impacts on the 2AP lifetime and fractional intensity distribution.

Figure 9A illustrates a potassium titration of 0–100mM KCl of the 22-base telomeric G-quadruplex sequence following changes in the circular dichroism signal at 295 nm. The titration shows that the transition from an unfolded state to a folded state is complete by 5 mM KCl.[72] CD is sensitive primarily to stacking of the G-quartet core and reflects global folding. Other probes, like 2AP fluorescence, provide additional information about local folding of other regions of the structure for a more complete description of the process.

The formation of the telomeric G-quadruplex structure of the 1AP and 13AP-substituted sequence was monitored by phasor representations at 93MHz upon the addition of potassium over the range of 0–100mM (Figure 9B). The phasor plots show multiphasic trajectories for both ODNs, with the first change occurring with large movements of the phasor point from 0 to 5 mM KCl, and a second lesser motion of the phasor point between 5 and 100 mM KCl. These phasor plots reveal subtleties in the folding process that are not evident in the titration monitored by CD. The initial phasor transition is the formation of G-quadruplex structure organized by the addition of potassium cations and stabilization of the G-tetrad core. This result parallels the data collected from circular dichroism measurements.[72] The binding of potassium ions was recently shown to be the largest driving force for telomeric G-quadruplex formation.[121] The time-resolved results appear to be sensitive to a second transition occurring at >5 mM KCl that was difficult to establish through CD measurements. The sensitivity of the phasor diagrams also shows differences between the two environments being probed by 1AP and 13AP substitutions. The 1AP potassium titration shows a sharp transition over the 5 to 100 mM KCl range, whereas the 13AP titration expands across a larger range of the phasor diagram. This indicates the two locations being probed are affected in different ways by the addition of potassium > 5 mM. The phasor representation of the 13AP-labeled ODN is illustrated in a 3D plot to clearly show the trajectory of the phasor points against increasing concentration of KCl (Figure 9C).
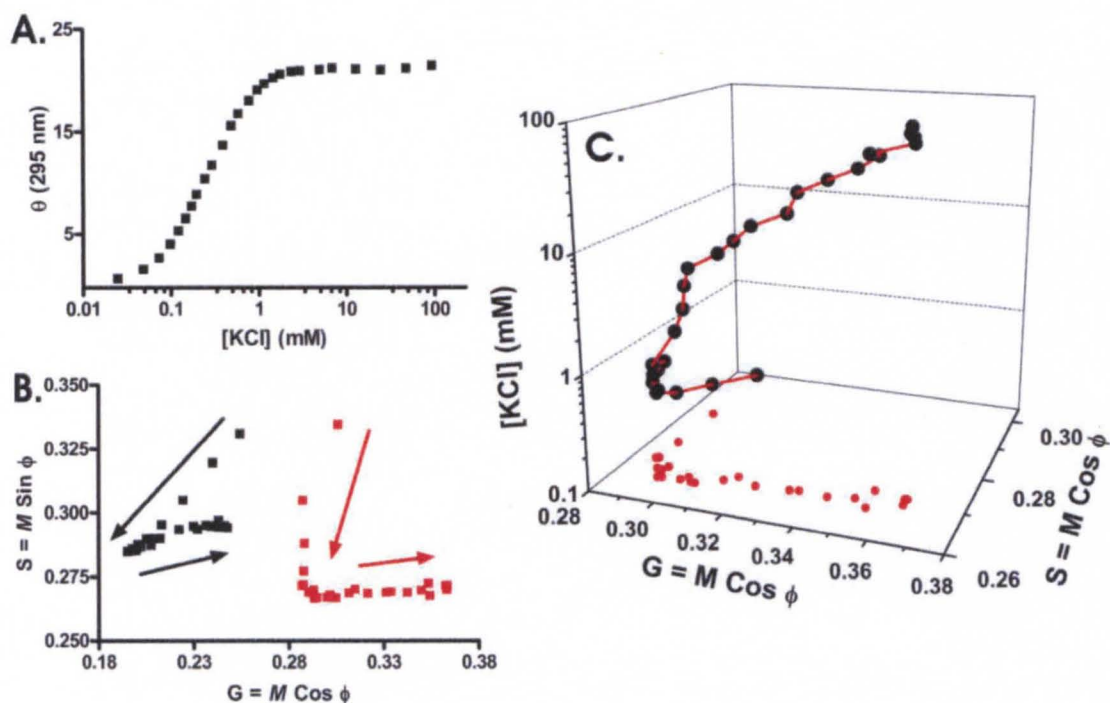
Figure 9. Monitoring changes to the 22-base G-quadruplex forming sequence upon the addition of potassium using circular dichroism and phasor plots. (A) Circular dichroism in millidegrees monitored at 295 nm increased dramatically upon the addition of potassium between 0.1 and 1 mM. (B) Phasor representations at 93MHz monitoring changes to the fluorophore environment for 1AP (black) and 13AP (red) G-quadruplex forming sequences. Arrows are used to trace the direction of increasing potassium concentrations. Both substitutions show a sharp and fast phasor change between 0 and 5 mM potassium, with a slow change occurring from 5 to 100 mM. The non-overlapping phasors for 1AP and 13AP correlate to differences in the environment of the 2AP, but show similar trends upon the addition of potassium. (C) Three-dimensional phasor representations of 13AP at 93MHz demonstrating the phasor points change direction > 1 mM potassium. Phasor representations are able to capture small changes during the folding of G-quadruplex DNA upon the addition of cation and can be used to locally monitor specific loop regions.

42

Digestion of G-quadruplex structure through Mung Bean Nuclease digestions. The single-strand specific mung bean nuclease offers an alternative method to probing the accessibility of G-quadruplex loop regions. Mung bean nuclease has been used to selectively digest loops in the study of DNA hairpin conformations,[122,123] suggesting it may also be a useful probe for intramolecular G-quadruplex loops. The endonucleolytic activity of mung bean nuclease has also been found to preferentially digest around adenosine residues.[124] Therefore, digestion of telomeric G-quadruplex structures should occur preferentially at the –TTA– loops. Evaluation of digestion kinetics for known G-quadruplex loops (double-chain reversal, edgewise, diagonal) may allow for conclusions to be drawn on the type of loops present in unknown G-quadruplex structures. Digestion kinetics can be used in conjunction with similar techniques such as fluorescence quenching, allowing for an accessibility profile of loop geometries to be established.

The digestion profiles of mung bean nuclease with the 13AP-labeled telomeric G-quadruplex forming sequence in the presence of potassium were evaluated using both steady-state and time-resolved fluorescence techniques. Steady-state digestion profiles of the 13AP-labeled telomeric G-quadruplex structure show a complex time course with a leading lag phase lasting ~3000 s (Figure 10A). Following the initial digestion lag phase, a sharp rise in the total fluorescent intensity occurs between 3000 and 15000 s, where the emission intensity reaches a plateau of 10 times the initial intensity.

The time-resolved phasor diagram shows two unique phasor trajectories during the digestion time course (Figure 5B). The initial phasor trajectory corresponds with the steady-state lag phase and is proposed to be an initial slow digestion of the telomeric G-quadruplex loops. The second phasor trajectory, which expands downward toward the universal circle,
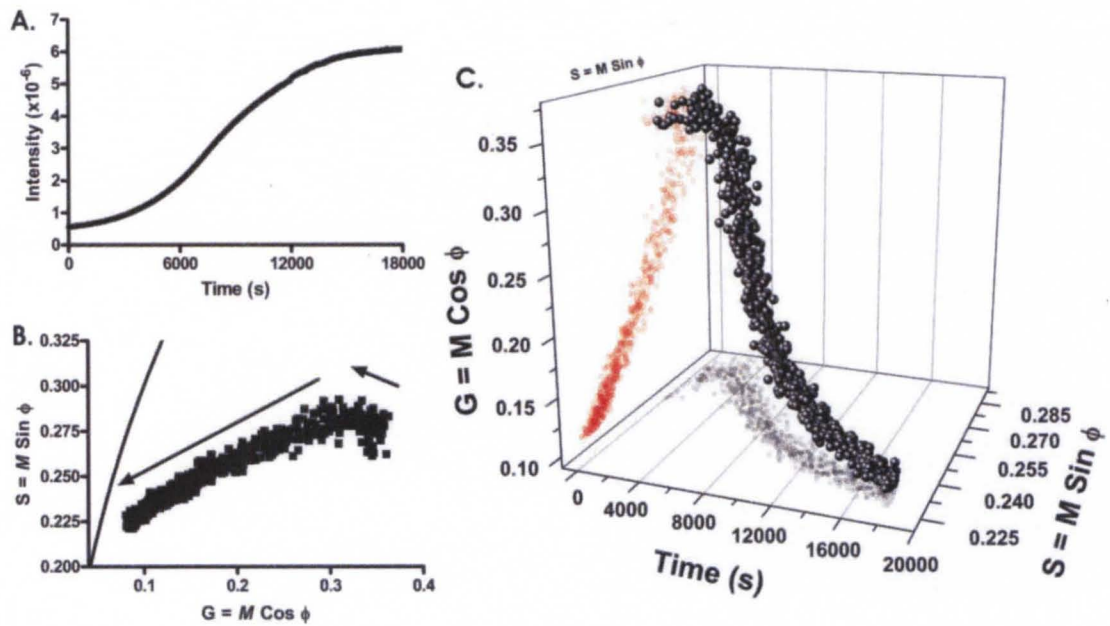
Figure 10. (A) Steady-state fluorescence monitoring of the mung bean nuclease digestions of

13AP shows an initial lag phase of ~3000 s following by an exponential digestion state before

leveling off after 15000 s. (B) Phasor representation at 93 MHz of the mung bean nuclease

digestion of 13AP demonstrates two unique phasor changes. Arrows indicate the direction of

increasing time. The universal circle is drawn as a solid arching line. The initial change of the

phasor is upward and continues for 3000–4000 s. The second phasor change continued until

16000 s at which point the phasor point remained nearly constant at G,S coordinates (0.11,

0.23). (C) Representation of the phasor diagram in a 3D plot emphasizing the change in (G,S)

against time.

corresponds to the large increase in the emission intensity seen from steady-state digestions. This portion of the digestion profile is proposed to be the digestion of single-stranded DNA after destabilization of the G-quadruplex structure has occurred, resulting from one or more nucleolytic nicks.

At digestion completion the phasor points remain stationary at approximately G,S = (0.11, 0.23), a phasor coordinate located inside of the universal circle. Although mung bean nuclease has been shown to digest DNA to the single nucleotide level,[124] the phasor diagram suggests that mung bean nuclease may not completely excise the 2AP from the nucleic acid polymer. If 2AP was completely excised from the nucleic acid polymer, this would result in a phasor point corresponding to a single lifetime similar to that of free 2AP measurements. Another possible explanation for failure of the phasor point to reach the universal circle after prolonged digestion is that the lifetime of 2AP may be influenced by the presence 5'-phosphate and 3'-hydroxyl groups. The impacts of these linkages on 2AP lifetime have not been established.

To emphasize the time course of the digestion profile, the G and S coordinates are plotted against time in 3D space (Figure 10C). When plotted against time, the phasor diagram shows the same digestion curve as the steady-state results, establishing the phasor diagram inherently contains the same information as the steady-state curve while allowing for more detailed conclusions to be drawn. These fluorescence based digestions show that phasors are able to monitor the degradation of G-quadruplexes and result in sensitive digestion profiles. For the scope of this manuscript, precise kinetic models were not evaluated using the resulting steady-state or time-resolved digestion profiles. Detailed kinetic analysis of G-quadruplex loops by mung bean nuclease will be investigated in future work with the potential for establishing accessibility profiles for known loop geometries.

Phasor interpretation of quencher accessibility. Another important fluorescence technique in the study of G-quadruplex structure is collisional quenching.[91] The accessibility of singly substituted 2AP residues to quenching by acrylamide has been used to probe the loops of human telomeric G-quadruplexes.[71] Stern–Volmer plots constructed for acrylamide quenching of the 13AP telomeric G-quadruplex show characteristic downward curvature, distinct from the quenching of high accessible 2AP free in solution (Figure 11A). Downward curvature of the Stern–Volmer quenching plots indicates heterogeneity in the 2AP environment, which can result from either multiple global G-quadruplex conformations or distinct localized position of the fluorophore within the G-quadruplex loop. The downward curvature of these Stern–Volmer plots is distinct from the quenching of 2AP nucleotides free in solution, which demonstrate a linear dependence to added quencher and were determined to have a diffusion limited quenching constant (Figure 11B).

A method for partitioning the number of components necessary for analyzing the downward curvature is to identify by time-resolved techniques multiple lifetime components.[74] Time-resolved fluorescence techniques offer sensitive measurements of multiple 2AP environments based on lifetimes, which can then be used to construct individual lifetime Stern–Volmer (τ_0/τ against quencher concentration) plots. The lifetime-based τ_0/τ plots can be used to establish Stern–Volmer quenching constants for individual components and relate directly to the type of quenching process.[74] The addition of acrylamide to the 13AP telomeric G-quadruplex sequence was used to construct an average τ_0/τ quenching plot for the time-resolved measurements and compared with steady-state intensity measurements (Figure 11A). The average lifetime plot shows the same downward curvature present from the steady-state plot, indicating the multiple fraction model accurately represents steady-state experiments. The differences between the two curves are due to the inability to accurately fit the mixture of
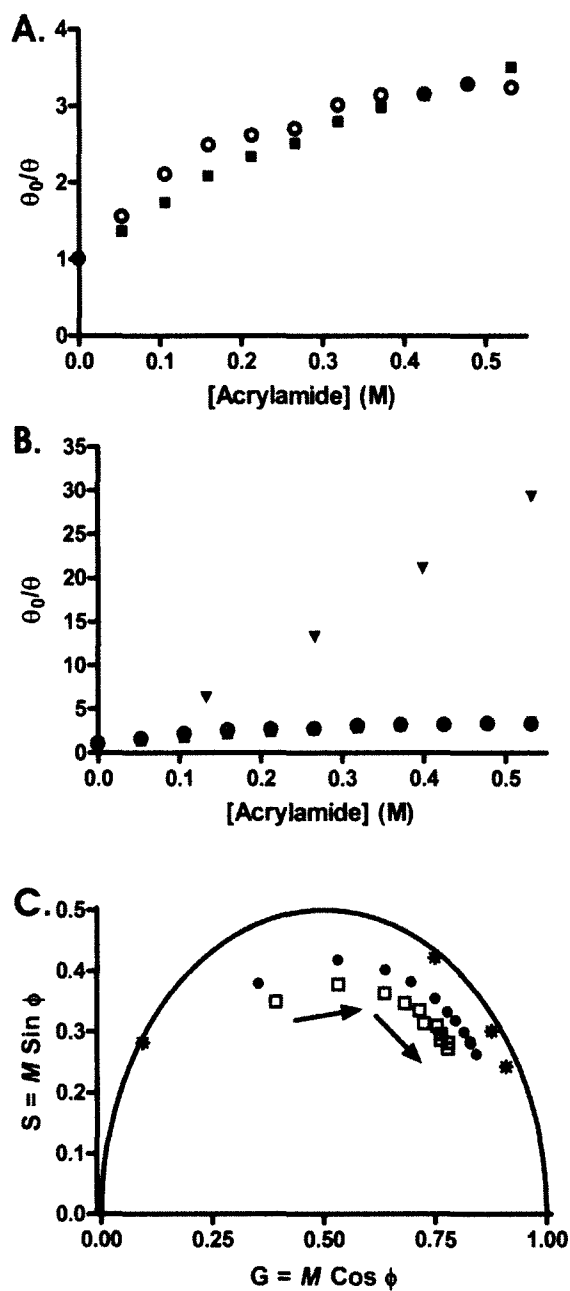
Figure 11. (A) Stern–Volmer quenching plots constructed from fluorescent intensity (F_0/F - filled squares) and average lifetime (τ_0/τ - open circles) for the addition of acrylamide to 13AP. The Stern–Volmer plots show downward curvature for both intensity and average lifetime indicating multiple accessible 2AP populations. (B) Stern–Volmer plots comparing acrylamide quenching of free 2AP (triangles) with the fluorescent intensity and averaged lifetime of 13AP. The

47

quenching of free 2AP shows diffusion limited accessibility to quenching by acrylamide. (C) Quenching of 1AP (filled circles) and 13AP (open boxes) by acrylamide monitored using phasors at 49 MHz. Arrows indicate the direction of increasing acrylamide concentration. Phasor diagram shows an initial fast change in phasor position followed by a slower moving phasor change. The quenching of free 2AP (stars) by acrylamide analyzed by phasor diagrams is located initially on the universal circle with slow deviation from the universal circle at high acrylamide concentrations. This is proposed to be a similar deviation from a single lifetime component as was previously reported in glycerol/water mixtures,[75] or small background interference being seen at high quencher concentrations.

lifetimes present, made difficult by the large number of components present with a wide range of lifetimes (10 ns to <1 ns lifetimes) along with the addition of quencher to this complex mixture. The ability to dissect the mixture of lifetimes into discrete components through model dependent methods becomes increasingly difficult with the addition of quencher. Because the lifetime mixture of 13AP contains very fast fluorescent decays (<1 ns), and the addition of quencher causes this species to act as though the lifetime has decreased, the sensitivity to analyze lifetime components this short decreases significantly.

Phasors monitor the quenching of the complex 2AP lifetime mixture without detailed analysis by model dependent methods. Acrylamide quenching of 1AP and 13AP telomeric G-quadruplex sequences in the presence of potassium presents arching phasor trajectories with increasing acrylamide concentration (Figure 11C). The arching phasor trajectories collected at a fixed modulation frequency signify decreasing lifetimes in solution.[101] The phasor points do not converge to the universal circle, implying the phasor method is sensitive to the multiple lifetimes present in solution, even at high acrylamide concentrations. The arching phasor trajectory is equivalent to that seen for the quenching of free 2AP in solution. Interestingly, the free 2AP acrylamide quenching trajectories show a gradual deviation from the universal circle at high acrylamide concentrations. This matches previous reports that the lifetime of 2AP increases in complexity from a single exponential decay upon quenching with acrylamide.[75] However, this earlier report used glycerol/water mixtures of 2AP and hence cannot be directly compared with our results. It is also possible that a small unquenched background fluorescence could explain the deviation of the phasor point from the universal circle at high acrylamide levels. Phasor diagrams result in the ability to track shortening lifetime mixtures that would otherwise be difficult to analyze by normal non-linear regression methods.

Evaluation of FRET by phasors. FRET has become an important aspect of G-quadruplex

solution studies because of the ability to determine molecular distance and interaction. FRET-

labeled G-quadruplex forming sequences have served to elucidate folding pathways and

kinetics, while single-molecule FRET studies have investigated G-quadruplex solution

heterogeneity.[113,114] FRET-labeled G-quadruplex structures can also be used in thermal

denaturation studies to determine the stability of the conformation, and can serve for high-

throughput analysis using fluorescence detecting thermal cyclers.[115,116] Although many of these

techniques rely on the fluorescence emission of either the acceptor or donor fluorophore to

monitor change, the use of FRET in time-resolved instrumentation is directly related to lifetime,

as the decrease in lifetime equates directly to the efficiency of energy transfer.[125]

Potassium and sodium titrations were completed using a FAM/TAMRA dual-labeled 22-

nt human telomeric G-quadruplex forming sequence where both the emission intensity and

frequency-domain lifetime measurements were collected (Figure 12). Filter sets were used for

the collection of fluorescence emission of only FAM. Total intensity measurements using the

FAM/TAMRA ODN show a decrease in the emission intensity of FAM upon the introduction of

cation. This decrease in fluorescence emission intensity is caused by the formation of G-

quadruplex, bringing the FAM/TAMRA pair closer together and allowing for increased efficiency

of energy transfer. This change is distinct from titrations conducted using donor-alone ODN.

When the 22-nt ODN is labeled with only FAM, the change in total intensity is <10% (Figure

12A). This small change with the FAM-only ODN indicates the change seen from the

FAM/TAMRA dual-labeled ODN corresponds to energy transfer and not an influence on the

fluorescence intensity due to the addition of cation. The potassium titration has a midpoint of 5

mM, similar to that found from the 2AP titrations, whereas the sodium titration has a
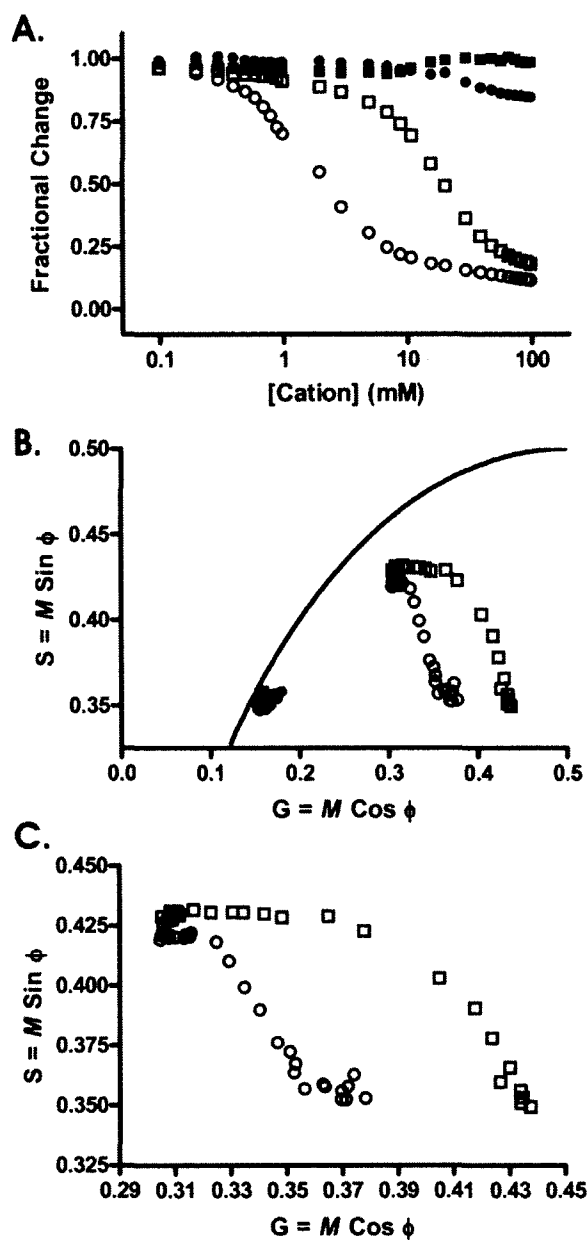
Figure 12. G-quadruplex folding monitored through changes in FRET efficiencies using donor alone (FAM) and donor-acceptor (FAM-TAMRA) labeled human telomeric 22-base sequences upon the addition of potassium (circles) or sodium (boxes). (A) G-quadruplex folding was monitored by total fluorescence intensity upon the addition of cation over a range of 0 – 100 mM. Both potassium and sodium show a single significant transition with midpoints at 5 and 25 mM, respectively. The background intensities from potassium (filled circles) and sodium (filled

boxes) titrations using donor alone show minimal changes in total intensity. (B) Phasor

diagrams using 93MHz frequency were prepared from the same titrations used for total

intensity. Solid arching line represents the universal circle. The donor alone samples (closed

circles and squares) show no movement and remain near the universal circle with increasing

cation concentration, whereas dual-labeled telomeric 22-base DNA shows phasor points distinct

from donor alone and significant trajectories with increasing cation concentration. (C) Enlarged

phasor diagrams emphasizing motion of the phasor point with cation titrations. Potassium

phasors (circles) show two unique trajectories with movement downward over the potassium

range 3 – 30 mM and movement right from 30 to 100 mM. Sodium (squares) shows only a

single trajectory with significant movement over the range 15 – 100 mM.

significantly higher midpoint at 25 mM. These results match previous telomeric G-quadruplex

folding studies.[72]

The FRET phasor diagrams track global conformational changes through the folding of

the G-quadruplex upon the addition of cation. The FAM-only 22-nt ODN sodium or potassium

titration phasor diagrams overlay near the universal circle and show no movement with

increasing concentration of cation (Figure 12B). This demonstrates the phasor diagrams are not

sensitive to the changes in G-quadruplex folding without the presence of the acceptor. The

FAM/TAMRA dual-labeled sodium titration phasor diagram shows a single trajectory across the

sodium concentration range. The phasor points move from an initial position of (0.31, 0.425) to

a final position of (0.43, 0.35) (Figure 11C). The phasor trajectory indicates only a single

conformation change over the sodium concentration range of 0 – 100 mM. Interestingly, the

potassium titration of the FAM/TAMRA dual-labeled ODN shows two trajectories similar to that

obtained from 2AP-labeled samples. An initial phasor trajectory is present upon the addition of

potassium from 3 to 30 mM, with significant downward movement of the phasor points. The

phasor points then shift outward during the range of 30 – 100 mM potassium, indicating

sensitivity to a secondary global conformational change from the time-resolved measurements.

The phasor points should remain stationary unless there is continued influence on the

fluorophore's environment, similar to the phasor diagram shown for sodium titrations. This

secondary transition was present in both the 2AP-labeled titrations and FRET-labeled titrations,

and seen only through the use of time-resolved methodologies.

Summary.

Phasor diagrams provided a simplified but highly informative representation of time-resolved fluorescence data. We show the utility of phasors for analysis of strategically substituted 2AP or FRET-labeled G-quadruplex studies. Phasor plot trajectories using these probes reveal subtle aspects of G-quadruplex conformational transitions that are often not evident in steady-state fluorescence, circular dichroism or absorbance measurements, thus providing novel insights into the underlying reaction mechanisms. We show the utility of phasor diagrams for providing signatures of different human telomere G-quadruplex conformations, monitoring cation-driven folding, evaluating nuclease digestion reactions and visualizing complex collisional quenching reactions. As is the case for most electronic spectroscopies (absorbance, CD, steady-state fluorescence), phasor diagrams cannot provide high-resolution structural details about G-quadruplexes, but are of significant value in monitoring changes in conformation driven by changes in solution conditions. Phasor diagrams have seen little, if any, use in time-resolved fluorescence studies of nucleic acids. We show here their value for revealing new information about G-quadruplex transitions not available by other spectroscopic tools.

CHAPTER 3

HYDRATION IS A MAJOR DETERMINANT OF THE G-QUADRUPLEX STABILITY AND CONFORMATION OF THE HUMAN TELOMERE 3' SEQUENCE OF D(AGGG(TTAGGG)₃).

The factors that determine the conformation and stability of G-quadruplex forming sequences remain poorly understood.[67,71,108,126] Here we demonstrate the influence of co-solvents on the conformation and stability of the human telomeric sequence d(AGGG(TTAGGG)₃) in both K^+ and Na^+ containing solutions using a combination of circular dichroism (CD), nuclear magnetic resonance (NMR), and thermodynamics. In the presence of certain monovalent cations, guanine-rich DNA of the general type $(G_{2+}N_n)_{4+Y}$ can form three-dimensional structures known as G-quadruplexes. These are made up of stacks of two or more square planar arrays of four Hoogsteen hydrogen bonded guanines called G-quartets. G-quadruplexes are known to be stabilized by coordinating a cation such as Na^+ or K^+ to the O6 of the guanines.[67] Although most often associated with the ends of telomeres, potential quadruplex-forming sequences have been found throughout the genome.[127]

Quadruplex formation is intrinsically complex. When factors such as strand orientation, loop type and arrangement, and glycosyl torsion angles are considered, even the simple human telomeric sequence d(GGGTTA)ₙ can theoretically fold into more than 200 unimolecular conformations.[13,108] Formation of tetramolecular and bimolecular structures of the G-quartets can lead to an even greater degree of polymorphism. Factors such as ion type, DNA concentration, the presence of other molecules such as organic solvents and various biological

molecules, and even annealing protocols can determine what quadruplex conformations are formed.[71,108,128,129]

Here we report the influence of solvation on the stability and conformation of the human telomeric quadruplex d(AGGG(TTAGGG)₃), which has strong implications for the state of this sequence in the conditions existing in the nucleus. Several structures based on this sequence have been reported, and the behavior, folding kinetics, and biophysical properties of this sequence have been extensively studied.[57,62,67,72,120,129] Recent studies concerning the role of small molecules in quadruplex stability have raised new questions about the properties of this sequence in solution particularly with regard to the stabilization/destabilization of the K^+ solution form and the species revealed by the crystal structure, which is not significantly populated in dilute aqueous solution.[67,71]

The crystallization procedure for the K+ form used PEG 400 as a precipitating agent.[57] Since PEG 400 is inconvenient for NMR and has a low dielectric constant, we chose an alternative solvent, acetonitrile, because it is completely miscible with water, is an excellent general solvent, has a high dielectric constant and low viscosity, does not participate in hydrogen bonding, and is readily available in deuterated form.

The CD of the oligonucleotide in the K^+ form changes dramatically in the presence of PEG or acetonitrile (Figure 13A and 13B), consistent with a change to a new conformation, that is very similar for the two co-solvents. In contrast, the sodium form, which has a radically different CD and structure from those in K^+ solution, showed only small changes on adding CH3CN (Figure 13C). This behavior was mirrored in the proton NMR spectra as shown in Figure 14. The NMR spectrum of the aqueous K^+ solution is the most complex (Figure 14A), which by integration comprises 23-24 iminoproton resonances. As the sequence contains 12 G residues, this is consistent with the presence of two major conformations. From the integration of

resolved resonances, these are approximately equally populated. This finding agrees with previously reported NMR data for this sequence under these conditions.[66] In contrast, in aqueous Na^+ solutions the NMR spectrum is much simpler (Figure 14C) comprising 11-12 narrow GN1H resonances, consistent with a single quadruplex structure that differs from the mix of species present in K^+ solution. Addition of CH_3CN to 50% (v/v) caused a large change in the NMR spectrum in K+ solution, to a single species containing approximately 11 GN1H resonances (Figure 14B). In contrast, the spectral changes from 50% (v/v) CH3CN in Na+ solution were much smaller (Figure 14D). These small differences may be attributable to the change in bulk susceptibility/dielectric constant.

The melting of the oligonucleotide measured by CD (Figure 15) shows that the mixture in the presence of potassium is more stable than the sodium form and that acetonitrile greatly increased the stability of both forms. At 298 K, the ΔG of folding of the sodium form increased from -3.74 to -7.1 kcal mol^{-1}. The stabilization of the K^+ form was substantially greater; the ΔG increased from -4.9 to -10.7 kcal mol^{-1} in 50% CH_3CN (Appendix A: Table A1). This is associated with a change in conformation (see above).

The NMR line widths and cross-relaxation rate constants determined from NOESY and truncated 1D NOE build-up curves showed that the rotational correlation time of this species (<2 ns at 293 K) implies that the quadruplex is monomeric (see Appendix A). NOESY data showed that the K^+-form in 50% CH_3CN contains G-quartets and that some nucleotides have glycosyl torsion angles different from those of the high anti range but are also not pure syn (Appendix A: Figures A1, A2, A3, A4, and Table A2). The latter suggests that the conformation is not identical to the all-parallel form reported for the crystal structure.[57] Similarly, assignment of the parallel structures in co-solvents such as PEG 200 based solely based on CD spectra[129] is not
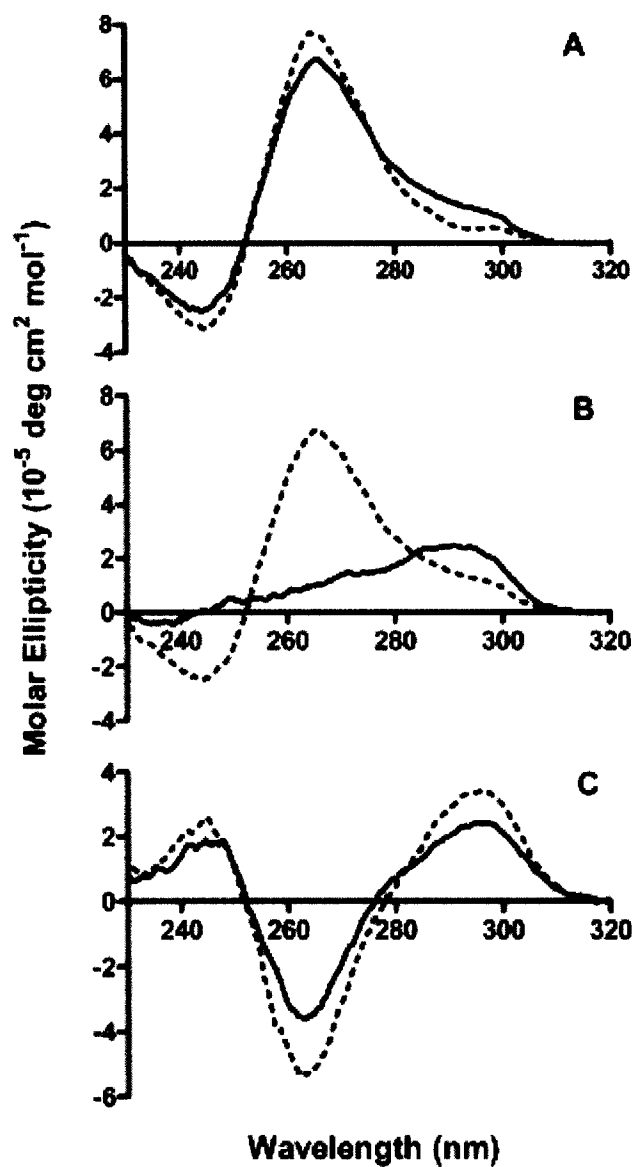
Figure 13. CD spectra of d(AGGG(TTAGGG)$_3$). Spectra were recorded on a JASCO J-810

spectropolarimeter at 293 K in buffers containing 25 mM KCl or NaCl and 10 mM phosphate

buffer, pH 7. (A) K$^+$ with 50% (v/v) acetonitrile (solid line), K$^+$ with PEG 400 (dashed line). (B) K$^+$

(solid line), K$^+$ with 50% (v/v) acetonitrile (dashed line). (C) Na$^+$ (solid line), Na$^+$ with 50% (v/v)

acetonitrile (dashed line).
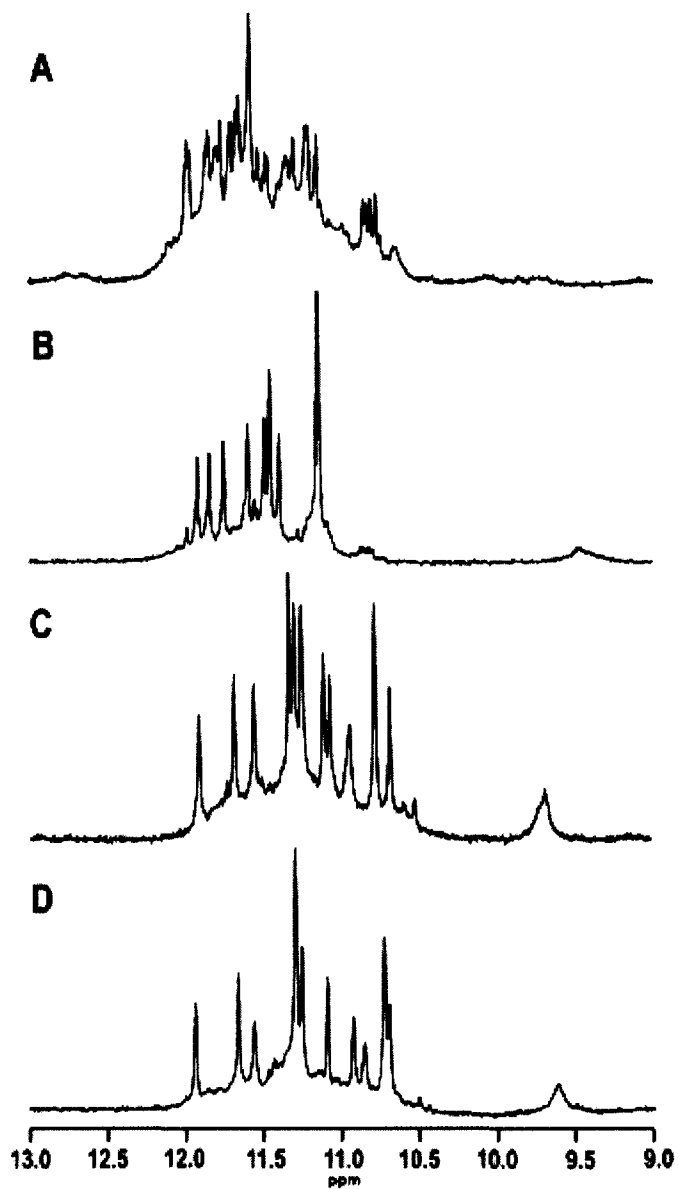
Figure 14. ^{1}H NMR of the imino protons of d(AGGG(TTAGGG)$_3$) under different solution conditions. NMR spectra were recorded at 293 K and 800 MHz on a Varian Inova spectrometer. (A) K$^+$; (B) K$^+$ with 50% CD$_3$CN; (C) Na$^+$; (D) Na$^+$ with 50% CD$_3$CN.
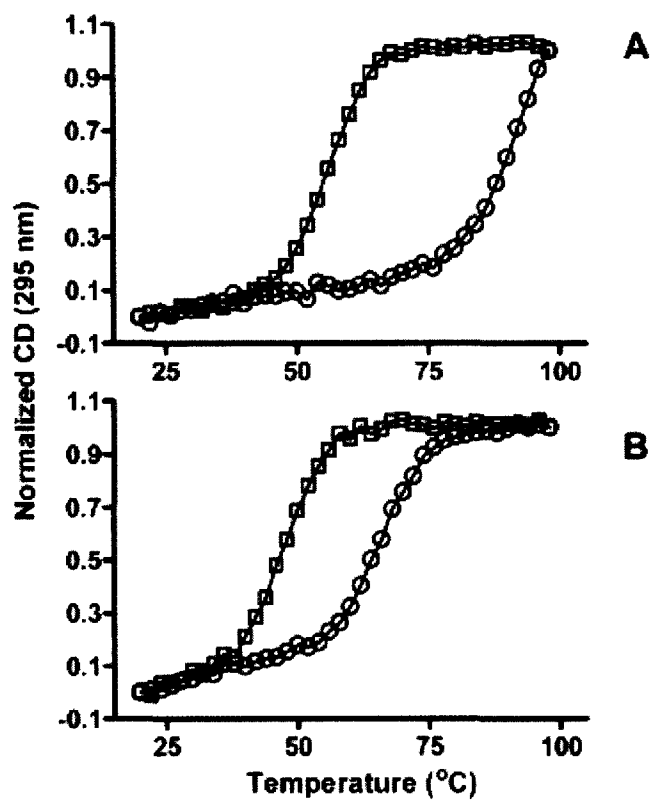
Figure 15. CD melting of d(AGGG(TTAGGG)$_3$) under different conditions. Squares: buffer; circles: 50% acetonitrile. (A) K$^+$ in the absence and presence of 50% CH$_3$CN. Data were fitted to a two-state model including sloping baselines.[138] T$_m$ = 372 K, ΔH = 53.7 kcal mol^{-1} (CH$_3$CN); T$_m$ = 329 K, ΔH = 51.9 kcal mol^{-1} (buffer). (B) Na$^+$ in the absence and presence of 50% CH$_3$CN. T$_m$ = 339 K, ΔH = 58.1 kcal mol^{-1} (CH$_3$CN); T$_m$ = 321 K, ΔH = 51.8 kcal mol^{-1} (buffer). Solid lines are nonlinear regression fits.

definitive for structural classification[126,130] because CD reports on base stacking interactions, not directly on strand polarity.[131] Conventional interpretation of quadruplex CD spectra holds that parallel quadruplexes generally display a negative peak near 240 nm and a positive peak near 265 nm, and that anti-parallel quadruplexes are characterized by a negative peak near 260 nm and a positive peak near 295 nm.[132-135] However, there are many exceptions to this rule[130,136,137] so no assignment to parallel or anti-parallel strand orientation is really possible without supporting structural data.

We recently reported that CD can hide the underlying complexity and polymorphism of quadruplex-forming sequences.[126] There were at least eight resolvable species present in solution, which collectively showed a CD spectrum that would be classified as being "parallel" according to this convention, yet individual components have a wide variety of CD spectra.[126] This illustrated the need for a high-resolution technique, such as NMR, to be used when investigating quadruplex topology.

There are three possible physical explanations for the effects of CD_3CN on stability and conformation: differential binding; effects on hydration; and steric crowding. Crowding effects require a sufficiently large difference in volume between the conformational states. Because the difference in excluded volumes among various folded isomers of the 22-mer are very small,[71] molecular crowding per se is unlikely to be important, and any steric crowding effect will favor the most compact state.[139] As the all-parallel form present in the crystal state[57] is less compact than other known structures[71] it seems unlikely that this structure can be produced by crowding. Simulations of the effect of non-interacting macro-solutes on the CH_3CN-induced conformational transition observed in K^+ solutions show that indeed the effect on the equilibrium is very small and in the direction that destabilizes the all-parallel form (Appendix A: Figure A5). A more plausible interpretation is that the species differ in their hydration

properties, as has been proposed for the thrombin aptamer.[129,140] In duplex DNA, the B-form is stabilized by solvation;[141] decreasing the water activity causes a transition to the A-form and, ultimately, to the strand state. In contrast G-quadruplex structures appear to be destabilized by hydration; the thermodynamic stability increases as the water activity is reduced,[140,142] and the surface area accessible to water is minimized. The transition between two compact molecular forms implies that there is a difference in solvent accessible surface area and/or new sites of different affinity are exposed or hidden during the transition. The solvent accessible surface areas, as calculated by Naccess,[143] for the known structures of the human telomere, i.e., the basket form (143D.pdb[62] 3749 Å^2), the parallel form (1KF1.pdb[57] 4182 Å^2), hybrid 1 (2JSM.pdb[67] 3814 Å^2, 2HY9.pdb[64] 3552 Å^2, 2GKU.pdb[65] 4066 Å^2), and hybrid 2 (2JPZ.pdb[66] 3581 Å^2, 2JSL. pdb[67] 3843 Å^2) when truncated to the d(GGG(TTAGGG)$_3$) quadruplex, are different and do not support the parallel form being the most compact. In an additional measure of compactness, the radius of gyration, as calculated by Hydropro,[144] for the basket form (143D.pdb 1.17×10^{-7} cm), the parallel form (1KF1.pdb 1.36×10^{-7} cm), hybrid 1 (2JSM.pdb 1.19×10^{-7} cm, 2HY9.pdb 1.16×10^{-7} cm, 2GKU.pdb 1.21×10^{-7} cm) and hybrid 2 (2JPZ.pdb 1.15×10^{-7} cm, 2JSL.pdb 1.23×10^{-7} cm) does not support the parallel form being the most compact.

To assess further experimentally the role of steric crowding on the conformational transition in K^+ solution, we have recorded the NMR spectrum of the oligonucleotide in K^+ buffer in the presence of 300 mg/mL bovine serum albumin (BSA) ($M_w = 66382$). At neutral pH BSA is negatively charged[145] and thus is unlikely to bind significantly to the negatively charged quadruplex. The protein contains no tryptophan residues and has no NMR resonances in the 10-12 ppm region that could overlap those of the DNA imino protons. Although the solution viscosity increases approximately 4-fold, thereby broadening the NMR spectrum, the NMR spectrum is more similar to that of the K^+ form in water than that observed in 50% CD_3CN

(Figure 14B and Appendix A: Figure A6). Even at this high cosolute concentration, the molar

concentration is only about 4.5 mM, and its effect on water activity is thus very small.[145,146]

Therefore in this system hydration effects are not significant, and therefore high concentrations

of BSA should influence equilibria only by steric crowding. As the spectral changes in the

presence of BSA do not resemble the formation of the species present in 50% CD_3CN, that

species is most likely stabilized by dehydration, and not by steric crowding. The steric crowding

effect is also small, as expected for an isomerization between two or more similarly compact

states (Appendix A: Figure A6).

In conclusion acetonitrile is an excellent solvent for nucleic acids and is suited for

detailed biophysical characterization of the solute. Compact G-quadruplexes appear to be

stabilized by dehydration, which has relevance to intracellular conditions where the absolute

water activity must be much lower than 55.5 M.[147]

CHAPTER 4

POLYETHYLENE GLYCOL ALTERS G-QUADRUPLEX STRUCTURE BY CONFORMATIONAL SELECTION VIA DIFFERENTIAL BINDING.

Introduction.

The highly conserved repetitive DNA sequence 5'-d(TTAGGG)$_n$ is found within the telomeres of all human chromosomes.[148] Thousands of copies of the sequence are found in the telomere, predominantly in duplex form, but with approximately 200 nt as a single-stranded 3' overhang.[149] This overhang is an unusual example of a functional single-stranded element within the human genome. In the presence of monovalent cations, at least four repeats of the TTAGGG sequence can fold into quadruplex structures comprising stacked G-quartets and a variety of loop topologies.[5,150-153] A designed quadruplex-specific ligand, 360A, binds preferentially to chromosomal terminal regions of both human normal and tumor cells,[154] supporting the hypothesis that quadruplex ligands can induce or stabilize quadruplex structures within telomeres of human cells. The vast literature on the chemistry and biology of human telomere DNA is summarized in a recent review.[155]

The telomeric TTAGGG repeat sequence can fold into numerous quadruplex structures depending on the cations present and on nucleotide additions to the ends of the sequence. In Na$^+$-containing solutions, an anti-parallel basket conformation was identified by NMR.[156] This structure features three stacked G-quartets, two lateral loops, and one diagonal loop. In K$^+$-containing crystals, a novel parallel-stranded "propeller" structure was observed,[57] featuring all-parallel strand segments and three stacked G-quartets linked by three chain-reversal "side"

64

loops. A variety of chemical and biophysical approaches has revealed that the propeller structure is not the predominant form in K^+-containing solutions.[157-161] Subsequently, two types of anti-parallel "hybrid" or "3+1" forms were determined by NMR.[65,162-164] These structures feature three stacked G-quartets, two lateral loops, and one side loop, where the location of the side loops differs in forms 1 and 2. Recently an anti-parallel basket form with only two stacked G-quartets was identified in K^+- solution by NMR.[68] Which of these structures exist in the cell is of great interest and a matter of some debate.

Quadruplex structures are very dependent on solution conditions. Addition of small molecule co-solvents or macromolecular co-solutes is commonly used to perturb conformational equilibria. There are four main mechanisms by which additives can influence conformation, namely macromolecular crowding (which acts on shape and size differences); osmotic stress (which acts by hydration differences); differential binding (general solvation effects); and by general electrostatics (e.g. dielectric effect on ionic activities). The propeller quadruplex conformation was found to form in solution under dehydrating conditions,[79,80] in high concentrations of polyethylene glycol (PEG),[80,129,158] and in a non-aqueous, high-viscosity deep eutectic solvent.[165] The cell and its compartments are certainly crowded, with macromolecules occupying approximately 30% of the total volume in eukaryotes.[166,167] The resultant excluded volume effects can drastically influence biochemical reactions by driving equilibria in the direction of species with the minimal volume.[168-172] Efforts to mimic the intracellular environment of quadruplexes have relied on solute crowding reagents, typically PEGs of various molecular weights.

Hydration is a fundamentally important aspect of biochemical reactions, as hydration significantly impacts macromolecule stability and conformation[173,174] and water is an active

participant in many biochemical reactions.[173,175-177] However, the properties of the aqueous environment in the cell are a matter of some debate.[173,177] The structure of water in the cell is inhomogeneous: the hydration layers around macromolecules differ from bulk water, and "free" water also differs in some properties, including viscosity and diffusion coefficients. Processes such as macromolecular folding that involve differentially hydrated conformations depend on the activity of the water, such that a reduction in water activity favors the less hydrated species.[174] However, the exact water activity throughout the cell is poorly defined. [178]

Crowding and hydration effects on reactions are distinct phenomena with different underlying physical principles, although an experimental separation of the effects can be complicated.[178] Experimental studies of hydration using osmotic stress[176,179] or of macromolecular crowding[180,181] both rely on the addition of neutral co-solutes intended primarily to alter either water activity or excluded volume, respectively. Because any co-solute affects both of these solution properties to variable degrees, separating the dominant influence poses a challenge. Typically, water activity is changed by the addition of small neutral organic compounds (e.g. alcohols, sugars, ethylene glycol) whereas excluded volume is altered by addition of large polymers (e.g. polyethylene glycols, dextrans, proteins). In both approaches, it is essential that the added solutes are inert and do not interact with either reactants or products of the reaction under study. By convention, properly designed hydration or crowding studies typically use a number of different solutes to show that any observed effects are independent of the chemical properties of the individual solute additives.[141]

One early example studied hydration and crowding effects on the stability of DNA triplex and duplex structures.[141] The small osmolytes ethylene glycol, glycerol, acetamide and sucrose were found to exert similar quantitative effects, decreasing water activity and

destabilizing both duplex and triplex DNA. The results showed that the helical forms are more hydrated than the denatured single-strands. In contrast, addition of the volume-excluding polymers PEG 400, 1000, 3400, and 10000 stabilized both duplex and triplex DNA. Quantitative analysis of a large set of thermal denaturation data demonstrated that the effects of the polymers could be correlated with geometric models for excluded volume and that their effects were due primarily to crowding. In this case, hydration changes and crowding effects could be clearly separated.

In contrast to DNA duplexes and triplexes, decreased water activity stabilizes quadruplexes and increases their melting temperature.[79,180,182-184] This indicates that folded quadruplexes are less hydrated than their unfolded single-strands, and that water is released upon folding. Initial quantitative estimates using PEG suggest 13-18 water molecules per G-quartet are released upon folding.[140,183] Decreased water activity also drives a conformational transition of the human telomeric quadruplex in potassium solutions from a hybrid to a propeller form,[79,80] indicating that these particular quadruplex conformations differ in their hydration.

Recent investigations have focused on the ability to perturb a variety of quadruplex equilibria through the addition of co-solvents. PEG 200 facilitates the folding of the sequence 5' $G_3(T_2AG_3)_3$ into an anti-parallel form in the absence of cation.[185,186] In addition, PEG drives the formation of a quadruplex from an initial duplex formed by $(TTAGGG)_4$ and its complementary strand.[187,188] Both of these cases are potentially consistent with a crowding mechanism which demands that the reaction be driven in the direction that reduces the total excluded volume of the system, yielding more compact species. In these cases, the starting reactants (single-strand, duplex) have larger hydrodynamic volumes than the final folded quadruplex product.

PEG has been reported to drive a transition between the anti-parallel hybrid and all-parallel propeller conformations of the human telomere sequence. This was first reported qualitatively in an attempt to rationalize why the solution structure differed from the structure observed in crystals.[158] More detailed studies followed,[129,186,189] including the determination of a high-resolution structure by NMR.[80] In none of these cases were the contributions from hydration and crowding clearly separated and quantitatively accounted for. A puzzling aspect of these studies is why crowding should drive the transition of the more compact folded hybrid form to the hydrodynamically larger propeller form[79] because this violates the fundamental basis of macromolecular crowding. Other crowding agents that also have minimal effects on water activity, including BSA, Ficoll 70, Ficoll 400 did not drive the hybrid to parallel quadruplex transition,[79,189,190] nor was the propeller conformation observed by CD or NMR in a *Xenopus laevis* egg extract, which represents perhaps the best mimic of the crowded intracellular environment currently available.[190] For a true crowding mechanism, the perturbation of the equilibrium should be independent of the chemical nature of the crowding agent. Macromolecular crowding therefore is an unlikely mechanism by which PEG alters quadruplex conformation and additional studies are needed to understand how it acts.

The use of PEG in macromolecular crowding studies has been discouraged.[171] A strict demand on a crowding reagent is that it be inert, with no binding to either the reactants or products of the reaction under study. Several studies have shown that the stabilization of protein by PEG cannot be described quantitatively in terms of excluded volume alone, but rather that there is an attractive interaction between PEG and nonpolar or hydrophobic side chains on the protein surfaces.[171] These attractive interactions make quantitative analysis of crowding effects difficult and compromise the use of PEG in crowding experiments. Further, PEG behaves

like a weak organic solvent, and not simply as an inert volume excluder.[191] For these reasons, it has been suggested PEG be avoided for crowding studies.[192]

Crowding studies using PEG have been used to support the existence of a parallel conformation for telomeric DNA *in vivo* and to argue that it is the most relevant conformation to serve as a target for drug discovery.[80,129,193] To assess these claims, it is important to understand the underlying mechanism by which PEG alters quadruplex conformation, and in particular to separate and quantify the contributions of hydration and crowding. An analytical strategy for the separation of preferential interactions and excluded volume effects was recently described.[194] The analysis requires a study of the effects of a full series of monomeric to polymeric ethylene glycols on the "m-value", the derivative of the observed free energy change for the reaction with respect to solute concentration. Such a systematic approach has not been previously used to study quadruplex structure and stability. The goals of this work are to evaluate the contributions of hydration, crowding, or other interactions on the conformational transition between the hybrid and propeller forms and to understand the mechanism by which PEG influences the transition. Our results will show that PEG is not acting as a macromolecular crowding agent, making suppositions about the conformation of the quadruplex in the cell based on studies using PEG suspect.

Methods and Materials.

Oligodeoxynucleotides and Co-solvent Preparation. The 22-nt human telomeric G-quadruplex forming oligodeoxynucleotide (ODN) 5`-d(AGGG(TTAGGG)$_3$) was purchased from Integrated DNA Technologies (Coralville, IA). DNA was purchased on a 10 µmol scale with standard desalting. Fluorescently tagged 6FAM- d(AGGG(TTAGGG)$_3$)-TAMRA was purchased from Sigma (St. Louis, MO) on a 10 µmol scale with standard desalting and HPLC purification. All DNA is quality checked by commercial sources through the use of either MALDI-TOF or Electrospray Ionization mass spectrometry. All purchased DNA was used without further purification. Fluorescent tag abbreviations are 6-carboxyfluorescein (6FAM) and tetramethylrhodamine (TAMRA). Prior to use, DNAs were diluted into folding buffer (10 mM tetrabutylammonium phosphate, 1 mM EDTA acid, pH 7.0) to a stock concentration of 1 mM and stored at 4 °C. The folding buffer is designed to prevent structure formation prior to the addition of cation. Acetonitrile and all PEGs were purchased from Sigma at the highest available grade. PEGs arrived as either a solid or premade 100% (w/v) or 50% (w/v) solutions. Acetonitrile and premade PEG solutions were used without further purification. Solutions of 50% (w/v) PEGs were prepared by dilution of 125 g PEG into 250 mL total volume using purified water. In some cases, gentle heating was required to aid in dissolving the PEG prior to dilution to final volume. Molar concentrations were calculated for all PEG solutions and densities were used to normalize all solutions to % (v/v) for pipette mixing of solutions.

Circular Dichroism and Absorbance Spectroscopies. Simultaneous circular dichroism (CD) and absorbance (Abs) spectra were collected using a Jasco J-815 spectropolarimeter equipped with a Peltier temperature controller. CD/Abs spectra were collected from 220 to 340 nm with 1 nm steps and standard sensitivity while maintained at a constant 20 °C. Scan rate

was set to 200 nm/min with a 2 second integration time. A total of four scans were collected and averaged. Prior to each scan, blank spectra were collected using the corresponding solvent conditions without the addition of DNA. For direct comparison of all collected CD spectra, scans were normalized to delta epsilon ($\Delta\varepsilon$) through the use of Equation 1, where θ_λ is the observed circular dichroism in millidegrees at wavelength λ, M is DNA strand concentration in molar, and L is pathlength in cm. Normalization was done using Microsoft Excel software and all plots were constructed using Origin 7.0 software.

$$\Delta\varepsilon_\lambda = \theta_\lambda/(32982*M*L) \quad \text{(Equation 5)}$$

Co-solvent Titrations. Isothermal co-solvent titrations were conducted using 18 individually prepared samples over a co-solvent range of 0 – 42.5% (v/v) measured at a constant temperature of 20 °C. DNA samples were prepared to a final volume of 2 mL by mixing of DNA, premade co-solvents, 10x concentrated buffer components, and purified water to a final buffer concentration of 25 mM KCl, 10 mM tetrabutylammonium phosphate, 1 mM EDTA, pH 7.0 (25 mM KCl-folding buffer) with 0 – 42.5% (v/v) co-solvent. Co-solvents were added based on volume through manual pipetting. Samples were prepared at 4 μM strand concentration and annealed by placement in 1L of boiling water for 20 minutes followed by reduction of heat to room temperature and slow cooling. Samples were allowed to cool slowly in the water bath overnight prior to use. CD/Abs scans were collected on all individually prepared samples and on buffered co-solvent solutions in the absence of DNA to ensure no significant background was present. Data was exported for all titration points as millidegrees and normalized to $\Delta\varepsilon$ using Equation 5.

Water Activities. Water activities for all co-solvents were obtained from published values.[195-197] Agreement between published values and solvents prepared in the laboratory was

checked when possible using a Vapro vapor pressure osmometer 5520 (Wescor Inc., Logan, UT).

Agreement between published values and experimentally obtained values was within at least

5% for the instrumental range of 3200 mmol/kg. Vapor pressures were converted to water

activities using Equation 2,[198] where mOsm is the vapor pressure measured in mmol/kg, α_w is

the water activity, and M_1 is the molecular weight of water.

$$mOsm \equiv -(10^6 \ln(\alpha_w))/M_1 \text{ (Equation 6)}$$

FRET Thermal Denaturations. Thermal denaturations were conducted using an Applied

Biosystems StepOnePlus Real-Time PCR instrument (Applied Biosystems, Carlsbad, CA).

Separate samples of the FRET-labeled FAM-5'-d(AGGG(TTAGGG)$_3$)-TAMRA were prepared

similarly to CD/Abs samples in 25 mM KCl-folding buffer with the addition of 0 – 50% (v/v)

acetonitrile over 11 points. Prior to melting experiments, samples were kept at room

temperature for 24 hours to ensure equilibration of the quadruplex species. Samples were

loaded in triplicate using 20 µL per well and scanned over the temperature range of 2.0 – 99.8

°C, with scans collected every 0.2 °C. Fluorescent emission was collected using all provided

instrumental cutoff filters simplifying data output to total emission intensity. Data was exported

as raw fluorescence emission against temperature. Nonlinear regression analysis was

conducted using *Mathematica 8.0* software.

Analytical Ultracentrifugation. AUC was carried out in a Beckman Coulter ProteomeLab

XL-A analytical ultracentrifuge (Beckman Coulter Inc., Brea, CA) at 20°C for 24 hours at 50,000

rpm in standard 2 sector cells. Data were analyzed using Sedfit (free software:

www.analyticalultracentrifugation.com) and DCDT+ (version 2.3.2, John Philo, Thousand Oaks

CA). Values for $S_{20,w}$ obtained by the two analysis methods agreed within 5% until 30% (v/v) PEG

was used, where DCDT+ failed to yield a result due to the very low value of the uncorrected

sedimentation coefficient in this high viscosity solution. Buffer density was determined on a Mettler/Paar Calculating Density Meter DMA 55A at 20 °C and buffer viscosity was determined using an Anton Parr Automated Micro Viscometer AMVn at 20 °C. For the calculation of frictional ratios, 0.55 mL/g was used for partial specific volume[199] and 0.3 g/g was assumed for the amount of water bound.[200]

Molecular Modeling. The propeller form of the human telomere (PDB entry 1KF1) and hybrid 1 form (PDB entry 2HY9) were used to examine PEG 200 and 600 binding. The hybrid 1 form was truncated to the same 22 bases as the parallel form (2HY9-t). Four PEG 200 or 600 molecules were manually docked to the groove regions of 1KF1 and in two separate systems to the groove regions and quartet backbone of 2HY9-t. The PEG 200 and 600 molecule partial charges were calculated using Antechamber and AM1-BCC methods and parameters from the GAFF force field. An explicitly solvated model and potassium counter ions were generated with the parm99SB.dat Amber force field using the following protocol: (i) two unsolvated potassium ions were placed between the G-quartet tetrads for stabilization (ii) the system was solvated by the addition of a rectangular box of TIP3P water at 15 angstroms, and (iii) neutralizing solvated potassium ions were added randomly around the quadruplex:PEG 200 or 600 structure using Amber 11 leap rules for counter ions. The system was heated slowly and equilibrated for 600 ps using the following protocol: minimize water holding the DNA:PEG 200 or 600 (100 kcal mol^{-1} Å^{-1}), minimize the complete system, 50 ps molecular dynamics (heating to 300 K) holding the DNA:PEG 200 or 600 fixed (100 kcal mol^{-1} Å^{-1}), (iv) unrestrained molecular dynamics for 300 ns. Simulations were performed in the isothermal isobaric ensemble (P = 1 atm, T = 300 K). Periodic boundary conditions and the Particle-Mesh-Ewald algorithm were used. A 2.0 fs time step was used with bonds involving hydrogen atoms frozen using SHAKE. For the equilibration steps and

the production steps, molecular dynamics calculations were carried out using AMBER 11 program sander and the cuda version of pmemd, respectively.

The 50% v/v acetonitrile system was investigated using the propeller form of the human telomere (PDB entry 1KF1) with 4199 acetonitrile and 8398 TIP3P water molecules using periodic box conditions. All atom parameters for acetonitrile were used from the Manchester Amber Parameter Database (http://www.pharmacy.manchester.ac.uk/bryce/amber#org). The 2HY9-t model was solvated in rectangular box of TIP3P water at 15 angstroms. The same protocol as above was used with restraints only on the quadruplex DNA for 60 ns to allow for equilibrium of the solvent mixture with production runs of 500 ns for the acetonitrile:water:propeller quadruplex and 100 ns for the water:hybrid 1 quadruplex. The final imaged snapshots of the production trajectories were used for solvation shell analysis.

Results and Analysis.

Acetonitrile causes a shift in quadruplex conformation in potassium solutions through changes in hydration. The conformational change of the 22-nt human telomeric quadruplex (hTel22) in the presence of potassium was monitored by changes in CD spectra over a range of 0 – 42.5% (v/v) acetonitrile at 20 °C (Figure 16A). In the absence of acetonitrile, hTel22 showed a major peak near 290 nm with a shoulder at 265 nm consistent with a hybrid conformation.[201] The addition of acetonitrile caused a shift in its CD spectrum to a major peak near 265 nm and a shoulder at 290 nm consistent with a parallel propeller conformation.[201] CD spectra were collected at intervals of 2.5% (v/v) acetonitrile and were used to construct a data matrix representing the spectral changes as a function of water activity. Changes in CD intensity at 265 and 290 nm can be used to monitor the appearance of the propeller form and the disappearance of the starting hybrid quadruplex, respectively (Figure 16B). Evaluation of the titration curves demonstrates 50% conversion of the hTel22 is obtained at $\ln(\alpha_w)$ of -0.12 (α_w=0.886) with complete conversion at -0.25 (α_w=0.778) and lower. The observation that decreased water activity favors the propeller form indicates that it is less hydrated than the hybrid form.

Singular value decomposition (SVD) was applied to the data matrix to evaluate the number of significant spectral species involved in the quadruplex conformational transition.[202] SVD results (not shown) show that only two significant spectral species were needed to describe the data matrix, indicating that the reaction is effectively two-state within the limitations of the information content of CD spectra. The presence of only two significant spectral species indicates that the conformational change can be modeled by a simple equilibrium between two
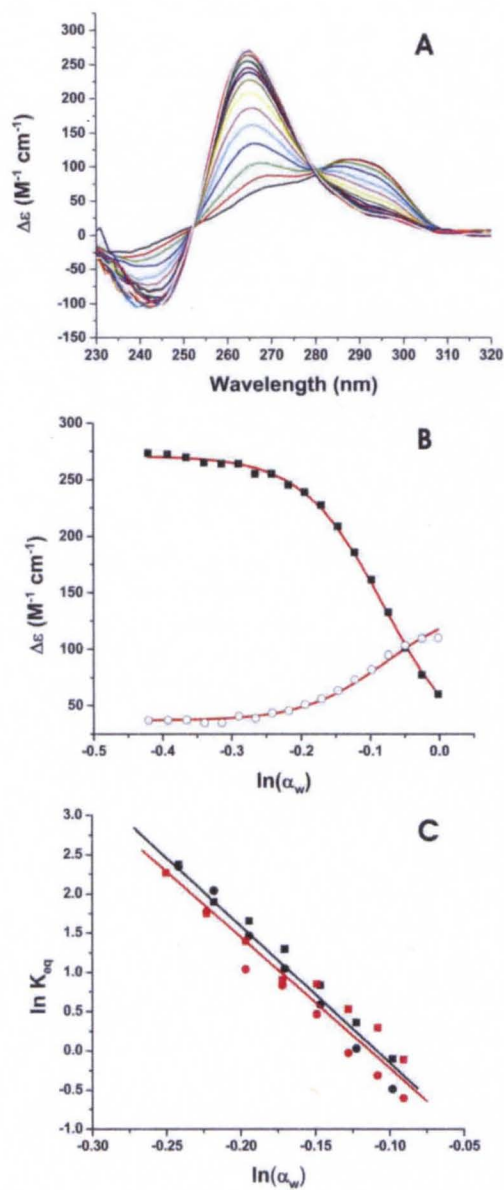
Figure 16. Conformational change of hTel22 driven by decreased water activity in the presence

of potassium at 20 °C. A) CD spectra obtained as a function of acetonitrile concentration.

Addition of acetonitrile causes a shift in the CD spectrum from a major peak at 290 nm to a

major peak at 265 nm. B) The conformational change of the quadruplex monitored at 290 nm

(circles) and 265 nm (solid squares) as a function of the natural logarithm of the water activity.

The 295 nm signal monitors the loss of initial hybrid structure while the 265 nm signal arises

from the appearance of parallel quadruplex. C) Data for acetonitrile (black) and ethylene glycol

(red) plotted according to Equation 5 for determination of $K_{eq,0}$ and Δn_w. Best-fit lines for

acetonitrile and ethylene glycol were determined to be $\ln(K_{eq}) = -1.77 - 17.0\ln(\alpha_w)$ and $\ln(K_{eq}) = -1.89 - 16.9\ln(\alpha_w)$, respectively.

hydrated quadruplex conformations.   Mechanism 1 is a minimal reaction scheme that

accounts for a shift in conformation from the initial hybrid form (Q_{Hyb}) to the converted

propeller form (Q_P) accompanied by a change in hydration ($\Delta n_w = x - y$).

$$Q_{Hyb} \cdot [H_2O]_x \longleftrightarrow Q_P \cdot [H_2O]_y + (x-y)[H_2O] \quad \text{(Mechanism 1)}$$
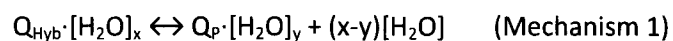
The two-state reaction mechanism defines an apparent equilibrium constant (K_{eq}) that can be

estimated at each acetonitrile concentration by Equation 7,

$$K_{eq} = \theta/(1-\theta) \quad \text{(Equation 7)}$$

where θ is the fraction of quadruplex converted. Equation 8 was used to calculate θ for each

acetonitrile concentration,

$$\theta = (\Delta\varepsilon_i - \Delta\varepsilon_0)/(\Delta\varepsilon_f - \Delta\varepsilon_0) \quad \text{(Equation 8)}$$

where $\Delta\varepsilon_0$ is the CD signal from the starting hybrid conformation, $\Delta\varepsilon_F$ is the signal from fully

converted quadruplex, and $\Delta\varepsilon_i$ is the CD signal at the i-th acetonitrile concentration. A linear

dependence of $\Delta\varepsilon_F$ on acetonitrile concentration was assumed to account for the slopes evident

in the ends of the titration curves in Figure 16B. Plots were constructed of $\ln(K_{eq})$ against $\ln(\alpha_w)$

resulting in a linear dependence of the equilibrium constant against varied water activity (Figure

16C).

$$\delta\ln(K)/\delta\ln(\alpha_w) = \Delta n_w \quad \text{(Equation 9)}$$

Linear regression was conducted simultaneously on results from 265 and 290 nm to give

an estimate of $\Delta n_w = -17.0 \pm 0.7$ based on Equation 9.[176] The negative sign of the Δn_w

demonstrates that the conformational change of the hTel22 is accompanied by a significant

decrease in hydration, with the release of approximately 17 waters per quadruplex. The

equilibrium constant for the conversion of the hybrid form to the propeller form in pure water, $K_{eq,0}$, was determined by extrapolation to be $K_{eq,0}$ = 0.17 ± 0.02. This indicates that 14.5±3 % of the quadruplex exists as the propeller structure in the absence of acetonitrile. The free energy change for the transition, $\Delta G_{25°C}$ = -RTln($K_{eq,0}$) ≈ +1.0 kcal mol^{-1} at 25 °C. We caution that these quantitative estimates of water release and $K_{eq,0}$ refer to the simplest possible equilibrium model that is allowed by the available CD data. Models that include added complexity (such as multiple hybrid forms converting to the propeller form by parallel pathways) require additional fitting parameters with no improvement of the fits. The parameter values obtained with the minimal models lead to slightly different numerical estimates of the equilibrium distribution and hydration number, but do not influence the overall conclusions.

Molecular dynamics simulations were performed to examine the water release from the hydrated hybrid 1 form to the acetonitrile hydrated propeller form. The hybrid 1 form had solvation shells at 4, 5, and 6 angstroms comprising of 369, 486, and 598 water molecules, respectively. The acetonitrile hydrated propeller form had solvation shells at 4, 5, and 6 angstroms comprised of 342, 463, and 568 water molecules and 68, 87, and 115 acetonitrile molecules, respectively. Therefore, the Δn_w values at 4, 5, and 6 angstroms are -27, -23, and -30, respectively, which is remarkably consistent with the experimental value. The acetonitrile hydrated propeller form molecular simulation also indicates that there is no preferential binding of acetonitrile with the quadruplex as the acetonitrile composition of the solvation shells is 16-17 % compared to 33% for the bulk solvent composition. This was achieved after the equilibrium period of the simulation.
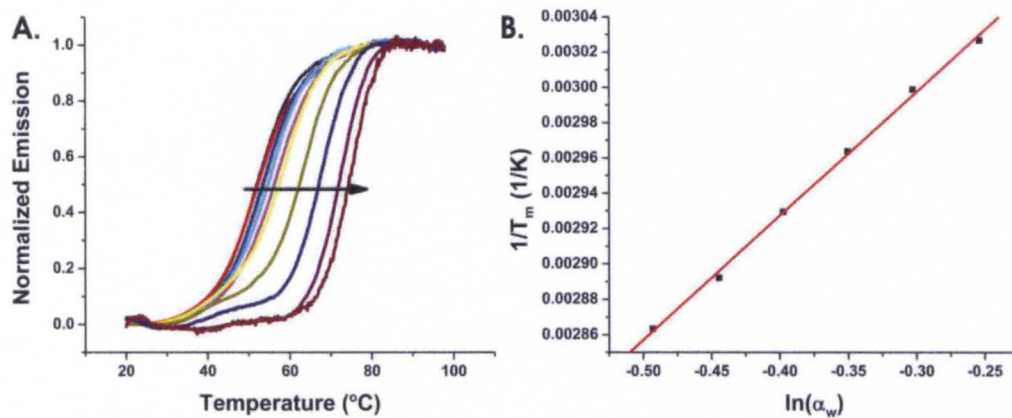
Figure 17. Thermal denaturation of hTel22 with increasing concentrations of acetonitrile. A) Normalized transition curves for FRET-labeled quadruplex in the presence of 25 mM KCl. The arrow indicates increasing acetonitrile concentration. B) The slope of T_m^{-1} versus $\ln(a_w)$ gives the hydration change (Δn_w) (Equation 10) for the denaturation of parallel hTel22. Red line was obtained by linear regression.

Conversion of the hTel22 by acetonitrile is accompanied by increased thermal stability. A FRET thermal denaturation assay was used to study changes in the thermal stability of the hTel22 quadruplex with increasing concentrations of acetonitrile. The conversion to the propeller form is essentially complete at ≥ 25% (v/v) ACN. Figure 17A shows that the quadruplex thermal stability increases with the concentration of acetonitrile beyond that point, indicating that the quadruplex is less hydrated than the unfolded single strand. These data were used to estimate the hydration change between the single-stranded state and folded propeller quadruplex. The fluorescent emission intensity curves were modeled as a two-state thermal denaturation process to determine melting temperatures (T_m). Best-fit parameters were used to construct plots of T_m^{-1} versus $\ln(\alpha_w)$ (Figure 17B) to evaluate the change in hydration upon conversion of the hTel22 using Equation 10.[179]

$$\Delta n_w = (-\Delta H/R)(\delta T_m^{-1}/\delta \ln(\alpha_w)) \quad \text{(Equation 10)}$$

A value of $\Delta n_w \approx$ +20 waters was determined for denaturation of the propeller quadruplex using an enthalpy of -56 kcal mol^{-1}. The positive sign indicating that more water is bound to the unfolded state. We caution that the exact value and meaning of Δn_w depends on the model assumed for the denaturation reaction, which may be more complex than a simple two-state process.

Acetonitrile, low MW PEGs, and high MW PEGs have distinct influences on the quadruplex conformational change. The conformational change of the hTel22 was studied by isothermal titrations using a set of increasing molecular weight (MW) glycols. The series included ethylene glycol (EG), diethylene glycol (diEG), triethylene glycol (triEG) and PEGs of MW 200, 400, 600, 1000, 1500, 3350, 8000, and 10000. Titrations over the range of 0-42.5 % (v/v) were monitored by circular dichroism. All glycols were found to drive the conformational

change of the hTel22. Analysis of spectral changes by SVD showed that only two significant spectral species were needed to describe the data matrix. At the highest glycol concentration used, a CD spectrum with a major peak at 265 nm and a shoulder at 290 nm is observed (Figure 18A). Conversion of the quadruplex by different MW glycols results in similar transition curves (Figure 18B). The magnitude of CD changes was found to be dependent on the MW, with EG, diEG, and triEG having lower values compared to the higher MW PEGs. This could result from either incomplete transitions or from actual differences in CD, perhaps resulting from binding interactions of the high MW PEGs with the quadruplex. The high MW PEGs showed $\Delta\varepsilon_{290}$ and $\Delta\varepsilon_{265}$ plateau regions of 28.6 ± 2.5 and 274.0 ± 21.9 M^{-1} cm^{-1}, respectively.

For comparison with acetonitrile titrations, the reduction in water activity caused by the addition of different MW PEGs was measured. The change in water activity required for the completed transition was found to decrease with increasing PEG MW (Figure 19A). For PEG 400 and larger, the transition midpoint is $\ln(\alpha_w) \approx -0.02$ ($\alpha_w \approx 0.98$) and the transition is complete at $\ln(\alpha_w) \approx -0.05$ ($\alpha_w \approx 0.95$). These values are dramatically different than the values seen for acetonitrile, and suggest that these PEGs are not greatly perturbing water activity and are not therefore acting by osmotic stress and dehydration. Figure 20 shows the dramatic difference between acetonitrile and PEG 600. The low MW glycols affect water activity to a greater extent (Figure 19A) and seem to act by a different mechanism than the high MW PEGs.

Ethylene glycol behaves very similar to acetonitrile. The data in Figure 18B can be transformed using Equations 3 and 4 to compare the effects of EG and acetonitrile, as shown in Figure 16C. Data for EG and acetonitrile are seen to superimpose within error, suggesting that both reagents are acting by the same mechanism to perturb quadruplex hydration. The EG data in Figure 16C yield an estimate of $\Delta n_w = -16.9 \pm 0.9$, in excellent agreement with the value found
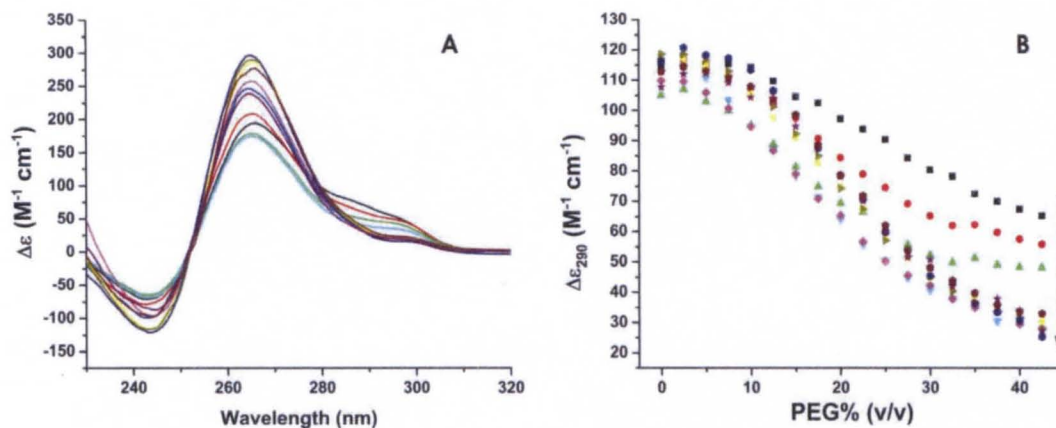
Figure 18. Isothermal titrations of hTel22 with a series of increasing MW PEGs. Colors correspond to ethylene glycol (black), diethylene glycol (red), triethylene glycol (green), PEG 200 (blue), PEG 400 (cyan), PEG 600 (magenta), PEG 1000 (yellow), PEG 1500 (dark yellow), PEG 3350 (dark blue), and PEG 8000 (purple), and PEG 10000 (wine). A) Comparison of hTel22 CD spectra obtained at the titration end point. These spectra demonstrate that all glycols cause conversion to a parallel conformation. B) CD titration curves for each glycol studied. Arrow indicates the direction of increasing MW.
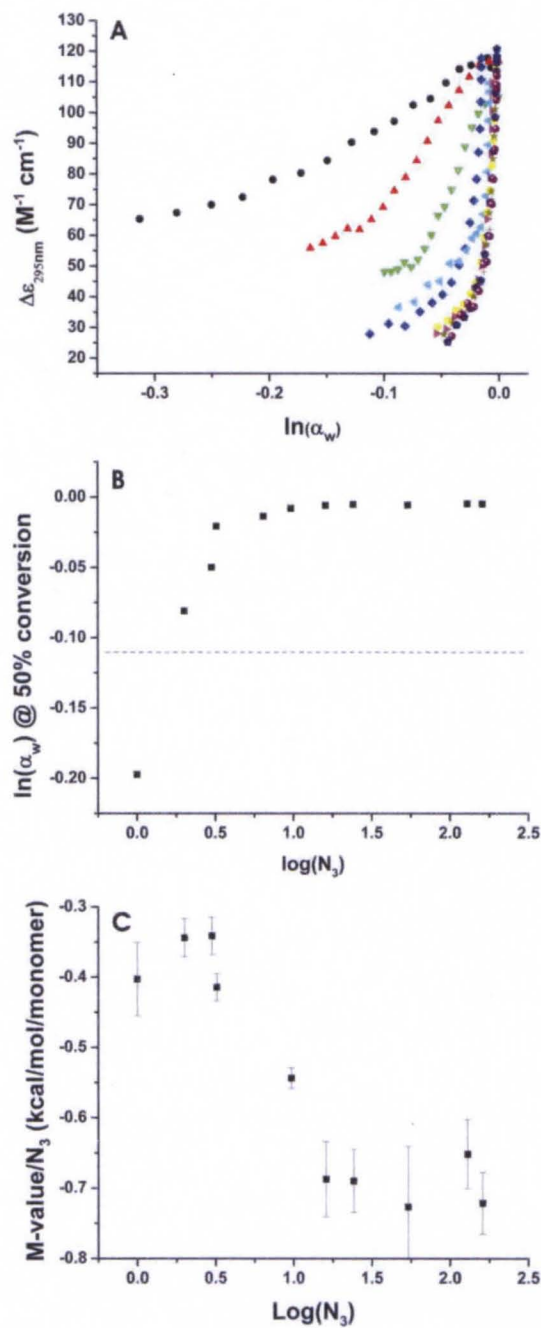
Figure 19. A) Transformed titration curves obtained from the primary data shown in Figure 18B.

B) Change in the transition midpoint as a function of the MW of PEG. C) M-value plot (see text)

for the conversion of the hTel22 demonstrates an increase in magnitude with a distinct shift at

PEG 600.
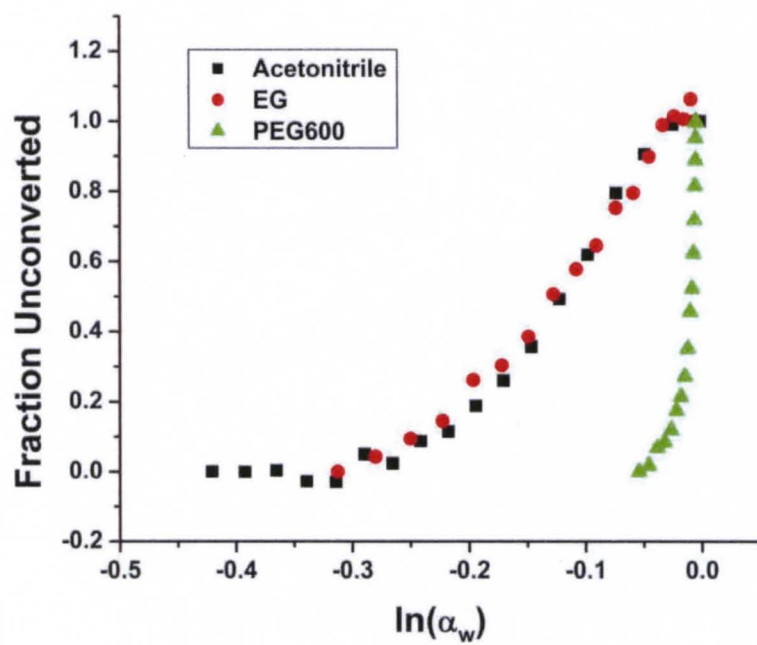
Figure 20. Comparison of the conversion of the hTel22 through addition of acetonitrile (black squares), ethylene glycol (red circles), and PEG 600 (green triangles).

for acetonitrile. EG and acetonitrile have different chemical properties yet yield similar quantitative estimates for the hydration change that accompanies the hybrid to propeller transition. This is expected if both compounds act primarily by osmotic stress and are appropriately inert. [176,179]

In contrast, the data for PEG 400 and greater (Figure 19A) yield estimates of Δn_w of about -1200, an unrealistically larger number of waters released. If a water molecule covers about 7-10 $Å^{2}$,[203] that value would indicate the loss of $\approx$12,000 $Å^{2}$ of solvent accessible surface upon conversion of the hybrid quadruplex to the propeller form. Because folded quadruplexes have solvent accessible surface areas of 3500-4000 $Å^{2}$,[79] this number is absurd.

Evaluating Preferential Interaction and Excluded Volume Effects. M-values were determined based on changes in the Gibbs free energy due to a shift in the quadruplex conformation with respect to increasing MW PEGs. M-value plots quantify solute effects on a process through the derivative of the standard Gibbs free energy change (ΔG_{obs}) with respect to solute (m_3) concentrations (Equation 11).[194]

$$\text{M-value} \equiv \delta\Delta G_{obs}/\delta m_3 \text{ (Equation 11)}$$

Titration results for each MW PEG were transformed to plots of Gibbs free energy change versus monomolar PEG concentration. Monomolar concentration refers to the molar concentration of equivalent monomer subunits in solution and normalizes the concentration with respect to the number of repeating ethylene glycol units (N_3) in order to diminish MW differences. The fraction of converted quadruplex was calculated using Equation 8, where the i-th component now refers to the i-th PEG concentration. Because of the variable end points of the low MW glycol titrations, the value for the fully converted CD spectra ($\Delta\varepsilon_f$) were fixed using the high MW PEG averages discussed above. The equilibrium constant was estimated using Equation 7 and

$\Delta G = -RTln(K)$ calculated at 25 °C. Linear regression of the Gibbs free energy change against monomolar PEG concentration results in a monomer M-value for each glycol studied (Figure 21). The resulting monomer M-values were plotted against $log(N_3)$ to evaluate distinct energy changes based on the MW of the PEG solution (Figure 19C). The conversion of the hTel22 was found to have monomer M-values that increase in magnitude with the MW of the glycols, nearly doubling for PEG 600 – PEG 10000. This behavior stands in sharp contrast to what was observed by the Record laboratory[194] for crowding effects on duplex stability, in which M-values changed from negative to positive over the same range of glycol MWs. Our data indicates that excluded volume effects are minimal for the hTel22 conformational transition driven by PEGs.

The propeller conformation is not the most compact hTel22 structure. hTel22 was determined to be an all-parallel "propeller" conformation in PEG solutions[80] that is similar to the conformation originally solved by X-ray crystallography.[57] The proposed mechanism for the conformational transition was molecular crowding, which requires that a reaction be driven to more compact states due to a decrease in available volume.[171] For molecular crowding to exert an effect on the hTel22 equilibrium, the propeller form must be more compact then the hybrid conformation. HYDROPRO software was used to calculate the hydrodynamic parameters of several hTel22 conformations based on their published structures. Bead models were constructed and the parameters describing the size and shape of the structures were calculated (Table 2). Evaluation of the hydrodynamic parameters demonstrates that the propeller conformation is not the most compact of the available solution conformations, having a larger Stokes' radius and radius of gyration than other known hTel22 conformations. According to the fundamental physical principles that govern molecular crowding, it cannot be the underlying mechanism that describes the effect of PEGs on the conversion of hTel22.
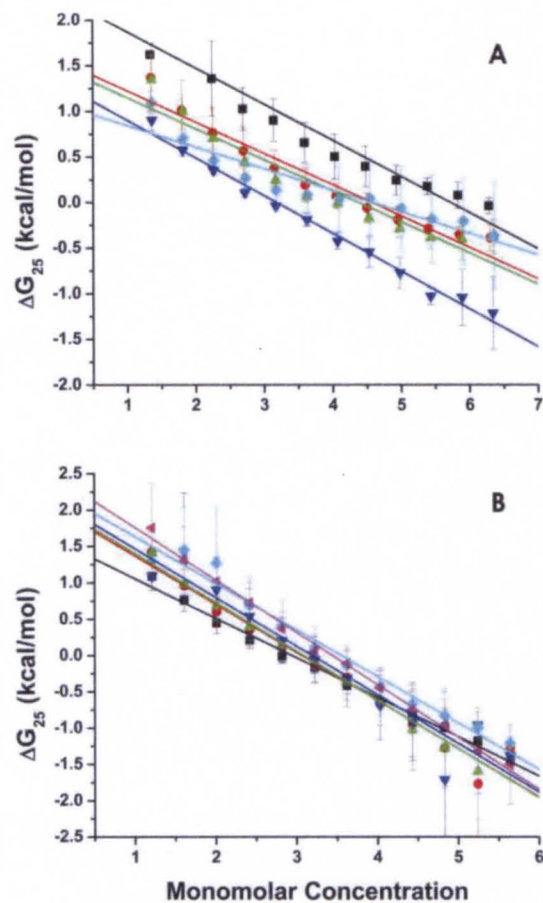
Figure 21. Determination of monomer M-values from linear regression of Gibbs free energy changes (ΔG_{25}) vs. monomolar PEG concentration. A) Ethylene glycol (black), diethylene glycol (red), triethylene glycol (green), PEG 200 (blue), and PEG 400 (cyan). B) PEG 600 (black), PEG 1000 (red), PEG 1500 (green), PEG 3350 (blue), PEG 8000 (cyan), and PEG 10000 (magenta).

Table 2. Hydrodynamic properties of hTel22 conformers calculated using HYDROPRO software.

	Stokes' Radius (10^{-7} cm)	Radius of Gyration (10^{-7} cm)	$S_{20,w}$ (Svedbergs)
propeller	1.602	1.322	1.744
hybrid 1	1.430	1.089	1.953
hybrid 2	1.515	1.172	1.843
basket	1.479	1.105	1.888

Aggregation of the hTel22 does not occur upon the addition of PEG. It is possible that a conformational transition of hTel22 coupled to aggregation upon the addition of PEG could explain its effect. For that to occur, the volume of the aggregate must be smaller than the sum of the monomer volumes. If such aggregation occurs during the conformational change, the PEG titration curves should be dependent on DNA concentration. The results of PEG titrations conducted at 10-fold higher hTel22 concentrations show identical behavior compared to low DNA concentrations (Figure 22), indicating the conversion of the quadruplex is independent of concentration in the range studied. Sedimentation velocity experiments were conducted in solutions of 0, 10, 20 and 30% (v/v) PEG 600 and show no aggregation upon conversion of hTel22 (Figure 23). It was determined from AUC experiments that the addition of PEG alters the hydrodynamic properties of hTel22, but that no higher-order species were formed (Table 3). $S_{20,w}$ values were found to decrease from 1.96 in the absence of PEG 600 to 1.75 in 30% (v/v) PEG 600. The $S_{20,w}$ values agree well with the HYDROPRO calculations for the hybrid and propeller conformations, respectively. The measured frictional ratios show that with addition of PEG 600 the hTel22 conformation shifts to a more elongated structure (increasing from 1.04 to 1.18) assuming 0.3 g/g for the amount of water bound.[200] The higher frictional ratio indicates a larger hydrodynamic volume in PEG, for which a molecular crowding mechanism is incompatible.

Differential binding of PEG to the propeller form drives the conformational change of hTel22. If molecular crowding or coupled aggregation cannot explain how PEGs alter the conformation of hTel22, what mechanism can? Evaluation of water activities clearly shows hydration is not a major determinant of the conformational change using high MW PEG solutions, and aggregation was not observed from AUC and DNA concentration dependent titrations. M-value investigations indicated a mechanism where preferential interactions

90
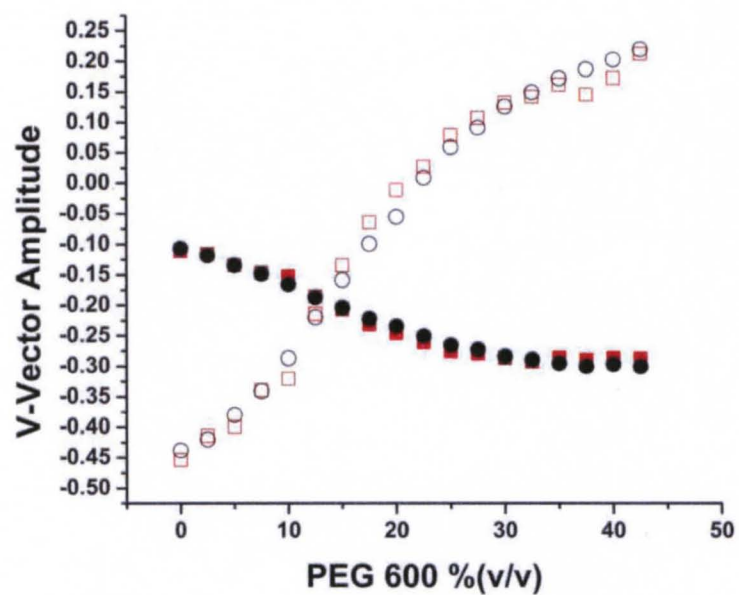
Figure 22. PEG 600 titrations conducted at 10X strand concentration monitoring a shift in CD during the conversion of the hTel22. Titrations were conducted at 4 µM (circles) and 40 µM (squares). Strand concentration does not influence the conversion of the hTel22 as monitored by two distinct SVD V-vectors corresponding to 290 nm (closed) and 265 nm (open).
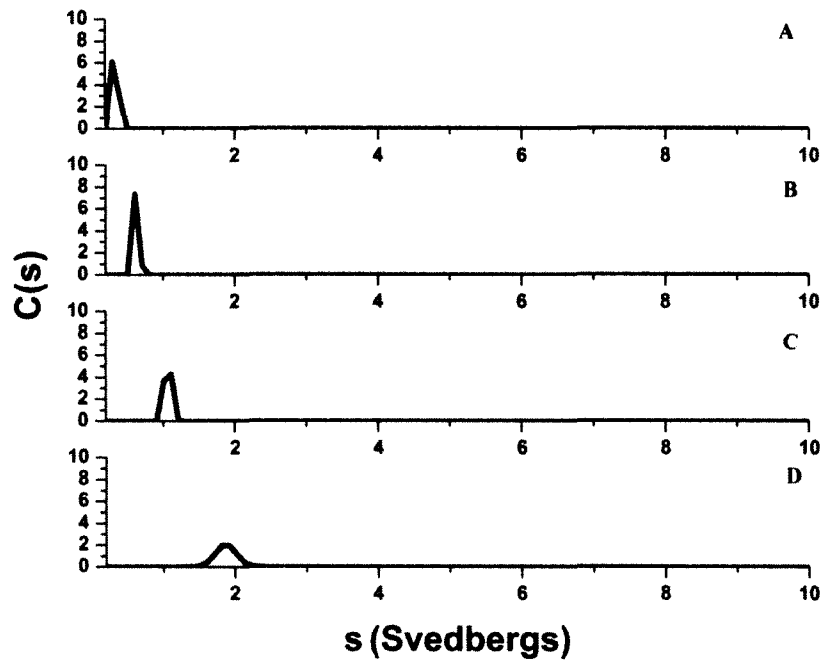
Figure 23. Effect of PEG concentration on hTel22 sedimentation coefficient. Results are plotted as the relative concentration (C(s)) versus uncorrected sedimentation coefficient for PEG concentrations of 0 (A), 10% (B) 20% (C) and 30% (D) in 0.4 M KCl-folding buffer. Data analysis was by Sedfit. Correcting for the measured density and viscosity of the PEG solutions yields the $S_{20,w}$ and f/f_o values shown in Table 3.

Table 3. Hydrodynamic parameters determined from sedimentation velocity experiments. *SE is the standard error estimated in the DCDT+ software report. The true molecular weight assuming 2 K^+ bound of 7045 was used for the calculation of frictional ratios assuming 0.3 g/g water bound.

PEG 600 %(v/v)	$S_{20,w}$ (Svedbergs)	Frictional Ratio	Molecular Weight ± SE*
0	1.96	1.040	7400 ± 40
10	1.87	1.106	6950 ± 40
20	1.87	1.101	6800 ± 100
30	1.75	1.118	7100 ± N/A

between PEG and the quadruplex seemed to be a strong driving force behind the conformational change. Differential binding of PEG to the propeller form (Q_P) compared to the hybrid form (Q_{Hyb}), can account for the action of PEGs. A plausible mechanism is that PEG binding is coupled to the hybrid to propeller transition:

$$Q_{Hyb} \longleftrightarrow Q_P + nPEG \longleftrightarrow Q_P \cdot PEG_n \text{ (Mechanism 2)}$$
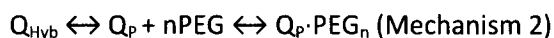
This mechanism requires an equilibrium between the propeller and hybrid quadruplex conformations in the absence of PEG ($K_{eq,0}$). Differential binding of PEG to the propeller conformation shifts the equilibrium to the preferred form, describing a conformational selection mechanism.[171] This model is a restricted form of the classic Monod-Wyman-Changeux model[204] and assumes that that the interaction of PEG with hybrid conformations is negligible and that microstates of hybrid bound by PEG are unpopulated (See Appendix B). Equation 12 is derived from the model and can be used to fit PEG titration data as shown in Figure 24.

$$\Delta\varepsilon_\lambda = \Delta\varepsilon_0 + \frac{K_{eq,0}(\Delta\varepsilon_f - \Delta\varepsilon_0)\left(\frac{[PEG]}{K_B} + 1\right)^n}{K_{eq,0}\left(\frac{[PEG]}{K_B} + 1\right)^n + 1} \qquad \text{(Equation 12)}$$

Data are fit to an equilibrium constant between the aqueous and PEG conformations in aqueous solution ($K_{eq,0}$), the moles of PEG bound (n), and a dissociation constant for the binding of PEG to the propeller conformation (K_B). A global analysis of titration results from 265 and 290 nm was conducted to improve analysis but required parameter constraints. The value of $K_{eq,0}$ was fixed at 0.17 based on the estimates obtained from osmotic stress experiments (Figure 16C). Error analysis using the method of Saroff [205] showed that the estimates of parameters K_B and n were highly correlated (Figure 25), and that only the lower limit of the value of n ≥ 2 can
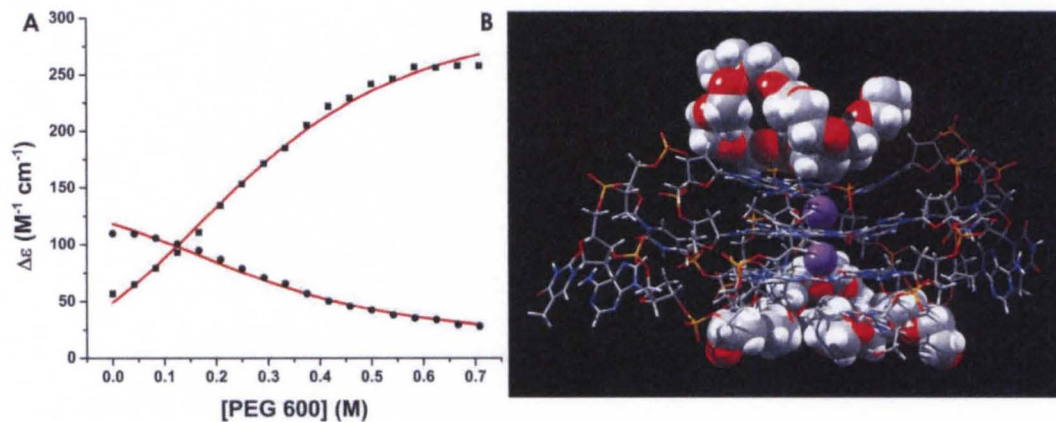
Figure 24. Conformational selection by differential binding of PEG to the propeller quadruplex.

A) Representative fit of titration data obtained for PEG 600 to Equation 12 (squares, 265 nm;

circles, 290 nm). Optimized parameters were $K_{eq,0}$ = 0.17, n = 4, and K_B = 0.43 ± 0.02 M for the

binding affinity of PEG 600. B) Results of molecular dynamics showing PEG 600 interacts

favorably with the planar G-tetrad faces of the propeller structure.
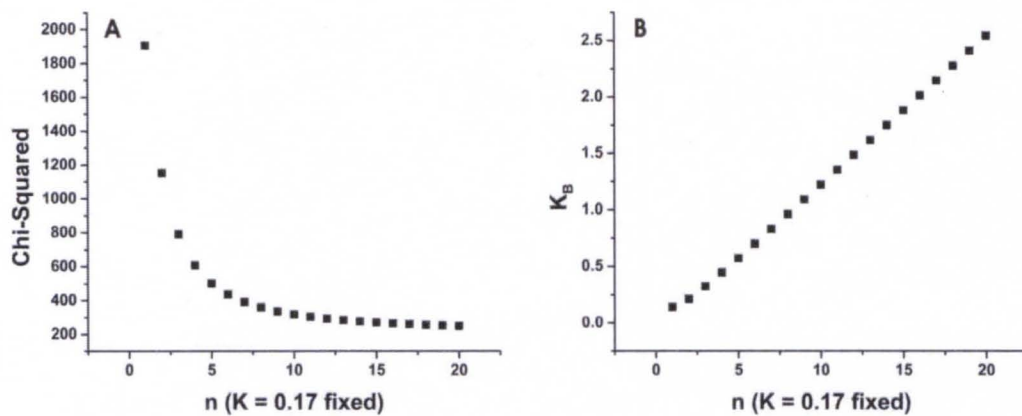
Figure 25. Analysis of parameter uncertainty and correlation for fits to Equation 12. A) Error space for estimates of the moles of PEG bound (n). Chi-squared values for the best fit at fixed n values are plotted according to the procedure of Saroff.[205] The plot defines the error space for the parameter n and shows that only a minimum value can be determined. B) Correlation of dissociation constant (K_B) and stoichiometry (n) estimates.

be defined. The number of bound molecules must be 2 or greater to properly match the sigmoidal nature of the titration. The Saroff analysis (Figure 25) and molecular modeling (see below) indicates a reasonable upper limit of 4. Model analysis results in an estimated dissociation constant for PEG 600 to be $K_B = 0.43 \pm 0.02$ M, with a fixed value of 4 PEG molecules bound per quadruplex (Figure 24A), as described below.

The results of the analysis were used to initiate a molecular dynamics study on the ability of PEG to interact with the propeller quadruplex conformation 1KF1 and the hybrid 1 form 2HY9-t. PEG 200 and 600 were initially placed in the side grooves of the propeller structure and in the grooves and quartet backbone regions of 2HY9-t. Propeller quadruplex form simulations show that after initial minimization both PEG 200 and 600 are unstable in the groove regions, but that they relocate to interact favorably with the planar region of the G-tetrad faces (Figure 24B) with long lived (greater than 300 ns) constant contact. In the hybrid 1 quadruplex form simulations, in contrast, both PEG 200 and 600 relocated to the bulk solvent with only transient interactions with the quadruplex over the 300 ns production trajectory. These simulations show differential binding interactions for PEG with the propeller and hybrid forms, consistent with the assumed mechanism described above. It should be noted that as the preferred binding site of the PEG molecules was the tetrad face, four PEG 200 and two PEG 600 could be accommodated in the preferred site of the propeller quadruplex. This is consistent with the lower limits of the binding stoichiometry estimated above, and provides a rationale for the value of n given in Figure 24A.

Discussion.

These results show that the conversion of the human telomere hybrid quadruplex to a propeller form by PEG does not result from "crowding" by excluded volume or by hydration changes resulting from osmotic stress, but instead most likely results from differential PEG binding coupled to the conformational transition between the two forms. Although the binding affinity of PEG for the propeller conformation is modest, it is sufficient to drive conformational selection at high PEG concentrations. These results raise doubts about claims based on studies using PEG that the propeller conformation is the "physiologically relevant" form because it is favored under crowding conditions that mimic the intracellular milieu. PEG is not "crowding" in this case, but instead is acting by a very different mechanism.

Dehydration favors the propeller quadruplex and increases its thermal stability.
Acetonitrile acts as an osmotic stress agent to decrease water activity without any appreciable excluded volume effects. Addition of acetonitrile drives the telomeric quadruplex from the anti-parallel hybrid form to the all-parallel propeller form. Since the propeller form is favored by decreased water activity, it must be less hydrated than the hybrid form. The data in Figure 16 can be used to quantify the effects of changes in water activity on the hybrid to propeller conformational transition. The transition midpoint is at $\alpha_w = 0.886$. The quadruplex conformational transition is linked to hydration changes, and 17-18 water molecules are released upon conversion of the hybrid form to the propeller form.

The data of Figure 1 can be used to estimate the equilibrium constant for the hybrid to propeller conformational change in aqueous solution, although the value cannot be determined with great precision. The best-fit line in Figure 1C yields a value of $\ln(K_{eq}) = -1.77 - 17.0\ln(\alpha_w)$ resulting in $K_{eq,0} = 0.17$ for the equilibrium constant at $\alpha_w = 1.0$ using acetonitrile results. This

value indicates that a finite amount of the propeller form exists in aqueous solution, although the hybrid form is thermodynamically favored. The free energy cost for the conversion of the hybrid to the propeller form is modest, with ΔG = -RTlnK $\approx$ +1.0 kcal mol^{-1} at 25°C. NMR studies of the two-repeat telomere sequence d(TAGGGTTAGGGT) showed that the bimolecular parallel and anti-parallel quadruplexes could coexist in K$^+$ solutions, with the anti-parallel form strongly favored at 20 °C.[61] A pulsed EPR study concluded that the four-repeat, unimolecular human telomere quadruplex existed in solution as a nearly 1:1 mixture of hybrid and propeller form.[206] Similarly, ^{125}I radio probing experiments[157,207] suggested that unimolecular basket, chair, hybrid and propeller conformations coexisted in K$^+$ solution. We caution that both of these studies were done at extremely low temperatures (40 K for EPR and -80 °C for the radioprobing) that would surely perturb both folding and conformational equilibria. Recent thermodynamic studies report a heat capacity change of about -300 cal mol^{-1} deg^{-1} for the folding of telomeric quadruplex, a value that indicates cold denaturation would complicated any studies done at extremely low temperatures.[208] NMR evidence[72,79] shows that the hTel22 sequence consists of two major non-parallel species in aqueous solution.

Conversion of the hybrid to the propeller conformation is essentially complete at 25% (v/v) acetonitrile concentration. The thermal denaturation experiments in Figure 17 show that increasing acetonitrile concentration beyond that point increases the thermal stability of the propeller quadruplex. The T_m increases from 57.4 to 76.2°C upon changing the acetonitrile concentration from 25 to 50% (v/v). This change in concentration corresponds to a decrease in water activity from 0.778 to 0.612. Qualitatively, the increase in T_m with decreased water activity indicates that the folded quadruplex is less hydrated than the unfolded single-strand. Water is taken up by the single strand upon denaturation, or, conversely, would be released upon folding of the single-strand to form a quadruplex. Figure 17B shows a quantitative analysis

of the effect, the result of which indicates that ≈20 water molecules are taken up by the single-strand upon denaturation. Our results are generally consistent with published reports[79,180,182-184] that decreased water activity increases the thermal stability of quadruplexes. We note also that the increase in hydration on unfolding is in line with expectation as the solvent accessible surface area increases on unfolding, unlike the anomalous hydration of duplex DNA.

Polyethyleneglycols do not alter quadruplex conformation by osmotic stress. Figures 18 and 19 show that the addition of polyethylene glycols effectively drives a transition from the hybrid to the propeller conformation. Inspection of these data shows that dehydration by osmotic stress cannot account for their effects. The acetonitrile data presented above show that a substantial decrease in water activity is needed to dehydrate the hybrid form to drive formation of the propeller form. The water activity at the transition midpoint is 0.886 and transition is complete at 0.778 water activity. Figure 19 shows that addition of PEGs 400 – 10000 reduces water activity at most to a value of 0.95, a change that is insufficient to completely drive the conformational transition by osmotic stress. This point is emphasized in Figure 20 that compares the effect of acetonitrile to PEG 600. The effect of PEG 600 on the transition is complete before the effect of acetonitrile has barely begun. The two agents must act by different underlying mechanisms.

In contrast, the small glycols (ethylene glycol, diethylene glycol and triethylene glycol) behave similarly to acetonitrile and probably act primarily by osmotic stress. Analysis of ethylene glycol data by Equations 7-9 yields an estimate of Δn_w of 17 water molecules released upon formation of the propeller form, values in good agreement with acetonitrile data. The result is consistent with ethylene glycol acting primarily through an osmotic stress mechanism, in contrast to the higher MW polyethylene glycols.

The conformational change induced by PEGs does not occur through a molecular crowding mechanism. Record and coworkers recently proposed an analytical method for separating preferential interaction and excluded volume effects on folding equilibria. [194] The approach requires data obtained for the concentration-dependent effects on the reaction of interest using a complete series of solutes differing in MW, from monomer to polymer. In our case, we used ethylene glycol through polymeric PEGs (Figures 18-19). The effect of solutes is quantified through "M-values", the derivative of the observed free energy changes with respect to solute concentration. For the case of duplex or hairpin DNA thermal denaturation, Record and coworkers observed a dramatic effect of solute MW on M-values, with a change in sign (from negative to positive) from the monomer ethylene glycol to the higher MW PEGs. The dramatic difference was attributed to the dominance of excluded volume effects ("crowding") for the high MW PEGs on duplex denaturation. We observe contrasting behavior for the quadruplex conformational change studied here. Figure 19C shows that the monomer M-values for hTel22 conversion become more negative with increased solute MW, behavior that is opposite than expected for a crowding mechanism.

Other observations are inconsistent with a crowding mechanism. The concept of molecular crowding requires the reaction be driven in the direction that reduces the total effective excluded volume of the system. [171] The size of the propeller conformation was calculated from HYDROPRO to be less compact than other human telomeric quadruplex conformations (Table 2). AUC experiments (Table 3) validate the HYDROPRO calculations and show that the propeller conformation has a smaller $S_{20,w}$ value than the hybrid form, with an increased frictional ratio. These observed changes in the hydrodynamic properties indicate an increased hydrodynamic volume in PEG, the exact opposite of what is required for a crowding effect.

PEG could act by a crowding effect if the hybrid to propeller conformational transition were coupled to aggregation of the propeller form. Crowding facilitates aggregation if the volume of the aggregate is less than the sum of the volumes of monomer units. However, aggregation was not detected from direct measurements of size through AUC measurements, eliminating this explanation as a possibility. Further, PEG titrations were also shown to be independent of quadruplex concentration over a 10-fold range (Figure 22). Higher quadruplex concentrations could not be evaluated due to potential self-aggregation of the human telomeric quadruplex.[209]

Differential binding of PEG can explain the hybrid to propeller conformational change. If high MW PEGs do not alter quadruplex conformation by hydration changes or by crowding, what is their mechanism of action? A mechanism in which PEG binding is coupled to the conformational change is plausible. The conformational selection mechanism is well known in protein binding,[210] less so in nucleic acid systems[211]. The PEG titration data (Figure 24) show a cooperative transition, indicating the binding mechanism requires more than one molecule of ligand per molecule of DNA.

Figure 24 shows that PEG titration data can be fit to a model in which PEG binding is coupled to the hybrid to propeller conformational transition with preferential binding to the latter form. The differential binding of PEG to the propeller conformation is a plausible driving force for the conversion of the human telomeric quadruplex. The energetic difference between the hybrid forms and propeller conformation was determined by isothermal titrations to be a moderate +1.0 kcal mol^{-1}. This energetic barrier between the hybrid and propeller conformations could be driven by a conformational selection model, where the preferential binding of PEG to the propeller conformation over the hybrid conformation shifts the

quadruplex solution population. The estimated binding energy of -0.5 kcal mol^{-1} for each PEG 600 bound to the quadruplex illustrates that binding at least two molecules can overcome the energetic differences between the two conformations (Diagram 2).

Although it is difficult to directly demonstrate the binding of PEG to the quadruplex, molecular dynamics simulations show a long lived preferred interaction of the PEG units with the G-tetrad face of the propeller structure. This interaction of the PEG units with hydrophobic regions of the nucleic acid structure is a similar result to previously published protein interactions with PEG. PEG is known to contain potential interactions with hydrophobic amino acid residues. The potential for this interaction with the exposed hydrophobic G-tetrads rationalizes a conformational selection model, where PEG units cause a shift in a conformation through an exposed binding site on the propeller quadruplex.
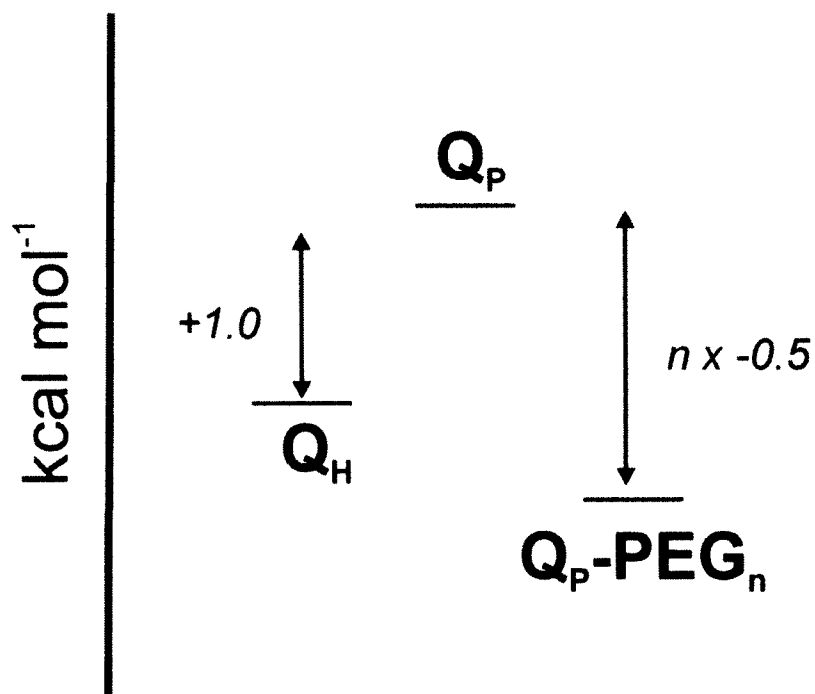
Diagram 2. Free energy diagram for the conformational selection mechanism. The differential binding of PEG to the propeller quadruplex drives the conversion from the hybrid form (Q_H) to the propeller form (Q_P). The hybrid and propeller quadruplex conformations have an unfavorable Gibbs free energy difference of approximately +1.0 kcal mol^{-1} in aqueous solution. Selective binding of PEG to the propeller form shifts the conformational equilibrium to the preferred form. The total free energy of binding (-0.5 kcal mol^{-1} per PEG) must overcome the energy barrier for the hybrid to propeller conversion.

Summary

These studies clarify the mechanism by which PEG acts to alter the conformation of the human telomeric quadruplex. Our data show that PEGs cannot act as a crowding or dehydrating agent, but rather act by selectively binding to the propeller conformation and perturbing the equilibrium between the hybrid and propeller forms. Such a conformational selection mechanism is common for protein interactions. While the effects of macromolecular crowding remain of great importance within the cell, inferences made about physiologically relevant quadruplex conformations based on an assumed crowding mechanism for PEG are unwarranted.

CHAPTER 5

POPULATED INTERMEDIATES IN THE THERMAL UNFOLDING OF THE HUMAN TELOMERIC QUADRUPLEX.

Introduction.

Telomeres are supramolecular structures consisting of repetitive DNA sequences situated at the ends of chromosomes that function in maintaining chromosomal integrity and preventing loss of genetic information during cell division.[212] The human telomere consists of consecutive repeats of the DNA sequence d(GGGTTA)n, 100–200 nucleotides of which are present in a single-stranded state at the 3' end of the chromosomal DNA.[148,149,213] Normal cells are unable to fully replicate telomeric DNA during mitosis because of low levels of the enzyme telomerase, which is essential for telomere replication. Thus, as the telomeric sequence erodes during mitosis, the number n of hexanucleotide repeats decreases with the cellular age. This erosion triggers apoptosis, and the cell self-destructs. Cancer cells frequently have elevated telomerase levels and hence overcome telomere-dependent cell death.[214]

Single-stranded telomeric DNA sequences generally do not form classical double helical structures but instead undergo cation-dependent folding into a variety topologies that commonly consist of stacks of planar G-quartets connected by loops. The G-quartets themselves are stabilized by Hoogsteen-type H-bonds, while the quadruplexes form stacked G-quartets stabilized by monovalent cation coordination to the guanine O6 atoms. The loop sequences connecting stacked quartets determine the overall topography (tertiary structure or

"fold") of the quadruplex. The topological complexity of G-quadruplex folding is more reminiscent of RNA and protein folding than of duplex formation.

NMR and X-ray crystallography have defined at least five topologies for human telomeric quadruplexes.[215] These topologies are characterized by the relative direction of the strand segments (parallel or anti-parallel), the geometry of the connecting loops (diagonal, lateral, or chain reversal), and the size of the grooves that define the external surface (wide, narrow or medium). Since only relatively small energetic differences exist between different topologies, the predominant arrangement depends on the number of stacked quartets, the loop sequences, the identity of the coordinating cation, and the presence of 5' or 3' flanking residues. Monomolecular topologies that have been identified include: basket (anti-parallel), chair (anti-parallel), mixed (three anti-parallel and one parallel), and all-parallel.

It is important to understand the thermodynamic factors that influence the relative stability of different conformers because this information is essential to understand the structure of the telomere and how switching between conformers could allow for regulation of telomere replication and to develop reagents that bind specifically to different conformers for use as drugs or structural probes.

Measurement of UV absorbance at a single wavelength is commonly used to monitor G-quadruplex thermal unfolding followed by analysis using a simple two-state model to extract thermodynamic parameters.[216-218] Representative thermodynamic parameters obtained by such methods have been tabulated and discussed in recent reviews.[81,108] However, use of the two-state model may mask complexity in the denaturation process and therefore must be justified by experimental tests. Indeed, a study using calorimetry and circular dichroism (CD) to monitor human telomeric quadruplex denaturation showed that at least one stable intermediate state

exists along the unfolding pathway.[219] A more recent study confirmed the existence of the intermediate state and provided a more sophisticated thermodynamic analysis of the process.[208]

Many unanswered questions remain about the mechanism of quadruplex denaturation, most importantly the number and nature of intermediate states along the unfolding pathway. Sugiyama and colleagues[220] used molecular dynamics simulations to explore possible folding pathways and proposed the existence of stable triple-helical intermediate states. Such states have yet to be identified and validated experimentally.

We undertook the current study as part of an overall goal to define the quadruplex energetic and kinetic landscapes using different spectroscopic probes to monitor the unfolding process in a site-specific manner to detect the presence of intermediates and to characterize their spectroscopic signatures to gain clues as to their structures. Here we use multi-wavelength CD, fluorescence emission, and FRET spectroscopies to map the thermal unfolding pathway of a human telomeric model oligonucleotide, the 22-mer d(AGGG(TTAGGG)$_3$) (Tel22) in KCl solutions. All of the spectroscopic techniques revealed at least one intermediate structure which predominated at 55–60 °C for Tel22 as well as a variety of similar quadruplex-forming sequences. The CD and fluorescence properties suggest that this structure has properties expected of a triplex structure that has been suggested as an intermediate in the cation-driven folding of quadruplexes.

Materials and Methods.

Oligonucleotides. The sequences and code names of the oligonucleotides used in this study are given in Table 4. All oligonucleotides except for the FRET derivative 6-FAM-Tel22-Tamra were obtained in desalted, lyophilized form from Integrated DNA Technologies, Inc., Coralville, IA. They were dissolved in 10 mm tetrabutylammonium phosphate, 1 mM EDTA, pH 7.0 (folding buffer) at a concentration of 0.5–1 mM, and stored at 4 °C. The FRET oligonucleotide 6-FAM-Tel22-Tamra was obtained in HPLC purified form from Sigma-Aldrich, St. Louis, MO. It was dissolved in folding buffer at a concentration of 250 µM and stored in an opaque container at 4 °C. Oligonucleotide concentrations were estimated from their absorbance at 260 nm using extinction coefficients supplied by the manufacturer. Other reagents were from Sigma-Aldrich.

Thermal Denaturation Experiments. Thermally induced quadruplex unfolding was followed either by CD or by fluorescence emission essentially as described previously.[202] Duplicate melts were carried out on the same sample on successive days; all results reported here are for the second melt. A third melt on the same sample generally gave results that were identical to the second melt. For CD experiments, samples were prepared by dilution of a concentrated stock solution of oligonucleotide at a previously determined concentration into folding buffer. Reference CD and absorbance spectra were recorded in a stoppered 1 cm path length cuvette at 20 °C with a Jasco J-810 spectropolarimeter equipped with a programmable Peltier thermostat and a magnetic stirrer. The desired salt concentration was then established (generally 25 mM KCl or 100 mM NaCl) by adding the requisite volume of 3 M salt solution. The instrumental parameters for the CD melting experiments were: 1.0 nm bandwidth, 2 s integration time, 200 nm/min scan rate, 4 °C/min ramp, ±0.05 °C equilibration tolerance, 60 s

Table 4. Sequences and Code Names of the Olignocleotides Used in This Study. *PDB

designation for the structure.[65,221]

Oligonucleotide	Sequence 5′ to 3′
Tel22	A GGGTTA GGGTTA GGGTTA GGGTTA GGG
Tel22AP1	**2AP** GGGTTA GGGTTA GGGTTA GGGTTA GGG
Tel22AP7	A GGGTT**2AP** GGGTTA GGGTTA GGGTTA GGG
Tel22AP13	A GGGTTA GGGTTA GGGTT**2AP** GGGTTA GGG
Tel22AP19	A GGGTTA GGGTTA GGGTTA GGGTT**2AP** GGG
2GKU*	TT GGGTTA GGGTTA GGGTTA GGGTTA GGG A
2AQY* + 6	GGGTTA GGGTTA GGGT + AGGGT
Tel-3G	TTA GGGTTA GGGTTA GGGTTA
6FAM-Tel22-Tamra	**6-FAM**-A GGGTTA GGGTTA GGGTTA GGGTTA GGG-**Tamra**

delay after equilibration. Three spectra were averaged at each temperature. CD data were normalized to strand concentration using the equation $\varepsilon L - \varepsilon R = \Delta\varepsilon = \theta/(32\ 980 \cdot c \cdot l)$, where θ is the observed ellipticity in millidegrees, c is the DNA strand concentration in mol L^{-1}, and l is the path length in cm.

Fluorescence melting profiles were obtained using 1 cm path length stopped fluorescence cuvettes in a Jasco FP-6500 fluorometer equipped with a programmable Peltier thermostat and a magnetic stirrer. Sample preparation was the same as described for the CD melts. Instrumental operating parameters were: 5 nm excitation and 10 nm emission bandwidths for 2AP derivatives and 5 nm/5 nm bandwidths for 6-FAM FRET experiments, 1 °C/min ramp, ±0.2 °C equilibration tolerance, 60 s delay after equilibration. Two spectra were averaged at each temperature. Excitation was at 305 or 490 nm for 2AP or 6-FAM FRET samples, respectively. Fluorescence and CD spectra were corrected by subtracting a blank (buffer only) spectrum.

Data Analysis by SVD. The CD or emission spectra obtained as a function of temperature T were assembled into a data matrix $D_{i,j}$ consisting of i rows of spectroscopic signal (e.g., CD or fluorescence intensity) at wavelengths λ_i over in 1 or 2 °C increments from 4 to 94 °C unless otherwise noted. In $D_{i,j}$, row i represents a melting profile at λ_i, and column j represents the optical spectrum at T_j. $D_{i,j}$ were analyzed by singular value decomposition (SVD) as previously described[202] to produce the S, U, and V matrices. S contains the singular values (the weights of the components comprising the data set), U contains the basis spectra (the spectral shapes) that make up the data set, and V contains the amplitudes of each component as a function of the experimental variable (temperature). SVD was carried out using the matrix algebra routines in the software suite Mathmatica 8.0 (Wolfram Research, Champaign, Il). The number of significant spectral components contained in $D_{i,j}$ was determined by evaluating the

relative magnitude of the singular values, their contribution to the total variance of the signal, and the values of the autocorrelation coefficients of the amplitude (V) vectors.

Mechanistic Analysis of Thermal Unfolding Data. Since the SVD analysis generally showed evidence for at least three spectroscopically distinct species over the experimental temperature range, we considered five pathways shown in Scheme 1 that might reasonably describe the temperature-dependent stability of quadruplex DNA. To analyze the data in terms of these mechanisms, the significant V vectors (defined as those vectors with autocorrelation coefficients ≥ 0.75) were globally fit by iterative nonlinear least-squares to the analytical expressions for thermal unfolding corresponding to each mechanism. These equations, which are based on van't Hoff analysis of multiple thermal transitions, were either taken from the literature[222,223] or derived specifically for this study; they are summarized in Appendix C : Analytical Models.

In brief, the spectral signal associated with each pure species in the mechanism (e.g F, I_n, and U) as well as the respective van't Hoff enthalpies ΔH and transition midpoint temperatures (T_m) for each were optimized by iterative nonlinear least-squares analysis using routines programmed in Mathmatica 8.0. The species designated F, I, and U most likely consist of ensembles of energetically equivalent conformers rather than a specific conformation. Fits to the individual mechanisms were compared by their respective $\chi 2$ values and residual plots. Uncertainties of the fitted parameters were derived by carrying out 1000 Monte Carlo simulations as described by Saroff.[205] Models with different numbers of adjustable parameters were compared using the Akaike information criterion (AIC) statistical analysis to account for the increased degrees of freedom associated with an increase in adjustable parameters.[224]

Mechanism 1: $F \leftrightarrow U$

Mechanism 2: $F \leftrightarrow I \leftrightarrow U$

Mechanism 3: $F_1 \leftrightarrow U \leftrightarrow F_2$

Mechanism 4: $F \leftrightarrow I_1 \leftrightarrow I_2 \leftrightarrow U$

Mechanism 5: $F_1 \leftrightarrow I \leftrightarrow U$

$\updownarrow$

F_2
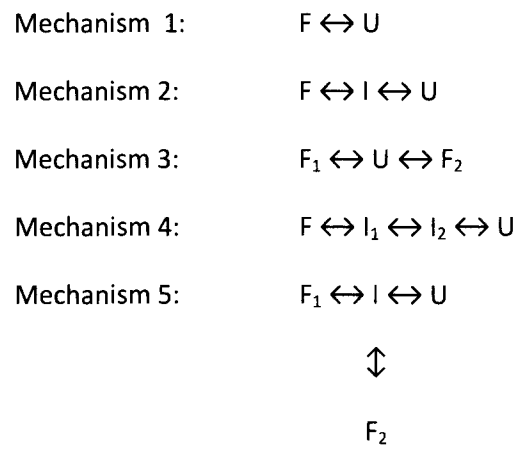
F = folded state(s)

I = intermediate state(s)

U = unfolded ensemble

Scheme 1. Mechanisms describing thermal unfolding of G-quadruplexes.

Results.

Melting of Tel22 Assessed by CD. We initially utilized temperature-dependent CD to investigate the unfolding pathway of Tel22 in 25 mM KCl. CD spectroscopy is sensitive to topological features of nucleic acid structure, particularly the orientation of guanine bases within the G-quartet stack. A family of spectra associated with a representative melt is shown in Figure 26A. The spectra of the folded state F at 4 °C have a maximum at ~289 nm and a minimum at ~235 nm with a prominent shoulder at ~270 nm. This spectrum is characteristic of a mixture of quadruplex conformers suggested by biophysical studies of this sequence in K^+-containing solutions.[61,63,72] As the temperature was increased, a new species appeared which has a maximum in the CD spectrum near 266 nm and a minimum near 242 nm; this signal is maximal at ~65 °C. The 266 nm species subsequently decayed to the unfolded state U as the temperature was increased to 94 °C. The unfolded ensemble has a relatively weak CD spectrum characteristic of a disordered polynucleotide.

A hallmark of a multistate denaturation process is that the progress unfolding curves measured at different wavelengths or determined with different spectral techniques differ.[225,226] This complexity is evident in Figure 26B which shows the temperature variation of the CD signal at 290 and 265 nm. The differences between the two melting profiles can be readily appreciated by comparing the first derivative curves. The inflection points in the derivative curves indicate the number of melting events as well as the midpoint temperature associated with each. As shown in Figure 26B, bottom panel, the 290 nm data exhibit a minimum at 50 °C, while the 265 nm data have a maximum at 46 °C and a minimum at 64 °C. From this, we conclude that the thermal unfolding pathway of Tel22 in 25 mM KCl cannot be described by the two-state model 1 in Scheme 1 but rather requires a more complex pathway. The three distinct
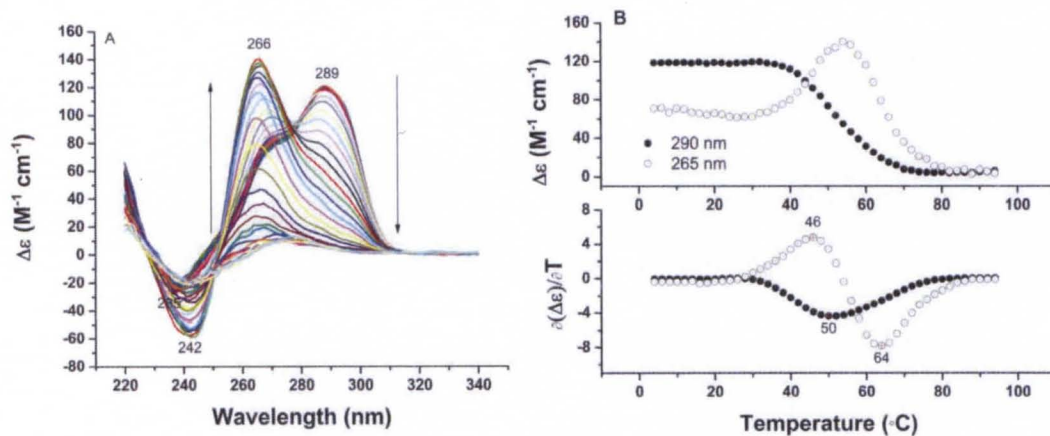
Figure 26. Temperature-dependent unfolding of Tel22 in 25 mM KCl as assessed by CD. (A) Family of CD spectra of Tel22 in 25 mM KCl as a function of temperature from 4 to 94 °C measured as described in Materials and Methods section. The arrows indicate the direction of the temperature change. (B, top panel) Single wavelength melting curves showing temperature dependence of the CD signal at 290 and 265 nm taken from the spectra of (A). (B, bottom panel) First derivative of the data in (B). The inflection points in the derivative curves (annotated in the figure) approximate the T_m values for the two processes.

inflection temperatures suggest contributions to the melting curves from at least four species as in models 4 and 5 in Scheme 1.

A rigorous procedure for determining the minimum number of spectroscopically distinct species in the unfolding experiments is to perform SVD on data sets that include all temperature-dependent spectra and then compare the singular values and the dependence of the V vectors (the amplitudes of each component) on temperature. The results of an SVD analysis of the data set in Figure 26A are summarized in Figure 27, which shows the temperature dependence of the four highest ranking V vectors ($V_1 - V_4$) and in Table 5, which summarizes the results of the analysis.

The SVD analysis suggests that at least three and possibly four species contribute significantly to the CD spectrum of Tel22 over the temperature range 4–94 °C. In Table 5, the three species with the highest singular values contribute 99.88% of the total variance. In addition, these species have V autocorrelation coefficients >0.9, showing the systematic variation in the amplitude vectors with temperature. Figure 27A,B shows that the four most significant amplitude vectors have temperature-dependent inflections as reflected in the relatively high values for their autocorrelation coefficients. The autocorrelation coefficient for V4 (0.78) was slightly less than 0.8, a value generally taken as a cutoff value for significance.[227] However, as described in the next section, including V_4 in the global least-squares analysis resulted in a significantly better fit of the data to a four species pathway compared to a three species pathway. This improvement in the quality of the fit led us to include V_4 in the least-squares analysis of the data.

Assignment of Thermal Unfolding Pathway for Tel22 in KCl Based on CD Melting.
Table 6 summarizes the statistics obtained by least-squares fitting of the data set corresponding to the melt in Figure 26A to the analytical expressions corresponding to the mechanisms in
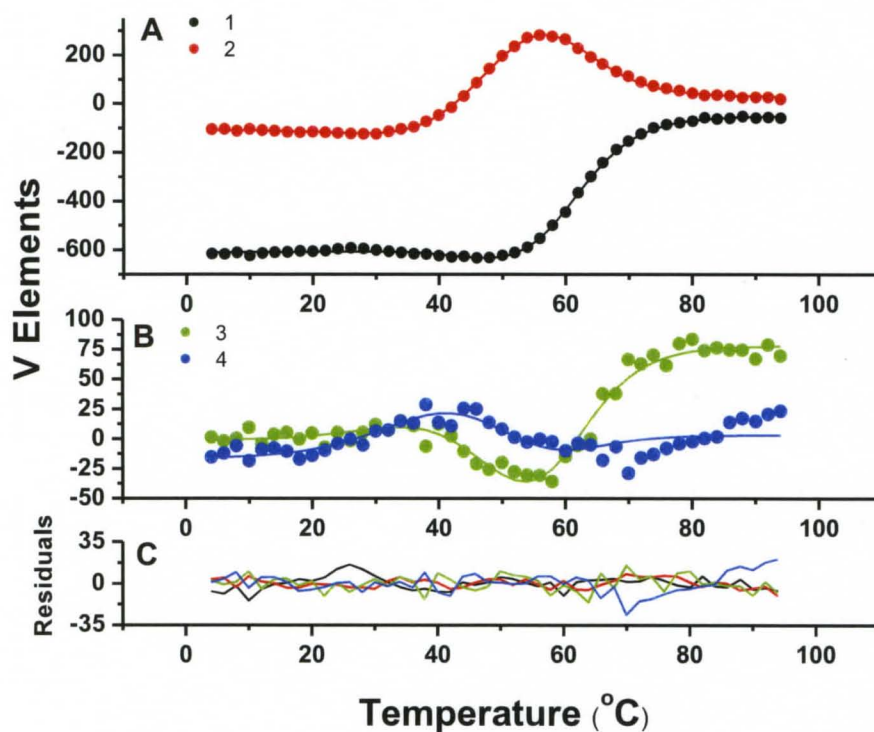
Figure 27. SVD analysis of thermal unfolding of Tel22 in 25 mM KCl. (A) Thermal profiles of vectors V_1 and V_2. (B) Profiles for V_3 and V_4 derived from the data set in Figure 26A. The lines illustrate the least squares fit of the data points to the three- step sequential mechanism 4 in Scheme 1. (C) Residual plots for the fits to the four vectors. The optimized parameters and the fitting equation are summarized in Table 6 and in Appendix C.

Table 5. Parameters from SVD analysis of melting data for Tel22 in Figure 26A.

Index	Singular Value	% of Total Variance	V Matrix Autocorrelation Coefficient
1	3303.96	92.3	0.9811
2	905.37	6.93	0.9789
3	278.22	0.65	0.9312
4	91.06	0.07	0.7774
5	37.28	0.01	-0.2056

Table 6. Model-dependent fitting statistics for thermal unfolding of Tel22 in 25 mM KCl

monitored by CD.

Model	Model	χ^2	AIC Value
Single Transition	1	720265.5	1544.1
Two Transitions	2	15235.4	846.6
Parallel Transitions	3	15235.4	846.6
Three Transitions	4	7759.5	734.5
Parallel Intermediate Transitions	5	7757.3	734.4

Scheme 1. It is clear from the χ^2 statistics that models 1–3 are inferior representations of the data, whereas models 4 and 5 (each with four species) give equally good fits by this criterion. The decreased AIC values for models 4 and 5 indicate that the enhanced fit with four rather than three species cannot be attributed to the increased number of adjustable parameters in the four-species model.

Since both models 4 and 5 of Scheme 1 gave equally good fits, we looked for another means to distinguish between them. To accomplish this we used the optimized values of the fitted thermodynamic parameters in Table 7 to estimate the standard free energy change ΔG for each species at 20 °C. We reasoned that the sequential, two-intermediate model 4 is consistent with the supposition that Tel22 in KCl is a mixture of at least two conformers present in roughly equal amounts.[61,63,206] Thus, the two structures must have approximately the same Gibbs free energy of folding. The calculated $\Delta\Delta G$ of 1.7 kcal mol^{-1} for the two conformers in model 5 requires an F_1:F_2 ratio of approximately $e^{-1.7/RT} \approx 22.5{:}1$, which is clearly larger than the expected 1:1 ratio.

CD Spectra of Intermediates in the Thermal Unfolding of Tel22 in KCl. Fitting a CD data matrix to an explicit unfolding mechanism allows calculation of the spectra of each of the significant species as well as their relative distributions along the unfolding pathway.[202,227] The calculated spectra and species distribution plots associated with model 4 are shown in Figure 28.

The spectra of the folded and unfolded ensembles obviously are very similar to the low- and high-temperature spectra in Figure 26A. The spectrum of I_1 resembles that of a folded quadruplex.[228] The spectrum of I_2 resembles that of Tel22 in KCl–acetonitrile, which is thought to be the all-parallel form.[79] The CD spectra the all-parallel form and the intermediate I_2 have a maximum at ~265 nm and a minimum at ~245 nm. However, the intensity ($\Delta\varepsilon$) of the 265 nm peak for I_2 is about three-fourths of that for the all-parallel form, suggesting that I_2 is not in the

Table 7. Optimal values of thermodynamic parameters for folding Tel22 in 25 mM KCl. ΔH and ΔG values are in kcal mol^{-1}, and Tm values are in °C. Standard free energy changes ΔG are calculated for 20 °C. *Standard deviations derived from 1000 Monte Carlo simluations.[205]

	Model 4	Model 5
ΔH_1	-21.9 ± 4.7*	-25.3
T_{m1}	35.2 ± 5.3	48.4
ΔH_2	-33.4 ± 4.2	-54.6
T_{m2}	47.6 ± 1.6	42.8
ΔH_3	-45.3 ± 4.2	-43.6
T_{m3}	61.4 ± 1.0	60.7
ΔG_1	-1.1 ±0.4	-2.2
ΔG_2	-2.9 ± 0.4	-3.9
ΔG_3	-5.6 ± 0.6	-5.3
ΔG_{Total}	-9.5 ± 0.7	-11.5
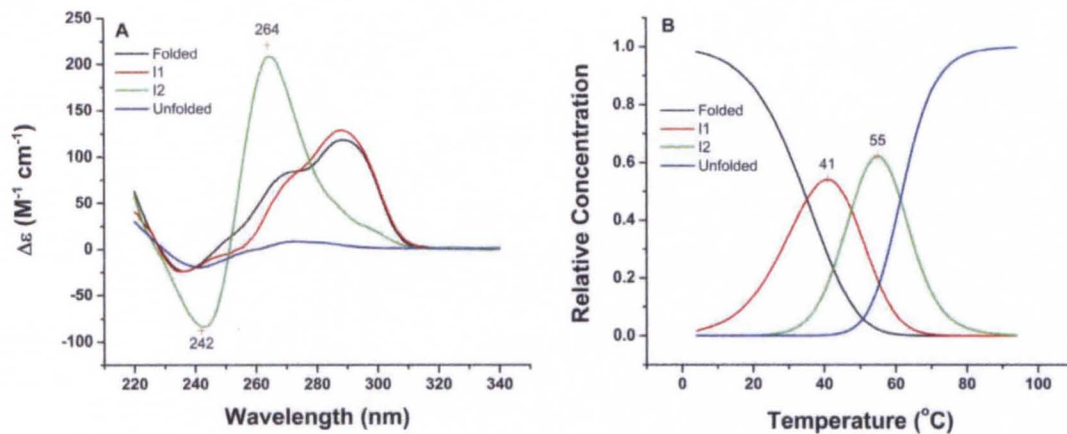
Figure 28. Calculated CD spectra and relative concentration profiles for the species derived by SVD analysis and fitting the data set in Figure 26A to the analytical expression associated with the sequential intermediate model 4. (A) CD spectrum of each significant species. (B) Relative concentrations of each as a function of temperature.

all-parallel state. In addition, as shown below for fluorescent analogs of Tel22 which form a similar intermediate at 55–65 °C, the fluorescence properties of the intermediate are inconsistent with those expected for the parallel conformation.

Melting of 2AP Derivatives of Tel22 Assessed by Fluorescence Quantum Yield. To further characterize the intermediate state(s) of Tel22, we determined the melting profiles of fluorescent derivatives of Tel22. 2-Aminopurine (2- AP) is a fluorescent analog of adenine whose fluorescence quantum yield is sensitive to various environmental and structural factors.[95,96,229,230] Serial substitution of 2AP for the 5′ terminal A and for the three-loop A residues provides a sensitive reporter of the local environment at these positions and thus potentially could indicate changes in loop geometry during melting independently of G-quartet melting.[76,91] Previous studies from our laboratory have shown that A → 2AP substitutions do not grossly perturb the structure of Tel22 in Na^+ or K^+ solutions as assessed by T_m or CD and therefore can be used to track folding as a function of cation concentration without perturbing the structure.[71,72] In addition, the quantum yield,[71] excited-state lifetimes,[76] and accessibility of the 2AP analogs of Tel22 to collisional quenchers, such as acrylamide, depends on the position of the 2AP residue within the sequence. Thus we expect that the melting profiles of the four different 2AP derivatives of Tel22 assessed by changes in emission intensity may reveal intermediates during thermally induced unfolding.

The temperature-dependent emission spectra of Tel22 with position-specific A → 2AP substitutions are given in Appendix C : Figures C1–C4. As expected, the emission maxima were invariant with temperature. However, a multiphasic response of emission intensity with respect to increasing temperature was noted for all four of the 2AP derivatives of Tel22 (Figure 29A). The derivative spectra in Figure 29B reveal that the loop variants AP7, AP13, and AP19 have inflections in their first derivative curves in temperature ranges of 43–52 and 64–70 °C, while
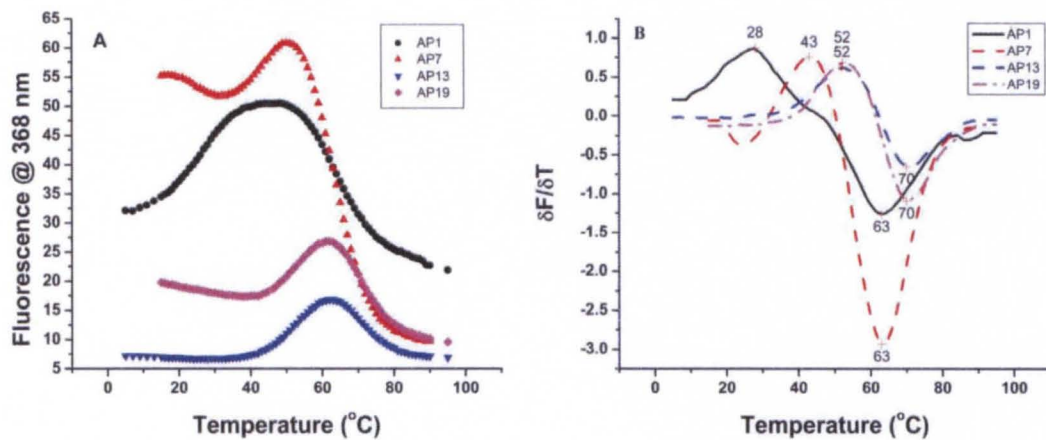
Figure 29. (A) Melting profiles for the temperature dependence of the emission intensity of the derivatives of Tel22 in which each of the four A residues is substituted serially with 2AP. The numbers in the legend on the figures show the position of the 2AP substitution in the sequence of Tel22. (B) First derivative curves of the data in (A). The numbers in (B) give the temperature associated with the maxima or minima in the curves. These temperatures are the approximate T_m of each step in the melting thermally induced unfolding process.

the terminal AP1 has inflections at 28 and 63 °C. The multiple inflections in the first derivative

curves show that the changes in 2AP emission intensity track at least three states during the

unfolding process. In contrast, the emission intensity of the Tel22 2AP derivatives in the

absence of KCl as well as that of a control oligonucleotide in 100 mM KCl (d(TT-2AP-GT)) were

only weakly temperature dependent (Appendix C : Figure C5). The sharp inflections seen in

Figure 29 above ~25 °C are therefore associated with quadruplex unfolding rather than

temperature effects on the intrinsic emission properties of 2AP.

SVD analysis of the fluorescence emission data sets (Appendix C: Figures C1–C4) for

melting of the 2AP derivatives of Tel22 shows that three species contribute 99.99% of the

variance and that all three of these species exhibit V matrix autocorrelation coefficients >0.75

(Table 8). Based on autocorrelation, the fluorescence studies with 2AP-Tel22 are consistent

with model 3 in which three species (F, I, and U) are populated in thermal unfolding of Tel22.

Although the AIC statistic suggests an improved fit with more species, the low value of the

autocorrelation coefficients for V_4 does not justify including a fourth species.

Emission Spectra and Thermodynamic Parameters for Melting of 2AP Derivatives of

Tel22. Fitting the 2AP data matrices corresponding to the data of Appendix C: Figures C1–C4 to

the analytical expression describing the two-step sequential mechanism allows determination of

the emission spectra as well as the T_m and ΔH_{vH} for melting of the F and I ensembles. These

spectra are collected in Figure 30, and the fitted parameters are in Table 9. Comparison of the

spectra shows at all positions; the emission of the intermediate is enhanced compared to both

the folded structures and the unfolded structures. Moreover, the emission intensity of the

intermediate structures is clearly different with AP7 > AP1 > AP19 > AP13.

Fluorescence Melting of 6-FAM-Tel22-Tamra. An additional spectroscopic probe to aid

in defining the thermal unfolding pathway is the FRET-labeled derivative of Tel22. Folding of

Table 8. Model-dependent fitting statistics and SVD analysis results for 2AP derivatives of Tel22 melting in 25 mM KCl.

Oligo/Index	Model Selection Statistics		SVD Analysis Statistics		
	χ^2	AIC	Singular Value	% of Total Variance	V Matrix Autocorrelation Coefficient
AP1					
1	557472.5	2461.7	3149.58	100	0.994
2	404.2	102.6	16.14	0	0.9666
3	404.3	102.6	5.67	0	0.8438
4	104.6	-328.9	2.6	0	0.3781
5	194.6	-125.2	1.94	0	0.1318
AP7					
1	37090.4	1482.4	3483.35	100	0.9895
2	3437.8	771.4	16.84	0	0.9576
3	3445.9	772.1	6.82	0	0.8801
4	3437.8	783.4	3.86	0	0.1726
5	74.3	-382.2	3.01	0	0.131
AP13					
1	39202.9	1591	820.25	99.96	0.9936
2	133.1	-261.9	16.58	0.04	0.9683
3	133.1	-261.9	1.83	0	0.4841
4	15.4	-956.3	1.43	0	-0.0361
5	15.4	-956.3	1.31	0	0.0156
AP19					
1	44883.5	1556.4	1503.22	99.98	0.9911
2	1763.8	571.5	23.25	0.02	0.9848
3	1763.8	571.5	3.17	0	0.8454
4	1763.8	583.5	1.72	0	0.0588
5	44883.5	1580.4	1.62	0	0.1135
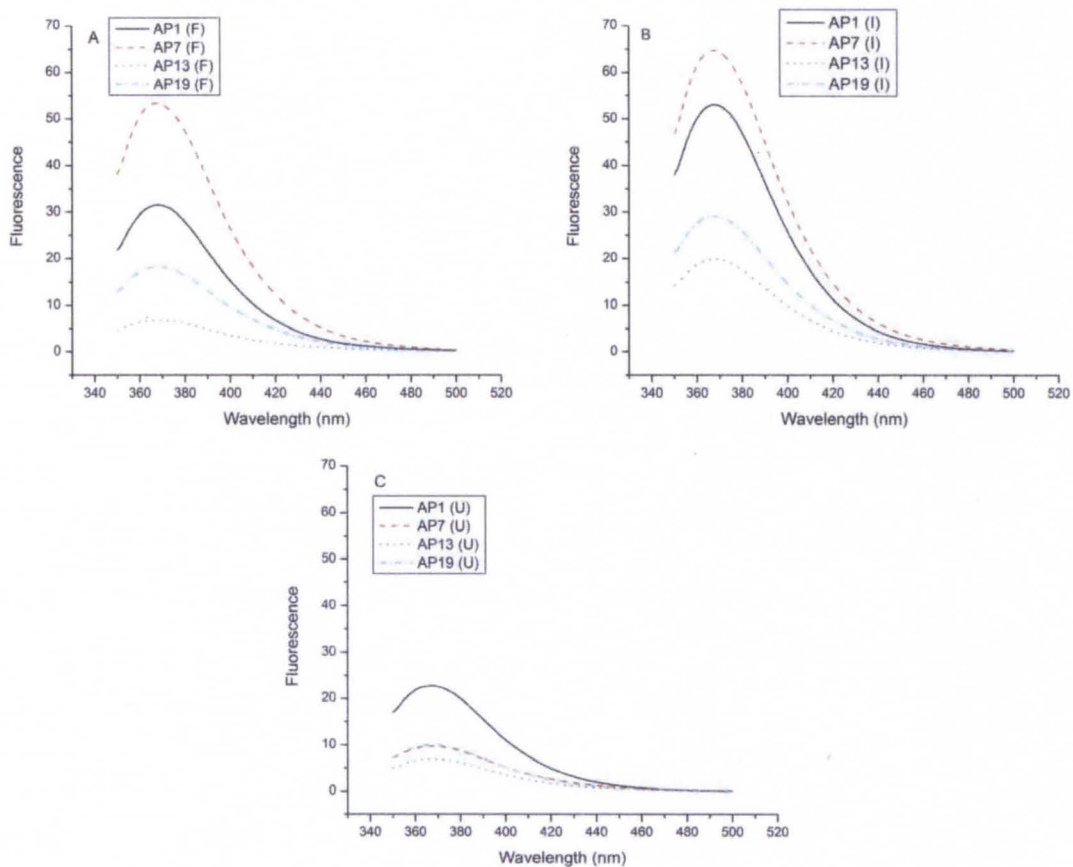
Figure 30. Emission spectra of 2AP derivatives of Tel22 in 25 mM KCl determined by SVD analysis of the thermal unfolding of Tel22 with serial substitutions of 2AP at each A residue. (A–C) Calculated spectra of the folded ensemble, the intermediate species, and unfolded ensemble, respectively. The spectra were calculated using model 2 and the optimized thermodynamic parameters in Table 9.

Table 9. Optimal thermodynamic parameters for folding 2AP derivatives of Tel22 in 25 mM KCl assuming model 2. ΔH and ΔG values are in kcal mol^{-1} and T_m values are in °C. Standard ΔG values are calculated for 20 °C.

	T_{m1}	ΔH_1	T_{m2}	ΔH_2	ΔG_1	ΔG_2	ΔG_{Total}
Tel22-AP1	26.5	-27.7	64.6	-37	-0.6	-4.9	-5.5
Tel22-AP7	45.4	-79.2	63.8	-49.1	-6.3	-6.4	-12.7
Tel22-AP13	53.6	-45.5	70.6	-51.9	-4.7	-7.6	-12.3
Tel22-AP19	54.1	-70.3	71.6	-53.9	-7.3	-8.1	-15.4

such models of the human telomere sequence has been investigated using oligonucleotides with suitable donor and acceptor fluorophores, such as 6-carboxyfluorescein and tetramethylrhodamine linked to the 5′ and 3′ ends of the oligonucleotide.[231,232] Since the termini of telomeric oligonucleotides are brought closer together during the folding process, folding should increase the efficiency of energy transfer, which would be apparent from quenching the 6-FAM fluorescence at 520 nm and enhancement of Tamra fluorescence at 585 nm. The oligonucleotide 6-FAM-Tel22-Tamra is expected to provide a sensitive test of the multistate folding process because appearance of a stable folding intermediate should manifest itself as a change in the distance and/or orientation properties of the FRET donor and acceptor which in turn will influence FRET efficiency.

Figure 31 illustrates the emission spectra of 6-FAM-Tel22-Tamra as a function of temperature in 25 mM KCl. Figure 31B,C shows the thermal unfolding profiles of 6-FAM-Tel22-Tamra at 520 and 585 nm, the emission maxima of 6FAM and Tamra, respectively. The intensity of the donor emission increases with temperature while that of acceptor concomitantly decreases as expected for an unfolding process which separates the donor and acceptor probes. The corresponding first derivative spectra $\partial F/\partial T$ (Figure 31D,E) reveal inflection points at ~45 and 65 °C.

Analysis of the data set in Figure 31 by SVD shows that four species contribute 99.99% of the variance and that the V profiles of the four species have autocorrelation coefficients >0.95 (Appendix C: Tables C1 and C2). Fitting the melting data to model 4 resulted in the species emission spectra shown in Figure 31C and the thermodynamic parameters summarized in Table 10. We conclude that the FRET melting data are consistent with a multistep melting process for Tel22.
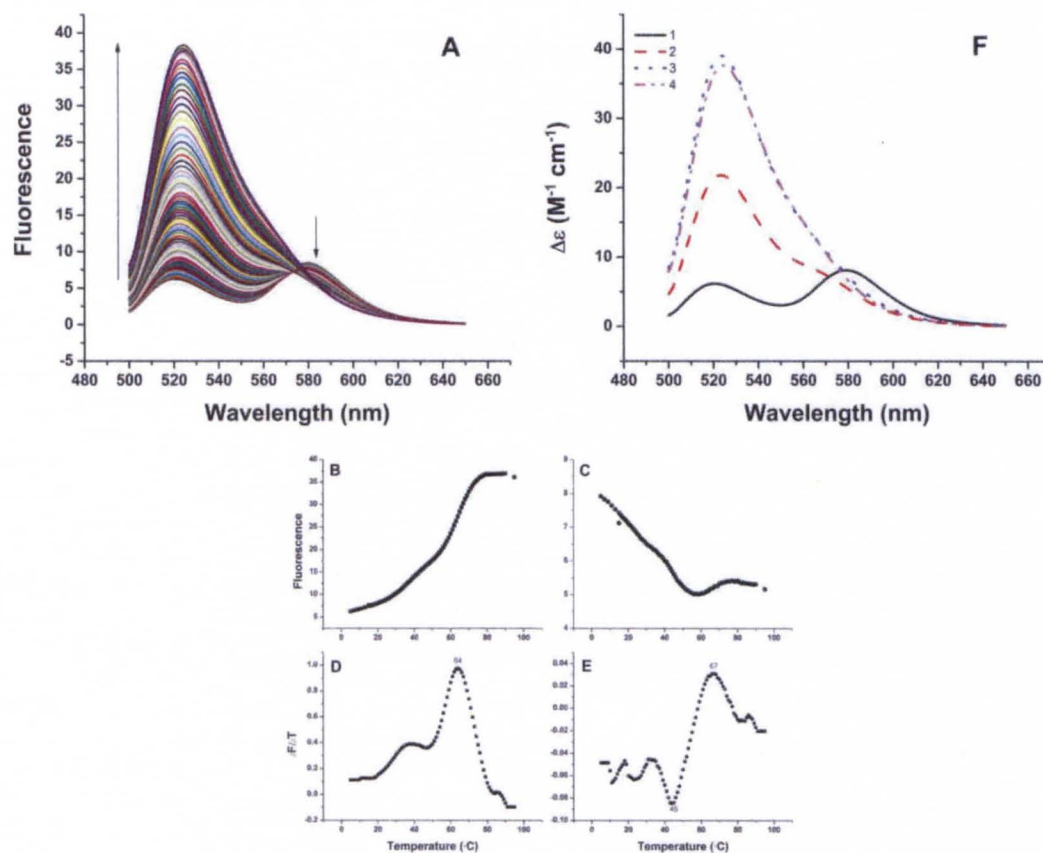
Figure 31. Thermal melting of 6-FAM-Tel22-Tamra in 25 mM KCl assessed by changes in FRET efficiency. (A) Family of emission spectra collected as a function of temperature between 5 and 95 °C. The arrows show the direction of the change (increase or decrease) in emission intensity with increasing temperature. (B,C) Changes in emission intensity at 520 nm (6-FAM) and 585 nm (Tamra) with increasing temperature derived from the spectra in (A). The first derivative spectra in (D,E) show at least two major inflections in the melting profiles at ~45 and ~65 °C. (F) Emission spectra of the four species obtained by fitting the data set to model 4 using the parameters in Table 10.

Table 10. Thermodynamic parameters for folding 6-FAM-Tel22-Tamra and quadruplex forming sequence variants. ΔH and ΔG values are in kcal mol^{-1} and T_m values are in °C. Standard ΔG values are calculated for 20 °C. *Standard deviations derived from 1000 Monte Carlo simulations.

	6FAM-Tel22-Tamra	2GKU		2AQY + 6	Tel-3G
Model (Scheme 1)	4	4		3	3
$T_{m,1}$	39.7	33.8	±6.6*	40.3	46.3
ΔH_1	-18.5	-21.1	±8.1	-36	-19.8
$T_{m,2}$	65.1	53.8	±1.9	57.9	70.4
ΔH_2	-45.7	-41.4	±6.2	-19.8	-21
$T_{m,3}$	86.5	60.8	±1.2	-	-
ΔH_3	-45.9	-55.8	±6.6	-	-
ΔG_1	-1.2	-0.9	±0.4	-2.3	-1.6
ΔG_2	-6.1	-4.3	±0.6	-2.3	-3.1
ΔG_3	-8.5	-6.8	±0.9	-	-
ΔG_{Total}	-15.8	-12	±1.0	-4.6	-4.7

Melting of Quadruplex-Forming Oligonucleotides with Variant Sequences. An

important question arising from characterization of thermal unfolding of Tel22 is whether the

heterogeneous pathway is unique. Is the complexity a consequence of the known topological

heterogeneity of the Tel22 system in KCl at room temperature? To answer this question, we

investigated the thermal unfolding of a variety of quadruplex-forming oligonucleotides of known

structure shown in Table 4. 2GKU was chosen because NMR shows that it forms predominantly

a single structure (the so-called hybrid-1 fold) in KCl.[65] 2AQY, a 16-mer with three G3 runs

(d(($G_3T_2A)_2G_3T$)), forms a quadruplex in 100 mM NaCl in the presence of a hexamer with a single

G3 run (d(AG_3T))[221], and Tel-3G (d(($T_2A(G_3T_2A)_3$))) has been reported to form a triplex in NaCl and

KCl.[233] In each of these cases, we observed a multistep unfolding process (data in Appendix C:

Figures C5–C7).

CD Spectra of the Intermediates Suggest a Common Structure. The CD spectra of the

melting intermediates are compared in Figure 32, and the corresponding spectra of the folded

and unfolded states are given in Appendix C: Figures C6–C8. The spectra of the intermediate

state for Tel22 and 2GKU (Figure 32) are similar with maxima at ~265 nm and minima at ~244

nm. The 2AQY system has a spectrally similar intermediate whose maximum is shifted slightly to

shorter wavelengths. The fourth spectrum in Figure 32 is that of a triplex-forming structure, Tel-

3G. It is noteworthy that the spectral shape and the positions of the maximum and minimum of

Tel-3G are similar to those of the quadruplex-forming oligonucleotides. These results show that

similar intermediates are produced from both oligonucleotides and that the melting mechanism

is not unique to Tel22.

The position and magnitude of the of the CD spectra of the intermediate resemble that

of the all-parallel topography, which generally has a maximum in the vicinity of 265 nm and a
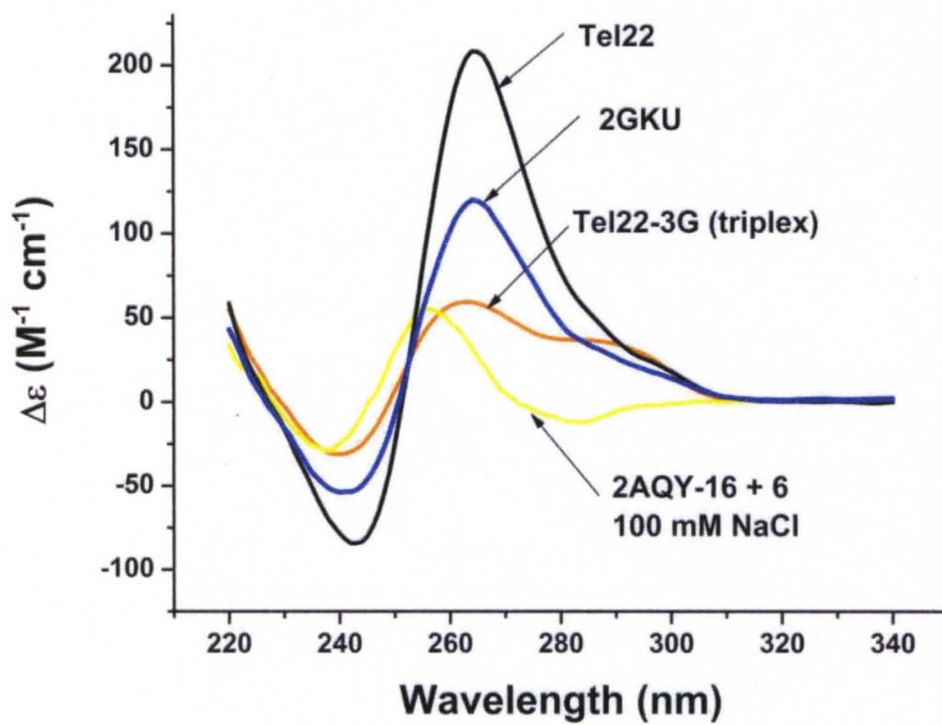
Figure 32. CD spectra of thermal unfolding intermediates derived by SVD analysis of the wavelength-dependent melting profiles and fitting the resulting significant V vectors of each data set to models 3 or 4 in Scheme 1.

smaller minimum near 245 nm.[228] However, an all parallel topography is not consistent with the

emission spectra of the 2AP and FRET analogues of Tel22 as discussed below.

Discussion.

We make four major points regarding the thermal unfolding of G-quadruplex forming oligonucleotides: (1) the thermal denaturation of quadruplex in KCl cannot be adequately described as a simple two-state unfolding process but rather requires three or four species to fit multiwavelength CD, fluorescence emission, or FRET spectral profiles over the temperature range of 4–90 °C; (2) the spectroscopic data support sequential models 2 and 4 outlined in Scheme 1; (3) the sequential pathway is common for two sequences (Tel22 and 2GKU) that exhibit differing degrees of structural heterogeneity at room temperature; and (4) taken together, the 2AP fluorescence and FRET melting experiments are inconsistent with the hypothesis that the intermediate state is an all-parallel folded state. The spectral properties of the intermediate state melting at 55–65 °C are consistent with its identification as a triple-helical structure. Our results and conclusions apply only to the human telomere quadruplex which is known to adopt a variety of structures. Whether other types of quadruplexes have multistate unfolding pathways is not known and requires investigation.

Implications of Multistate Melting. Published studies often have assumed that quadruplex melting follows a simple two-state reaction in Scheme 1, thus allowing estimation of van't Hoff enthalpy changes and associated (derived) thermodynamic parameters, such as $\Delta G_{folding}$ from a single melting curve. This is in spite of the fact that numerous calorimetric studies have shown that some quadruplex thermograms derived by DSC require more than two species in equilibrium during the entire course of the melt.[208,219] The assumption of two-states should be justified by showing, for example, that monitoring unfolding at two or more wavelengths or by two independent methods gives overlapping results for the unfolding profile.[202,225,226,234,235] By these criteria, the unfolding of Tel22 in KCl and its sequence

homologues are certainly not two-state. Without verification of the model, the derived thermodynamic parameters may be invalid.

What are the consequences of neglecting the complexity of the quadruplex unfolding mechanism? Analysis using single wavelength data and the two-state assumption lead to underestimates of the thermodynamic parameters of the denaturation reaction and to possible misconceptions about quadruplex stability. This can be illustrated by using the data in Figure 26B. The melting curve obtained as recommended[217] at 295 nm appears to be consistent with a two-state process and can be fit to yield T_m = 52.6 and ΔH = 30.4 kcal mol^{-1} for the denaturation reaction, from which $\Delta G_{20°C}$ = −2.6 kcal mol^{-1} (for folding) can be obtained. These values are fully consistent with literature reports for the stability of this quadruplex.[81] But this free energy estimate is much smaller in magnitude than that obtained using the sequential mechanism demanded by our analysis (Table 7) and gives the erroneous impression of a less stable folded form. The origins of the problem may arise from (at least) two possible sources: First, by using a single wavelength to monitor quadruplex denaturation, one may focus on the melting of only a particular region of the molecule and miss the contributions of other regions to the unfolding process. Note that the thermodynamic parameters obtained by the two-state analysis using only the 295 nm data most closely resemble those that characterize the second step in the sequential mechanism, which we assign to the $I_1 \rightarrow I_2$ transition. Alternatively, a melting curve obtained at a single wavelength may not fully resolve into multiple transitions, appearing instead as a single broad transition. Since the van't Hoff enthalpy determines the slope and width of the transition, an underestimate of the true underlying overall enthalpy would result.

Reliable thermodynamic parameters are necessary for any attempt to predict quadruplex stability. Recently a novel Bayesian prediction framework based on Gaussian process regression was used to predict the thermodynamic stability of previously unmeasured

G-quadruplexes from sequence information alone.[25] The algorithm relied primarily on experimentally measured T_m values for a quadruplex training set, a perfectly reasonable starting point but a rather primitive measure of stability nonetheless. Such predictive algorithms could be greatly improved by use of more complete and more accurate thermodynamic parameters obtained from more realistic models of the denaturation process.

Support for Sequential Unfolding Mechanism. A useful aspect of using SVD to analyze multi-wavelength data sets is that it provides an estimation of the relative contribution of each species to the temperature profile. All spectrally visible species contribute to the overall signal as a function of temperature and wavelength according to the magnitude of their respective singular values. Thus the melting profile of each species during the melt can be visualized by plotting the magnitude of the corresponding V vector as a function of temperature. V profiles that vary in a nonrandom manner with temperature can be assigned to individual spectroscopic species in a model independent manner. A quantitative means of establishing whether a particular V profile varies periodically or randomly is to calculate the autocorrelation coefficient of the successive elements of the V vector. When this test of all the data sets in this study was carried out over the complete temperature range, the V vectors for all (except one) had autocorrelation coefficients ≥ 0.75 for the three V vectors with the highest S value, indicating a high degree of correlation for these vectors. The remaining vectors have autocorrelation coefficients <0.75 (see Appendix C). The exception is the data set for 6-FAM-Tel22-Tamra which by this criterion has four significant V vectors. However, the Tm associated with V_4 is ~86 °C; this high temperature reflects a process that occurs post melting of the intermediate species. Since none of the other constructs exhibited such a high T_m species, it seems likely that this process is associated with the donor–acceptor pair rather than quadruplex unfolding per se.

A criterion for distinguishing among models with differing numbers of adjustable parameters (such as model 4 from 5) is the AIC value. For all of the data sets, there was generally improvement in AIC when additional steps were included. However, it was not always possible to distinguish between models with the same number of steps because these models gave approximately the same AIC. Although including an additional step in the model generally gave improved fits, we excluded models that had V vector autocorrelations <0.75 since this would indicate temperature variation in the amplitude vector that was inconsistent with the variation expected for a melting phenomenon. Because the autocorrelation coefficients of V_4 for the 2AP data sets did not meet this criterion, the data were fit to the two-step model 3.

Three-step models 4 and 5 gave equivalent fits to the data sets in terms of the χ^2 and AIC criteria. We prefer the three-step sequential model for the following reasons. First, both Tel22 and 2GKU gave equally good fits to this model, even though Tel22 in KCl is known to consist of at least two conformers in approximately equal amounts (e.g., the ΔG of conversion between the two species ≈ 0), while 2GKU is >90% a single conformer. The F state likely consists of an ensemble of closely related, nearly isoenergetic conformations that convert to the I states and then to the U ensemble rather than states F_1 and F_2 unfolding to the U state without passing through an intermediate state that is present in a detectable concentration.

Fluorescence Emission Spectra of the Thermal Intermediates Are Inconsistent with the Properties Expected of the Parallel Fold. Since the A residues located within the TTA loops are in equivalent environments in the symmetrical all-parallel topology, we expect the fluorescence emission properties of the 2AP analogs to be similar if the thermal intermediate is an all-parallel structure. A similar strategy has been previously used to distinguish the solution structure of Tel22 from the crystal structure; this reference contains a detailed structural justification for the assertion of topological equivalence of the loop adenine residues in the all-parallel structure.[71]

Figure 30A-C compares the calculated emission spectra for the F and thermally unfolded U ensembles of the four 2AP derivatives of Tel22 in 25 mM KCl. The spectra are similar to those previously published.[71] Figure 30B depicts the emission spectra deduced for the I state in the thermally induced unfolding of the Tel22 derivatives. Comparison of the spectra in Figure 30A-C shows two features: First, the quantum yield is larger for 2AP emission at all positions when compared to the corresponding F and U states. Second, the quantum yield is nearly the same for 2AP in positions 1 and 7, suggesting that these residues experience fewer stacking interactions with other residues of the folded structure. In other words, these residues, both of which are located in the first arm of the folded quadruplex, experience more freedom of motion in the intermediate than in either the folded or unfolded states.

The emission spectrum of 6-FAM-Tel22-Tamra in the I state is also inconsistent with that expected for the parallel conformation. The intermediate spectrum in Figure 31D shows a decrease in the FRET efficiency compared to the unfolded state. The decreased FRET efficiency is not as strong as that for the U state, implying that I retains structure but that the distance between donor and acceptor and/or their relative orientations is significantly altered compared to both the F and U ensembles. Examination of a molecular model of Tel22 in the parallel conformation shows that the 5' and 3' ends of the molecule, where the donor and acceptor are attached, are relatively close together (6.9 Å calculated from the crystal structure).[57] Thus there should be significant donor to acceptor energy transfer which was not observed. In fact, the degree of end-to-end separation is nearly as great as for the fully unfolded state.

Thermal Intermediate May Be a Triplex. Several groups have suggested that cation-driven quadruplex folding may proceed through a hairpin duplex that in a stepwise manner adds consecutive strands to form a triplex intermediate consisting of three stacks of three H-bonded G residues, followed by association of the fourth G_3 run to form the nascent quadruplex.[69,120,236-]

[238] Subsequent slow relaxations result in positioning the loops with respect to the G-stack and adjustment of the loop topology to ultimately form the thermodynamically most stable structure(s). For a quadruplex with four G_3 runs such as those in this study, either the first or fourth G_3 run could be added last; thus, there could be two triplex structures depending on which run was added last.

Until recently, this folding pathway was speculative. However, Mashimo et al.[220] published a rigorous molecular dynamics analysis of these pathways and thereby provided a detailed visualization and energetic analysis of the triplex folding pathway that suggested that triplex structures are energetically feasible, i.e., G-triplets were nearly as stable as G-quartets (-50.6 vs   -67.8 kcal mol^{-1} G). More recently, Koirala et al.,[233] building on previous studies by others, investigated the structure of d(TTA(GGGTTA)3) (Tel-3G), a sequence that contains three tandem G_3 repeats. Based on physical studies (optical tweezers and CD), molecular dynamics and footprinting,

they demonstrated that in the presence of 100 mM NaCl, this sequence forms a triplex structure. The published CD spectrum of the triplex in NaCl is similar to the shape of the CD spectrum of the 55–60 °C intermediate observed in our studies. This similarity, as well as the fluorescence properties of the corresponding melting intermediate observed in this study, suggests the possibility that it is a mixture of triplex states of the human telomeric DNA oligonucleotide.

Comparison with Calorimetry. A melting intermediate was previously identified for the denaturation of the human telomere quadruplex (TTAGGG)$_4$ by a combined CD and DSC study.[219] A very recent calorimetric/CD study used a sophisticated global analysis method obtained the thermodynamic parameters for the denaturation of the human telomere quadruplex d(AGGG(TTAGGG)$_3$) in K$^+$ solutions.[208] A single intermediate was identified,

corresponding to our model 3 (Scheme 1). Our data suggest an additional intermediate is necessary. Our results complement and extend this recent study and fully agree about the need to include intermediate states along the denaturation pathway, a conclusion arrived at by independent physical methods. There is also agreement that one intermediate state may be a triplex form. The strength and novelty of our approach are that the spectroscopic properties of the intermediates are defined, providing a deeper characterization of their physical properties.

An important difference between the CD/DSC study and ours is that Bončina et al.[208] were able to determine a heat capacity change (ΔC_p) for quadruplex unfolding, which we are not able to do with confidence. ΔC_p values are best determined by calorimetry[239,240] and are notoriously difficult to obtain from fits of spectroscopic thermal denaturation data alone without more elaborate protocols that combine chemical and thermal denaturation experiments.[241,242] Inclusion of ΔC_p values in our analysis would introduce many additional parameters that would tax our nonlinear fitting methods. However, we can estimate the consequences of neglecting ΔC_p values on some of our parameters. Bončina and co-workers report an overall ΔC_p value of 310 cal mol^{-1} K^{-1} for the unfolding of the human telomere quadruplex in K$^+$ solution. (A much larger calorimetric value of $\approx$1600 cal mol^{-1} K^{-1} was reported for the unfolding of the same quadruplex in Na$^+$ solution,[243] a value that seems unreasonably large but which attests to the difficulty of obtaining accurate ΔC_p values even by calorimetry.) For the final step in the unfolding process, Bončina and co-workers report an enthalpy value of 28.8 kcal mol^{-1} at 25 °C compared to our value for a similar step of 45.3 kcal mol^{-1} at 61.4 °C. If we correct our value using the reported ΔC_p value, we estimate an enthalpy of 33.3 kcal mol^{-1} at 25 °C, in good agreement (within about 15%) with the calorimetric value. For the first step in the unfolding process, a negligible heat capacity change was reported. The reported enthalpy change of 26.7 ± 1.0 kcal mol^{-1} at 25 °C is in good agreement with our value of 21.9 ± 4.7 kcal

mol^{-1} at 35.2 °C, although the calorimetrically determined value is clearly more precise. Where comparison is possible, the values obtained in our study agree well with the results of Bončina and co-workers. Neglect of ΔC_p would lead to an overestimate of the magnitudes of our free energy changes at 20 °C (Table 7). If we assume the ΔC_p of Bončina and co-workers, our estimate of ΔG_{Total} would decrease from -9.5 to -8.3 kcal mol^{-1}. A variety of calorimetric studies is underway in our laboratory to address the issue of heat capacity changes, but these are incomplete and beyond the scope of this report.

One reviewer questioned if neglect of heat capacity changes might have influenced our choice of the most probable mechanism and suggested that if ΔC_p parameters were included we might choose the three-state model (model 3, Scheme 1) over the four-state model (model 4, Scheme 1). That is not likely to be the case. A key element in our selection of a mechanism comes from SVD analysis which enumerates in a model-independent fashion the number of significant spectral species that must be accounted for. There is little ambiguity in the SVD analysis of CD data for Tel22 (Table 5) that four species are needed, as was subsequently confirmed by the statistical comparison of fitting models (Table 6). The model-free SVD analysis places valuable limits on the minimal number of species that must be include in any mechanism, four in this case.

Conclusions.

Our data show that thermal denaturation of the human telomere quadruplex is not a simple two-state process but rather proceeds along a sequential pathway with multiple populated intermediate states. One of these states is most probably a triple helical form, as proposed by recent molecular dynamics simulations.

CHAPTER 6

THERMODYNAMIC CHARACTERIZATION OF HUMAN TELOMERE QUADRUPLEX UNFOLDING.

Introduction.

The biological relevance of G-quadruplexes, which have potential functions at each step of the central dogma, has become an important research topic.[24,244] For example, enzymes involved in DNA replication interact with quadruplex structures whose unwinding is critical for efficient replication through guanine rich regions of the genome.[38,42] Cells deficient in the quadruplex unwinding enzyme REV1 demonstrate a loss of repressive markers and epigenetic instability due to uncoupled histone recycling during DNA synthesis.[41] In addition, proper maintenance of the telomeric overhang requires the processive unwinding of quadruplex structures by the POT1-TPP1 complex.[245] Quadruplexes also have important regulatory functions in gene transcription and translation. Formation of quadruplex structures in oncogenic promoters modulate gene expression,[246] while quadruplex forming sequences in untranslated regions of mRNA regulate protein production.[43] Quadruplexes offer unique structures for targeting by small molecules that can potentially modulate biological functions, gene expression, and protein production.

Studies often demand high-resolution structures for development of ligands with high selectivity for distinct quadruplex conformations.[59,60] Quadruplex structures nominally contain two or more stacked G-tetrads stabilized by the coordination of monovalent cations within the central core. Quadruplex topologies can differ in molecularity, glycosidic bond angles,

144

participating strand orientation, and connecting loop regions.[5] These variations in structure

inherently complicate quadruplex systems through development of heterogeneous solution

conformations. Conformational heterogeneity is exemplified by human telomeric quadruplex

(ht-quadruplex) systems. Human telomeres consists of repetitive DNA sequences in which 100-

200 bases are found as a single-stranded 3'-overhang.[148] Sequences taken from the human

telomeric repeat, 5'-d(TTAGGG), are capable of forming quadruplex structures *in vitro*.[11]

Through the use of high-resolution techniques, five distinct monomeric ht-quadruplex

topologies have been resolved. Although the monomeric ht-quadruplexes contain the same

core 21-nt sequence (5'-d(GGG(TTAGGG)$_3$)), the addition of flanking residues significantly

perturb the quadruplex systems into distinct folding topologies.

The first high-resolution study of ht-quadruplexes evaluated a 22-nt sequence in the

presence of sodium ions. Under these conditions, this sequence adopted a monomeric anti-

parallel "basket" topology with a three-stacked G-tetrad core with loops ordered 5' to 3' of

edgewise, diagonal, edgewise.[62] This was followed by an X-ray crystallographic study that

demonstrated the same 22-nt sequence formed an all-parallel "propeller" conformation in

potassium solution.[57] The propeller conformation has three stacked G-tetrads with all double-

chain reversal loops, but has been shown to have unfavorable stability and is unlikely a

predominant conformation in pure aqueous solutions.[71]

After elucidation of the basket and propeller conformations, three distinct "hybrid"

conformations have been resolved using the core 21-nt sequence with different 5' and 3'

flanking residues. Hybrid-1 and hybrid-2 conformations exhibit a unique (3+1) strand-

orientation, in which three strands are oriented parallel with the fourth strand running anti-

parallel, but differ in the orientation of the connecting loop regions. Hybrid-1 has loop regions

ordered 5' to 3' of double-chain reversal, edgewise, edgewise,[63-65] while the hybrid-2

conformation has the reversed 5' to 3' order of connecting loop regions.[66,67] The hybrid-3

conformation does not have the (3+1) strand orientation, but instead displays an anti-parallel

topology resembling the original 22-nt conformation established in sodium with loops ordered

5' to 3' of edgewise, diagonal, edgewise.[68,69] Hybrid-3 conformations have the unique property

of a two stacked G-tetrad core instead of the three G-tetrads encountered with all previous ht-

quadruplex conformations.

Telomeric quadruplex systems are also known to be influenced by the addition of

dehydrating co-solvents[79] and high concentrations of polyethylene glycols.[180] The transition was

established to be a change from a hybrid to a propeller conformation in high concentrations of

PEG.[80] The mechanism of the conformational transformation has been proposed to be driven by

molecular crowding, although the propeller conformation occupies a larger molecular volume

than the hybrid conformation, in opposition to the precept that molecular crowding favors more

compact structures. An analytical method for partitioning excluded volume and preferential

interactions[194] was recently employed to investigate the mechanism of co-solvent conversion.

Findings suggest that the conversion in acetonitrile is driven primarily by hydration effects,[79]

whereas preferential interactions significantly contribute to the conformational change driven

by large MW PEGs. Therefore, the mechanism of PEG induced conformational changes has been

proposed to be conformational selection of the propeller form through a linked equilibrium with

PEG binding (Chapter 4).

Quadruplexes from the human telomere repeat constitute a unique set of diverse

topologies that can be used for a systematic study of thermodynamic differences based on

conformation and sequence variation. Such a unique set has been used for hydrodynamic bead

modeling to evaluated hydrodynamic differences in size, shape, radius of gyration and sedimentation coefficients between the distinct ht-quadruplex conformations.[247] Hydrodynamic characterization through computation methods is capable of accurate estimation of solution properties, bridging the gap between high and low-resolution methodologies. However, many unanswered questions still remain as to how quadruplex structures unfold. Computational studies have shown that triple-helix intermediates are energetically stable compared with quadruplex structures.[248] Recent spectroscopic investigations demonstrated that a 22-nt ht-quadruplex unfolds by a complex mechanism requiring multiple intermediate states, and proposed unfolding contains a triple-helix intermediate.[208,249] Understanding the thermodynamics involved in quadruplex unfolding can aid in the rational design of quadruplex-interactive compounds and improve predictive algorithms used to evaluate quadruplex formation.[25,250]

Investigated here are the unfolding thermodynamics of eight sequences derived from the human telomeric repeat known to form monomeric quadruplexes. Quadruplexes are evaluated in both aqueous and 50% (v/v) acetonitrile. Singular value decomposition is employed for determination of the number of spectral components necessary to describe the unfolding mechanism. This work includes a detailed thermodynamic characterization of ht-quadruplex unfolding, putting to use recently established ht-quadruplex unfolding mechanisms.[249] The results of the thermodynamic characterization are used to rationalize potential intermediate states and elucidate similarities between distinct quadruplex topologies. Finally, differential scanning calorimetry is used to evaluate enthalpy changes for comparison with spectral studies.

Methods.

Oligodeoxynucleotides. Nucleic acid sequences, length, extinction coefficients and corresponding PDB names for the ht-quadruplexes are given in Table 11. The eight sequences correspond to high-resolution NMR studies demonstrating a predominant quadruplex topology in potassium solutions. Oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA) on a 1-μmol scale with standard desalting. Lyophilized DNA was dissolved at a concentration of 1 mM in 10 mM tetrabutylammonium phosphate, 1 mM EDTA, pH 7.0 (folding buffer) and stored at 4 °C. All concentrations were calculated from 260 nm absorbance data using nearest-neighbor extinction coefficients.[251] Because the extinction coefficients reflect the single-stranded state, normalized CD values may contain error estimated to be ±15% due to DNA concentration.[252] When stated, annealing was completed by placement of DNA into pre-boiling water for 20 minutes, followed by the removal of heat and cooling to room temperature over night. All reagents were purchased from Sigma-Aldrich (Sigma) at the highest available purity. All plots were prepared in Origin 7.0 software.

Circular Dichroism and Thermal Denaturation Protocol. Thermal denaturation of quadruplexes was monitored using a Jasco J-810 spectropolarimeter (JASCO, USA) equipped with a programmable Peltier thermostatted cell holder and electronic stirrer. Sample preparation and CD parameters were described previously.[249] Background CD and absorbance spectra were recorded using 25 mM KCl in folding buffer prepared by addition of a 1M KCl-folding buffer solution. The relatively low potassium concentration was chosen to ensure quadruplex structures are fully folded in solution,[121] while allowing the full thermal denaturation profile of the quadruplex system to be investigated.

Table 11. Human telomeric quadruplex forming deoxyoligonucleotides. The ht-quadruplexes studied in this work are given with their PDB designation and corresponding sequence. The hybrid conformations are grouped based on loop arrangement into hybrid-1, hybrid-2, and hybrid-3. *Tel22 sequence corresponds to K^+-crystal structure 1KF1 and Na^+ structures 143D. Tel22 is believed to be a mixture of anti-parallel quadruplexes in aqueous K^+ solutions.[72,79]

PDB ID	Sequence (5' to 3')	Type	ε_{260} (M^{-1} cm^{-1})	Length
2GKU[65]	TTGGGTTAGGGTTAGGGTTAGGGA	hybrid-1	244300	24
2HY9[64]	AAAGGGTTAGGGTTAGGGTTAGGGAA	hybrid-1	278200	26
2JSM[67]	TAGGGTTAGGGTTAGGGTTAGGG	hybrid-1	236500	23
2JPZ[66]	TTAGGGTTAGGGTTAGGGTTAGGGTT	hybrid-2	261200	26
2JSL[67]	TAGGGTTAGGGTTAGGGTTAGGGTT	hybrid-2	253100	25
2KF8[68]	GGGTTAGGGTTAGGGTTAGGGT	hybrid-3	223500	22
2KKA[69]	AGGGTTAGGGTTAGGGTTAGGGT	hybrid-3	237000	23
Tel22*[56,61]	AGGGTTAGGGTTAGGGTTAGGG	Mixture	228500	22

Thermal denaturation samples were prepared at 3-4 uM in 25 mM KCl-folding buffer using a 1 cm path length quartz cuvette. CD spectra were collected using instrumental parameters: 220 < λ < 350 nm, 1.0 nm step size, 200 nm/min scan rate, 1.0 nm bandwidth, 2 s integration time, with 4 total scans averaged. For melting experiments, samples were equilibrated in the cuvette holder at 4 °C prior to starting the melt. Melts were collected in duplicate on successive days with data presented from the second melt. Melting experiments were collected with the additional thermal parameters: 4 °C/min ramp, 0.05 °C equilibration with a 60 s delay prior to acquisition. CD data were normalized to $\Delta\varepsilon$ based on DNA concentration using Equation 5.

Isothermal CD samples were prepared at 3-4 uM in 25 mM KCl-folding buffer followed by annealing. For experiments that contained acetonitrile, samples were prepared by dilution into 50% (v/v) acetonitrile in 25 mM KCl-folding buffer. All acetonitrile containing samples were annealed prior to the first melting experiment. Reference scans were collected using 50% (v/v) acetonitrile 25 mM KCl-folding buffer in the absence of DNA.

NMR sample preparation and spectroscopy. NMR samples were prepared by dilution of lyophilized DNA to 0.4-0.5 mM using folding buffer. Once fully dissolved samples were brought to 25 mM KCl and annealed. D_2O and DSS were added to final concentrations of 10% and ≈30 μM, respectively. Samples were loaded into 5 mm Shigemi NMR tubes. NMR spectra were recorded at 20 °C using a 5 mm inverse triple resonance (HCN) probe on a Varian Inova spectrometer at 14.1 T using a cold probe. Diluted NMR samples gave CD spectra matching low concentration samples, indicating the high concentration used did not influence conformation.[253]

Singular Value Decomposition. Thermal denaturation results were used to construct three-dimensional data matrices as described previously.[202] The three-dimensional data matrix contains elements $D_{i,j}$, where each row i corresponds to a melting profile at λ_i and each column j represents a CD spectra collected at temperature T_j. The matrix **D** was analyzed by SVD, generating three additional matrices such that $\mathbf{D} = \mathbf{USV}^T$.[83] Matrix **S** is a diagonal matrix whose elements represent the weight of each spectral component. **U** is a matrix of basis spectra for each spectral component. **V** is a matrix of amplitude values corresponding to a spectral component's change with temperature. SVD was calculated using built-in algorithms in *Mathematica 8.0* (Wolfram Research, Champaign, IL).

SVD represents a model-free interpretation of the spectroscopic results capable of extracting information on the presence of significant spectral species prior to the application of analytical models. Autocorrelations values were calculated for all **V** matrices using Equation 13, where $X_{j,i}$ and $X_{j,i+1}$ are the *ith* and *ith+1* row elements from column j.

$$C(X_i) = Sum(X_{j,i}X_{j,i+1}) \quad \text{(Equation 13)}$$

V-vector autocorrelations (V-autocorrelations) are a measure of the apparent noise of each V-vector. A decrease in V-autocorrelation magnitude indicates a loss in confidence that the V-vector is significant to the spectral changes observed.

Non-linear Regression Analysis. Five unique quadruplex unfolding mechanisms were recently investigated.[249] These mechanisms included both sequential unfolding and parallel-species unfolding with an increasing number of intermediate species. The mechanism proposed to best described the unfolding of ht-quadruplexes was a four-species sequential mechanism, in which a transition from folded to unfolded occurs through two intermediate states (Scheme 1,

Mechanism 4). The present study additionally investigates the thermodynamics involved in the melting of ht-quadruplexes in the presence of 50% (v/v) acetonitrile. A kinetic and thermodynamic study of the unfolding of ht-quadruplexes under cosolvents conditions proposed three significant species are required for accurate modeling.[189] A mechanistic investigation was conducted using a three-state sequential mechanism (Scheme 1, Mechanism 3) and parallel unfolding mechanism (Scheme 1, Mechanism 2).

Each mechanism was used to describe the unfolding process by analytical models (Appendix C, Analytical Models). The analytical models were fit using non-linear regression that included the simultaneous minimization of four S-value weighted V-vectors. Analysis resulted in a set of signal amplitude values for each V-vector with globally determined melting temperatures (T_m1-3) and enthalpy changes (ΔH1-3). Signal amplitude values are used to construct generated CD spectra for each species involved in the thermal denaturation process. Generated spectra are mechanism-dependent such that $\mathbf{CF}^T = \mathbf{D} = \mathbf{USV}^T$, where $\mathbf{C}$ is a matrix whose columns contain the isolated spectra of each species, and $\mathbf{F}$ is a matrix of titration profiles from a pure species recovered from non-linear regression analysis.[83] Best-fit parameters and χ^2 values were incorporated into Monte Carlo simulations as described by Saroff for investigation of parameter uncertainty.[205]

Differential Scanning Calorimetry. DSC was conducted using a MicroCal VP-DSC (MicroCal info). The calorimeter was thermally equilibrated through sequential scanning of buffer blanks prior to sample scans. Samples were prepared at 0.25-0.3 mM strand concentration in 25 mM KCl-folding buffer with annealing. DSC sample CD spectra were found to match low concentration samples. Samples were scanned successively from 4 – 98 °C using instrumental parameters of 30 °C/hr, 10 min pre-scan equilibration, with medium gain.

Thermograms were exported to Origin 7.0 software for baseline corrections and further

analysis.

Reproducible Basis Set of Human Telomeric G-quadruplexes. All ht-quadruplex sequences used in this work have been shown to form intramolecular (single-stranded) structures. The sequences can be grouped based on the connecting loop regions into three distinct topologies: hybrid-1 (2GKU, 2HY9, 2JSM), hybrid-2 (2JPZ, 2JSL), and hybrid-3 (2KF8, 2KKA). Figure 33 shows concentration-normalized CD spectra for each ht-quadruplex recorded at 20 °C. The CD spectra have been reproducibly observed using multiple syntheses of the deoxyoligonucleotides, but are sensitive to changes in potassium concentration and temperature. Ht-quadruplex CD spectra are consistent with hybrid conformations,[254] but demonstrate a variety of subtle differences. All ht-quadruplexes give CD spectra with two regions of positive ellipticity near 265 and 290 nm, with the higher intensity at 290 nm. Although three distinct topologies may be present in the set of quadruplexes studied, CD spectra do not contain distinguishable differences to evaluate quadruplex conformation.

The reproducible CD spectra along with their ^{1}H NMR are shown in Figure 34. The ^{1}H NMR spectra can be used as an informative measurement of quadruplex conformation. The ht-quadruplex imino-region ^{1}H NMR agree with published high-resolution topologies for each sequence. Broad peak widths and the presence of smaller intensity peaks indicate the presence of minor amounts of additional conformations in solution, accounting for approximately 15-20% of the total signal. Interestingly, this heterogeneity may reflect the sensitivity of quadruplex conformations to temperature, as will be shown through CD thermal denaturation experiments.
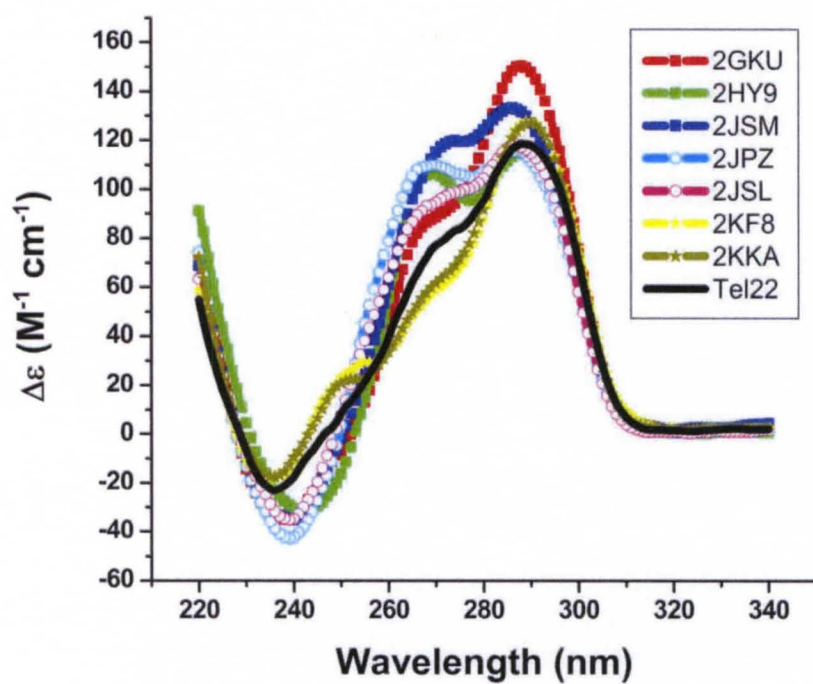
Figure 33. Concentration-normalized CD spectra for the eight ht-quadruplex sequences. CD

spectra were collected in the presence of 25 mM KCl in folding buffer held constant at 20 °C. CD

spectra are grouped by conformation: hybrid-1 (squares), hybrid-2 (circles), and hybrid-3 (stars).
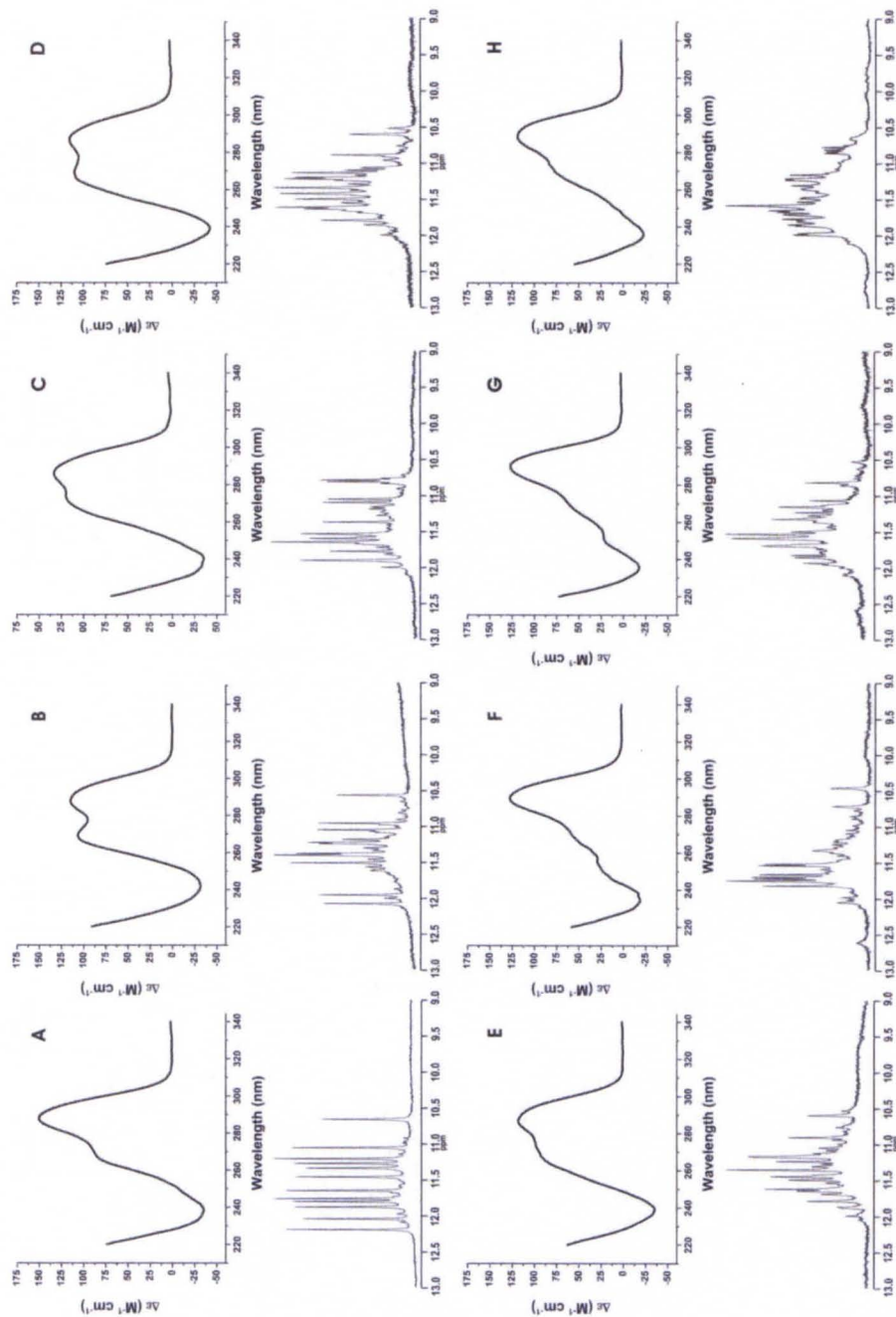
Figure 34. Concentration-normalized CD spectra and corresponding ^{1}H NMR for the eight ht-quadruplexes. Normalized CD spectra and ^{1}H NMR spectra were reproducibly obtained using the low-potassium slow-annealing procedure described. Letter designations are (A) 2GKU, (B) 2HY9, (C) 2JSM, (D) 2JPZ, (E) 2JSL, (F) 2KF8, (G) 2KKA, (H) Tel22.

Unfolding of ht-quadruplexes requires a non-two-state mechanism. Unfolding of ht-quadruplexes was monitored using CD spectra collected over the temperature range 4 - 94 °C in 25 mM KCl-folding buffer. A representative set of primary CD results for the thermal denaturation of the hybrid-2 sequence 2JSL is shown in Figure 35A. Temperature-dependent CD spectra demonstrate a decrease in 290 nm intensity concurrent with increased intensity at 265 nm prior to denaturation to a CD spectrum resembling G-rich single-stranded DNA. This is visualized by monitoring the two major peaks, 265 and 290 nm, against temperature (Figure 35B). Multi-wavelength CD intensity demonstrates a non-linear correlation (Figure 36), indicating unfolding is not a two-state process.[254] Derivative plots of the change in $\Delta\varepsilon$ with respect to temperature for 260 and 295 nm show three unique maximum/minimum values related to distinct melting transitions (Figure 35C). The presence of multiple inflection points recovered from the melting profile indicates the presence of distinct folding states during the thermal denaturation process. The derivative-derived melting temperatures for the ht-quadruplex are summarized in Table 12, and were used as initial estimates for non-linear regression analysis.

SVD Spectral Species and V-Autocorrelations. Thermal denaturation of ht-quadruplexes was used to construct individual three-dimensional matrices of circular dichroism signal intensity based on changes in temperature and wavelength. The matrices were subjected to SVD to investigate the number of significant spectral species required to accurately model the unfolding process. V-autocorrelations were calculated and compared for each ht-quadruplex (Figure 37). V-autocorrelations are greater than 0.9 for the first three spectral species of all ht-quadruplexes. A slight drop in autocorrelation is seen for species 4; however, the loss in magnitude is less than 0.15 for six of the eight quadruplexes (Tel22, 2GKU, 2HY9, 2JSL, 2KF8, 2KKA) with all but one quadruplex (2JSM) having a V-autocorrelation magnitude greater than
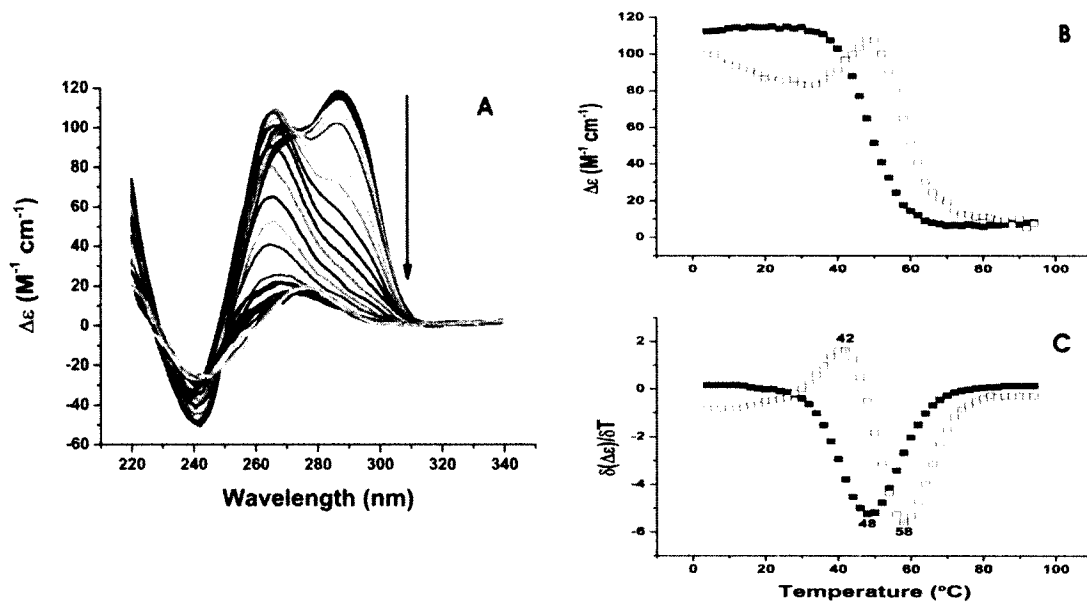
Figure 35. Three-dimensional thermal denaturation primary results of the hybrid-2 G-quadruplex forming sequence 2JSL. (A) Primary CD spectra. (B) Single wavelength analysis of 290 nm (filled squares) and 265 nm (open squares) demonstrate a non-two state behavior. (C) Derivative plots allow for the estimation of melting temperatures prior to SVD analysis and represent accurate starting values for non-linear regression.
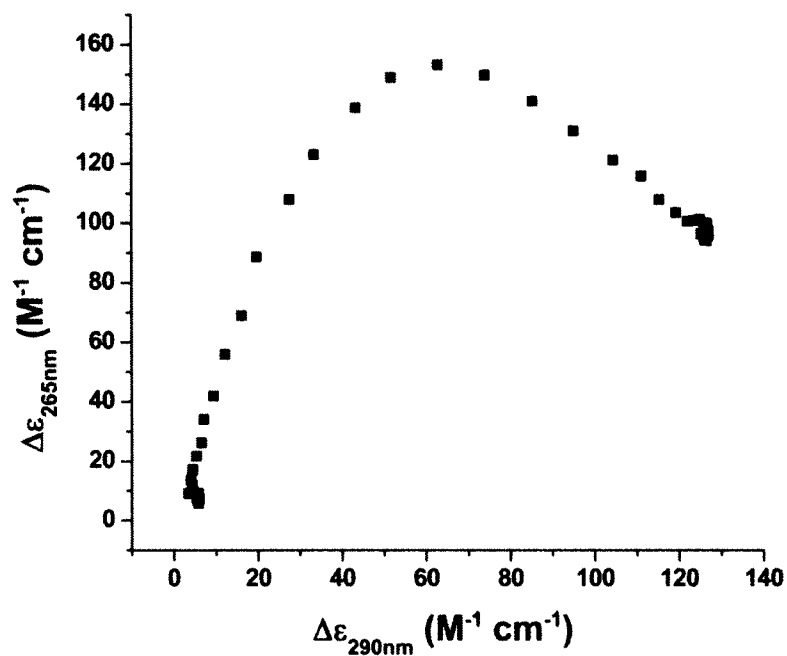
Figure 36. Non-linearity of 290 nm vs. 265 nm for the thermal denaturation of the hybrid-2

quadruplex forming sequence 2JSL. Non-linearity of two wavelengths collected simultaneously

during the thermal denaturation of the quadruplex structure indicates a complex non-two-state

model is required for accurately modeling the unfolding process.

Table 12. Melting temperature estimates from first derivative plots using 260 and 295 nm. Estimated melting temperatures are used as starting parameters during global non-linear regression analysis of the four-state sequential melting algorithm.

	T_m1 (°C)	T_m2 (°C)	T_m3 (°C)
Tel22	46	50	64
2GKU	44	54	62
2HY9	34	46	56
2JSM	44	50	62
2JPZ	38	46	56
2JSL	42	48	58
2KF8	48	62	66
2KKA	44	54	64
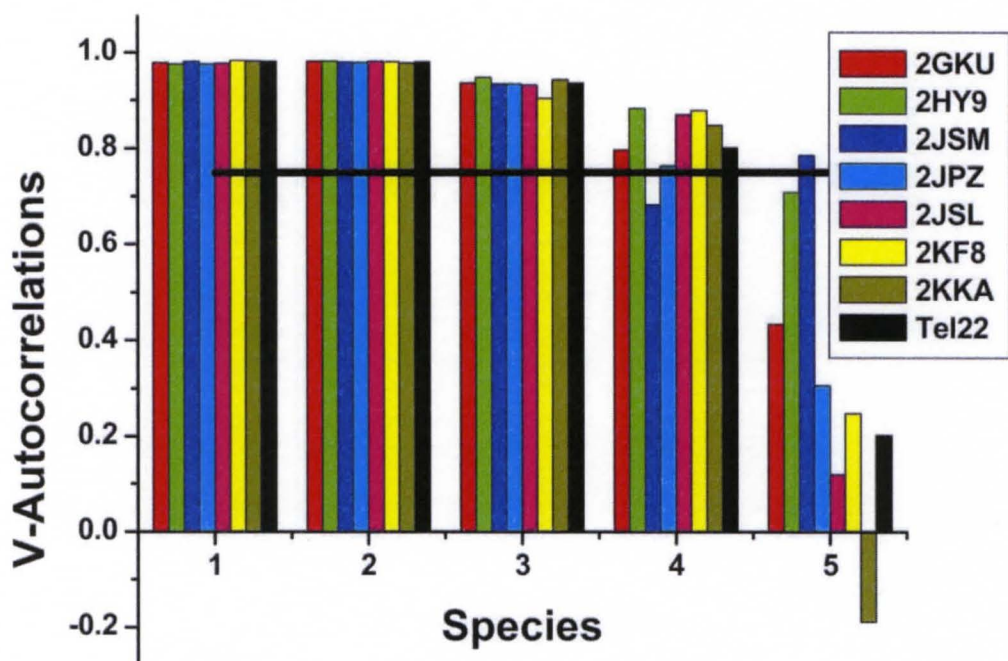
Figure 37. V-autocorrelations used to evaluate the presence of spectral species. A cutoff value of 0.75 (solid black line) was established for which a singular value species is above the noise threshold of the system. Comparison of V-autocorrelations for all eight ht-quadruplex forming sequences demonstrates four species are required for the accurate modeling of the unfolding process.

0.75. The fifth spectral species shows a drop in V-autocorrelation of 0.35 or greater for six of the eight quadruplexes study, indicating species beyond the first four do not contribute significantly to the spectral signal. V-autocorrelations from quadruplex thermal denaturations indicate the unfolding mechanism requires four unique spectral species.

Human telomere quadruplex unfolding thermodynamics. A four-state sequential mechanism, in which transition from a folded to unfolded state occurs through two distinct intermediate states, was recently proposed for the unfolding of ht-quadruplexes (Mechanism 4, Scheme 1).[249] The analytical model (Appendix C) was fit by simultaneous analysis of the first four V-vectors for accurate determination of thermodynamic parameters. Weights were implemented based on the significance of the V-vectors using their corresponding singular values (Figure 38). The best-fit residuals of global non-linear regression for the hybrid-2 forming G-quadruplex 2JSM are shown in Figure 39. Fit residuals demonstrate the first four V-vectors are well determined by the analytical fitting algorithm. Small spikes in the fit residuals are seen for V-vectors 3 and 4, corresponding to a high temperature melting species. However, because of the rigorous requirements discussed for species and model determination, there is no basis for introduction of a fifth species in an attempt to improve fit residuals.

Resulting best-fit parameters and Monte Carlo obtained error estimates for each ht-quadruplex are shown in Table 13. The eight ht-quadruplexes were accurately modeled by the global analysis of the first four weighted V-vectors, with only one quadruplex (2JPZ) displaying a large parameter uncertainty. Unfolding of the ht-quadruplexes occurs with three transitions, corresponding to melting temperature ranges of 33.8 - 44.2, 43.7 - 58.5, and 50.7 - 61.7 °C for the first, second, and third transitions, respectively. Although the sequences are treated with the same mechanism, global non-linear regression results define unique thermodynamic
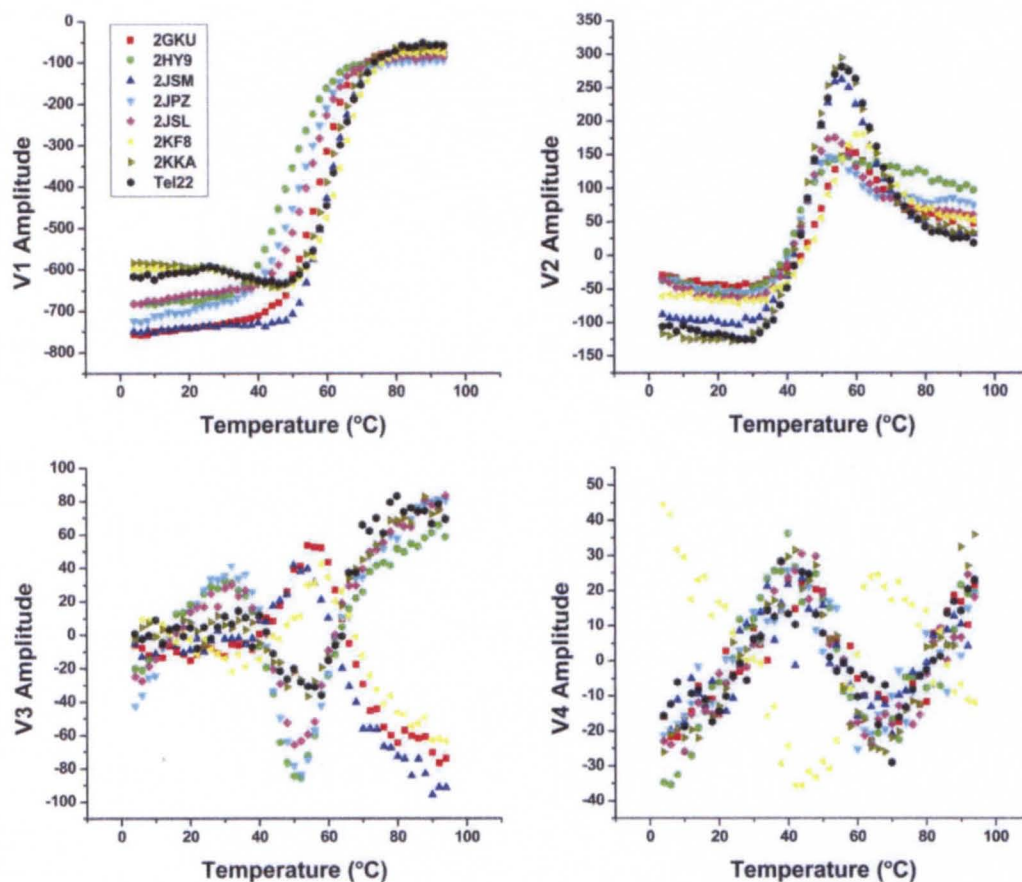
Figure 38. S-value weighted V-vectors for the first four spectral species of ht-quadruplex thermal denaturations. The first V-vector is weighted by the highest singular valued spectral species and resembles a thermal denaturation profile recovered from monitoring a single wavelength. The additional three V-vectors contain information necessary for extracting thermodynamic parameters for modeling three thermal transitions.
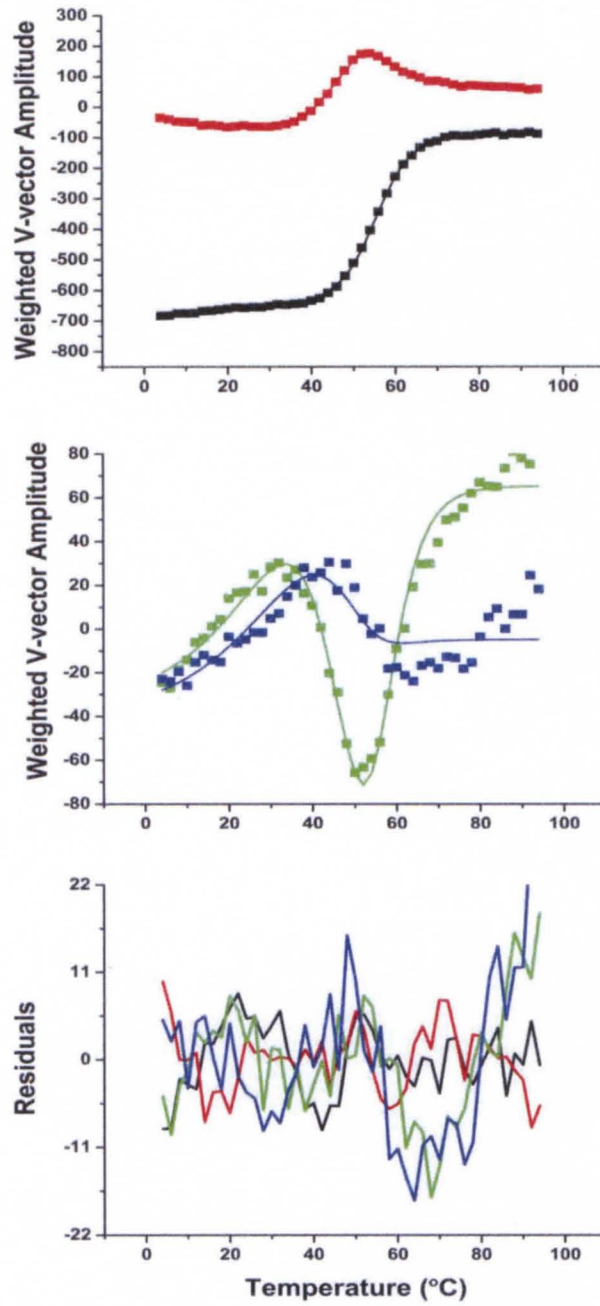
Figure 39. Best-fit residuals from the global analysis of four S-value weighted V-vectors. Best-fit

lines are shown for the first (black), second (red), third (green), and fourth (blue) V-vectors.

Table 13. Thermodynamic values for the thermal denaturation of eight distinct ht-quadruplex forming sequences. All thermodynamic parameters were determined from thermal denaturations conducted in 25 mM KCl TBAP buffer using equilibrated samples. All ΔH values are in kcal mol^{-1} and T_m values are in °C.

	ΔH_1	T_m1	ΔH_2	T_m2	ΔH_3	T_m3
2GKU	-21.1 ± 8.1	33.8 ± 6.6	-41.4 ± 6.2	53.8 ± 1.9	-55.8 ± 6.6	60.8 ± 1.2
2HY9	-12.6 ± 2.6	35.7 ± 8.6	-28.2 ± 4.0	45.7 ± 3.3	-29.8 ± 4.2	50.7 ± 3.7
2JSM	-19.1 ± 2.2	40.7 ± 7.3	-29.2 ± 2.1	48.8 ± 2.8	-41.6 ± 3.6	58.4 ± 1.3
2JPZ	-7.8 ± 1.8	37.5 ± 22.5	-36.4 ± 4.7	43.7 ± 2.9	-39.9 ± 4.8	54.3 ± 1.7
2JSL	-13.3 ± 2.6	39.2 ± 8.8	-30.7 ± 3.3	47.4 ± 3.2	-38.8 ± 4.4	52.9 ± 2.3
2KKA	-24.9 ± 4.2	42.0 ± 3.7	-22.9 ± 3.5	58.5 ± 5.2	-44.8 ± 4.8	61.7 ± 2.0
2KF8	-22.5 ± 3.8	44.2 ± 6.7	-31.6 ± 3.4	46.0 ± 2.7	-48.7 ± 4.3	60.8 ± 1.0
Tel22	-21.9 ± 4.7	35.2 ± 5.3	-33.4 ± 4.2	47.6 ± 1.6	-45.3 ± 4.2	61.4 ± 1.0

differences between the ht-quadruplexes.   The Gibbs free energy changes for the three

observed transitions are summarized in Table 14.   The initial transition demonstrates a relatively

low Gibbs free energy change (-0.4 to -1.7 kcal mol^{-1}) for all ht-quadruplexes.   The second and

third transitions occur with Gibbs free energy changes ranging from -2.26 to -4.30 kcal mol^{-1} and

-2.86 to -6.80 kcal mol-^{1}, respectively.   The average Gibbs free energy change for ht-quadruplex

unfolding is -8.8 kcal mol^{-1}, with most ht-quadruplexes being separated by less than 2.0 kcal

mol^{-1}.

Ht-quadruplexes have a unified unfolding mechanism with a common intermediate.

SVD analysis and Monte Carlo results using best-fit parameters were used to reconstruct the

melting process based on the four-state sequential unfolding mechanism. Analytical models

resulted in signal values used to generate CD spectra for each species of the unfolding process

(Figure 40). Reconstructed CD spectra were used for generalized inspection of quadruplex

conformational changes, but cannot be used to rationalize absolute quadruplex topologies.

Generated CD spectra show the starting and first intermediate species resemble hybrid-like

quadruplex conformations, displaying the highest ellipticity at 290 nm.[10] The second

intermediate species resembles the spectrum recently obtain for a G-rich triple-helical

structure.[255] The generated spectra for the second intermediate species of each ht-quadruplex

are shown in Figure 41. The recovery of similar CD spectra from all ht-quadruplexes indicates a

common intermediate state exists during the unfolding of ht-quadruplexes, regardless of

starting conformation. Although high-resolution structures have not been established for the

intermediate spectra, recent biophysical investigations suggest the CD spectra results from an

ensemble of triple-helix intermediate states.[249]

Table 14. Gibbs free energy changes for the thermal denaturation of ht-quadruplexes. All Gibbs free energy changes were calculated at 20 °C assuming no heat capacity change. All ΔG values were calculated at 20 °C in kcal mol^{-1}.

	ΔG_1	ΔG_2	ΔG_3	ΔG_{Tot}
2GKU	-0.9 ± 0.4	-4.3 ± 0.6	-6.8 ± 0.9	-12.0 ± 1.0
2HY9	-0.6 ± 0.2	-2.3 ± 0.5	-2.9 ± 0.7	-5.7 ± 0.5
2JSM	-1.2 ± 0.4	-2.6 ± 0.3	-4.8 ± 0.5	-8.7 ± 0.5
2JPZ	-0.4 ± 0.5	-2.7 ± 0.5	-4.2 ± 0.7	-7.3 ± 0.6
2JSL	-0.8 ± 0.3	-2.6 ± 0.3	-3.9 ± 0.7	-7.3 ± 0.6
2KKA	-1.7 ± 0.2	-2.6 ± 0.4	-5.6 ± 0.8	-9.9 ± 0.8
2KF8	-1.7 ± 0.4	-2.6 ± 0.4	-6.0 ± 0.6	-10.2 ± 0.7
Tel22	-1.1 ± 0.4	-2.9 ± 0.4	-5.6 ± 0.6	-9.5 ± 0.7
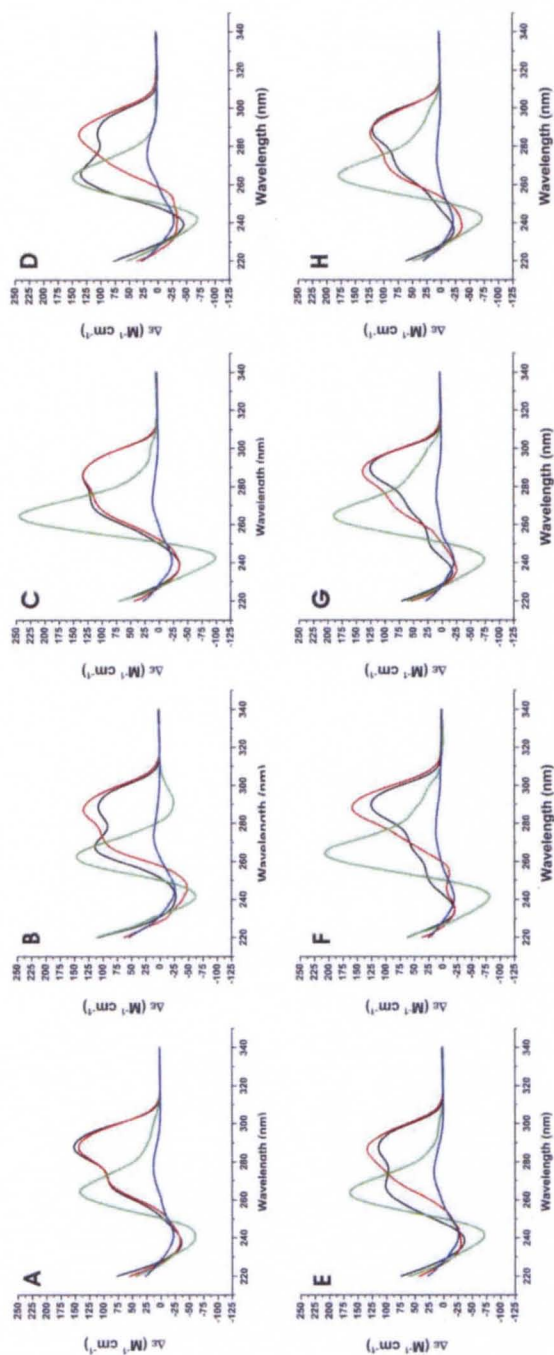
Figure 40. Generated spectra of the four species involved in the unfolding of human telomere quadruplexes in 25 mM KCl-folding buffer. Generated CD spectra correspond with the starting (black), first intermediate (red), second intermediate (green), and unfolded (blue) species. Letter designations are A) 2GKU, B) 2HY9, C) 2JSM, D) 2JPZ, E) 2JSL, F) 2KF8, G) 2KKA, H) Tel22.
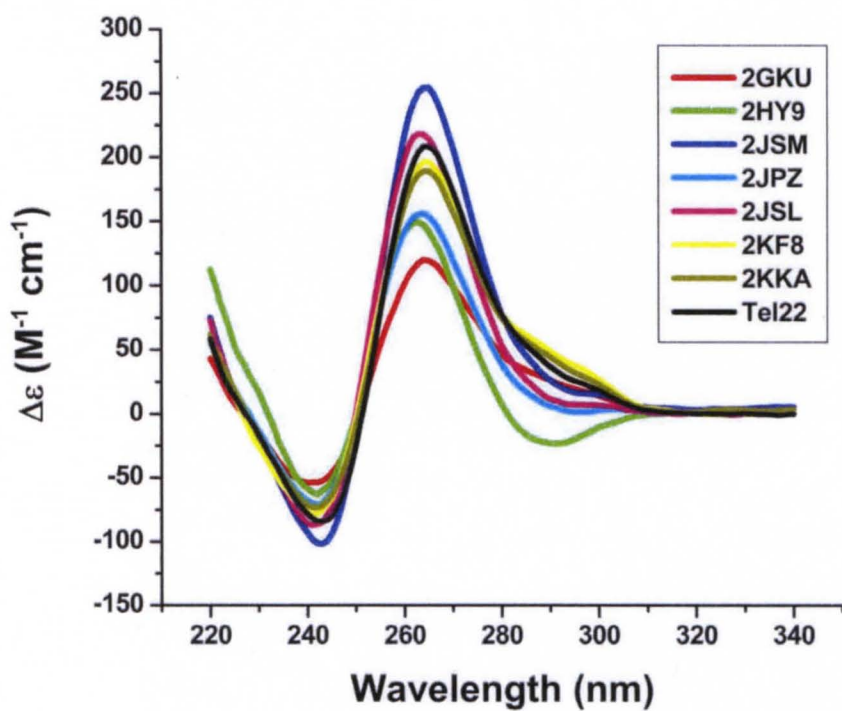
Figure 41. Generated spectra of the second unfolding intermediate recovered from mean values of Monte Carlo analysis. Generated spectra retain similar features independent of quadruplex conformation, having peak maximum and minimum of 265 and 245 nm, respectively. The largest uncertainly in the generated spectra are at the 290 nm shoulder (data not shown).

The relative concentration of each species and their dependence on temperature is shown in Figure 42. Modulation of relative concentrations for each species with increasing temperature is similar for all ht-quadruplexes. The relative concentration of each species was used to define an unfolding diagram based on the system of ht-quadruplexes investigated (Figure 43). The relative concentration diagram illustrates the unfolding of ht-quadruplexes, starting from an ensemble of quadruplex species that unfold through the presence of two distinct intermediate states. The average relative concentration diagram demonstrates a potential cause of quadruplex solution heterogeneity at 20 °C observed by NMR. The diagram contains nearly 15% of the first intermediate species at 20 °C, with almost equal concentrations of starting and first intermediate species near 37 °C. This heterogeneity results from two species with similar CD spectra (Figure 40). In summary, unfolding of ht-quadruplexes in the presence of 25 mM potassium requires two distinct intermediate species, with the second intermediate having a common CD spectra for all ht-quadruplexes.

Thermal denaturation of quadruplexes in 50% (v/v) acetonitrile. Ht-quadruplexes undergo a conformational change upon the addition of 50% (v/v) acetonitrile resulting in CD spectra with maximum and minimum intensities at 265 and 245 nm, respectively (Figure 44). The resulting CD spectra match recent investigations of ht-quadruplexes under dehydrating conditions,[79] and was recently determined by NMR to be a propeller topology in the presence of high concentrations of polyethylene glycol.[80] The results of thermal denaturation experiments for each ht-quadruplex in 50% acetonitrile were analyzed by SVD and V-autocorrelations were calculated (Figure 45). V-autocorrelations for 50% (v/v) acetonitrile melts suggest that the unfolding of the dehyrated quadruplex is simplified, having only three species with magnitudes greater than 0.75. Six of the eight quadruplexes demonstrate a 0.25 or greater drop in V-autocorrelation magnitude for species four and higher.

170
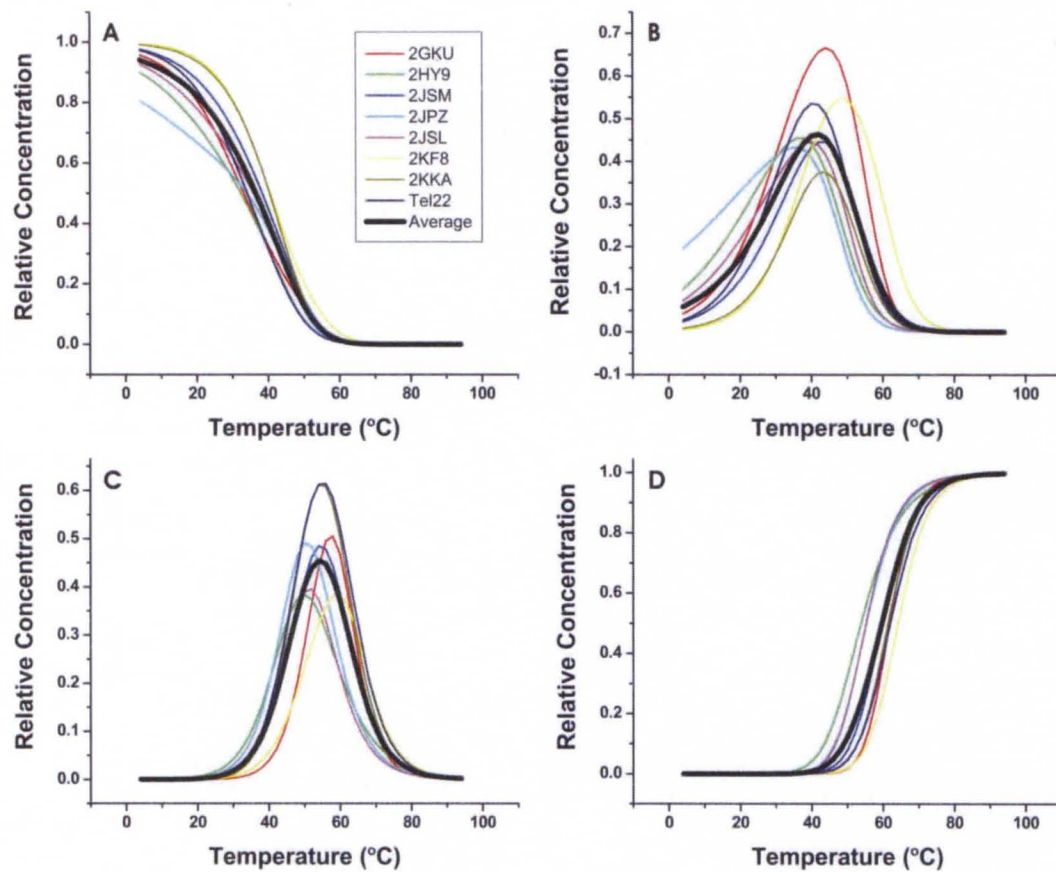
Figure 42. Relative concentrations of the A) first, B) second, C) third, and D) fourth species during the unfolding of the ht-quadruplexes. The relative concentrations from each quadruplex studied are shown behind the averaged curve (bold black line).
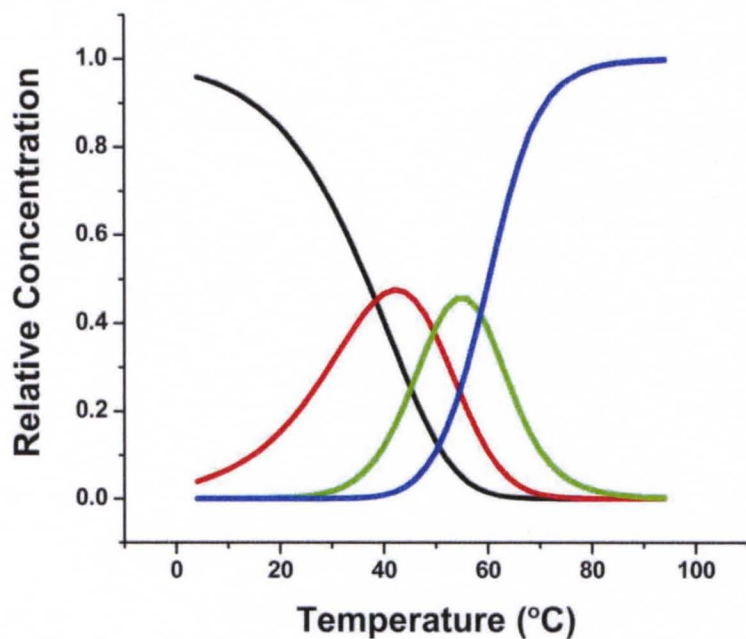
Figure 43. Relative concentration of each species present during the unfolding of human telomere quadruplexes. Concentrations are based on the average of eight human telomere quadruplex forming sequences used in the study. Diagram shows the relative concentration of each species with respect to temperature, illustrating the unfolding mechanism of human telomere quadruplexes. Lines represent the starting (black), first intermediate (red), second intermediate (green), and unfolded (blue) species.
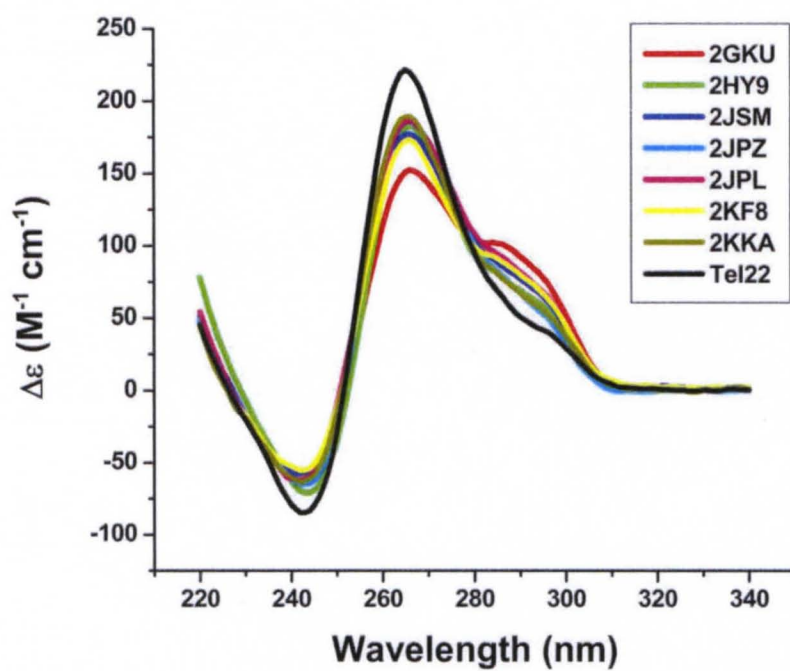
Figure 44. Comparison of quadruplex CD spectra in 50% (v/v) acetonitrile 25 mM KCl-folding

buffer. Conversion of the quadruplex conformation occurs upon addition of 50% (v/v)

acetonitrile to a CD spectra with a maximum peak at 265 nm and minimum at 245 nm.
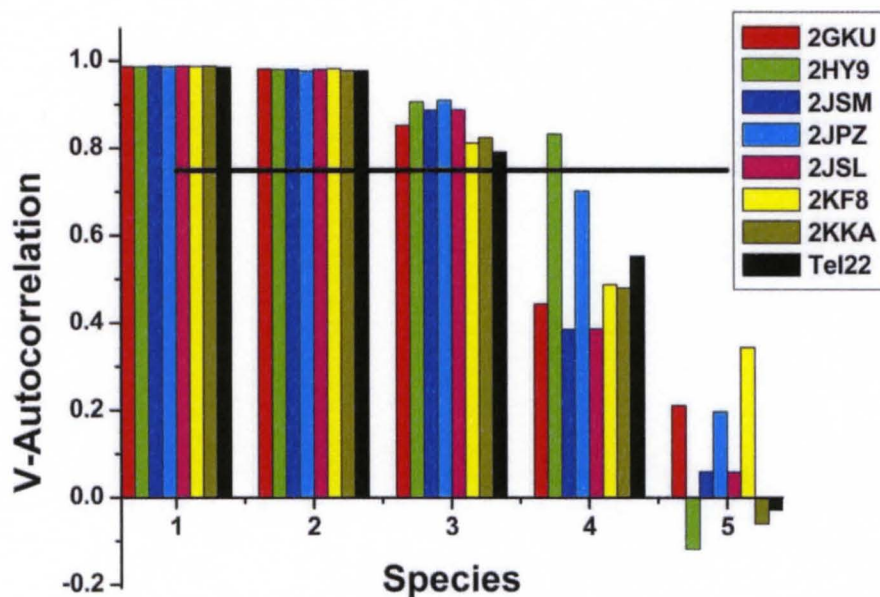
Figure 45. V-autocorrelations from SVD analysis of quadruplex thermal denaturations in the presence of 50% (v/v) acetonitrile. The solid black line represents the cutoff of 0.75. Autocorrelation analysis contains a dramatic drop in autocorrelation values for 4 or more species for all but one sequence (2HY9). This result demonstrates only three significant spectral species are required to accurately model the thermal unfolding of quadruplexes in the presence of acetonitrile.

Two mechanisms exist for the unfolding of the converted ht-quadruplexes in 50% acetonitrile: Mechanisms 2 and 3 (Scheme 1) describe the three-species sequential and parallel models, respectively. The parallel model differs from the sequential model by asserting multiple starting conformations are capable of formation, which melt independently to a common unfolded species. Table 15 evaluates the Gibbs free energy change calculated at 20 °C for each model. Both models give accurate fits to the data, but differ in average free energy changes by -10 kcal mol^{-1}. Because each model resulted in identical χ^2 values and have equivalent degrees of freedom, it is not possible to statistically differentiate the models based on F-statistic or Akaike Information Criterion.

We propose that the three-species sequential model (Mechanism 3, Scheme 1) best represents dehydrated ht-quadruplex unfolding based on two criteria. First, NMR studies indicate the presence of homogenous samples in the presence of co-solvents,[63] with sharp imino-region peaks with minimal background intensity. Second, the Gibbs free energy changes from the sequential model demonstrate a -4 kcal mol^{-1} increase in stability compared to aqueous results. Use of the parallel model would require an additional -14 kcal mol^{-1} energy be accounted for upon conversion of the ht-quadruplexes, a value that cannot be rationalized based on changes in conformation. Therefore, it is proposed that the unfolding of dehydrated ht-quadruplexes occurs through a three-state sequential model.

The thermodynamic parameters describing the unfolding process are shown in Table 16. Melting occurs by denaturation of the starting quadruplex structure to form a stable intermediate species. The intermediate CD spectra are similar regardless of sequence and resembles the second intermediate found from experiments done in purely aqueous environments (Figure 46). The first transition has a melting temperature range of 44.3 - 52.4 °C,

Table 15. Comparison of Gibbs free energy changes determined from a three-state sequential model and a three-state parallel model. All ΔG values were calculated at 20 °C in kcal mol^{-1}. *Results could not be accurately fit for a three-species parallel model.

	Sequential Model			Parallel Model		
	ΔG_1	ΔG_2	ΔG_{Tot}	ΔG_1	ΔG_2	ΔG_{Tot}
2GKU	-4.1	-9.7	-13.7	-13.7	-9.7	-23.4
2HY9	-3.5	-8.2	-11.7	-11.7	-8.2	-19.8
2JSM	-4.5	-8.6	-13.1	-13.1	-8.6	-21.7
2JPZ	-4.3	-8.2	-12.5	-12.5	-8.2	-20.8
2JSL	-4.5	-8.6	-13.1	-13.1	-8.6	-21.7
2KKA	-4.3	-9.5	-13.9	-13.9	-9.5	-23.4
2KF8	-4.4	-8.6	-13.0	-13.0	-8.6	-21.6
Tel22	-3.0	-7.4	-10.5	4.2*	-7.7*	-3.5*

Table 16. Thermodynamic parameters describing the unfolding of human telomeric

quadruplexes in the presence of 50% (v/v) acetonitrile. Parameters were obtained using a

three-state sequential unfolding model. All ΔH values are in kcal mol^{-1} and T_m values are in °C.

	ΔH_1	T_m1	ΔH_2	T_m2
2GKU	-42.9	50.7	-49.2	91.7
2HY9	-42.6	45.9	-45.3	84.8
2JSM	-54.2	46.2	-45.2	88.8
2JPZ	-55.9	44.3	-46.4	83.4
2JSL	-54.2	46.2	-45.2	88.8
2KKA	-43.6	52.4	-46.5	95.4
2KF8	-55.1	45.5	-42.1	95.5
Tel22	-39	44.6	-35.3	98.4
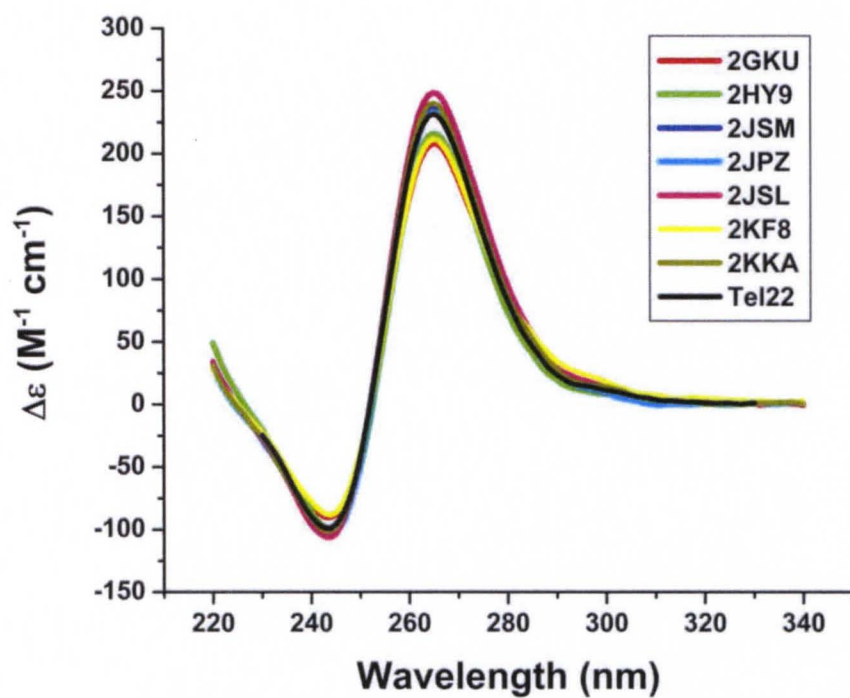
Figure 46. Comparison of generated CD spectra for the melting intermediates determined from

thermal denaturation of the ht-quadruplexes in 50% (v/v) acetonitrile.

similar to that of the first intermediate found from aqueous investigations, while the second

transition demonstrates elevated melting temperatures ranging from 83.4 - 98.4 °C. This

indicates the increase in ht-quadruplex stability may be through stabilization of the intermediate

state rather than the fully-folded quadruplex.

Assessing human telomere quadruplex heat capacity changes. The thermal stability of

human telomere quadruplexes was studied through the use of DSC to assess the heat capacity

changes associated with ht-quadruplexes. Each DSC thermogram was normalized for strand

concentration and baseline corrected using a cubic regression technique (Figure 47). Baseline

correction routines are often arbitrary, with techniques varying based on how endpoints are

determined and the order of the polynomial chosen for extrapolation. Baseline-corrected DSC

thermograms were integrated to evaluate total calorimetric enthalpy changes. Calorimetric

enthalpies are found to be much lower in magnitude compared to CD-derived van't Hoff

enthalpy changes (Table 17). If the chosen mechanism accurately reflects quadruplex unfolding,

calorimetric and van't Hoff enthalpy changes should agree. The significant disagreement

between the two techniques indicates a correction is required in the handling of DSC

thermograms and that heat capacity changes may significantly influence the thermodynamic

parameters.

To investigate this possibility, thermodynamic parameters determined from CD melting

were used to simulate a DSC thermogram for ht-quadruplex unfolding, and the difference

between the generated thermogram and experimentally collected thermogram was calculated

(Figure 48). The difference curve reflects the baseline required to achieve agreement between

CD and DSC results and contains a complex multi-phase behavior with three distinct changes in

magnitude. We propose that this difference curve reflects unique heat capacity changes that
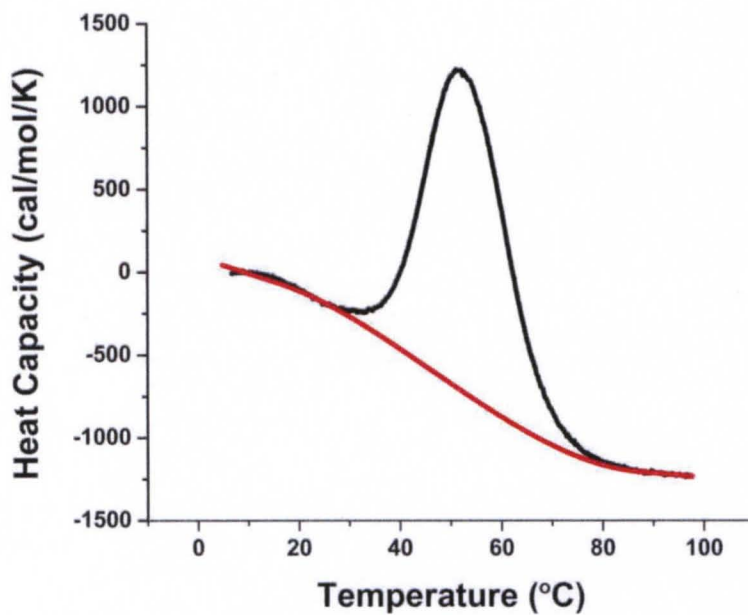
Figure 47. DSC baseline correction method. Buffer-blanked DSC results (black) are fit for a

baseline curve (red) by connecting denaturation endpoints using a cubic-spline function.

Table 17. Calorimetric and van't Hoff enthalpy changes for the unfolding of human telomeric quadruplexes.

	ΔH_{cal} (kcal/mol)	ΔH_{vH} (kcal/mol)
2GKU	-34.2	-100.6
2HY9	-23.8	-118.3
2JSM	-36.3	-82.9
2JPZ	-32.5	-70.7
2JSL	-37.4	-84.1
2KF8	-44.1	-89.9
2KKA	-36.4	-92.6
Tel22	-34.4	-102.8
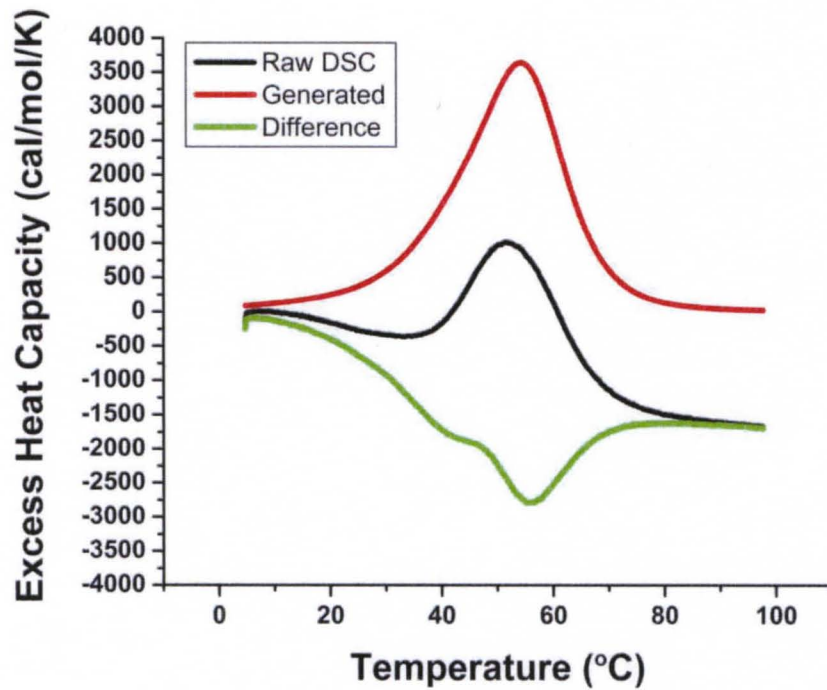
Figure 48. Comparison of buffer-blanked DSC thermogram for 2JSL and thermogram generated from global analysis of CD thermal denaturations. The difference curve reflects the baseline necessary to achieve agreement between the DSC and CD results, and may indicate the presence of three-distinct heat capacity changes occurring during the unfolding of ht-quadruplexes.

correspond to the three transitions required to accurately model ht-quadruplex unfolding.

Distinct heat capacity changes would require baseline corrections for each species, rather than

treating DSC thermograms for a single baseline correction. It is possible that heat capacity

changes could be obtained for each species involved in the unfolding mechanism, which could

lead to agreement between the thermodynamic parameters recovered from spectroscopic and

calorimetric techniques.

Discussion.

The thermodynamic characterization of ht-quadruplex thermal unfolding investigated here compliments the mechanistic study recently conducted by our laboratory.[249] Circular dichroism has been established as a key method capable of capturing the intermediate states involved in ht-quadruplex unfolding. SVD analysis of spectrally relevant components necessitates the use of a multi-state mechanism for accurate modeling of the unfolding mechanism. A four-state sequential mechanism is found to describe the unfolding of each of the ht-quadruplexes. This complexity is reduced in the presence of acetonitrile. This study represents the first detailed thermodynamic analysis of a set of diverse quadruplex topologies by complex unfolding mechanisms, expanding upon the recent hTel22 characterization.[249] The study highlights unique similarities upon the unfolding of distinct ht-quadruplex topologies, while parsing the free energy contributions to illustrate differences in the overall stability of each sequence. The thermodynamic parameters describing the unfolding of ht-quadruplexes can aid in the design of more specific quadruplex-interactive compounds and improve predictive algorithms for quadruplex stability.

Ht-quadruplexes represent a system of diverse topologies. The human telomeric system of quadruplexes contains a rich history of high-resolution investigations that includes distinct monomeric topologies in potassium solutions.[63-69] Investigated here was a set of eight quadruplex forming sequences adopted from the human telomeric repeat. The set of ht-quadruplexes represents a model system for characterizing similarities and differences between distinct quadruplex topologies. Spectroscopic techniques were employed to evaluate differences between the ht-quadruplex forming sequences under standardized conditions. CD signatures were reproducibly obtained for the ht-quadruplexes (Figure 33), but cannot be used

to establish solution structure. Resulting CD signatures from equivalent hybrid topologies display small variations in CD at two predominant wavelengths, 265 and 290 nm, with intensity that depends on sequence. The differences in peak intensities are reflections of both quadruplex conformation (differential stacking of loop bases) and the presence of additional flanking residues to the quadruplex forming sequence. CD variations may also reflect influences of unique base-pair or potential A-triplet capping structures.[64] The reproducible CD signatures were found to correlate with ^{1}H NMR spectra that match published high-resolution topology studies (Figure 34). Collected NMR spectra demonstrate minimal solution heterogeneity (<15%) for each of the sequences studied at 20 °C. No influence of solution structure was seen for strand concentrations up to 0.4-0.5 μM, although high strand concentration has been proposed to influence solution structure.[253]

Ht-quadruplexes are best modeled by a four-state sequential unfolding mechanism.
Thermal denaturation experiments were analyzed by multiple methods to establish a multi-state pathway that was required to accurately describe the unfolding mechanism.[202,225,254,256] Simultaneously collected wavelength data show non-linear correlations (Figure 36), while derivative curves ($\delta\Delta\varepsilon/\delta T$) demonstrated the presence of multiple unique minimum and maximum values (Figure 35). These two analysis techniques indicate that ht-quadruplex unfolding cannot be described by a two-state mechanism. The application of SVD, a model-independent deconvolution technique, resulted in V-autocorrelations indicating four significant spectral components are necessary to describe ht-quadruplex unfolding (Figure 37).

A four-species sequential mechanism was recently proposed for ht-quadruplex unfolding. Mechanisms containing two, three or four species unfolding sequentially or with formation of multiple starting species (parallel) were rigorously investigated for the Tel22 and

2GKU sequences.[249] The four-species mechanism accurately described the thermal denaturation process while resulting in minimized error surfaces based on Akaike Information Criterion (AIC) to account for increasing numbers of adjustable fitting parameters. Although AIC could not be used to determine statistical differences between the sequential and parallel mechanisms, evaluation of the Gibbs free energy changes for the parallel mechanism resulted in differences between multiple starting species that deviates from the proposed 1:1 equilibrium between hybrid conformations for the Tel22 sequence.[79,119] Thus, monitoring of ht-quadruplex unfolding through collection of CD intensity is likely not sensitive to the formation of multiple distinct quadruplex conformations, but instead senses changes in the topology of an ensemble of quadruplexes that occurs due to thermal disruption of the system.

The four-state sequential mechanism is proposed to be sensitive to the "breathing" of the quadruplex structure. Unlike previously proposed mechanisms where multiple quadruplex conformations interconvert rapidly in solution, the sequential mechanism asserts that quadruplex conversion can occur only upon reaching an intermediate or unfolded state. Therefore, the equilibrated solutions obtained by annealing refer to a preformed ensemble of quadruplexes, where the CD signatures reflect a weighted average of the quadruplex conformations. The first unfolding transition, $F \leftrightarrow I_1$, corresponds to a shift in state rather than multiple solution topologies. The first transition is likely a relaxing of the quadruplex conformation, with the generated CD spectra from SVD analysis reflecting changes in the stacking of the G-tetrad core. This change from a rigid to a relaxed state affects the chirality of the stacked G-tetrad core and potential weak interactions such as would be found in loop regions or capping structures. Disrupting the chirality results in small shifts in the CD signatures of the system.[10] This matches the changes seen from generated spectra that resemble hybrid quadruplex CD signatures for the initial folded and first intermediate states.

The second transition, $I_1 \leftrightarrow I_2$, was recently proposed to be a change from a quadruplex to a triple-helix conformation.[249] Spectroscopic investigation using CD and fluorescent techniques demonstrated the second intermediate state (I_2) is best characterized by a triple-helical conformation, with a recovered CD spectra that resembles a G-rich triple-helix. The second intermediate state for all ht-quadruplexes investigated is found to give CD signatures that resemble the triple-helical state (Figure 41), indicating the unfolding of ht-quadruplexes may contain similar triple-helix intermediates independent of starting conformation. However, the rigorous biophysical investigation conducted using the Tel22 to characterize the populated intermediate state has not been done for any of the additional ht-quadruplexes used in this study.

Thermodynamic characterization of ht-quadruplexes. A standard solution condition of 25 mM KCl-folding buffer, pH 7.0 was used to evaluate energetic differences between the eight ht-quadruplexes. Although the ht-quadruplexes have been characterized to adopt similar conformations, thermodynamic analysis resulted in a wide range of stabilities using the four-state sequential mechanism (Table 14). Overall, the ht-quadruplexes were found to have similar energetic contributions from the three transitions that occur upon unfolding, with each transition having an increase in the Gibbs free energy change magnitude such that $\Delta G_1 << \Delta G_2 < \Delta G_3$. The low energetic contribution of the first transition ($\Delta G_{1,ave}$ = -1.0 kcal mol^{-1}) indicates the ht-quadruplexes are unlikely to adopt fully-stacked rigid conformations, with small changes in temperature inducing thermal disruption to a more relaxed state. The second transition was found to have an increased free energy change ($\Delta G_{2,ave}$ = -2.8 kcal mol^{-1}), where removal of one of the terminal strands causes formation of the triple-helical state. The free energy change from the initial folded state to the second intermediate triple-helix state (F $\leftrightarrow$ I_2) contains a low total free energy change ($\Delta G_{1+2,ave}$ = -3.8 kcal mol^{-1}), agreeing with computational investigations that

the triple-helix state may have similar stability to the quadruplex.[248] Unfolding of the second intermediate requires the largest energetic change ($\Delta G_{3,ave}$ = -4.9 kcal mol^{-1}), having a larger magnitude than the first two transitions combined. This large energetic change corresponds with the total disruption of structured DNA, where the folded triple-helical state unfolds to a random-coil single stranded state.

Although the ht-quadruplexes contain similarities in their energetic contributions, the overall stability of the quadruplexes show significant variations. The ht-quadruplex Gibbs free energy changes ordered from lowest to highest are 2HY9 < 2JPZ < 2JSL < 2JSM < Tel22 < 2KKA < 2KF8 < 2GKU. The ht-quadruplexes have an average Gibbs free energy change of -8.8 kcal mol^{-1}. Differences in the total Gibbs free energy changes are proposed to arise from unique flanking residues and differences in stabilizing interactions such as capping structures. The total Gibbs free energy change correlates well with sequence length, with longer sequences having decreased total Gibbs free energy changes. The longest sequence (2HY9) is highly destabilized in comparison with the other ht-quadruplexes, and contains the addition of five adenine residues to the core 21-nt quadruplex forming sequence. This result agrees with a study that showed addition of adenine residues to the 3'-end of the Tel22 decrease stability.[257]

Cosolvent-converted ht-quadruplexes display reduced unfolding complexity. The ht-quadruplexes were also studied in the presence of 50% (v/v) ACN to evaluate changes to the unfolding thermodynamics caused by dehydration. All ht-quadruplexes are found to convert to a CD spectrum having a maximum intensity near 260 nm. This conversion has been shown to be a change in conformation to a propeller quadruplex topology in the presence of PEG.[80] SVD analysis of thermal denaturation profiles resulted in V-autocorrelations demonstrating reduced unfolding complexity, with only three significant components required to describe the unfolding

process. The reduction in unfolding complexity is associated with a loss of the first heterogeneous step discussed for the four-state sequential mechanism, with the converted ht-quadruplexes not displaying low-temperature variations in conformation.

A three-state sequential unfolding mechanism was chosen to describe the unfolding of ht-quadruplexes in the presence of acetonitrile based on Gibbs free energy changes (Table 15). The sequential mechanism demonstrates an average Gibbs free energy change of -12.7 kcal mol^{-1}, a difference of -4 kcal mol^{-1} from results determined in the absence of acetonitrile. The initial transition occurs with an average Gibbs free energy change of -4.1 kcal mol^{-1}, similar to the -3.8 kcal mol^{-1} recovered from combination of the first two transitions in an all aqueous environment. The stabilization of the ht-quadruplexes is through a large increase in Gibbs free energy change for the intermediate state. The intermediate state is found to have an average Gibbs free energy change of -8.6 kcal mol^{-1}, compared to the -4.9 kcal mol^{-1} recovered from aqueous results. Although rigorous biophysical investigations have not been conducted, the intermediate CD spectra recovered from SVD analysis is similar to the recovered CD spectra from aqueous investigations that is proposed to be a triple-helical conformation. Overall, ht-quadruplexes are found to have increased stability upon dehydration by acetonitrile with a significant contribution to the increased stability coming from the intermediate state.

Heat capacity change of ht-quadruplexes. Estimates of the heat capacity change for ht-quadruplex unfolding are required for a full thermodynamic characterization. DSC was employed to evaluate the heat capacity changes of each of the ht-quadruplexes, as calorimetric techniques are most sensitive to heat capacity effects.[240] Evaluation of DSC thermograms using a common baseline correction technique resulted in a large discrepancy in the total enthalpy

change compared to the values recovered by global SVD analysis (Table 17). This is a consequence of the uncertainty in heat capacity changes for ht-quadruplex systems.

Recent combined calorimetric and spectroscopic studies employed global CD and DSC fits to estimated the total heat capacity change for the folding of Tel22 to be -310 cal mol^{-1} K^{-1}.[208] Problems with this methodology are that baseline corrections are done prior to global analysis, where baselines are chosen based on end point values and regression done typically through the use of polynomial splines (Figure 47). If heat capacity changes are present for each transition in the unfolding mechanism, it may be that the use of polynomial regressions for baseline correction influences the estimated total heat capacity change. It may also be true that each transition has a unique heat capacity change, requiring the baseline correction to contain multiple distinct transitions as is observed by evaluating the difference curves between CD and DSC recovered in this study (Figure 48). At this time, the heat capacity change for ht-quadruplex system is still uncertain, with estimates ranging from 310 to 1600 cal mol^{-1} K^{-1}.[208,243] While agreement is being reached that ht-quadruplex unfolding must be analyzed for multi-state transitions, it will be necessary for detailed calorimetric studied to be employed in an attempt to accurately assess heat capacity changes.

Conclusions.

Human telomeric quadruplex thermal unfolding cannot be described by a two-state mechanism. Under aqueous and dehydrated conditions, sequential mechanisms with one or two intermediate species accurately describe the unfolding of three unique hydrated and one dehydrated topology. This indicates sequential unfolding mechanisms may be common for biologically-relevant unimolecular quadruplexes, although deconvolution of melting profiles for other distinct quadruplexes systems is required. Additionally, thermal deconvolutions indicate a common intermediate spectral species during human telomeric quadruplex unfolding. Similar CD spectra recovered from eight distinct human telomeric sequences, under both hydrated and dehydrated conditions, indicate these intermediate are relevant in quadruplex unfolding. Recent biophysical characterization of Tel22 suggests this intermediate is an ensemble of triple-helical conformations. The thermodynamic description of these intermediate states improves the interpretation of the dynamic quadruplex equilibrium exhibited by telomeric sequences. Whether such intermediates could be of biological relevance or used as small-molecule targets remains to be seen.

CHAPTER 7

DISSERTATION CONCLUSIONS.

Telomere quadruplexes are capable of forming at least five distinct unimolecular conformations. Multiple ht-quadruplex conformations suggest tertiary structure may have potential effects on biological functions. There exist several proteins with interactions specific for quadruplexes that demonstrate regulatory effects. These include helicases involved in maintaining efficient replication through quadruplex-rich motifs[36-40] and translation initiation complexes that interact with quadruplex-containing IRES[48]. It has also recently been shown that frequently used origins of replication are associated with DNA sequences capable of quadruplex formation, with over 90% of the origins corresponding with PQSs.[258] This indicates that quadruplex structures may be important for origin recognition in human cells.[259] Unique quadruplex tertiary structures may therefore potentially serve biological functions. Developing a biophysical description of quadruplex structures is critical for understanding potential biological influences.

This dissertation provides a multi-faceted biophysical description of human telomeric quadruplexes. It contains the first biophysical description of the influence of co-solvents on quadruplexes and demonstrates conformation can be influenced through linked equilibria with binding. These results are biologically relevant because they serve as a model of protein induced conformational changes that may induce regulation of quadruplex functions. This dissertation also provides the first in-depth spectroscopic deconvolution of the thermal unfolding of human telomeric quadruplexes. Complex thermal unfolding pathways are

192

revealed and help to obtain accurate estimates of the stability of quadruplex structures and the

influence of sequence and tertiary structure. The collection of energetic values for complex

unfolding mechanisms also advance studies evaluating helicase activity on quadruplex

structures and how quadruplexes can be maintained in single-stranded states by protein

complexes. The models developed in this dissertation also result in unique intermediate states

that may become populated during enzymatic unfolding and could serve as important

recognition motifs. These newly clarified intermediate states may also be used as new potential

structures for general targeting of human telomere quadruplexes. The work conducted here

contains essential biophysical information for the description of human telomeric quadruplexes.

The dissertation was broken down into three specific aims addressing the central hypothesis:

> G-quadruplexes are polymorphic in solution. A dynamic equilibrium exists
> among an ensemble of structures that is thermodynamically linked to solvent
> composition and ligand binding. Investigation of G-quadruplexes through
> biophysical techniques will provide essential thermodynamic information about
> G-quadruplex polymorphism.

Specific Aim I included the investigation of quadruplex conformation through time-

resolved fluorescent methodologies. Time-resolved fluorescence is sensitive to conformational

transitions and can be simplified by the model-independent phasor diagram technique (Chapter

2). Phasor diagrams are capable of the sensitive detection of both quadruplex folding and

unfolding through time-resolved measurements of the complex multiple-exponential lifetime of

2-aminopurine. This specific aim was addressed using phasor diagrams to evaluate human

telomeric quadruplex cation-dependent structural changes and folding mechanisms.

Quadruplex folding monitored by phasor diagrams indicated a unique conformational transition

at high potassium concentrations. This type of transition may indicate differences in folding

topologies that occur through additional potassium binding sites. Differences in bound

potassium ions may play a role in stabilizing larger quadruplex structures, as can be adopted by eight or more repeats of the human telomere sequence. Such structures may be important for stabilizing telomere ends and protecting chromosomal termini from degradation.

Probing the structural characteristics of quadruplexes can be further improved by the application of additional time-resolved techniques. Methodologies such as time-resolved FRET may be capable of capturing the flexibility of quadruplex structures and provide potential resolution of multiple quadruplex conformations, while time-resolved anisotropy measurements could be used to improve the biophysical description of quadruplex structures in solution. Quadruplex structural characteristics can be further defined by the application of additional time-resolved fluorescence methods.

A significant focus of this dissertation was the ability of co-solvents to perturb the human telomeric quadruplex equilibrium. The cellular interior is highly crowded, with macromolecules accounting for approximately 30% of the total volume in eukaryotic cells,[166,167] and presents an environment where water activity less than unity. This dissertation provides thermodynamic studies mimicking cellular effects to address the energetic contributions that define the dynamic behavior of quadruplexes. Specific Aim II showed that the propeller conformation has an unfavorable free energy in aqueous environments and is less hydrated than the hybrid conformations (Chapters 3 and 4). Additionally, the conversion of ht-quadruplexes by PEG was shown to be inconsistent with a molecular crowding mechanism. Results suggest a conformational selection model that includes the differential binding of PEG to the propeller form. This specific aim clarified the role of co-solvents in ht-quadruplex conversion making the recent postulate that the propeller conformation is the most-likely biological form[80] based on PEG studies unmerited.

The work done in co-solvent solutions demonstrates the sensitivity of quadruplex structures to solvent composition, with dehydration by osmotic stress or differential PEG binding causing the equilibrium to favor the propeller conformation. Specific Aim II addressed the influence of solution composition on quadruplex conformation and helps define the energetic contributions that could influence quadruplex structures under biological settings. The results of this specific aim show that quadruplex conformation is strongly influenced by binding and solution conditions. Conformational changes may be important for cell-cycle dependent effects where cellular conditions may influence the tertiary structure of quadruplexes. Additionally, the ability to influence quadruplex conformation through PEG binding can serve as a model for protein interactions providing energetic contributions for conformational selection. Conformational selection mechanisms may serve important regulatory functions. Conversion of quadruplex structures to the all-parallel form may serve as a mechanisms for condensation of the telomeric overhang, where tight packing of propeller conformations serves to protect chromosomal termini or act as a signaling mechanism.

This dissertation also resulted in a thermodynamic description of the thermal unfolding of ht-quadruplexes. Specific Aim III evaluated the thermal unfolding of ht-quadruplexes monitored by several biophysical techniques to obtain an essential thermodynamic description of quadruplex stability. A mechanistic study based on SVD deconvolution of the denaturation spectra was used to show that ht-quadruplex thermal unfolding is best modeled by a four-species sequential mechanism in aqueous solutions (Chapter 5). This mechanism was found to accurately predict the thermal unfolding of ht-quadruplexes for eight human telomere DNA sequences known to form three distinct quadruplex topologies (Chapter 6). It was also demonstrated that the unfolding of ht-quadruplexes in 50% (v/v) acetonitrile has reduced complexity, requiring only a three-species sequential model. These models must be extended to

other quadruplex systems, such as promoter and RNA quadruplexes, to evaluate if unfolding complexity can be generalized for quadruplex systems. Providing accurate mechanisms for folding is essential for defining quadruplex stability. Specific Aim III was address by proving important standards for the energetic stability of quadruplexes that is necessary for defining how enzymatic systems act to destabilize and maintain quadruplex structures in single-stranded states. Such mechanisms are critical for enzymatic actions such as the sequestering of TTAGGG repeats by the DNA-binding protein POT1, whose action is believed to promote activity of the enzyme telomerase.[53]

Specific Aim III also provided a biophysical characterization of the unique spectral species apparent during the thermal denaturation process. One particular intermediate species was found through CD, fluorescent emission, and FRET measurements to be best characterized by an ensemble of triple-helical structures (Chapter 5). CD thermal denaturation results based on an array of ht-quadruplex sequences suggested this intermediate species may be common for the thermal unfolding of ht-quadruplexes. Similar CD spectra for the intermediate species were generated from denaturation of eight ht-quadruplex sequences under both aqueous and dehydrated conditions (Chapter 6). The thermodynamic characterization of ht-quadruplexes has provided essential information on its polymorphic nature and has suggested exciting new intermediate species. Such intermediate species could serve as small-molecule targets for ligands specific for a broad range of quadruplexes. It also remains to be seen if such intermediates serve biological relevance, as many of the potential roles of quadruplexes include unfolding or enzyme recognition. Models developed in the thermodynamic studies conducted in this dissertation may provide insight into unfolding mechanisms driving quadruplex-specific helicase activity.

REFERENCES

(1)   Watson, J. D.; Crick, F. H. *Nature* **1953**, *171*, 737.

(2)   Watson, J. D.; Crick, F. H. *Nature* **1953**, *171*, 964.

(3)   Yatsunyk, L. A.; Bryan, T. M.; Brad Johnson, F. *Biochimie* **2012**.

(4)   Gellert, M.; Lipsett, M. N.; Davies, D. R. *Proceedings of the National Academy of Sciences of the United States of America* **1962**, *48*, 2013.

(5)   Yang, D.; Okamoto, K. *Future medicinal chemistry* **2010**, *2*, 619.

(6)   Bates, P.; Mergny, J. L.; Yang, D. *EMBO reports* **2007**, *8*, 1003.

(7)   Campbell, N. H.; Neidle, S. *Metal ions in life sciences* **2012**, *10*, 119.

(8)   Randazzo, A.; Spada, G. P.; Silva, M. W. *Topics in current chemistry* **2012**.

(9)   Phan, A. T. *The FEBS journal* **2010**, *277*, 1107.

(10)   Phan, A. T.; Kuryavyi, V.; Burge, S.; Neidle, S.; Patel, D. J. *Journal of the American Chemical Society* **2007**, *129*, 4386.

(11)   Webba da Silva, M. *Chemistry* **2007**, *13*, 9738.

(12)   Davis, J. T.; Spada, G. P. *Chemical Society reviews* **2007**, *36*, 296.

(13)   Huang, Y. C.; Sen, D. *Journal of the American Chemical Society* **2010**, *132*, 2663.

(14)   Devetak, M., Masiero, S., Pieraccini, S., Spada, G. P., Čopič, M., & Olenik, I. D *Applied Surface Science* **2010**, *256*, 2038.

(15)   Guo, L.; Nie, D.; Qiu, C.; Zheng, Q.; Wu, H.; Ye, P.; Hao, Y.; Fu, F.; Chen, G. *Biosensors & bioelectronics* **2012**, *35*, 123.

(16)   Li, T.; Wang, E.; Dong, S. *Analytical chemistry* **2010**, *82*, 7576.

(17)   Zhang, L.; Zhu, J.; Ai, J.; Zhou, Z.; Jia, X.; Wang, E. *Biosensors & bioelectronics* **2013**, *39*, 268.

(18)   Bates, P. J.; Laber, D. A.; Miller, D. M.; Thomas, S. D.; Trent, J. O. *Experimental and molecular pathology* **2009**, *86*, 151.

(19)   Zavyalova, E.; Golovin, A.; Reshetnikov, R.; Mudrik, N.; Panteleyev, D.; Pavlova, G.; Kopylov, A. *Current medicinal chemistry* **2011**, *18*, 3343.

(20)   Bates, P. J.; Kahlon, J. B.; Thomas, S. D.; Trent, J. O.; Miller, D. M. *The Journal of biological chemistry* **1999**, *274*, 26369.

(21)   Reyes-Reyes, E. M.; Teng, Y.; Bates, P. J. *Cancer research* **2010**, *70*, 8617.

(22)   Lipps, H. J.; Rhodes, D. *Trends in cell biology* **2009**, *19*, 414.

(23)   Bochman, M. L.; Paeschke, K.; Zakian, V. A. *Nature reviews. Genetics* **2012**, *13*, 770.

(24)   Stegle, O.; Payet, L.; Mergny, J. L.; MacKay, D. J.; Leon, J. H. *Bioinformatics* **2009**, *25*, i374.

(25)   Capra, J. A.; Paeschke, K.; Singh, M.; Zakian, V. A. *PLoS computational biology* **2010**, *6*, e1000861.

(26)   Paeschke, K.; Simonsson, T.; Postberg, J.; Rhodes, D.; Lipps, H. J. *Nature structural & molecular biology* **2005**, *12*, 847.

(27)   Schaffitzel, C.; Berger, I.; Postberg, J.; Hanes, J.; Lipps, H. J.; Pluckthun, A. *Proceedings of the National Academy of Sciences of the United States of America* **2001**, *98*, 8572.

(28)   Granotier, C.; Pennarun, G.; Riou, L.; Hoffschir, F.; Gauthier, L. R.; De Cian, A.; Gomez, D.; Mandine, E.; Riou, J. F.; Mergny, J. L.; Mailliet, P.; Dutrillaux, B.; Boussin, F. D. *Nucleic acids research* **2005**, *33*, 4182.

(29)   Collie, G. W.; Parkinson, G. N. *Chemical Society reviews* **2011**, *40*, 5867.

(30)   Baral, A.; Kumar, P.; Halder, R.; Mani, P.; Yadav, V. K.; Singh, A.; Das, S. K.; Chowdhury, S. *Nucleic acids research* **2012**, *40*, 3800.

(31)   Mathad, R. I.; Hatzakis, E.; Dai, J.; Yang, D. *Nucleic acids research* **2011**, *39*, 9023.

(32)   Siddiqui-Jain, A.; Grand, C. L.; Bearss, D. J.; Hurley, L. H. *Proceedings of the National Academy of Sciences of the United States of America* **2002**, *99*, 11593.

(33)   Brooks, T. A.; Hurley, L. H. *Genes & cancer* **2010**, *1*, 641.

(34)   Paeschke, K.; McDonald, K. R.; Zakian, V. A. *FEBS letters* **2010**, *584*, 3760.

(35)   Schuldt, A. *Nature reviews. Molecular cell biology* **2011**, *12*, 402.

(36)   London, T. B.; Barber, L. J.; Mosedale, G.; Kelly, G. P.; Balasubramanian, S.; Hickson, I. D.; Boulton, S. J.; Hiom, K. *The Journal of biological chemistry* **2008**, *283*, 36132.

(37)   Sanders, C. M. *The Biochemical journal* **2010**, *430*, 119.

(38)   Paeschke, K.; Capra, J. A.; Zakian, V. A. *Cell* **2011**, *145*, 678.

(39)   Lopes, J.; Piazza, A.; Bermejo, R.; Kriegsman, B.; Colosio, A.; Teulade-Fichou, M. P.; Foiani, M.; Nicolas, A. *The EMBO journal* **2011**, *30*, 4033.

(40)   Sarkies, P.; Reams, C.; Simpson, L. J.; Sale, J. E. *Molecular cell* **2010**, *40*, 703.

(41)   Sarkies, P.; Murat, P.; Phillips, L. G.; Patel, K. J.; Balasubramanian, S.; Sale, J. E. *Nucleic acids research* **2012**, *40*, 1485.

(42)   Bugaut, A.; Balasubramanian, S. *Nucleic acids research* **2012**, *40*, 4727.

(43)   Arora, A.; Maiti, S. *The journal of physical chemistry. B* **2009**, *113*, 10515.

(44)   Huppert, J. L.; Bugaut, A.; Kumari, S.; Balasubramanian, S. *Nucleic acids research* **2008**, *36*, 6260.

(45)   Kumari, S.; Bugaut, A.; Huppert, J. L.; Balasubramanian, S. *Nature chemical biology* **2007**, *3*, 218.

(46)   Shahid, R.; Bugaut, A.; Balasubramanian, S. *Biochemistry* **2010**, *49*, 8300.

(47)   Morris, M. J.; Negishi, Y.; Pazsint, C.; Schonhoft, J. D.; Basu, S. *Journal of the American Chemical Society* **2010**, *132*, 17831.

(48)   Xu, Y. *Chemical Society reviews* **2011**, *40*, 2719.

(49)   Hayflick, L.; Moorhead, P. S. *Experimental cell research* **1961**, *25*, 585.

(50)   Zhou, X.; Xing, D. *Chemical Society reviews* **2012**, *41*, 4643.

(51)   Baumann, P.; Cech, T. R. *Science* **2001**, *292*, 1171.

(52)   Zaug, A. J.; Podell, E. R.; Cech, T. R. *Proceedings of the National Academy of Sciences of the United States of America* **2005**, *102*, 10864.

(53)   Griffith, J. D.; Comeau, L.; Rosenfield, S.; Stansel, R. M.; Bianchi, A.; Moss, H.; de Lange, T. *Cell* **1999**, *97*, 503.

(54)   Wei, C.; Price, M. *Cellular and molecular life sciences : CMLS* **2003**, *60*, 2283.

(55)   Xu, Y.; Sato, H.; Sannohe, Y.; Shinohara, K.; Sugiyama, H. *Journal of the American Chemical Society* **2008**, *130*, 16470.

(56)   Parkinson, G. N.; Lee, M. P.; Neidle, S. *Nature* **2002**, *417*, 876.

(57)   Petraccone, L.; Spink, C.; Trent, J. O.; Garbett, N. C.; Mekmaysy, C. S.; Giancola, C.; Chaires, J. B. *Journal of the American Chemical Society* **2011**, *133*, 20951.

(58)   Campbell, N.; Collie, G. W.; Neidle, S. *Current protocols in nucleic acid chemistry / edited by Serge L. Beaucage ... [et al.]* **2012**, *Chapter 17*, Unit17 6.

(59)   Chen, Y.; Yang, D. *Current protocols in nucleic acid chemistry / edited by Serge L. Beaucage ... [et al.]* **2012**, *Chapter 17*, Unit17 5.

(60)   Phan, A. T.; Patel, D. J. *Journal of the American Chemical Society* **2003**, *125*, 15021.

(61)   Wang, Y.; Patel, D. J. *Structure* **1993**, *1*, 263.

(62)   Ambrus, A.; Chen, D.; Dai, J.; Bialis, T.; Jones, R. A.; Yang, D. *Nucleic acids research* **2006**, *34*, 2723.

(63)   Dai, J.; Punchihewa, C.; Ambrus, A.; Chen, D.; Jones, R. A.; Yang, D. *Nucleic acids research* **2007**, *35*, 2440.

(64)   Luu, K. N.; Phan, A. T.; Kuryavyi, V.; Lacroix, L.; Patel, D. J. *Journal of the American Chemical Society* **2006**, *128*, 9963.

(65)   Dai, J.; Carver, M.; Punchihewa, C.; Jones, R. A.; Yang, D. *Nucleic acids research* **2007**, *35*, 4927.

(66)   Phan, A. T.; Kuryavyi, V.; Luu, K. N.; Patel, D. J. *Nucleic acids research* **2007**, *35*, 6517.

(67)   Lim, K. W.; Amrane, S.; Bouaziz, S.; Xu, W.; Mu, Y.; Patel, D. J.; Luu, K. N.; Phan, A. T. *Journal of the American Chemical Society* **2009**, *131*, 4301.

(68)   Zhang, Z.; Dai, J.; Veliath, E.; Jones, R. A.; Yang, D. *Nucleic acids research* **2010**, *38*, 1009.

(69)   Phan, A. T.; Luu, K. N.; Patel, D. J. *Nucleic acids research* **2006**, *34*, 5715.

(70)   Li, J.; Correia, J. J.; Wang, L.; Trent, J. O.; Chaires, J. B. *Nucleic acids research* **2005**, *33*, 4649.

(71)   Gray, R. D.; Petraccone, L.; Trent, J. O.; Chaires, J. B. *Biochemistry* **2010**, *49*, 179.

(72)   Noomnarm, U.; Clegg, R. M. *Photosynthesis research* **2009**, *101*, 181.

(73)   Laws, W. R.; Contino, P. B. *Methods in enzymology* **1992**, *210*, 448.

(74)   Bharill, S.; Sarkar, P.; Ballin, J. D.; Gryczynski, I.; Wilson, G. M.; Gryczynski, Z. *Analytical biochemistry* **2008**, *377*, 141.

(75)   Buscaglia, R.; Jameson, D. M.; Chaires, J. B. *Nucleic acids research* **2012**, *40*, 4203.

(76)   Stefl, M.; James, N. G.; Ross, J. A.; Jameson, D. M. *Analytical biochemistry* **2011**, *410*, 62.

(77)   James, N. G.; Ross, J. A.; Stefl, M.; Jameson, D. M. *Analytical biochemistry* **2011**, *410*, 70.

(78)   Miller, M. C.; Buscaglia, R.; Chaires, J. B.; Lane, A. N.; Trent, J. O. *Journal of the American Chemical Society* **2010**.

(79)   Heddi, B.; Phan, A. T. *Journal of the American Chemical Society* **2011**, *133*, 9824.

(80) Chaires, J. B. *The FEBS journal* **2010**, *277*, 1098.

(81) Gray, R. D.; Buscaglia, R.; Chaires, J. B. *Journal of the American Chemical Society* **2012**, *134*, 16834.

(82) Hendler, R. W.; Shrager, R. I. *Journal of biochemical and biophysical methods* **1994**, *28*, 1.

(83) Hwang, H.; Buncher, N.; Opresko, P. L.; Myong, S. *Structure* **2012**, *20*, 1872.

(84) Ward, D. C.; Reich, E.; Stryer, L. *The Journal of biological chemistry* **1969**, *244*, 1228.

(85) Reha-Krantz, L. J. *Methods Mol Biol* **2009**, *521*, 381.

(86) Tang, G. Q.; Anand, V. S.; Patel, S. S. *Journal of molecular biology* **2011**, *405*, 666.

(87) Sarkar, K.; Nguyen, D. A.; Gruebele, M. *RNA* **2010**, *16*, 2427.

(88) Ballin, J. D.; Prevas, J. P.; Bharill, S.; Gryczynski, I.; Gryczynski, Z.; Wilson, G. M. *Biochemistry* **2008**, *47*, 7043.

(89) Degtyareva, N. N.; Petty, J. T. *Methods in enzymology* **2011**, *492*, 213.

(90) Gray, R. D.; Petraccone, L.; Buscaglia, R.; Chaires, J. B. *Methods Mol Biol* **2010**, *608*, 121.

(91) Sarkar, K.; Meister, K.; Sethi, A.; Gruebele, M. *Biophysical journal* **2009**, *97*, 1418.

(92) Guest, C. R.; Hochstrasser, R. A.; Sowers, L. C.; Millar, D. P. *Biochemistry* **1991**, *30*, 3271.

(93) Fogarty, A. C.; Jones, A. C.; Camp, P. J. *Physical chemistry chemical physics : PCCP* **2011**, *13*, 3819.

(94) Rachofsky, E. L.; Osman, R.; Ross, J. B. *Biochemistry* **2001**, *40*, 946.

(95) Jean, J. M.; Hall, K. B. *Proceedings of the National Academy of Sciences of the United States of America* **2001**, *98*, 37.

(96) Jean, J. M.; Hall, K. B. *Biochemistry* **2002**, *41*, 13152.

(97) Brochon, J. C. *Methods in enzymology* **1994**, *240*, 262.

(98) Cole, K. S.; Cole, R. H. *J Chem Phys* **1941**, *9*, 341.

(99)   Clayton, A. H.; Hanley, Q. S.; Verveer, P. J. *Journal of microscopy* **2004**, *213*, 1.

(100)   Redford, G. I.; Clegg, R. M. *Journal of fluorescence* **2005**, *15*, 805.

(101)   Jameson, D. M.; Gratton, E.; Hall, R. D. *Appl. Spectrosc. Rev.* **1984**, *20*, 55.

(102)   Weber, G. *J. Phys. Chem.* **1981**, *85*, 949.

(103)   Hanley, Q. S. *J. R. Soc. Interface* **2009**, *6*, S83.

(104)   Digman, M. A.; Caiolfa, V. R.; Zamai, M.; Gratton, E. *Biophysical journal* **2008**, *94*, L14.

(105)   Wouters, F. S.; Esposito, A. *HFSP journal* **2008**, *2*, 7.

(106)   Leray, A.; Spriet, C.; Trinel, D.; Blossey, R.; Usson, Y.; Heliot, L. *Cytom. Part A* **2011**, *79A*, 149.

(107)   Lane, A. N.; Chaires, J. B.; Gray, R. D.; Trent, J. O. *Nucleic acids research* **2008**, *36*, 5482.

(108)   Huppert, J. L. *Chemical Society reviews* **2008**, *37*, 1375.

(109)   Kimura, T.; Kawai, K.; Fujitsuka, M.; Majima, T. *Chem Commun (Camb)* **2004**, 1438.

(110)   Kimura, T.; Kawai, K.; Fujitsuka, M.; Majima, T. *Tetrahedron* **2007**, *63*, 3585.

(111)   Petraccone, L.; Garbett, N. C.; Chaires, J. B.; Trent, J. O. *Biopolymers* **2010**, *93*, 533.

(112)   Okamoto, K.; Sannohe, Y.; Mashimo, T.; Sugiyama, H.; Terazima, M. *Bioorganic & medicinal chemistry* **2008**, *16*, 6873.

(113)   Shirude, P. S.; Balasubramanian, S. *Biochimie* **2008**, *90*, 1197.

(114)   De Cian, A.; Guittat, L.; Kaiser, M.; Sacca, B.; Amrane, S.; Bourdoncle, A.; Alberti, P.; Teulade-Fichou, M. P.; Lacroix, L.; Mergny, J. L. *Methods* **2007**, *42*, 183.

(115)   De Cian, A.; Guittat, L.; Shin-ya, K.; Riou, J. F.; Mergny, J. L. *Nucleic Acids Symp Ser (Oxf)* **2005**, 235.

(116)   Barbieri, B.; Terpetschnig, E.; Jameson, D. M. *Analytical biochemistry* **2005**, *344*, 298.

(117)   Beechem, J. M.; Gratton, E. *Flourescence spectroscopy data analysis environment a second generation global analysis program.*, 1988; Vol. 909.

(118) Gray, R. D.; Chaires, J. B. *Nucleic acids research* **2008**, *36*, 4191.

(119) Gray, R. D.; Li, J.; Chaires, J. B. *The journal of physical chemistry. B* **2009**, *113*, 2676.

(120) Gray, R. D.; Chaires, J. B. *Biophysical chemistry* **2011**, *159*, 205.

(121) Baumann, U.; Frank, R.; Blocker, H. *European journal of biochemistry / FEBS* **1986**, *161*, 409.

(122) Xodo, L. E.; Manzini, G.; Quadrifoglio, F.; van der Marel, G.; van Boom, J. *Nucleic acids research* **1991**, *19*, 1505.

(123) Sung, S. C.; Laskowski, M., Sr. *The Journal of biological chemistry* **1962**, *237*, 506.

(124) Klostermeier, D.; Millar, D. P. *Biopolymers* **2001**, *61*, 159.

(125) Dailey, M. M.; Miller, M. C.; Bates, P. J.; Lane, A. N.; Trent, J. O. *Nucleic acids research* **2010**, *38*, 4877.

(126) Huppert, J. L.; Balasubramanian, S. *Nucleic acids research* **2005**, *33*, 2908.

(127) Vorlickova, M.; Chladkova, J.; Kejnovska, I.; Fialova, M.; Kypr, J. *Nucleic acids research* **2005**, *33*, 5851.

(128) Xue, Y.; Kan, Z. Y.; Wang, Q.; Yao, Y.; Liu, J.; Hao, Y. H.; Tan, Z. *Journal of the American Chemical Society* **2007**, *129*, 11185.

(129) Dapic, V.; Abdomerovic, V.; Marrington, R.; Peberdy, J.; Rodger, A.; Trent, J. O.; Bates, P. J. *Nucleic acids research* **2003**, *31*, 2097.

(130) Masiero, S.; Trotta, R.; Pieraccini, S.; De Tito, S.; Perone, R.; Randazzo, A.; Spada, G. P. *Organic & biomolecular chemistry* **2010**, *8*, 2683.

(131) Balagurumoorthy, P.; Brahmachari, S. K.; Mohanty, D.; Bansal, M.; Sasisekharan, V. *Nucleic acids research* **1992**, *20*, 4061.

(132) Giraldo, R.; Suzuki, M.; Chapman, L.; Rhodes, D. *Proceedings of the National Academy of Sciences of the United States of America* **1994**, *91*, 7658.

(133)   Hazel, P.; Huppert, J.; Balasubramanian, S.; Neidle, S. *Journal of the American Chemical Society* **2004**, *126*, 16405.

(134)   Kypr, J.; Kejnovska, I.; Renciuk, D.; Vorlickova, M. *Nucleic acids research* **2009**, *37*, 1713.

(135)   Dapic, V.; Bates, P. J.; Trent, J. O.; Rodger, A.; Thomas, S. D.; Miller, D. M. *Biochemistry* **2002**, *41*, 3676.

(136)   Jing, N.; Rando, R. F.; Pommier, Y.; Hogan, M. E. *Biochemistry* **1997**, *36*, 12498.

(137)   Asensio, J. L.; Lane, A. N.; Dhesi, J.; Bergqvist, S.; Brown, T. *Journal of molecular biology* **1998**, *275*, 811.

(138)   Minton, A. P. *Biophysical journal* **2005**, *88*, 971.

(139)   Miyoshi, D.; Karimata, H.; Sugimoto, N. *Journal of the American Chemical Society* **2006**, *128*, 7957.

(140)   Spink, C. H.; Chaires, J. B. *Biochemistry* **1999**, *38*, 496.

(141)   Olsen, C. M.; Gmeiner, W. H.; Marky, L. A. *The journal of physical chemistry. B* **2006**, *110*, 6962.

(142)   Hubbard, S. J.; Thornton, J. M. 1993.

(143)   Garcia De La Torre, J.; Huertas, M. L.; Carrasco, B. *Biophysical journal* **2000**, *78*, 719.

(144)   Courtenay, E. S.; Capp, M. W.; Anderson, C. F.; Record, M. T., Jr. *Biochemistry* **2000**, *39*, 4455.

(145)   Kiyosawa, K. *Biophysical chemistry* **2003**, *104*, 171.

(146)   Hancock, R. *Biology of the cell / under the auspices of the European Cell Biology Organization* **2004**, *96*, 595.

(147)   Moyzis, R. K.; Buckingham, J. M.; Cram, L. S.; Dani, M.; Deaven, L. L.; Jones, M. D.; Meyne, J.; Ratliff, R. L.; Wu, J. R. *Proceedings of the National Academy of Sciences of the United States of America* **1988**, *85*, 6622.

(148)   Wright, W. E.; Tesmer, V. M.; Huffman, K. E.; Levene, S. D.; Shay, J. W. *Genes Dev* **1997**, *11*, 2801.

(149)   Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. *Nucleic acids research* **2006**, *34*, 5402.

(150)   Lane, A. N.; Chaires, J. B.; Gray, R. D.; Trent, J. O. *Nucleic Acids Res* **2008**, *36*, 5482.

(151)   Patel, D. J.; Phan, A. T.; Kuryavyi, V. *Nucleic acids research* **2007**, *35*, 7429.

(152)   Phan, A. T.; Mergny, J. L. *Nucleic acids research* **2002**, *30*, 4618.

(153)   Granotier, C.; Pennarun, G.; Riou, L.; Hoffschir, F.; Gauthier, L. R.; De Cian, A.; Gomez, D.; Mandine, E.; Riou, J. F.; Mergny, J. L.; Mailliet, P.; Dutrillaux, B.; Boussin, F. D. *Nucleic Acids Res* **2005**, *33*, 4182.

(154)   Xu, Y. *Chem Soc Rev* **2011**, *40*, 2719.

(155)   Wang, Y.; Patel, D. J. *Structure* **1993**, *1*, 263.

(156)   He, Y.; Neumann, R. D.; Panyutin, I. G. *Nucleic acids research* **2004**, *32*, 5359.

(157)   Li, J.; Correia, J. J.; Wang, L.; Trent, J. O.; Chaires, J. B. *Nucleic Acids Res* **2005**, *33*, 4649

(158)   Qi, J.; Shafer, R. H. *Nucleic acids research* **2005**, *33*, 3185.

(159)   Redon, S.; Bombard, S.; Elizondo-Riojas, M. A.; Chottard, J. C. *Nucleic acids research* **2003**, *31*, 1605.

(160)   Ying, L.; Green, J. J.; Li, H.; Klenerman, D.; Balasubramanian, S. *Proceedings of the National Academy of Sciences of the United States of America* **2003**, *100*, 14629.

(161)   Ambrus, A.; Chen, D.; Dai, J.; Bialis, T.; Jones, R. A.; Yang, D. *Nucleic Acids Res* **2006**, *34*, 2723.

(162)   Phan, A. T.; Luu, K. N.; Patel, D. J. *Nucleic Acids Res* **2006**, *34*, 5715.

(163)   Xu, Y.; Noguchi, Y.; Sugiyama, H. *Bioorganic & medicinal chemistry* **2006**, *14*, 5584.

(164)   Lannan, F. M.; Mamajanov, I.; Hud, N. V. *Journal of the American Chemical Society* **2012**.

(165)   Ellis, R. J. *Trends Biochem Sci* **2001**, *26*, 597.

(166)   Ellis, R. J. *Curr Opin Struct Biol* **2001**, *11*, 114.

(167)   Hall, D.; Minton, A. P. *Biochim Biophys Acta* **2003**, *1649*, 127.

(168)   Minton, A. P. *J Mol Recognit* **1993**, *6*, 211.

(169)   Minton, A. P. *Curr Biol* **2006**, *16*, R269.

(170)   Zhou, H. X.; Rivas, G.; Minton, A. P. *Annu Rev Biophys* **2008**, *37*, 375.

(171)   Zimmerman, S. B.; Minton, A. P. *Annu Rev Biophys Biomol Struct* **1993**, *22*, 27.

(172)   Ball, P. *Chem Rev* **2008**, *108*, 74.

(173)   Parsegian, V. A. *Int Rev Cytol* **2002**, *215*, 1.

(174)   Ball, P. *Chemphyschem* **2008**, *9*, 2677.

(175)   Parsegian, V. A.; Rand, R. P.; Rau, D. C. *Meth. Enzymol.* **1995**, *259*, 43.

(176)   Chaplin, M. *Nature reviews. Molecular cell biology* **2006**, *7*, 861.

(177)   Lane, A. N. *Biochimie* **2012**, *94*, 277.

(178)   Rozners, E. *Current protocols in nucleic acid chemistry / edited by Serge L. Beaucage ... [et al.]* **2010**, *Chapter 7*, Unit 7 14.

(179)   Miyoshi, D.; Fujimoto, T.; Sugimoto, N. *Topics in current chemistry* **2012**.

(180)   Petraccone, L.; Pagano, B.; Giancola, C. *Methods* **2012**.

(181)   Miyoshi, D.; Karimata, H.; Sugimoto, N. *Nucleosides Nucleotides Nucleic Acids* **2007**, *26*, 589.

(182)   Olsen, C. M.; Gmeiner, W. H.; Marky, L. A. *The journal of physical chemistry* **2006**, *110*, 6962.

(183)   Smirnov, I. V.; Shafer, R. H. *Biopolymers* **2007**, *85*, 91.

(184)   Kan, Z. Y.; Yao, Y.; Wang, P.; Li, X. H.; Hao, Y. H.; Tan, Z. *Angewandte Chemie* **2006**, *45*, 1629.

(185)   Zhou, J.; Wei, C.; Jia, G.; Wang, X.; Tang, Q.; Feng, Z.; Li, C. *Biophysical chemistry* **2008**, *136*, 124.

(186)   Kan, Z. Y.; Lin, Y.; Wang, F.; Zhuang, X. Y.; Zhao, Y.; Pang, D. W.; Hao, Y. H.; Tan, Z. *Nucleic acids research* **2007**, *35*, 3646.

(187)   Zhou, J.; Wei, C.; Jia, G.; Wang, X.; Feng, Z.; Li, C. *Biochimie* **2009**, *91*, 1104.

(188)   Petraccone, L.; Malafronte, A.; Amato, J.; Giancola, C. *The journal of physical chemistry. B* **2012**, *116*, 2294.

(189)   Hansel, R.; Lohr, F.; Foldynova-Trantirkova, S.; Bamberg, E.; Trantirek, L.; Dotsch, V. *Nucleic acids research* **2011**, *39*, 5768.

(190)   Hirano, A.; Shiraki, K.; Arakawa, T. *Biopolymers* **2012**, *97*, 117.

(191)   Elcock, A. H. *Curr Opin Struct Biol* **2010**, *20*, 196.

(192)   Neidle, S.; Parkinson, G. N. *Biochimie* **2008**, *90*, 1184.

(193)   Knowles, D. B.; LaCroix, A. S.; Deines, N. F.; Shkel, I.; Record, M. T., Jr. *Proceedings of the National Academy of Sciences of the United States of America* **2011**, *108*, 12699.

(194)   Cohen, J. A.; Highsmith, S. *Biophysical journal* **1997**, *73*, 1689.

(195)   French, H. T. *J Chem Thermodyn* **1987**, *19*, 1155.

(196)   Ninni, L.; Camargo, M. S.; Meirelles, A. J. A. *Thermochim Acta* **1999**, *328*, 169.

(197)   Zhang, W.; Capp, M. W.; Bond, J. P.; Anderson, C. F.; Record, M. T., Jr. *Biochemistry* **1996**, *35*, 10506.

(198)   Hellman, L. M.; Rodgers, D. W.; Fried, M. G. *Eur Biophys J* **2010**, *39*, 389.

(199)   Fernandes, M. X.; Ortega, A.; Lopez Martinez, M. C.; Garcia de la Torre, J. *Nucleic acids research* **2002**, *30*, 1782.

(200)   Randazzo, A.; Spada, G. P.; Silva, M. W. *Top Curr Chem* **2012**.

(201)   Gray, R. D.; Chaires, J. B. *Current protocols in nucleic acid chemistry / edited by Serge L. Beaucage ... [et al.]* **2011**, *Chapter 17*, Unit17 4.

(202)   Colombo, M. F.; Rau, D. C.; Parsegian, V. A. *Science* **1992**, *256*, 655.

(203)   Monod, J.; Wyman, J.; Changeux, J. P. *Journal of molecular biology* **1965**, *12*, 88.

(204)   Saroff, H. A. *Analytical biochemistry* **1989**, *176*, 161.

(205)   Singh, V.; Azarkh, M.; Exner, T. E.; Hartig, J. S.; Drescher, M. *Angewandte Chemie* **2009**, *48*, 9728.

(206)   Gaynutdinov, T. I.; Neumann, R. D.; Panyutin, I. G. *Nucleic acids research* **2008**, *36*, 4079.

(207)   Boncina, M.; Lah, J.; Prislan, I.; Vesnaver, G. *Journal of the American Chemical Society* **2012**, *134*, 9657.

(208)   Renciuk, D.; Kejnovska, I.; Skolakova, P.; Bednarova, K.; Motlova, J.; Vorlickova, M. *Nucleic acids research* **2009**, *37*, 6625.

(209)   Changeux, J. P.; Edelstein, S. *F1000 Biol Rep* **2011**, *3*, 19.

(210)   Chaires, J. B. *ACS chemical biology* **2008**, *3*, 207.

(211)   Chan, S. R. W. L.; Blackburn, E. H. *Phil. Trans. Roy. Soc. London. Ser. B: Biol. Sci.* **2004**, *359*, 109.

(212)   Cross, S. H.; Allshire, R. C.; McKay, S. J.; McGill, N. I.; Cooke, H. J. *Nature* **1989**, *338*, 771.

(213)   Shay, J. W.; Wright, W. E. *Sem. Cancer Biol.* **2011**, *21*, 349.

(214)   Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. *Nucleic Acids Res.* **2006**, *34*, 5402.

(215)   Mergny, J. L.; Lacroix, L. *Oligonucleotides* **2003**, *13*, 515.

(216)   Mergny, J. L.; Lacroix, L. *Curr. Protoc. Nucleic Acid Chem.* **2009**, *Chapter 17*, Unit 17 1.

(217)   Mergny, J. L.; Phan, A. T.; Lacroix, L. *FEBS Lett.* **1998**, *435*, 74.

(218)   Antonacci, C.; Chaires, J. B.; Sheardy, R. D. *Biochemistry* **2007**, *46*, 4654.

(219) Mashimo, T.; Yagi, H.; Sannohe, Y.; Rajendran, A.; Sugiyama, H. *J. Am. Chem. Soc.* **2010**, *132*, 14910.

(220) Zhang, N.; Phan, A. T.; Patel, D. J. *J. Am. Chem. Soc.* **2005**, *127*, 17277.

(221) Dignam, J. D.; Qu, X.; Chaires, J. B. *J. Biol. Chem.* **2001**, *276*, 4028.

(222) Greenfield, N. J. *Nat. Protoc.* **2006**, *1*, 2527.

(223) Motulsky, H. C., A *Fitting Models to Biological Data Using Linear and Nonlinear Regression*; GraphPad Software Inc.: San Diego, CA, 2003.

(224) Lumry, R.; Biltonen, R. *Biopolymers* **1966**, *4*, 917.

(225) Wallimann, P.; Kennedy, R. J.; Miller, J. S.; Shalongo, W.; Kemp, D. S. *J. Am. Chem. Soc.* **2003**, *125*, 1203.

(226) Henry, R.; Hofrichter, J. In *Methods Enzymol.*; Brand, L., Johnson, M., Eds.; Academic Press: New York, 1992; Vol. 210, p 129.

(227) Karsisiotis, A. I.; Hessari, N. M.; Novellino, E.; Spada, G. P.; Randazzo, A.; Webba da Silva, M. *Angew. Chem. Int. Ed.* **2011**, *50*, 10645.

(228) Jean, J. M.; Krueger, B. P. *J. Phys. Chem. B* **2006**, *110*, 2899.

(229) O'Neill, M. A.; Barton, J. K. *J. Am. Chem. Soc.* **2002**, *124*, 13053.

(230) Juskowiak, B. *Anal. Chim. Acta* **2006**, *568*, 171.

(231) Juskowiak, B.; Takenaka, S. *Methods Mol. Biol.* **2006**, *335*, 311.

(232) Koirala, D.; Mashimo, T.; Sannohe, Y.; Yu, Z.; Mao, H.; Sugiyama, H. *Chem. Commun. (Camb.)* **2012**, *48*, 2006.

(233) Ginsburg, A.; Carroll, W. R. *Biochemistry* **1965**, *4*, 2159.

(234) Haq, I.; Chowdhry, B. Z.; Chaires, J. B. *European biophysics journal : EBJ* **1997**, *26*, 419.

(235) Dai, J.; Carver, M.; Yang, D. *Biochimie* **2008**, *8*, 1172.

(236)   Mashimo, T.; Sannohe, Y.; Yagi, H.; Sugiyama, H. *Nucleic Acids Symp. Ser. (Oxf.)* **2008**, 409.

(237)   Su, D. G.; Fang, H.; Gross, M. L.; Taylor, J. S. *Proc. Natl. Acad. Sci. U. S. A.* **2009**, *106*, 12861.

(238)   Becktel, W. J.; Schellman, J. A. *Biopolymers* **1987**, *26*, 1859.

(239)   Privalov, P. L.; Khechinashvili, N. N. *Journal of molecular biology* **1974**, *86*, 665.

(240)   Pace, C. N.; Laurents, D. V. *Biochemistry* **1989**, *28*, 2520.

(241)   Talla-Singh, D.; Stites, W. E. *Proteins* **2008**, *71*, 1607.

(242)   Majhi, P. R.; Qi, J.; Tang, C. F.; Shafer, R. H. *Biopolymers* **2008**, *89*, 302.

(243)   Bochman, M. L.; Paeschke, K.; Zakian, V. A. *Nat Rev Genet* **2012**.

(244)   Hwang, H.; Buncher, N.; Opresko, P. L.; Myong, S. *Structure* **2012**.

(245)   Brooks, T. A.; Kendrick, S.; Hurley, L. *The FEBS journal* **2010**, *277*, 3459.

(246)   Le, H. T.; Buscaglia, R.; Dean, W. L.; Chaires, J. B.; Trent, J. O. *Topics in current chemistry* **2012**.

(247)   Mashimo, T.; Yagi, H.; Sannohe, Y.; Rajendran, A.; Sugiyama, H. *Journal of the American Chemical Society* **2010**, *132*, 14910.

(248)   Gray, R. D.; Buscaglia, R.; Chaires, J. B. *Journal of the American Chemical Society* **2012**.

(249)   Payet, L.; Huppert, J. L. *Biochemistry* **2012**, *51*, 3154.

(250)   Cantor, C. R.; Warshaw, M. M.; Shapiro, H. *Biopolymers* **1970**, *9*, 1059.

(251)   Li, J.; Center, T. U. o. M. M. *Biophysical Studies of Human Telomeric DNA*; University of Mississippi Medical Center, 2006.

(252)   Abu-Ghazalah, R. M.; Rutledge, S.; Lau, L. W.; Dubins, D. N.; Macgregor, R. B., Jr.; Helmy, A. S. *Biochemistry* **2012**, *51*, 7357.

(253)   Wallimann, P.; Kennedy, R. J.; Miller, J. S.; Shalongo, W.; Kemp, D. S. *Journal of the American Chemical Society* **2003**, *125*, 1203.

(254)   Koirala, D.; Mashimo, T.; Sannohe, Y.; Yu, Z.; Mao, H.; Sugiyama, H. *Chem Commun (Camb)* **2012**, *48*, 2006.

(255)   Haq, I.; Chowdhry, B. Z.; Chaires, J. B. *European biophysics journal : EBJ* **1997**, *26*, 419.

(256)   Ren, J.; Qu, X.; Trent, J. O.; Chaires, J. B. *Nucleic acids research* **2002**, *30*, 2307.

(257)   Besnard, E.; Babled, A.; Lapasset, L.; Milhavet, O.; Parrinello, H.; Dantec, C.; Marin, J. M.; Lemaitre, J. M. *Nature structural & molecular biology* **2012**, *19*, 837.

(258)   Gilbert, D. M. *Nature structural & molecular biology* **2012**, *19*, 740.

APPENDIX A

CHAPTER 3 SUPPLEMENTAL MATERIAL.

Methods.

Preparation of circular dichroism samples. The oligonucleotide d(AGGG(TTAGGG)$_3$) used in this study was purchased from Integrated DNA Technologies (Coralville, Iowa) and delivered as dry, desalted pellets. The oligonucleotide was shown to be a single chemical species by HPLC and electrospray MS. DNA stock solutions were dissolved in ddH$_2$O and dialyzed against 2 L ddH$_2$O for 2 separate 24 h periods at room temperature resulting in DNA stock solutions of 4 mM. Buffer stocks were prepared at 4x concentration as either potassium or sodium (in parentheses) buffer containing 100 mM KCl (NaCl), 25 mM K$_2$HPO$_4$ (Na$_2$HPO$_4$) (pH=7.0). Samples for CD were made by diluting DNA to an absorbance of approximately 0.9 (4 μM) in 1x sodium or potassium buffer, with or without acetonitrile (25 mM KCl (NaCl), 5 mM K$_2$HPO$_4$ (Na$_2$HPO$_4$), pH=7.0, or 50% acetonitrile, 25 mM KCl (NaCl), 5 mM K$_2$HPO$_4$ (Na$_2$HPO$_4$), pH=7.0). Each sample was annealed by submersion in boiling water for 10 minutes followed by quenching the sample on ice for 10 minutes. Annealing protocols were finalized based on determination that sample CD signatures were independent of fast- or slow-cooling methods.

Circular dichroism spectrum acquisition and thermal denaturations. CD data were collected using a Jasco J-810 spectropolarimeter. Annealed d(AGGG(TTAGGG)$_3$) samples, prepared as described above, were placed in a 4 mL sealable quartz cuvette with stirring (see above for buffer conditions). Spectra were collected from 220 to 340 nm, 0.2 nm spacing, at a rate of 50 nm min^{-1} and 2 second signal integration times, where the average of four scans is reported. Buffer blanks were recorded prior to sample scans. Thermal denaturation curves were collected using the same spectral acquisition, a temperature range of 293 to 371 K, a scan rate of 30 K h^{-1}, pre-scan incubation times of 30 s and data collection every 2 K. The spectra shown were recorded at 293 K.

NMR sample preparation and data collection. NMR samples were freshly annealed

DNA samples which had been diluted into fresh buffer as described above and D_2O added to

10%. Sample concentrations varied from 1 to 2 mM. Samples without acetonitrile were loaded

into 5 mm Shigemi NMR tubes. Samples containing acetonitrile were loaded into 5 mm NMR

tubes. Acetonitrile samples were prepared first by dialyzing the sample dissolved in K^+ buffer as

described above, followed by lyophilization of 0.65 ml and redissolution in 0.65 mL 50%

CD_3CN:50% H_2O or D_2O. NMR spectra were recorded on Varian Inova spectrometers at either

14.1 T or 18.8 T. All spectra were referenced to internal DSS-d6. 1D spectra were recorded with

an acquisition time of 1.5 s and a recycle time of 4 s. 2D spectra were recorded with acquisition

time of 0.256 s (H_2O) or 0.5 s (D_2O) in t2 and 0.0.38 s in t1 and a recycle time of 2 s. 1D spectra

were processed with a single zerofilling an unshifted Gaussian function and a 1 Hz line

broadening exponential function. The following NMR spectra were recorded of the K^+ form in

50% CD_3CN:50% H_2O: NOESY with mixing times of 300 ms and 100 ms. In 50% CD_3CN: 50% D_2O:

NOESY with mixing times of 50, 150 ms, DQF-COSY, TOCSY with a B1 field strength of 8 kHz for a

duration of 50 ms. 1D truncated NOE experiments with saturation of resolved H1' resonances

(interleaved) for 50, 100,150 and 200 ms.

Rotational correlation times were determined from the cross-relaxation rate constant

by fitting the 1D NOE time course for the H1'-H2" pairs (r = 2.35 ± 0.05 Å) as previously

described.[1] At 293 K, assuming a spherical shape, with a monolayer of hydration the rotational

correlation time in D2O would be ca. 3-3.5 ns. For M = 6800 the Stokes-Einstein formula gives

the rotational correlation time:

$\tau = \eta V / RT = 4 \pi \eta (r+ 2.8)^3 / 3RT$

The radius, r, for an anhydrous sphere with v = 0.55 ml/g is 11.4 Å.[2]

$\eta (D_2O)$ at 293 K = 0.012 P

Thus τ = 3.55 ns.

In 50% acetonitrile, the viscosity is some 10 - 15% less (even though the viscosity of pure acetonitrile is only 0.38 cP[3], so that the expected τ_c would be 2.5 - 3 ns, for full solvation. This may be less in acetonitrile.

At 293 K, NMR at 800 MHz, using both NOESY (50 ms mixing time) and 1D NOE build up curves of htel22 in 50% D_2O, 50% ACN-d3 gave τ_c of the order 2 ±.5 ns, which is comparable to a monomer. A dimer would have multiple correlation times, one twice that of the monomer, the other 4.67 times that of the monomer.[4] Together with the observation of only 12 GN1H resonances, this is best accounted for by a monomeric species.

Table A1. Thermodynamic data generated from CD. *These values are estimates. The melting point of this structure is too high to be accurately determined. The lack of an upper melting baseline made determination of T_m, ΔH, and ΔG_{298K} estimated values for these conditions.

Cation	CH_3CN % (v/v)	T_m (K)	ΔH (kcal mol^{-1})	ΔG_{298K} (kcal mol^{-1})
K^+	0	392.2 ± 0.2	-52 ± 2.0	-4.9 ± 0.2
K^+	50	384*	-53.7*	-10.7*
Na^+	0	321.2 ± 0.4	-52.0 ± 3.5	-3.8 ± 0.3
Na^+	50	339.4 ± 0.4	-58.1 ± 3.7	-7.1 ± 0.4

Table A2. Line widths in different solvent conditions. Resonance widths at half height of the

GN1H were measured from spectra recorded at 293 K, 14.1 T with no line broadening.

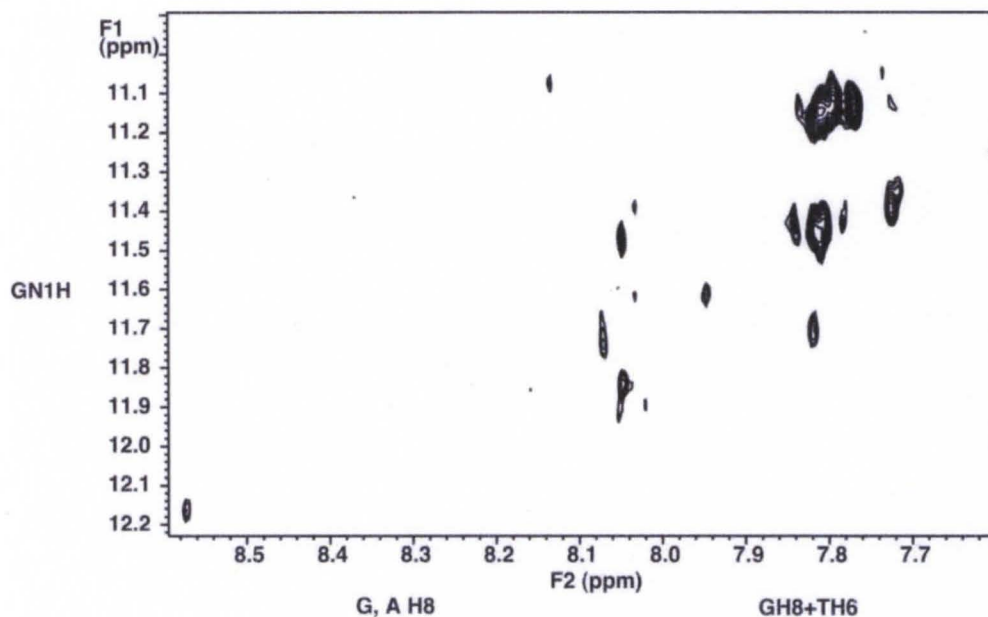| Condition	Mean linewith (Hz)	sd (Hz)
K^+, water	6.7	0.4
K^+, 50% CH_3CN	8.4	0.7
Na^+, water	9.7	0.8
Na^+, 50% CH_3CH	8.7	0.6

Figure A1. NOESY spectrum of d(AGGG(TTAGGG)$_3$) in 50% CD$_3$CN 50% H$_2$O showing GN1H-

H8/H6. The spectrum was recorded at 14.1 T, 293 K using a Watergate for solvent suppression.

The acquisition times were 0.256 s in t$_2$ and 26 ms in t$_1$. The mixing time was 300 ms. The

GN1H-GH8 NOE cross-peaks are characteristic of G-quartets, and are found only for the G
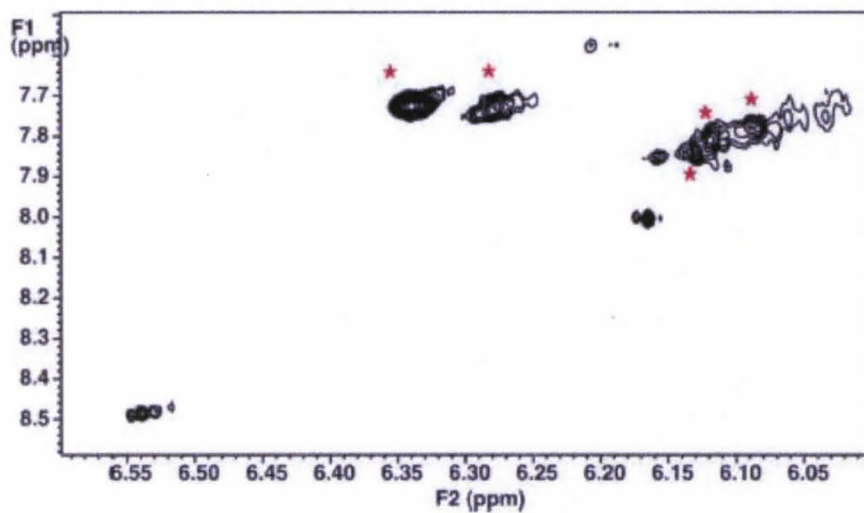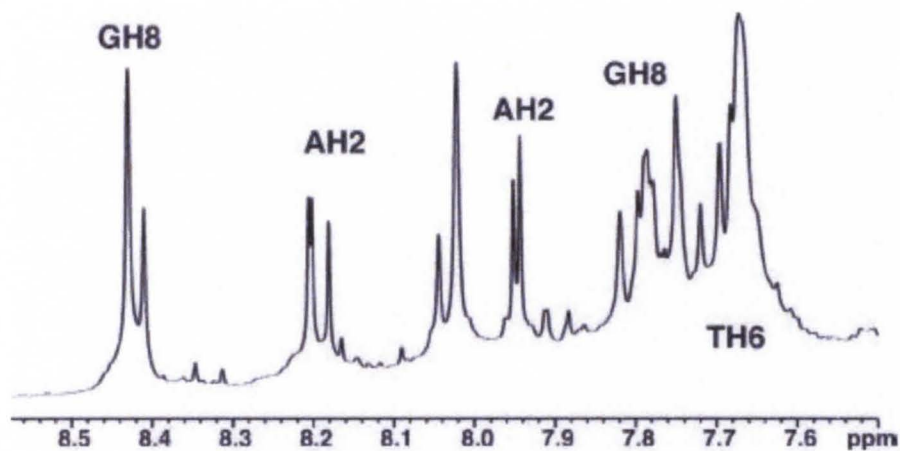
residues involved in quartets (and not the A, T in loops).[5]

Figure A2. NMR spectra of d(AGGG(TTAGGG)₃) in 50% CD_3CN 50% D_2O. The spectra were recorded at 18.8 T, 293 K.

Top. 1D ^1H NMR spectrum showing the protons of the nucleobases: TH6, GH8, AH8 and AH2. The acquisition time was 1.5 s. A 1 Hz line broadening exponential was applied. The chemical shifts of the AdeH2 were found from 1D and NOESY in D_2O and the TH6 resonances were from the 4-bond scalar coupling in 2-quantum filtered COSY and TOCSY spectra (Figure B4). GH8 positions were found from intraquartet NOEs between GN1H and GH8 (Figure B2).

220

Bottom. H8-H1' region of a NOESY spectrum. Acquisition times were 0.512 s in t_2 and 38 ms in t_1 and the mixing time was 50 ms. The cross-peaks correspond to intraresidue NOEs, of which there should be up to 22, but only 8 are significant at this short mixing time. The most intense resonances (denoted by *) are not pure anti with a 'S' sugar pucker (as observed in the DQF-COSY spectra, Figure B4), and correspond to GH8-H1' based on residue type assignments as described above (and Figures B1 and B3). The H1'-H8 cross-peaks were volume integrated and normalized to the H1'-H2" volumes, giving H1'-H8 distances significantly shorter than expected for high anti glycosyl torsions as found in the all-parallel conformation. In regular B DNA with anti glycosyl torsion angles (c ≈ -100-120°) and 'S'-type sugar puckers the intranucleotide NOESY cross-peak volumes are in the order H8-H2'>H8-H1'≈H8-H3' whereas for a syn nucleotide (c ≈ +70°) 'S'-type sugar puckers the order becomes H8-H1'>H8-H2'≈H8-H3'.[6-7]
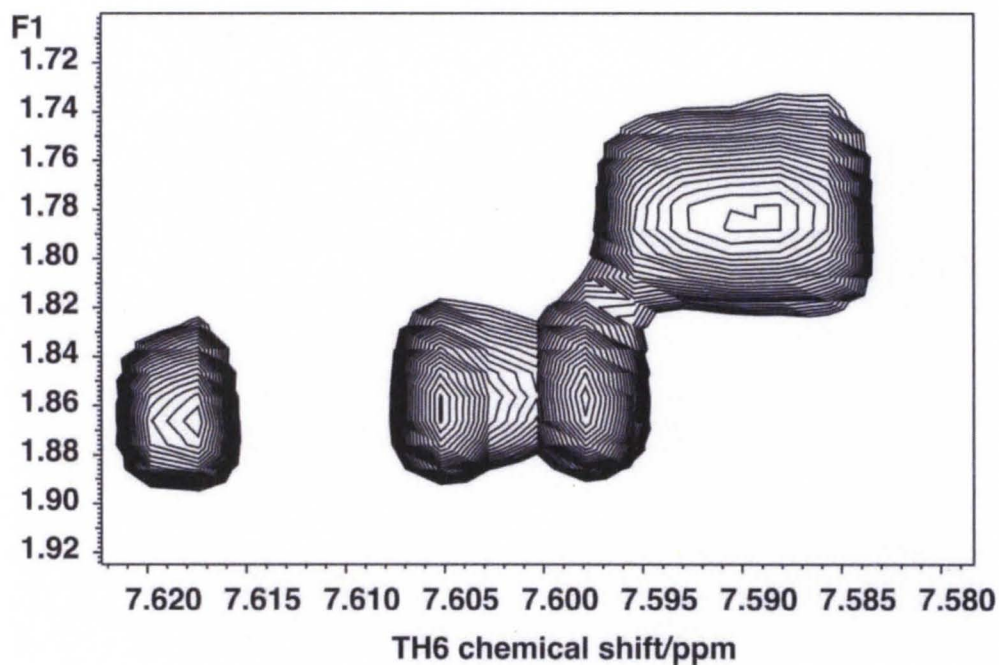
Figure A3. TOCSY spectrum of htel22 in 50% D_2O:50% CD_3CN. The spectrum was recorded at 293 K at 800 MHz using a spin lock field strength of 8 kHz for 50 ms. The region shown corresponds to the 4-bond scalar coupling between the H6 and CH_3 of the T residues. This therefore identifies the position in the spectrum of the TH6 resonances (Figures B1 and B2).

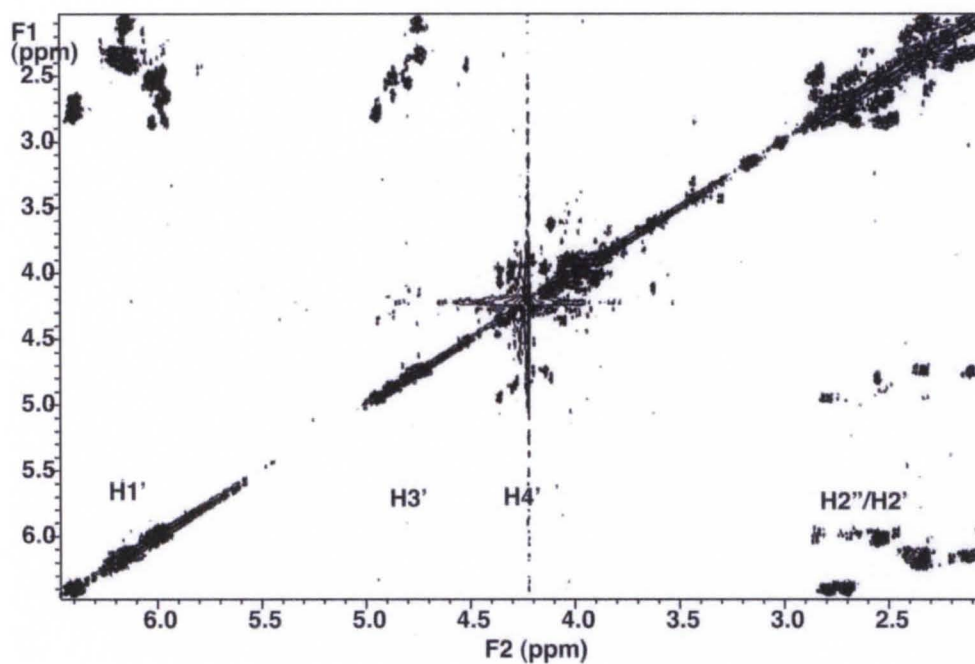**Figure A4.** 2-quantum filtered COSY spectrum of htel22 in 50% D_2O:50% CD_3CN. The spectrum was recorded at 800 MHz, 293 K with acquisition time of 0.512 s in t_2 and 36.6 ms in t_1. The spectrum was transformed into 16 k by 4 k points using unshifted Gaussians in both dimensions, and a single linear prediction in t_1. The regions displayed correspond to the 3-bond scalar couplings among sugar protons, as designated.
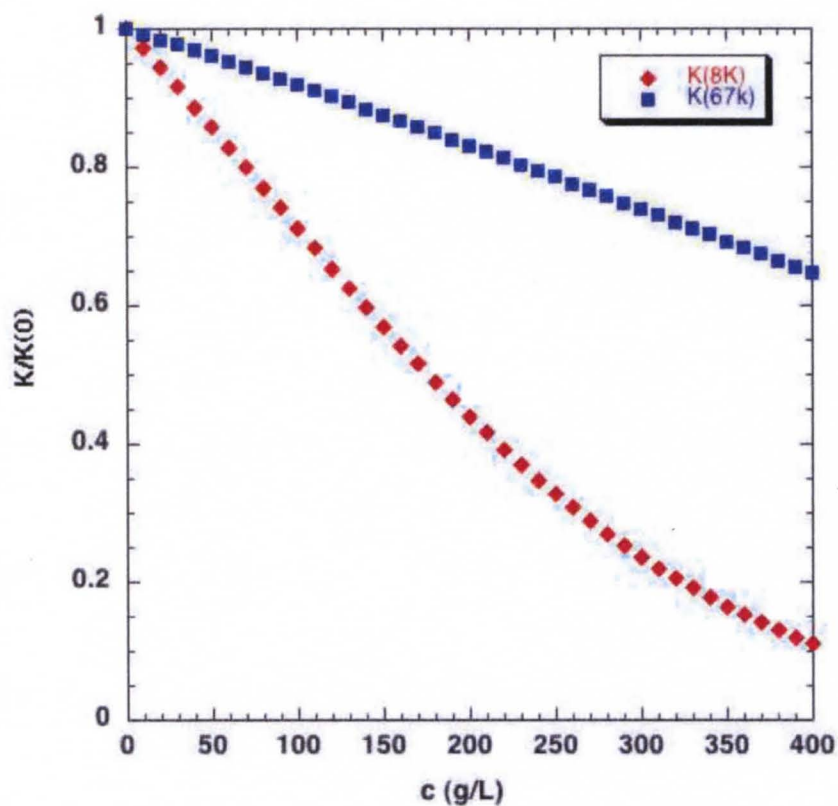
Figure A5. Simulated equilibrium constant for isomerization of htel to parallel state. The

system was approximated as comprising quadruplex in a compact state and the all-parallel

crystal structure state. Form the observed and calculated sedimentation coefficient of the

telomere in H_2O^8 the effective radii were calculated as 13.5 Å (water) and 16.8 Å (crystalline

parallel form). Spherical proteins of molecular weights 8.4 kDa and 67 kDa, with Stokes' radii of

15 and 30 Å, respectively were used to calculate the apparent activity coefficient of the

quadruplex due solely to steric crowding by the protein, as a function of the cosolute

concentration. The second and third virial coefficients were calculated according the

MacMillan-Mayer theory for noninteracting particles, using the particle number density

according to the concentration c and molecular weight M.

$K/K_0 = \exp[\ln(\gamma_2) - \ln(\gamma_1)]$

With $\ln\gamma = -\Sigma(k+1)B_{k+1}\rho^k/k$

For k = 1,2

$B_2 = 2\pi a^3/3$ and $B_3 = 5B_2^2/8$ [9]

Where a is the sum of the hard sphere radii of the DNA and crowding proteins.

    In aqueous K^+ buffer the equilibrium constant favors the anti-parallel forms by >100:1 at 293 K. Note that steric crowding favors the more compact forms over the parallel form.
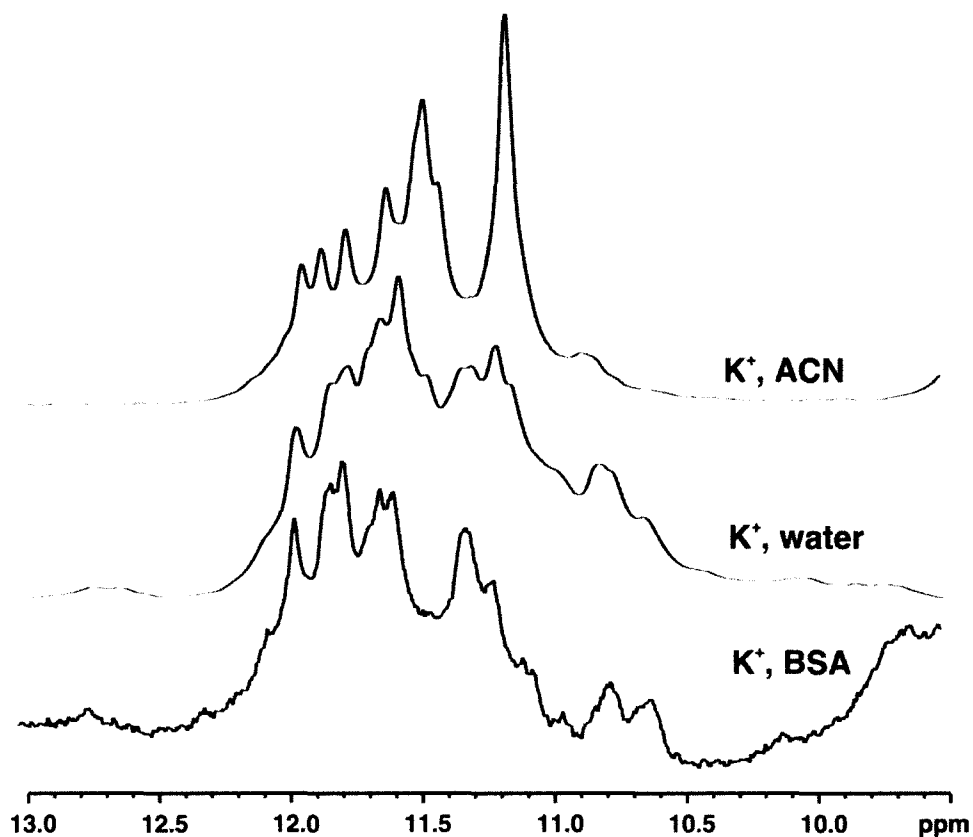
Figure A6. Effect of BSA on the NMR spectrum of htel22 NMR spectra were recorded at 18.8 T

293 K using a Watergate TOCSY sequence with a spin lock strength of 9 kHz for 10 ms to

suppress the water and the excess intensity from the BSA solution. The acquisition time was 1.5

s and the recycle time was 3 s. The spectrum in the presence of BSA was processed using an

unshifted Gaussian and 4 Hz line broadening. The control spectra were processed using an

unshifted Gaussian and 25 Hz line broadening to simulate the effect of the 4-fold increase in

solution viscosity.[10] Increasing the viscosity 4 fold increases the line widths on average from 7 Hz

to 28 Hz, greatly reducing the resolution. In 30% BSA, the NMR spectrum of K[+] htel22 is very

broad. Nevertheless, the spectrum in BSA is more similar to the (broadened) spectra in H_2O

than to that in 50% ACN.

Appendix A References.

(1) Lane, A. N. In *Nuclear Magnetic Resonance and Nucleic Acids*; Academic Press, Inc.: San Diego, 1995; Vol. 261, p 413-435.

(2) Hellman, L.; Rodgers, D.; Fried, M. *Eur. J. Biophys.* **2009**.

(3) J. Will Thompson; Theodore J. Kaiser; Jorgenson, J. W. *J. Chromatog. A* **2006**, *1134*, 201-209.

(4) Schurr, J. M.; Babcock, H. P.; Fujimoto, B. S. *Journal of Magnetic Resonance Series B* **1994**, *105*, 211-224.

(5) Patel, D. J.; Phan, A. T.; Kuryavyi, V. *Nucleic Acids Research* **2007**, *35*, 7429-7455.

(6) Lane, A. N.; Jenkins, T. C.; Brown, D. J. S.; Brown, T. *Biochemical Journal* **1991**, *279*, 269-281.

(7) Feigon, J.; Koshlap, K. M.; Smith, F. W. In *Nuclear Magnetic Resonance and Nucleic Acids* 1995; Vol. 261, p 225-255.

(8) Li, J.; Correia, J. J.; Wang, L.; Trent, J. O.; Chaires, J. B. *Nucleic Acids Research* **2005**, *33*, 4649-4659.

(9) Hill, T. L. *An Introduction to Statistical Thermodynamics*; Dover: Toronto, 1986.

(10) Monkos, K. *Intl. J. Biol. Macromol.* **1996**, *18*, 61-68.

APPENDIX B

CHAPTER 4 SUPPLEMENTAL MATERIAL.

Discussion of Equation 12 as a restricted form of the MWC model. The well-known Monod-Wyman-Changeux model[1] as generalized by Eigen[2] posits a preexisting equilibrium between two states A and B, with equilibrium constant L (here denoted K_0), with ligand independent binding constant to each site that differ between the two states (denoted by K_A and K_B), i.e. a three parameter function. The binding sites are assumed to be independent (i.e. no binding cooperativity within a state). Here we consider the possibility that the number of potential binding sites in states A and B may be different, denoted by n and m, respectively. This might arise from an increase in surface area in state B. The general form of the equilibrium constant for this model is:

$$K = K_0(1+x/K_B)^n/(1+x/K_A)^m$$

where x is the activity of the ligand. As the titration data can be adequately fitted even to a simple Hill equation, it is clear that the general mechanism of Equation 12 is underdetermined. We consider physical simplifications that reduce to the minimum number of parameters needed to fit the experimental data adequately. If $K_A = K_B$, i.e. there is only differential binding stoichiometries, the above equation reduces to

$$K = K_0(1+x/K_A)^{n-m}$$

In this instance the apparent stoichiometry is the difference between the two states, analogous to the model of differential hydration. If the value of K_A is very high compared with the maximum value of x, then

$$K = K_0(1+x/K_B)^m$$

This is exclusive binding, and the stoichiometry is the number of ligand molecule binding to state B. IF n=m, the form reduces to the conventional MWC model.

Appendix B References.

(1) Monod, J.; Wyman, J.; Changeux, J. P. *Journal of molecular biology* **1965**, *12*, 88.

(2) M. Eigen, in S. Claesson (Ed.), *Fast Reactions and Primary Processes in Chemical Kinetics*, Nobel Symposium No. 5, Almqvist and Wiksell, Stockholm, 1967, p. 333.

APPENDIX C

CHAPTER 5 SUPPLEMENTAL MATERIAL.

Analytical Models. Equations used for iterative nonlinear least squares fitting melting data to the five mechanisms outlined in Scheme 1. $dHi = \Delta H$(folding) for step i; Tmi = mid-point temperature for step i; Sni = optical signal for folded conformer i; Si = optical signal for intermediate species i; Su = optical signal of the unfolded ensemble; $R = 1.987$ cal/$^\circ$K/mol.

Single Transition (Mechanism 1)

$$S_{obs_j} = \frac{S_{n_j} + S_{u_j}\, e^{\frac{-dH_1}{R}\left(\frac{1}{Tm1} - \frac{1}{T}\right)}}{1 + e^{\frac{-dH_1}{R}\left(\frac{1}{Tm1} - \frac{1}{T}\right)}}$$

Parallel Transitions (Mechanism 2)

$$S_{obs_j} = \frac{S_{n1_j}\, e^{\frac{dH_1}{R}\left(\frac{1}{Tm1} - \frac{1}{T}\right)} + S_{n2_j}\, e^{\frac{dH_2}{R}\left(\frac{1}{Tm2} - \frac{1}{T}\right)} + S_{u_j}}{1 + e^{\frac{dH_1}{R}\left(\frac{1}{Tm1} - \frac{1}{T}\right)} + e^{\frac{dH_2}{R}\left(\frac{1}{Tm2} - \frac{1}{T}\right)}}$$

Two Transitions (Mechanism 3)

$$S_{obs_j} = \frac{S_{n_j} + S_{1_j}\, e^{\frac{-dH_1}{R}\left(\frac{1}{Tm1} - \frac{1}{T}\right)} + S_{u_j}\, e^{\frac{-dH_1}{R}\left(\frac{1}{Tm1} - \frac{1}{T}\right)} e^{\frac{-dH_2}{R}\left(\frac{1}{Tm2} - \frac{1}{T}\right)}}{1 + e^{\frac{-dH_1}{R}\left(\frac{1}{Tm1} - \frac{1}{T}\right)} + e^{\frac{-dH_1}{R}\left(\frac{1}{Tm1} - \frac{1}{T}\right)} e^{\frac{-dH_2}{R}\left(\frac{1}{Tm2} - \frac{1}{T}\right)}}$$

Three Transitions (Mechanism 4)

$$S_{obs_j} = \frac{S_{n_j} + S_{1_j}\, e^{\frac{-dH_1}{R}\left(\frac{1}{Tm1} - \frac{1}{T}\right)} + S_{2_j}\, e^{\frac{-dH_1}{R}\left(\frac{1}{Tm1} - \frac{1}{T}\right)} e^{\frac{-dH_2}{R}\left(\frac{1}{Tm2} - \frac{1}{T}\right)} + S_{u_j}\, e^{\frac{-dH_1}{R}\left(\frac{1}{Tm1} - \frac{1}{T}\right)} e^{\frac{-dH_2}{R}\left(\frac{1}{Tm2} - \frac{1}{T}\right)} e^{\frac{-dH_3}{R}\left(\frac{1}{Tm3} - \frac{1}{T}\right)}}{1 + e^{\frac{-dH_1}{R}\left(\frac{1}{Tm1} - \frac{1}{T}\right)} + e^{\frac{-dH_1}{R}\left(\frac{1}{Tm1} - \frac{1}{T}\right)} e^{\frac{-dH_2}{R}\left(\frac{1}{Tm2} - \frac{1}{T}\right)} + e^{\frac{-dH_1}{R}\left(\frac{1}{Tm1} - \frac{1}{T}\right)} e^{\frac{-dH_2}{R}\left(\frac{1}{Tm2} - \frac{1}{T}\right)} e^{\frac{-dH_3}{R}\left(\frac{1}{Tm3} - \frac{1}{T}\right)}}$$

Parallel Intermediate Model (Mechanism 5)

$$S_{obs_j} = \frac{e^{\frac{-dH_3}{R}\left(\frac{1}{Tm1}-\frac{1}{T}\right)}\left(S_{n1_j}e^{\frac{dH_1Tm_2(T-Tm_1)+dH_3Tm_1(T-Tm_2))}{R\cdot T\cdot Tm_1\cdot Tm_2}} + S_{n2_j}e^{\frac{(dH_1+dH_3)(T-Tm_2)}{R\cdot T\cdot Tm_2}} + S_{1_j}e^{\frac{dH_3\left(\frac{1}{Tm2}-\frac{1}{T}\right)}{R}} + S_u\right)}{1+e^{\frac{dH_1}{R}\left(\frac{1}{Tm1}-\frac{1}{T}\right)}+e^{\frac{dH_2}{R}\left(\frac{1}{Tm2}-\frac{1}{T}\right)}+e^{\frac{-dH_3}{R}\left(\frac{1}{Tm3}-\frac{1}{T}\right)}}$$

233

Table C1. Parameters from SVD analysis of melting data for 6-FAM-143D-Tamra in Figure 31 (main text).

Index	Singular Value	% of Total Variance	**V**-Matrix Autocorrelation Coefficient
1	1276.9	98.36	0.985
2	164.5	1.63	0.974
3	8.3	0.004	0.897
4	3.6	0.0008	0.914
5	0.7	2.59E-05	0.709

Table C2. Model-dependent fitting statistics for 6-FAM-143D-Tamra.

	Model #	χ^2	AIC Value
Single Transition	1	3658.3	813.05
Two Transitions	2	106.1	-336.28
Parallel Transitions	3	106.1	-336.28
Three Transitions	4	47.2	-589.98
Parallel Intermediate Transitions	5	53.9	-546.28

Table C3. Parameters from SVD analysis of melting data for 2GKU in Figure C6.

Index	Singular Value	% of Total Variance	**V**-Matrix Autocorrelation Coefficient
1	3724.5	97.74	0.978
2	495.4	1.73	0.981
3	252.7	0.45	0.936
4	94.2	0.06	0.796
5	37.3	0.001	0.433

Table C4. Model-dependent fitting statistics for 2GKU.

	Model #	χ^2	AIC Value
Single Transition	1	146810.2	1251.48
Two Transitions	2	12833.9	815.06
Parallel Transitions	3	14103.3	832.42
Three Transitions	4	6950.9	714.23
Parallel Intermediate Transitions	5	5606.8	674.69

Table C5. Parameters from SVD analysis of melting data for 2AQY + 6 in Figure C7.

Index	Singular Value	% of Total Variance	$\mathbf{V}$-Matrix Autocorrelation Coefficient
1	1504.6	95.26	0.972
2	309.0	4.02	0.951
3	100.5	0.42	0.909
4	56.6	0.14	0.452
5	39.9	0.07	0.464

Table C6. Model-dependent fitting statistics for 2AQY + 6.

Model	Model #	χ^2	AIC Value
Single Transition	1	96018.8	958.36
Two Transitions	2	28060.9	793.21
Parallel Transitions	3	28060.9	793.21
Three Transitions	4	22878.1	775.81
Parallel Intermediate Transitions	5	23092.7	777.15

Table C7. Parameters from SVD analysis of melting data for Tel-3G in Figure C8.

Index	Singular Value	% of Total Variance	**V**-Matrix Autocorrelation Coefficient
1	1645	81.65	0.957
2	731.6	16.15	0.97
3	196.8	1.17	0.786
4	113.4	0.39	0.06
5	92.4	0.26	-0.2

Table C8. Model-dependent fitting statistics for Tel-3G.

Model	Model #	χ^2	AIC Value
Single Transition	1	15748.6	840.7
Two Transitions	2	6316.5	684.6
Parallel Transitions	3	6316.5	684.6
Three Transitions	4	6316.5	696.6
Parallel Intermediate Transitions	5	4794.2	645.9
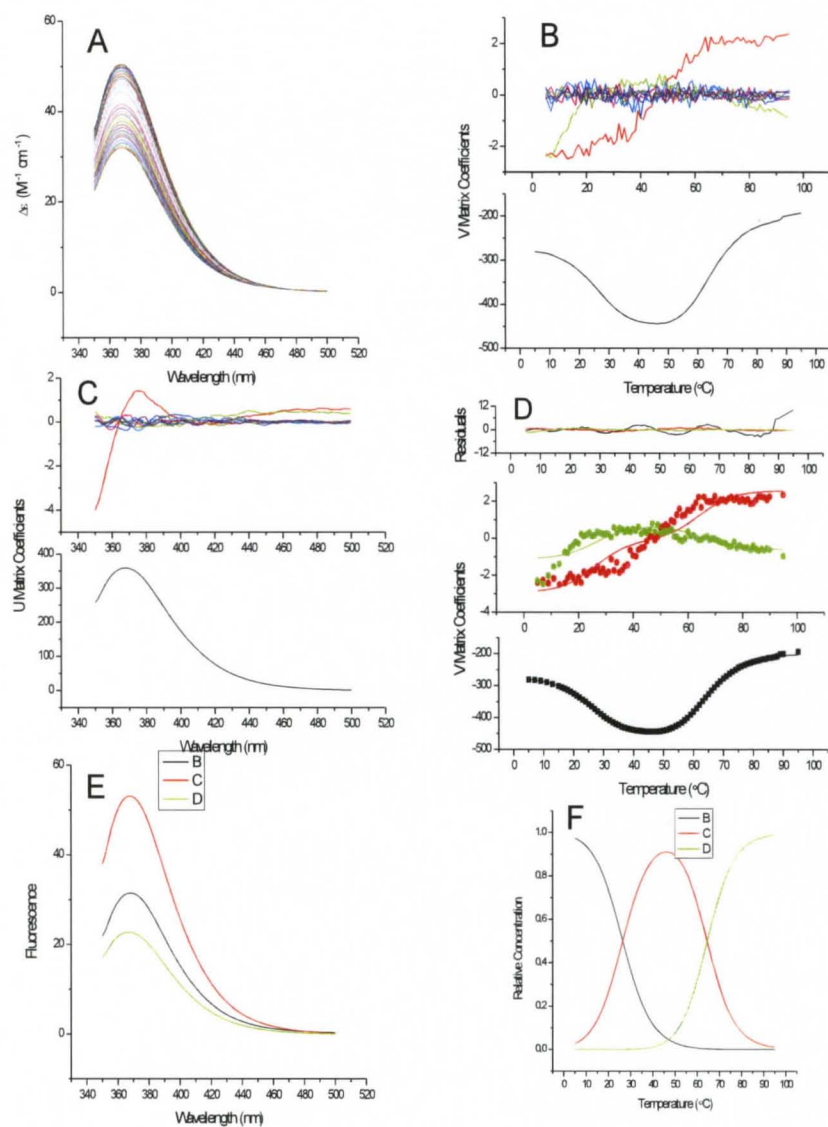
Figure C1. Data for temperature-induced unfolding of Tel22AP1 in 25 mM KCl, pH 7.0. (A)

Fluorescence emission spectra of Tel22AP1 as a function of temperature between 5 °C and 95

°C. (B) **V** matrix coefficients for amplitude vectors V_2-V_5 (top) and V_1 (bottom). (C) **U** matrix

coefficients for vectors U_2-U_5 (top) and U_1 (bottom). (C) SVD analysis of the data set indicated

three significant **V** vectors; thus V_1-V_3 were fit to the analytical expression associated with model

4 (see text). The fit to V_1 (bottom panel) and V_2 and V_3 (top panel) matrix coefficients is

illustrated with the points indicating the elements of each vector and the lines showing the least squares fit. Residuals are plotted at the top. (E) Calculated emission spectra of folded (black), intermediate (red), and unfolded ensembles (green). (F) Fractional distribution of starting (black), intermediate (red) and unfolded ensembles (green) as a function of temperature. [143D-AP1] = 1.7 μM.
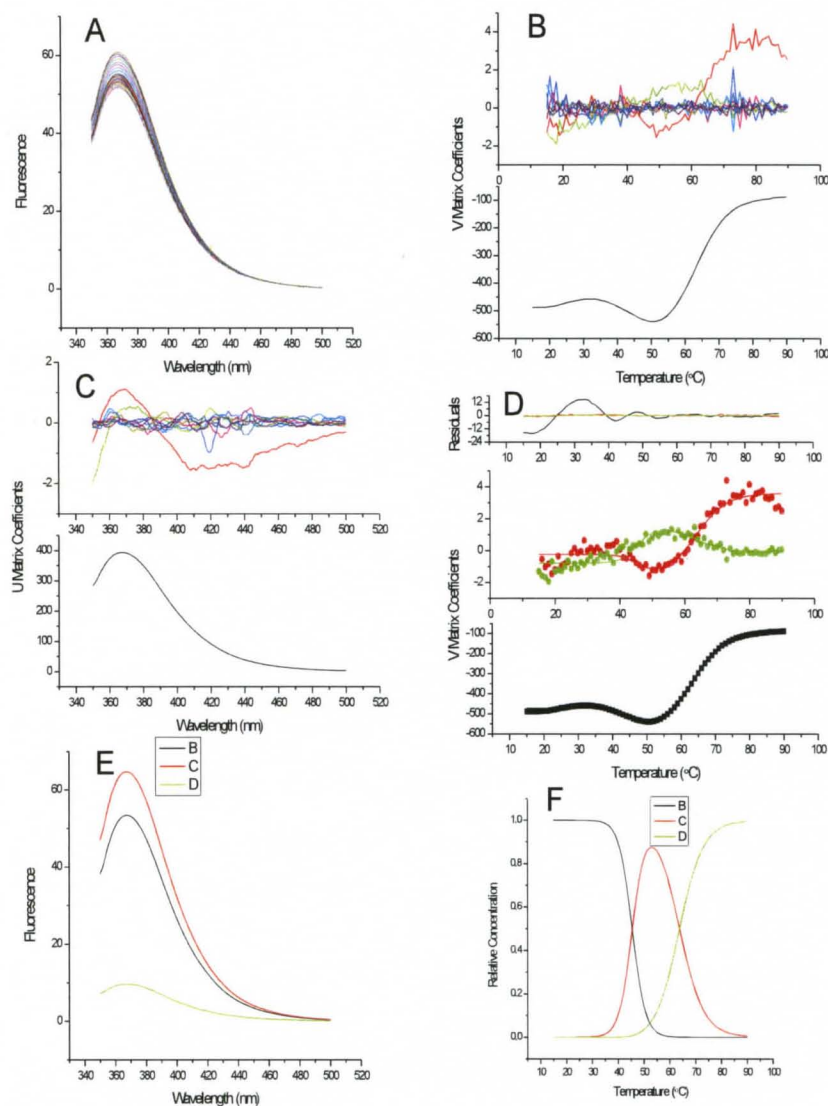
Figure C2. Data for temperature-induced unfolding of Tel22AP7 in 25 mM KCl, pH 7.0. (A)
Fluorescence emission spectra of Tel22AP7 as a function of temperature between 15 °C and 90
°C. (B) V matrix coefficients for vectors V_2-V_5 (top) and V_1 (bottom). (C) U matrix coefficients for
vectors U_2-U_5 (top) and U_1 (bottom). (C) SVD analysis of the data set indicated three significant
V vectors; thus V_1-V_3 were fit to the analytical expression associated with model 4 (see text).
The fit to V_1 (bottom panel) and V_2 and V_3 (top panel) matrix coefficients is illustrated with the

points indicating the elements of each vector and the lines showing the least squares fit. Residuals are plotted at the top. (E) Calculated emission spectra of folded (black), intermediate (red), and unfolded ensembles (green). (F) Fractional distribution of starting (black), intermediate (red) and unfolded ensembles (green) as a function of temperature. [143D-AP7] = 1.0 μM.
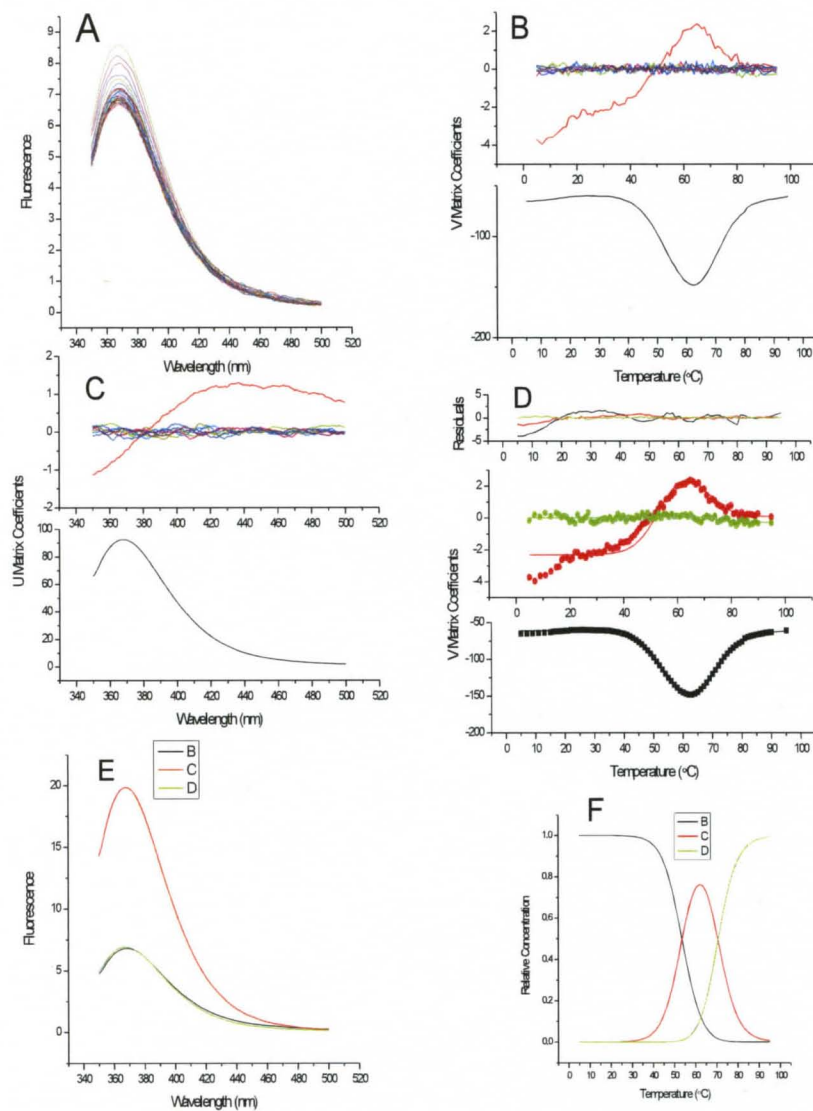
Figure C3. Data for temperature-induced unfolding of Tel22AP13 in 25 mM KCl, pH 7.0. (A) Fluorescence emission spectra of Tel22AP13 as a function of temperature between 5 °C and 95 °C. (B) **V** matrix coefficients for vectors V_2-V_5 (top) and V_1 (bottom). (C) **U** matrix coefficients for vectors U_2-U_5 (top) and U_1 (bottom). (C) SVD analysis of the data set indicated three significant **V** vectors; thus V_1-V_3 were fit to the analytical expression associated with model 4 (see text). The fit to V_1 (bottom panel) and V_2 and V_3 (top panel) matrix coefficients is illustrated with the

points indicating the elements of each vector and the lines showing the least squares fit.

Residuals are plotted at the top. (E) Calculated emission spectra of folded (black), intermediate

(red), and unfolded ensembles (green). (F) Fractional distribution of starting (black),

intermediate (red) and unfolded ensembles (green) as a function of temperature. [143D-AP13]
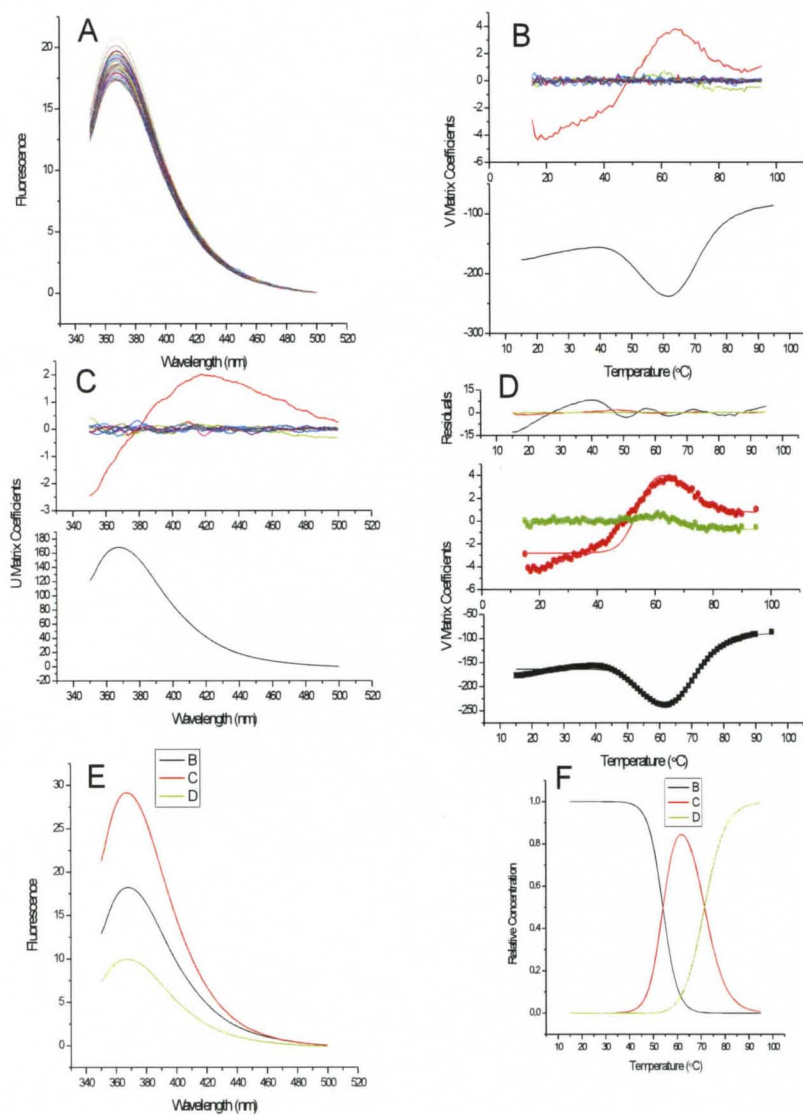
= 1.0 μM.

Figure C4. Data for temperature-induced unfolding of Tel22AP19 in 25 mM KCl, pH 7.0. (A)

Fluorescence emission spectra of Tel22AP19 as a function of temperature between 15 °C and 95

°C. (B) V matrix coefficients for vectors V_2-V_5 (top) and V_1 (bottom). (C) U matrix coefficients for

vectors U_2-U_5 (top) and U_1 (bottom). (C) SVD analysis of the data set indicated three significant

V vectors; thus V_1-V_3 were fit to the analytical expression associated with Mechanism 4 (see

text). The fit to V_1 (bottom panel) and V_2 and V_3 (top panel) matrix coefficients is illustrated with the points indicating the elements of each vector and the lines showing the least squares fit. Residuals are plotted at the top. (E) Calculated emission spectra of folded (black), intermediate (red), and unfolded ensembles (green). (F) Fractional distribution of starting (black), intermediate (red) and unfolded ensembles (green) as a function of temperature. [143D-AP19] = 1.1 μM.
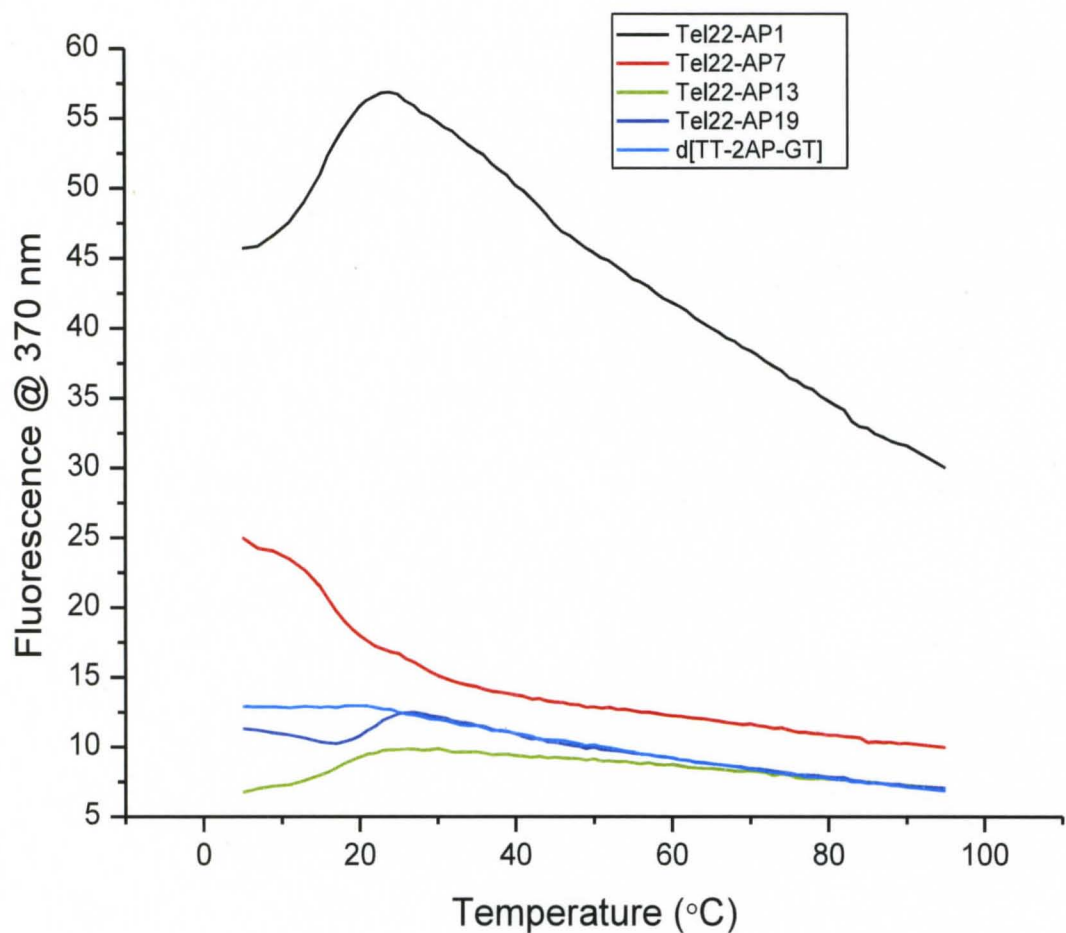
Figure C5. Effect of temperature on fluorescence intensity of Tel22-2AP and control

oligonucleotide d(TT-2AP-GT) in absence of added KCl. Conditions: 1 μM 143D-2AP

oligonucleotide or 0.8 μM control oligonucleotide in 10 mm *t*-butylammonium phosphate, 1 mm

EDTA, pH 7.0. There was no appreciable effect of 100 mM KCl on the temperature-dependence

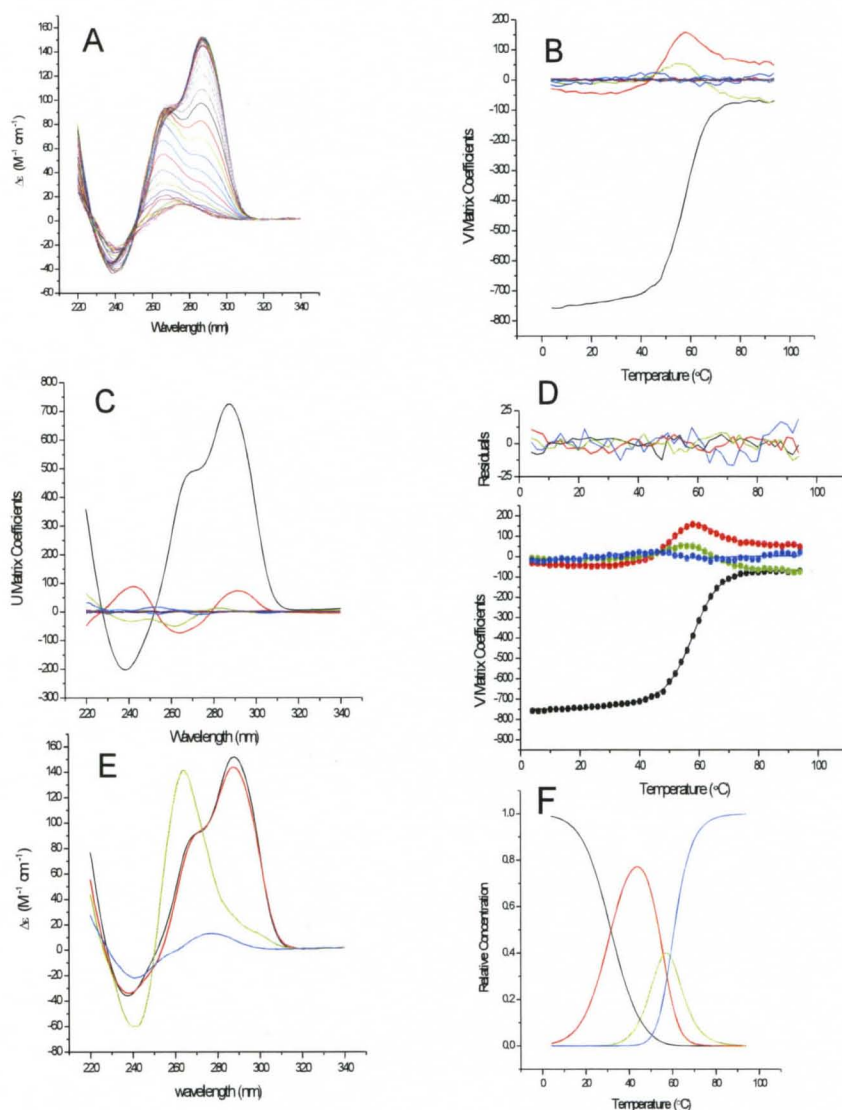of 2AP emission of the control oligonucleotide.

Figure C6. Data for temperature-induced unfolding of 2GKU in 25 mM KCl, pH 7.0. (A) CD

spectra of 2GKU oligonucleotide as a function of temperature between 4 °C and 95 °C. (B) V

matrix coefficients for vectors V_2-V_5 (top) and V_1 (bottom). (C) U matrix coefficients for vectors

U_1-U_5. (C) SVD analysis of the data set indicated four significant V vectors; thus V_1-V_4 were fit to

the analytical expression associated with model 4 (see text). The fit to V_1 (black) and V_2 (red), V_3

(green) and V_4 (blue) matrix coefficients is illustrated with the points indicating the elements of

each vector and the lines showing the least squares fit. Residuals for each are plotted at the top. (E) Calculated CD spectra of starting (black), I_1 (red), I_2 green) and unfolded ensembles (blue). (F) Fractional distribution of starting (black), I_1 (red), I_2 (green) and unfolded ensembles (blue) as a function of temperature.
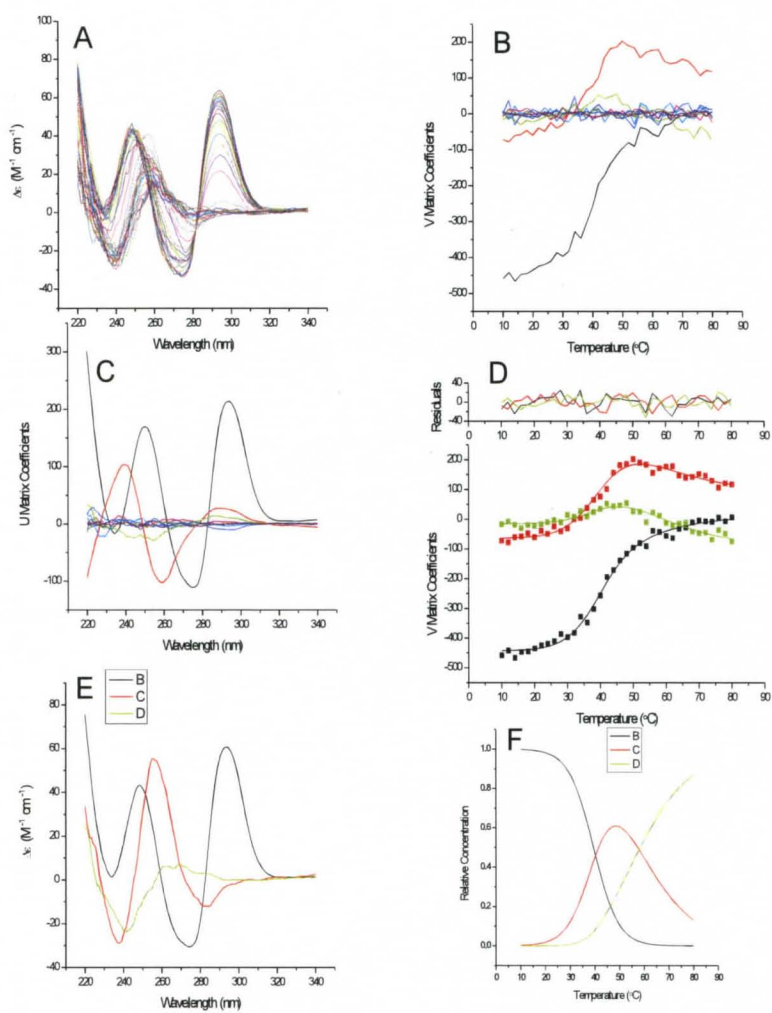
Figure C7. Data for temperature-induced unfolding of 2AQY + 6 in 100 mM NaCl, pH 7.0. (A) CD spectra of 2AQY + 6 oligonucleotides as a function of temperature between 4 °C and 95 °C. (B) V matrix coefficients for vectors V_2-V_5 (top) and V_1 (bottom). (C) U matrix coefficients for vectors U_1-U_5. (C) SVD analysis of the data set indicated three significant V vectors; thus V_1-V_3 were fit to the analytical expression associated with model 4 (see text). The fit to V_1 (black) and V_2 (red) and V_3 (green) matrix coefficients is illustrated with the points indicating the elements of each vector and the lines showing the least squares fit. Residuals are plotted at the top. (E) Calculated CD spectra of folded (black), intermediate (red), and unfolded ensembles (green). (F)

Fractional distribution of starting (black), intermediate (red) and unfolded ensembles (green) as a function of temperature.

Figure C8. Data for temperature-induced unfolding of Tel-3G in 100 mM KCl, pH 7.0. (A) CD

spectra of Tel-3G as a function of temperature between 4 °C and 95 °C. (B) V matrix coefficients

for vectors V_2-V_5 (top) and V_1 (bottom). (C) U matrix coefficients for vectors U_1-U_5. (D) SVD

analysis of the data set indicated three significant V vectors; thus V_1-V_3 were fit to the analytical

expression associated with model 4 (see text). The fit to V_1 (black) and V_2 (red) and V_3 (green)

matrix coefficients is illustrated with the points indicating the elements of each vector and the

lines showing the least squares fit. Residuals are plotted at the top. (E) Calculated CD spectra of folded (black), intermediate (red), and unfolded ensembles (green). (F) Fractional distribution of starting (black), intermediate (red) and unfolded ensembles (green) as a function of temperature.
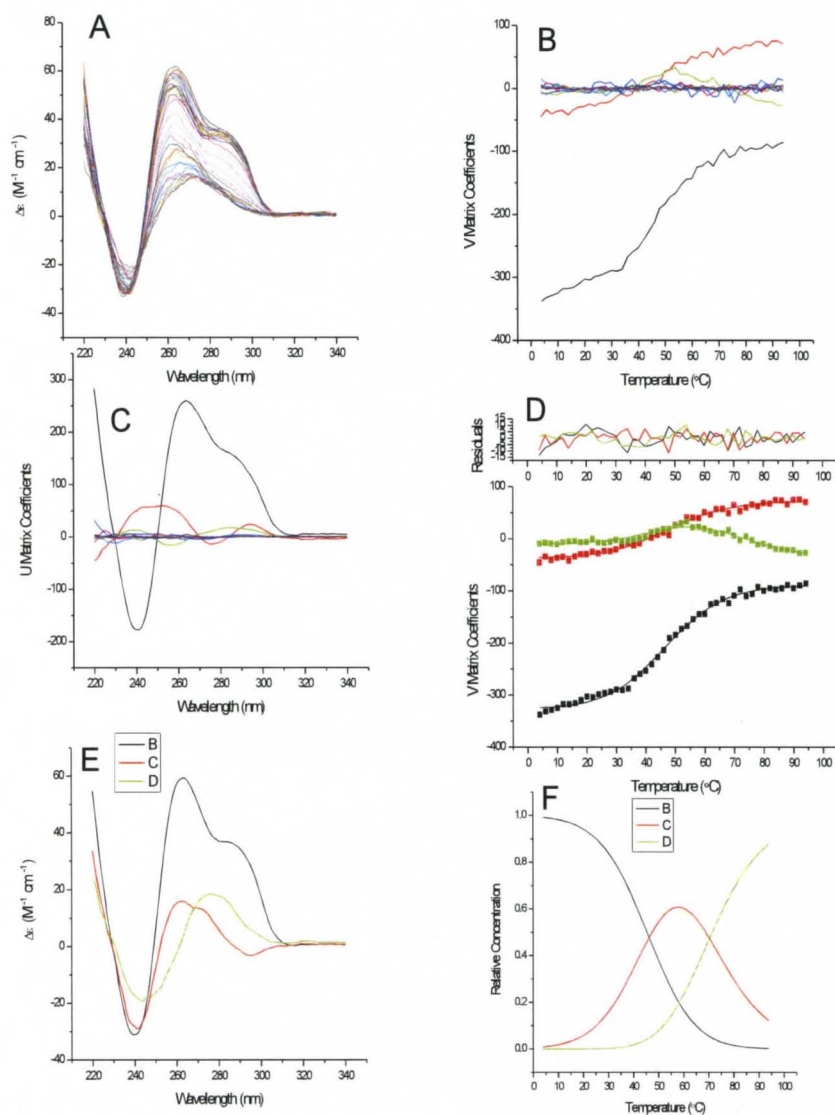
APPENDIX D

PUBLICATIONS, REPRINT PERMISSIONS, AND CONTRIBUTIONS.

Chapter 2

Buscaglia, R., Jameson, D. M., & Chaires, J. B. (2012). G-quadruplex structure and stability illuminated by 2-aminopurine phasor plots. *Nucleic acids research*, *40*(9), 4203–15. doi:10.1093/nar/gkr1286

Permission for use was obtained from Oxford Journals through RightsLink.

RB contributed the collection and analysis of all fluorescence data, production of all

figures and results, and preparation of manuscript. JBC and DMJ provided intellectual novelty

and supervision of all conducted research and manuscript submission.

Chapter 3

Miller, M. C., Buscaglia, R., Chaires, J. B., Lane, A. N., & Trent, J. O. (2010). Hydration Is a Major Determinant of the G-Quadruplex Stability and Conformation of the Human Telomere 3' Sequence of d(AG(3)(TTAG(3))(3)). *Journal of the American Chemical Society*, *3*, 17105–17107. doi:10.1021/ja105259m

Permission for use was obtained from ACS Publications through RightsLink.

MCM prepared manuscript and provided intellectual novelty. RB prepared all circular

dichroism results and performed data analysis. ANL conducted all NMR experimentation and

analysis. JBC provided intellectual novelty and supervision of manuscript preparation. JOT

conducted all computational investigations, provided intellectual novelty, and supervision of

manuscript preparation.

Chapter 4

Buscaglia, R., Miller, M. C., Dean, W. L., Gray, R.D., Lane, A. N., Trent, J. O., & Chaires, J. B. (2012) Polyethylene glycol alters G-quadruplex structure by conformational selection via differential binding. *Journal of the American Chemical Society*, Submitted.

RB performed all circular dichroism spectroscopy data collection and analysis and was

responsible for manuscript preparation. MCM provided help with data collection. WLD

conducted analytical ultracentrifugation and analysis. JOT conducted computation

investigations, provided intellectual novelty, and supervised manuscript submission. RDG, ANL,

and JBC provided intellectual novelty and supervision of manuscript submission.

Chapter 5

Gray, R. D., Buscaglia, R., & Chaires, J. B. (2012). Populated intermediates in the thermal unfolding of the human telomeric quadruplex. *Journal of the American Chemical Society*, *134*(40), 16834–44. doi:10.1021/ja307543z

Permission for use obtained from ACS Publications through RightsLink.

RDG conducted spectroscopic investigations, applied analytical models, and prepared

manuscript for submission. RB was involved in spectroscopic data collection and analysis,

developed analytical models and statistical analysis, and was part of manuscript preparation.

JBC provided intellectual novelty and supervision of manuscript preparation.

Chapter 6

Buscaglia, R., Gray, R. D., & Chaires, J. B. (2012) Thermodynamic Characterization of Human

Telomere Quadruplex Unfolding. *Biochemistry*. In Preparation.

RB conducted spectroscopic and calorimetric investigations, developed and applied

analytical models, and prepared manuscript for submission. RDG was involved in spectroscopic

data collection and analysis, provided intellectual novelty, and supervised manuscript

preparation. JBC oversaw preparation of manuscript and provided intellectual novelty.

CURRICULUM VITAE

Robert Buscaglia

James Graham Brown Cancer Center

University of Louisville

505 S. Hancock St.

Louisville, KY 40202

Home Address:

1820 Cimmaron Trail

Louisville, KY 40223

Telephone: 502-296-3637 (cell)

Email: r0busc01@louisville.edu

Education

Northern Arizona University

Degrees:

B.S. Chemistry, ACS Biochemistry Emphasis, May 2007

B.S. Mathematics, Applied Mathematics Emphasis, May 2007

Graduated Summa Cum Laude with Honors

Professional memberships

Biophysical Society, 2008-Current

American Chemical Society, 2007-Current

Honors and awards

University of Louisville School of Medicine IPIBS Fellowship, 2007-2009

Northern Arizona University President's Prize, 2007

Northern Arizona University Golden Axe Award, 2007

Northern Arizona University Outstanding Senior in Mathematics and Statistics, 2007

USA Today's Third Team All-USA College Academic Team, 2006

Arnold and Mabel Beckman Foundation Scholar, 2006

V. M. Slipher Award in Mathematics, 2006

V. M. Slipher Award in Chemistry and Biochemistry, 2006

Barry M Goldwater Scholar, 2005-2006

Raytheon Corporation Scholarship, 2005

Russ Powell Scholarship, 2005

Northern Arizona University Outstanding Mathematics Achievement at the Senior Level, 2005-2007

Hooper Undergraduate Research Award, 2004-2007

Northern Arizona University Dean's List 2003-2007

Training

Spring 2008 - Advanced Analytical Ultracentrifugation Workshop. Boston Biomedical Research Institute, Boston, MA.

Spring 2007 - Principles of Fluorescence Techniques. University of Chicago, Chicago, IL.

Campus Activities

Attended Annual James Graham Brown Cancer Center Retreats, University of Louisville, Louisville, KY, 2008-2011

Attended 12th-14th Annual Institute of Molecular Diversity and Drug Design (IMD^3) Symposium, University of Louisville, Louisville, KY. 2008-2012

Attended University of Louisville Biochemistry and Molecular Biology Department Retreats, Louisville, KY. 2009, 2011.

Guest lecture on Fluorescence in Biomolecular Interactions graduate level course. University of Louisville, 2010.

Participated in National Science Foundation sponsored programs "Bridges to the Baccalaureate" and "Research Experience for Undergraduates" helping train Native American Students in research laboratory settings. Flagstaff, AZ. 2004-2005

Invited to give a speech on pursuing a college education to high school seniors. Highland High School, Gilbert, AZ. 2005

Managed undergraduate research projects in Dr. Edwin A. Lewis' laboratory. 2004-2007

Presentations

56th Annual Biophysical Society Meeting. (2012) "Phasor Diagram Analysis of Telomeric G-Quadruplex Stability and Heterogeneity." Poster Presentation in San Diego, CA.

24th Annual Gibbs Conference on Biothermodynamics. (2010) "Alteration of Human Telomere G-Quadruplex Conformation and Stability by Acetonitrile." Poster Presentation in Carbondale, IL

2nd International Meeting on G-Quadruplex DNA. (2009) "Acrylamide Quenching Data Analysis and Fluorescent Lifetimes of 2-Aminopurine Labeled G-Quadruplexes." Poster Presentation in Louisville, KY.

58th Southeastern Regional ACS Meeting, invited 20 minute oral presentation. (2006) "Calorimetric and Spectroscopic Evidence for the Binding of TMPyP4 to a 23-mer c-MYC i-Motif Mutant." Augusta, GA.

20th Annual Gibbs Conference on Biothermodynamics. (2006) "Influences of temperature, ionic strength and pH on the stability of a mutant c-MYC i-Motif." Poster Presentation in Carbondale, IL.

Hurley Quadruplex Retreat, invited 25 minute oral presentation. (2006) "Binding of TMPyP4 to the mutant c-MYC i-motif sequence: solution and stability effects." Carefree, AZ.

19th Annual Gibbs Conference on Biothermodynamics. (2005) "Nonlinear Regression Fitting Algorithms for Multiple Independent Binding Equilibria, Monte Carlo Error Analysis,   and Saroff Distributions." Poster Presentation in Carbondale, IL.

61st Southwest and 57th Southeast Joint Meeting of the American Chemical Society, (2005). "Conformational Equilibria of NHE-III1 c-MYC Promoter Quadruplex/i-Motif Structures." Oral presentation in Memphis, TN.

Hurley Quadruplex Retreat, invited 20 minute oral presentation. (2005) "DSC Investigations of the Structural Stability of the NHE-III1 Promoter Element." Scottsdale,AZ.

Hurley Quadruplex Retreat, invited 15 minute oral presentation. (2004) "Thermodynamic Study of Interactions Between TMPyP4 and the Complete c-MYC Silencer Element." Sedona, AZ.

Publications

1. **Buscaglia, R.**, Miller, M.C., Dean, W.L., Gray, R.D., Lane, A.N., Trent, J.O., and Chaires, J.B. (2012) "Polyethylene glycol alters G-quadruplex structure by conformational selection via differential binding." *J. Am. Chem. Soc.* Submitted.

2. Le, V., **Buscaglia, R.**, Chaires, J.B., and Lewis, E.A. (2012) "Modeling complex equilibria in isothermal titration calorimetry experiments: estimation of thermodynamic parameters for multiple competitive binding processes." *Anal. Bioc.* Submitted.

3. Le, H.T., Miller, M.C., **Buscaglia, R.**, Dean, W.L., Holt, P.A., Chaires, J.B., and Trent, J.O. (2012) "Not all G-quadruplexes are created equally: an investigation of the structural polymorphism of the c-Myc G-quadruplex-forming sequence and its interaction with the porphyrin TMPyP4." *Org Biomol Chem.* 10(47), 9393-404.

4. Gray, R. D., Buscaglia, R., & Chaires, J. B. (2012). Populated intermediates in the thermal unfolding of the human telomeric quadruplex. *J. Am. Chem. Soc.* 134(40), 16834–44.

5. Le, H.T., **Buscaglia, R.**, Dean, W.L., Chaires, J.B., and Trent, J.O. (2012) "Calculation of hydrodynamic properties of G-quadruplex nucleic acid structures from in silico bead models." *Top. Curr. Chem.* doi:10.1007/128

6. **Buscaglia, R.**, Jameson, D.M., and Chaires, J.B. (2012) "G-quadruplex structure and stability illuminated by 2-aminopurine phasor plots." *Nucleic Acids Res.* 40(9), 4203-15.

7. Bhattacharjee, A.J., Ahluwalia, K., Taylor, S., Jin, O., Nicoludis, J.M., **Buscaglia, R.**, Chaires, J.B., Kornfilt, D.J., Marquardt, D.G., and Yatsunyk, L.A. (2011) "Induction of G-quadruplex DNA structure by Zn(II) 5,10,15,20-tetrakis(N-methyl-4-pyridyl)porphyrin." *Biochemie.* 93(8), 1297-309.

8. Holt, P.A., **Buscaglia, R.,** Trent, J.O., and Chaires, J.B. (2011) "A discovery funnel for nucleic acid binding drug candidates." *Drug Dev. Res.* 72(2), 178-186.

9. Dettler, J.M., **Buscaglia, R.,** Le, V.H., and Lewis, E.A. (2011) "DSC deconvolution of the structural complexity of c-MYC P1 promoter G-quadruplexes." *Biophys. J.* 100(6), 1517-25.

10. Miller, M.C., **Buscaglia, R.,** Chaires, J.B., Lane, A.N., and Trent, J.O. (2010) "Hydration is a major determinant of the G-quadruplex stability and conformation of the human telomere 3' sequence of d(AG(3)(TTAG(3))(3))." *J. Am. Chem. Soc.*

11. Dettler, J.M., **Buscaglia R.,** Cui, J., Cashman, D., Blynn, M., and Lewis, E.A. (2010) "Biophysical characterization of an ensemble of intramolecular i-motifs formed by the human c-MYC NHE III1 P1 promoter mutant sequence." *Biophys. J.* 99(2), 561-7.

12. Nagesh, N., **Buscaglia, R.,** Dettler, J.M., and Lewis, E.A. (2010) "Studies on the site and mode of TMPyP4 interactions with Bcl-2 promoter sequence G-quadruplexes." *Biophys. J.* 98(11), 2628-33.

13. Gray, R.D., Petraccone, L., **Buscaglia, R.,** and Chaires, J.B. (2010) "2-aminopurine as a probe for quadruplex loop structures." *Methods Mol. Biol.* 608, 121-36.

14. Cashman, D.J., **Buscaglia, R.,** Freyer, M.W., Dettler, J., Hurley, L.H., and Lewis, E.A. (2008) "Molecular modeling and biophysical analysis of the c-MYC NHE-III1 silencer element." *J. Mol. Model.* 14(2), 93-101.

15. Freyer, M.W., **Buscaglia, R.,** Hollingsworth, A., Ramos, J., Blynn, M., Pratt, R., Wilson, W.D., and Lewis, E.A. (2007) "Break in the heat capacity change at 303 K for complex binding of netropsin to AATT containing hairpin DNA constructs." *Biophys. J.* 92(7), 2516-22.

16. Freyer, M.W., **Buscaglia, R.,** Kaplan, K., Cashman, D., Hurley, L.H., and Lewis, E.A. (2007) "Biophysical studies of the c-MYC NHE III1 promoter: model quadruplex interactions with cationic porphyrin." *Biophys. J.* 92(6), 2007-15.

17. Freyer, M.W., **Buscaglia, R.,** Cashman, D., Hyslop, S., Wilson, W.D, Chaires, J.B., and Lewis, E.A. (2007) "Binding of netropsin to several DNA constructs: evidence for at least two different 1:1 complexes formed from an –AATT– containing ds-DNA construct and a single minor groove binding ligand." *Biophys. Chem.* 126(1-3), 186-96.

18. Freyer, M.W., **Buscaglia, R.,** Nguyen, B., Wilson, W.D., and Lewis, E.A. (2006) "Binding of netropsin and 4,6-diamidino-2-phenylindole to an A2T2 DNA hairpin: a comparison of biophysical techniques." *Anal. Biochem.* 355(2), 259-66.