8-2014

Biophysical and computational investigations into G-quadruplex structural polymorphism and interaction with small molecules.

Huy Tuan Le
University of Louisville

Follow this and additional works at: https://ir.library.louisville.edu/etd

Part of the Biochemistry Commons, and the Molecular Biology Commons

www.manaraa.com

BIOPHYSICAL AND COMPUTATIONAL INVESTIGATIONS INTO
G-QUADRUPLEX STRUCTURAL POLYMORPHISM AND
INTERACTION WITH SMALL MOLECULES

By

Huy Tuan Le
B.S., University of Louisville, 2008
M.S., University of Louisville, 2013

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

August 2014

BIOPHYSICAL AND COMPUTATIONAL INVESTIGATIONS INTO
G-QUADRUPLEX STRUCTURAL POLYMORPHISM AND
INTERACTION WITH SMALL MOLECULES

By

Huy Tuan Le
B.S., University of Louisville, 2008
M.S., University of Louisville, 2013

A Dissertation Approved on

June 16, 2014

by the following Dissertation Committee

_____
Dissertation Director
John O. Trent

_____
Jonathan B. Chaires

_____
Paula J. Bates

_____
Douglas S. Darling

_____
William L. Dean

ii

DEDICATION

This dissertation and all that it represents are dedicated to

my parents

Mr. Dai-Chi Thanh Le and Mrs. Lan Kim-Mai Le

and my grandparents

Mr. Hiep Phai Le and Mrs. Ngoc-Anh Thi Dao (paternal)

and

Mr. Khanh Kim Mai and Mrs. My-Luoc Thi Nguyen (maternal)

who were the first teachers in my life, who imparted on me the importance of education,

and whose continued encouragement and sacrifices

have inspired me to be where I am today.

Uống nước nhớ nguồn
Làm con phải hiếu.
Ai ơi hãy nhớ năm xưa
Những ngày còn thơ
Công ai nuôi dưỡng.

Công đức sinh thành
Người ơi! đừng quên.
Công cha như núi Thái Sơn
Nghĩa mẹ như nước
Trong nguồn chảy ra.

-Ơn Nghĩa Sinh Thành (Vietnamese Folksong)

ACKNOWLEDGMENTS

This dissertation would not be possible without the support and encouragement of many people. First and foremost, I would like to acknowledge the members of my dissertation committee. I am grateful to John O. Trent, my Mentor and Dissertation Director, for his guidance and support. My pursuit of this doctoral degree is the direct result of his encouragement. In addition, I would like to thank Jonathan B. Chaires for his financial support of my dissertation research. As a collaborator, I am grateful for his advice, encouragement, and the occasional admonishment. I thank William L. Dean for his continuing friendship. As a longtime member of the Department of Biochemistry and Molecular Biology, he has observed my growth from undergraduate studies and I am thankful for his support and encouragement along the way. Last but not least, I thank Paula J. Bates and Douglas S. Darling for agreeing to serve on the dissertation committee and for their discussion, criticism, and support of my research.

In addition to the members of my dissertation committee, I would like to thank all the friends, colleagues, and collaborators in the James G. Brown Cancer Center, Department of Biochemistry and Molecular Biology, and the University of Louisville M.D./Ph.D. Program. In especially, I would like to thank Jonathan M. Maguire for his hard work in maintaining the computers needed for the many computational studies. Thank you to M. Clarke Miller for sharing his expertise in chromatographic techniques. Thank you to Robert D. Gray for being available to discuss G-quadruplex whenever I had questions.

Thank you to Lynn W. DeLeeuw for all her help with experiments. Thank you to Barbara J. Clark and Stephen R. Ellis for their guidance. I thank Robert Buscaglia, Rafael G. Del Villar, Indira Chaudury, Nikhil A. Shukla, Nichola C. Garbett, and Alfred B. Jenson for their support and encouragement. In addition, I was fortunate to work with several highly productive summer students, Michael T. Miao, John D. Gettelfinger, Nicholas Siow, and Yidi Huang, who helped generate the data for several manuscripts in preparation. I want to remember Michael W. Gordon who passed away too young. You will be missed dearly.

I reserve my deepest gratitude and appreciation for my family, without whose support and encouragement, this accomplishment would not be possible. My first thank you is to my beautiful wife, Saasha H. Le, for all her help in proofreading this dissertation. Thank you for always standing by me. I love you and I look forward our life together. Thank you to my parents, Dai-Chi T. Le and Lan K.M. Le, for your support and encouragement. You have given up everything to move to a new and foreign country so that I might have the opportunity to succeed. This dissertation is as much your accomplishment as it is mine. Your sacrifices, both mentally and physically, remain an inspiration to me. Thank you to my sister, Kim-Mai T. Le, who has grown from a precocious little girl to a beautiful young lady. Thank you for allowing me the privilege of being one of your role models. Thank you to my grandparents, Khanh K. Mai and My-Luoc T. Nguyen, for your love and support. I would like to thank my uncle, Cuong K. Mai, for his support and encouragement. Lastly, I want to acknowledge my cousin, Hung L.T. Mai, who looks up to me as a big brother. I hope that you will take this dissertation as an encouragement for your own educational journey. Remember that with hard work, dedication, and perseverance, anything is possible.

ABSTRACT

BIOPHYSICAL AND COMPUTATIONAL INVESTIGATIONS INTO

G-QUADRUPLEX STRUCTURAL POLYMORPHISM AND

INTERACTION WITH SMALL MOLECULES

Huy Tuan Le

16 June 2014

In the cell, guanine-rich nucleic acids can self-assemble into unique four stranded tertiary structures known as G-quadruplexes. G-quadruplex formation in the telomere leads inhibits telomerase, an enzyme activated in cancer cells to maintain the telomere and allowing for cancer cells to achieve immortality. G-quadruplex formation in the promoters and 5'-untranslated regions regulates the expression of many oncogenes. Furthermore, G-quadruplex formation during cellular replication promotes genomic instability, a characteristic which enables tumor development. Because of their implication in cancer, G-quadruplex structures have emerged as attractive drug targets for anti-tumor therapeutics. In the current dissertation work, we present three experimental approaches to investigate G-quadruplex structures, biophysical properties, small molecule interaction, and the thermodynamics of G-quadruplex formation. Current approaches to study G-quadruplex structures often employ sequence modifications or changes to the experimental condition, as a way of resolving the structural polymorphism associated with many

G-quadruplex-forming sequences, to select for a single conformation for high-resolution structural studies. Our strategy for resolving G-quadruplex structural polymorphism is superior in that the experimental approaches do not result in drastic perturbation of the system. In the first approach, we employed size exclusion chromatography to separate a mixture of G-quadruplex structures formed from a G-quadruplex-forming sequence. We demonstrated that it is possible to isolate distinct species of G-quadruplex structures for further biophysical studies. In the second approach, we employed hydrodynamic bead modeling to study the structural polymorphism of a G-quadruplex-forming sequence. We showed that properties calculated from models agreed with experimentally determined values and could be used to predict the folding of G-quadruplex-forming oligonucleotides whose high-resolution structures are ambiguous or not available. In our third approach, we presented a virtual screening platform that was successful in identifying a new G-quadruplex-interacting small molecule. The results of the virtual screen were validated with extensive biophysical testing. Our target for the virtual screen was a G-quadruplex structure generated *in silico*, which represents one approach to receptor-based drug discovery when high-resolution structures of the binding site are not available. Taken together, our three approaches represent a new paradigm for drug discovery from guanine-rich sequence to anti-cancer drugs.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

Within the cell, the predominant tertiary structures of nucleic acid molecules are double helices consisting of two complementary antiparallel strands associating through hydrogen bond interaction between purine and pyrimidine base pairs, adenine pairs with thymine or uracil and guanine pairs with cytosine (Ghosh and Bansal, 2003). Under certain conditions, however, guanine-rich nucleic acids can self-assemble to form G-quadruplexes, non-canonical tertiary structures consisting of four strands (each contains a run of two or more guanines) associating through G-quartet formation (Figure 1). In the G-quartet arrangement, the N1 and N2 nitrogen atoms of one guanine molecule act as the hydrogen bond donors for the O6 oxygen and N7 nitrogen atoms, respectively, of the next guanine molecule while its O6 and N7 atoms, in turn, serve as the hydrogen bond acceptors for the N1 and N2 atoms, respectively, of a previous guanine molecule. The result of this hydrogen bonding pattern is a square planar arrangement of four guanine molecules, which can stack onto one another to form a quadruple helix. Over the past 25 years, G-quadruplexes have emerged from "frequent nuisance[s] in the laboratory" (Guschlbauer *et al.*, 1990) into biologically relevant nucleic acid structures with significant implications in the regulation of cellular processes and the pathogenesis of

cancer and other diseases (Wu and Brosh, 2010, Bryan and Baumann, 2011, Collie and Parkinson, 2011, Zhou *et al.*, 2011, Bochman *et al.*, 2012, Bugaut and Balasubramanian, 2012, Baral *et al.*, 2013, Tarsounas and Tijsterman, 2013, Murat and Balasubramanian, 2014). For example, G-quadruplex formation in the telomere inhibits telomerase, an enzyme responsible for maintaining the length of the telomere (Blackburn, 1991). In over 90% of cancers, the reactivation of telomerase allow the cells to evade senescence and attain immortality (Shay and Bacchetti, 1997), one hallmark of cancer cell transformation (Hanahan and Weinberg, 2000, Hanahan and Weinberg, 2011). In addition to enabling replicative immortality, G-quadruplex formation is also implicated in six other hallmarks of cancer (Table 1). Furthermore, G-quadruplex formation has also been shown to contribute to genomic instability and mutation, an enabling characteristic in tumor development. The implication of G-quadruplex structures in cancer has led to an interest in these structures as potential drug targets for the development of novel anti-cancer drugs. However, the structural polymorphism associated with G-quadruplex-forming sequences has severely hindered investigation of G-quadruplex biophysical properties, small molecule interaction, and the thermodynamics of formation. A critical limitation of current strategies for investigating G-quadruplex structures is the artificial reduction of structural polymorphism through modifications of the G-quadruplex-forming sequence or alterations of experimental conditions. The untested assumption is that such approaches select for an existing topology from the ensemble of G-quadruplex species. However, the very possible and unintended consequence is that the topology selected is new and either might not be representative of or may not be a part of the native ensemble and can have significant implication on what can be claimed as biologically relevant.

2

Figure 1. The structure of a G-quartet
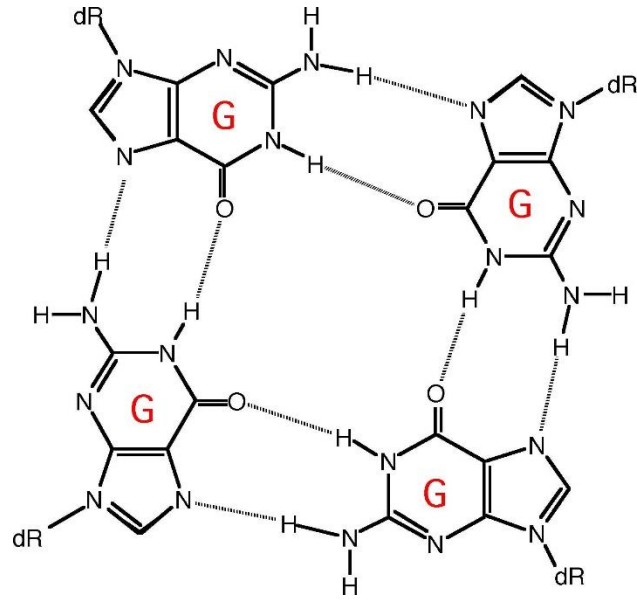
Table 1. G-Quadruplex Formation and the Hallmarks of Cancer

Hallmark of Cancer	G-Quadruplex Involvement
• Sustaining Proliferative Signaling	✓ QF in *c-Myc* promoter (Ambrus *et al.*, 2005) ✓ QF in *NRAS* mRNA (Bugaut *et al.*, 2010)
• Evading Growth Suppressors	✓ QF in *RB* mRNA(Xu and Sugiyama, 2006) ✓ QF in *TP53* mRNA (Marcel *et al.*, 2011)
• Resisting Cell Death	✓ QF in *Bcl-2* promoter (Dai *et al.*, 2006)
• Enabling Replicative Immortality	✓ QF in the telomere ✓ QF in *hTERT* promoter (Lim *et al.*, 2010)
• Inducing Angiogenesis	✓ QF in *VEGF* promoter (Sun *et al.*, 2005) ✓ QF in *hVEGF* mRNA (Morris *et al.*, 2010)
• Activating Invasion and Metastasis	✓ QF in *Wnt1* promoter (Wang *et al.*, 2014)
• Deregulating Cellular Energetics	✓ QF in *HIF-1α* promoter (De Armond *et al.*, 2005)
Emerging Characteristic of Cancer	G-Quadruplex Involvement
• Genome Instability and Mutation	✓ Polymerase stalls at QF in replication fork (Paeschke *et al.*, 2011) ✓ QF sites overlap with DNA breakpoints (De and Michor, 2011)

[*]G-Quadruplex Formation (QF)

Recognizing the current limitation in G-quadruplex research, this dissertation work presents three examples of experimental approaches, which can be used to probe the conformational space surrounding G-quadruplex formation in solution without significant disruption or perturbation of the system. The remainder of this introduction aims to highlight the significance of G-quadruplex nucleic acids by providing a historical review of G-quadruplex research and summarizing the proposed biological functions of G-quadruplex structures. The introductions that accompany Chapters II, III, and IV will discuss in further details G-quadruplex formation and stability. This introduction concludes with a summary of the current dissertation work and surveys the accomplishments stemmed from this work. Other applications of G-quadruplex structures with regard to aptamer-based therapeutics, biosensor development, supramolecular chemistry, and nanotechnology are outside the scope of the current work and will not be discussed in this introduction. For a summary of these topics, the readers are referred to several published reviews (Davis, 2004, Collie and Parkinson, 2011, Düchler, 2012, Lv *et al.*, 2012, Ruttkay-Nedecky *et al.*, 2013)

From Base Pairs to Quartets: The Discovery of G-Quadruplex Structures

The discovery of the molecular structure of DNA is arguably one of the most important scientific achievements of the 20th century. The works of Watson, Crick, Wilkins, Franklin and colleagues led to a proposed model for DNA structure: a right-handed double helix consisting of two strands of DNA in antiparallel orientation associated through hydrogen bond interaction between complimentary purine/pyrimidine, i.e. adenine with thymine (A/T) and guanine with cytosine (G/C), base pairs (Watson and

Crick, 1953a, Wilkins *et al.*, 1953, Franklin and Gosling, 1953c). The discovery of the double helix structure provided the foundation for modern day molecular biology (Watson and Crick, 1953b) and its significance extends far beyond the scientific community as the term "double helix" has come to be widely associated with, and in some cases is synonymous with, nucleic acid structures among the lay population. Indeed, the "B-form" double helix proposed in 1953 is the most common tertiary structure adopted by genomic DNA within the cell (Ghosh and Bansal, 2003). However, under certain conditions or when containing certain sequences, nucleic acids can also assume other tertiary structures (Rich, 1993). For example, the presence of the 2'-OH group prevents double stranded RNA from forming the "B-form" double helix, an observation that was made by Watson and Crick (1953a) in their original paper. Consequently, the structure of the RNA double helix reported by Rich and Davies (1956) was observed to resemble the "A-form" double helix that was previously observed by Franklin and Gosling for dehydrated DNA (Franklin and Gosling, 1953a, Franklin and Gosling, 1953b, Franklin and Gosling, 1953d). This "A-form" double helix has also been proposed as a structure for DNA:RNA hybrids formed during transcription and replication (Rich, 1960, Wang *et al.*, 1982, Egli *et al.*, 1992). A third class of double helix, the left-handed "Z-form", has been observed under high salt conditions for sequences with a high G/C content (Mitsui *et al.*, 1970, Pohl and Jovin, 1972, Wang *et al.*, 1979). Four years after the discovery of the double helix, research into the structures of double stranded RNA by Felsenfeld *et al.* (1957) yielded the first observation of a triple helix, consisting of one poly(rA) and two poly(rU) strands, which formed in solutions containing magnesium. In the model proposed, the poly(rA) and poly(rU)

strands associate by Watson-Crick hydrogen bonds, which is indicated in text by a dot, $\cdot$. The addition of $MgCl_2$ to the solution drives the binding of a second poly(rU) strand to the major groove of the poly(rA)$\cdot$poly(rU) double helix. This third strand interacts with the poly(rA) strand by Hoogsteen hydrogen bonds, which is indicated in text by an asterisk, * (Hoogsteen, 1959, Hoogsteen, 1963). Aside from the U*A$\cdot$U triple helix, other combinations include (in chronological order of discovery) C*G$\cdot$C and G*G$\cdot$C (Lipsett, 1964), T*A$\cdot$T (Arnott and Selsing, 1974), A*A$\cdot$U (Broitman *et al.*, 1987), and A*A$\cdot$T (Howard *et al.*, 1995). In addition, triple helix DNA/RNA hybrids have also been reported (Morgan and Wells, 1968, Roberts and Crothers, 1992). Synthetic triplex-forming oligonucleotides (TFO) can also inhibit gene transcription in the cell by inducing triple helix formation through anti-sense binding at sequence-specific sites in the genome forming the underlying basis for the anti-gene strategy for controlling gene expression. (Hélène, 1991, Praseuth *et al.*, 1999, Knauert and Glazer, 2001, Duca *et al.*, 2008).

The present day interest in G-quadruplex structures is the consequence of a series of findings published in the late 1980's, which discovered that oligonucleotides containing guanine-rich telomeric repeats associate to form G-quadruplex structures in solution (Henderson *et al.*, 1987, Williamson *et al.*, 1989, Sundquist and Klug, 1989). The origin of G-quadruplex research, however, began much earlier. Levene and Jacobs (1909) reported that solutions of guanine riboside or guanosine form highly viscous gels at millimolar concentrations. A similar observation was made by Bang (1910) for solutions of guanosine monophosphate (GMP). Of the five main bases present in DNA and RNA, the gel-forming property appears to be unique to guanine (Chantot and Guschlbauer, 1972), which contains two hydrogen bond donor groups. Gel formation is

9

not associated with solutions containing nucleosides and nucleotides of the other four bases, i.e. adenine, cytosine, thymine, and uracil, which only contain one hydrogen donor group each. This unique property of GMP nucleotides severely hindered initial efforts in the 1950's to synthesize poly(rG) polynucleotides for structural studies. It is in this context that poly(rI) polynucleotides were produced. Inosine is a close analogue of guanosine and did not form gel in solution (Guschlbauer *et al.*, 1990). The base of inosine is hypoxanthine, which is the deaminated derivative of guanine. Rich (1958) performed X-ray diffraction of poly(rI) fibers and reported the presence of a helical structure. The diffraction pattern was consistent with both a triple helix and a quadruple helix. However, the quality of the image at the time did not allow for the conclusive determination of the correct structure. Four years later, Gellert *et al.* (1962) performed X-ray diffraction of 3'-GMP and 5'-GMP nucleotides fibers and reported the presence of helical structures containing a novel substructure now known as the guanine quartets or G-quartets. Davies (2005) remarked on this serendipitous discovery, "Marie [Lipsett] originally thought that she had been able to make poly(G), but was then disappointed to discover that what she had was unpolymerized GMP that was forming a viscous solution that looked just like DNA, [the same observation made by Bang (1910) 50 years prior]. As soon as she told me this I rushed over and pulled some fibers that gave diffraction patterns that could be explained by the formation of G-quartets."

Subsequently, Iball *et al.* (1963) reported that 2-deoxyguanosine (the DNA counterpart of guanosine) and 2-deoxyguanosine monophosphate can also form quartets. In fact, quartet formation have also been observed for inosine (Brahms and Sadron, 1966), 8-bromoguanosine, 2',3'-*O*-diacetylguanosine (Tougard *et al.*, 1973), and over 20

guanosine derivatives (Guschlbauer *et al.*, 1990). Having reported the successful synthesis of poly(rG) (Fresco and Su, 1962) a year prior, Fresco and Massoulie (1963) performed spectroscopic studies of poly(rG) solutions and confirmed the presence of multi-stranded helices. However, the structure of which would not be determined until over a decade later when Zimmerman *et al.* (1975) perform X-ray diffraction of poly(rG) fibers and reported the first structure of the guanine quadruple helix (G-quadruplex). A year prior, Arnott *et al.* (1974) had repeated the X-ray diffraction of poly(rI) fibers and described the structure of the I-quadruplex. Miles and Frazier (1978) later performed spectroscopic study of poly(rI) solutions and noted significant differences in the melting profiles of different quadruple helices formed in the presence of different metal ions. From these findings, it was proposed that certain alkali metal ions can occupy the central channel of the helix and coordinate with the O6 atoms of the quartets. The coordination of metal ions was later shown to be essential for G- and I-quadruplex formation by helping to stabilize the stacking of individual quartet units (Howard and Miles, 1982).

The Biological Functions of G-Quadruplex Nucleic Acids

G-Quadruplex Formation in the Human Telomere

The interest in G-quadruplex following its discovery was limited, in part due to the difficulty of obtaining poly(dG) and poly(rG), as synthesis remains a challenge. Of the many potential sites for G-quadruplex formation in the cell, the first to be investigated was the telomere. All organisms with linear chromosomes contain telomeres, DNA sequences comprise of short (2-8 bases) polynucleotide repeats that cap the ends of

11

chromosomes and allow the cell to distinguish between normal chromosomal ends and double strand breaks (Blackburn, 1991). During an investigation into the structure of the telomere from different organisms, Henderson *et al.* (1987) performed non-denatured gel electrophoresis of several guanine-rich oligonucleotides and reported that many of these sequences contained a second band with increased electrophoretic mobility. The intensity of the second band increased at lower electrophoretic temperature. Using thermal denaturation analysis by the absorbance method, the authors demonstrated that these faster migrating species displayed hyperchromicity at high temperature, consistent with helix to coil transformation. Upon further inspection by ^{31}P-NMR spectroscopy, the authors concluded that the faster migrating species are intramolecular folded double stranded hairpin structures containing non-standard guanine/guanine base pairs. Subsequently, Williamson *et al.* (1989) and Sundquist and Klug (1989) performed footprinting experiments on the same sequences, reported that all the guanine N7 sites were protected from methylation by chemical agent. In addition, Williamson *et al.* (1989) reported that the protection of N7 sites only occurred in gels containing Na$^+$, K$^+$, and Cs$^+$ but not in gels containing Li$^+$ or no added salts. Taken together, these findings were consistent with formation of G-quartets and G-quadruplex structures.

The discovery of G-quadruplex formation in the telomere renewed interest in these structures. Structural investigations of the sequence d(GGGGTTTTGGGG) from the telomere of *Oxytricha* resulted in the first single crystal analysis and X-ray crystals structure of a G-quadruplex by Kang *et al.* (1992) which was followed shortly by the first NMR study and solution structure by Smith and Feigon (1992). The first high-resolution G-quadruplex structure of a human sequence, d(AGGGTTTAGGGTTAGGGTTAGGG)

12

or hTel22 in sodium, was first solved by Wang and Patel (1993). Parkinson *et al.* (2002) later report a crystal structure of the same sequence, which was the first reported high-resolution structure of a human sequence in potassium. Within the cell, G-quadruplex formation is thought to occur at the distal 3' end of the telomere, which contains a single-stranded guanine-rich overhang of approximately 100 to 200 bases (Wright *et al.*, 1997). The stabilization of these G-quadruplex structures by a small molecule inhibitor is currently being investigated with the aim of bringing a novel class of anti-cancer therapeutic to the market (Han and Hurley, 2000, Neidle and Read, 2000, White *et al.*, 2001, Saretzki, 2003, Ou *et al.*, 2008, De Cian *et al.*, 2008, Balasubramanian and Neidle, 2009). Small molecules that interact with telomeric G-quadruplexes *in vitro* were also able to inhibit the activity of telomerase in an *ex vivo* assay. Since telomerase is activated in more than 90% of all cancers to maintain the length of the telomere thereby conferring immortality upon cancer cells (Shay and Bacchetti, 1997), G-quadruplex-based anti-telomerase therapy could be an attractive strategy for the development of anti-cancer therapeutics. Treating cells with these G-quadruplex-interacting and telomerase-inhibiting agents have shown to promote telomere dysfunction in the cell resulting in cellular senescence and cell death. (Riou *et al.*, 2002, Cuesta *et al.*, 2003, De Cian *et al.*, 2008, Lopes *et al.*, 2011, Rodriguez *et al.*, 2012).

G-Quadruplex Formation in Oncogene Promoters

G-quadruplex formation is not just limited to the guanine-rich telomere. Putative G-quadruplex-forming sequences (PQS) are found within 1,000 bases of the transcription start sites (TSS) (Du *et al.*, 2008, Eddy and Maizels, 2008, Du *et al.*, 2009). In particular,

13

PQS are overrepresented in the promoter regions of many oncogenes, such as *c-Myc* (Ambrus *et al.*, 2005), *c-Kit* (Hsu *et al.*, 2009, Phan *et al.*, 2007a), *Bcl-2* (Dai *et al.*, 2006), *VEGF* (Sun *et al.*, 2005), *HIF-1α* (De Armond *et al.*, 2005), *RET* (Tong *et al.*, 2011), and *hTERT* (Lim *et al.*, 2010), whereas the occurrence of PQS in tumor suppressor genes tends to be much lower (Eddy and Maizels, 2006). The concentration of PQS near the promoter regions appear to be a conserved trait between several species including humans (Rawal *et al.*, 2006, Du *et al.*, 2007, Zhao *et al.*, 2007, Verma *et al.*, 2008). A recently study reported that the enrichment of PQS in these positions is the result of positive selection suggesting that these sequences conferred an evolutionary advantage to the organism (Smith, 2010). A proposed function of G-quadruplex structures in these promoter regions is thought to be transcriptional regulation (Brooks *et al.*, 2010, Balasubramanian *et al.*, 2011). For example, in the case of the *c-Myc* oncogene, a 27 base-pair guanine-rich sequence located in the nuclease hypersensitivity element III region 1 (NHE-III$_1$) upstream of the P1 promoter has been demonstrated to form G-quadruplex structures in solution (Simonsson *et al.*, 1998). The NHE-III$_1$ region is thought to be the predominant regulator of *c-Myc* transcription (Brooks and Hurley, 2010). In a study employing a cell-based luciferase reporter system, the formation of G-quadruplex structures in NHE-III$_1$ was proposed as a mechanism for regulating *c-Myc* expression (Siddiqui-Jain *et al.*, 2002). Sequence mutations that destabilize G-quadruplex formation *in vitro* resulted in an increase of *c-Myc* expression in the cells, while actions that promote G-quadruplexes formation, e.g. treating the cell with the G-quadruplex stabilizing compound *meso*-Tetra (N-methyl-4-pyridyl)porphine (TMPyP4) (Grand *et al.*, 2002) among others (Brown *et al.*, 2011), led to a decrease in *c-Myc* expression. In

14

addition, overexpression of the protein nucleolin also resulted in a similar decrease in *c-Myc* expression (González *et al.*, 2009). The binding of nucleolin to the G-quadruplex structure in NHE-III$_1$ has been proposed as a mechanism for normal control of *c-Myc* expression. These findings are significant as *c-Myc* gene is a particularly important oncogene, whose altered expression has been associated with many types of cancer, including but not limited to breast, lung, prostate and hematological cancers (Nesbit *et al.*, 1999, Nilsson and Cleveland, 2003).

Other Biological Functions of DNA G-Quadruplex Structures

In addition to regulation of transcription, DNA G-quadruplex structures have also been linked to genomic instability and loss of epigenetic control. The first evidence of G-quadruplex-mediated genomic instability came from the investigation of Dog1 helicase in *Caenorhabditis elegans* worms. In Dog1 deficient cells, massive genome-wide deletions were observed at guanine-rich sites (Cheung *et al.*, 2002). Subsequent investigation revealed that only sites containing PQS motifs were deleted (Kruisselbrink *et al.*, 2008). In fact, sites containing guanine-rich sequences not capable of forming G-quadruplex structures (e.g. sites containing CG repeats) remained unaffected. Dog1 was later discovered to be the worm homologue of the human FANCJ helicase (Youds *et al.*, 2008). FANCJ is one of sixteen genes linked to Fanconi Anemia, a rare genetic disease characterized by defective DNA repair resulting in hematologic failures. Consequently, because of the defective DNA repair, Fanconi Anemia patients often end up developing cancers (Rosenberg *et al.*, 2003). In Fanconi Anemia patients with FANCJ deficiency, similar genome-wide deletion of PQS sites were observed (London *et al.*, 2008). Consistent with the theory that FANCJ resolves G-quadruplex-mediated genomic

15

instability, FANCJ was shown to unwind G-quadruplex DNA with a 5'-to-3' polarity (Wu *et al.*, 2008). Investigations of the interaction between the protein Pif1 and G-quadruplex structures provided additional evidence of G-quadruplex-mediated genomic instability (Paeschke *et al.*, 2013). Pif1 is a DNA helicase that binds to and unwinds G-quadruplex structures (Sanders, 2010). In Pif1 deficient *Saccharomyces cerevisiae* yeast cells, replication fork progression stalls at sites containing PQS motifs leading to double strand breaks (Paeschke *et al.*, 2011). Replication fork stalling, in addition to impeding the successful transfer of genetics information from parent to daughter cells, can also affect the transfer of epigenetic control by interfering with histone recycling from parent cell to daughter cells (Whitehouse and Smith, 2013). A class of proteins known as Y-family translesion polymerase is responsible for restarting stalled replication fork (Edmunds *et al.*, 2008). REV1 is a Y-family translesion polymerase. A study carried out in DT40 chicken cells showed that REV1 deficient cells have distinct transcriptional profiles compared to REV1 proficient cells with genes being both up-regulated and down-regulated (Sarkies *et al.*, 2010). Moving a PQS motif into a silent locus leads to its de-repression in REV1 deficient cells but not in REV1 proficient cells. The authors concluded that, in REV1 deficient cells, DNA synthesis is uncoupled from histone recycling, resulting in localized loss of repressive chromatin through biased incorporation of newly synthesized histones. The loss of epigenetic controls is worsened in cells with mutations in FANCJ (Sarkies *et al.*, 2012).

An Introduction to RNA G-Quadruplex Structures

In addition to DNA, G-quadruplex formation can also occurs with RNA nucleic acids. In fact, in about 20% of human genes, there is an overrepresentation of PQS in

16

non-coding regions that get transcribed into messenger RNA (mRNA) (Huppert *et al.*, 2008). Compared to DNA, RNA G-quadruplex formation is more favorable. A recent assessment of the stability of G-quadruplex structures compared to double helix concluded that in the presence of a complementary strand, duplex formation is highly favored over G-quadruplex formation (Lane, 2012). During certain aspect of cellular function in which the DNA duplex is transiently unwound into two single strands, such as during replication or transcription, G-quadruplex formation might be favored as the folded G-quadruplex conformation reduces the size of the solvent accessible surface area and allows for better shielding of the hydrophobic base from the aqueous environments of the cell. Messenger RNA is traditionally single stranded and, once transported out of the nucleus, there is no complementary strand. Thus, it follows that an mRNA sequence would more likely be able to adopt a G-quadruplex conformation compared to its DNA counterpart. In addition, compared to G-quadruplex structures formed from DNA oligonucleotides, G-quadruplex structures formed from RNA oligonucleotides of the same sequence are more stable (Arora and Maiti, 2009, Joachimi *et al.*, 2009, Zhang *et al.*, 2010a). Lastly, for any given genes in which a PQS is present in both the DNA and RNA forms, there are only two maximum potential DNA G-quadruplex structures (one in each allele) while the number of potential RNA G-quadruplex structures is in the thousands. This difference in the number of structures could have significant implication in the design of sequence specific small molecule inhibitors. While RNA G-quadruplex structures present several advantages over the DNA counterpart, there is one serious drawback. DNA G-quadruplex structures are highly polymorphic and have been shown to fold into many different topologies, all known structures of RNA G-quadruplex displayed

17

the same parallel topology (Liu *et al.*, 2000, Matsugami *et al.*, 2001, Deng *et al.*, 2001, Liu *et al.*, 2002c, Liu *et al.*, 2002a, Liu *et al.*, 2002b, Uesugi *et al.*, 2003, Xu *et al.*, 2008a, Xu *et al.*, 2008b, Martadinata and Phan, 2009, Lipay and Mihailescu, 2009, Martadinata and Phan, 2013). The lack of structural polymorphism regarding RNA G-quadruplex could have significant implication on the ability of a small molecule to distinguish between G-quadruplex structures from different genes.

Biological Functions of RNA G-Quadruplex Structures

One function of G-quadruplex structures in mRNA is thought to be translational repression (Halder *et al.*, 2009, Arora and Suess, 2011). The stabilization of G-quadruplex structures located in the 5'-untranslated region of the *NRAS* mRNA by a bisquinolone compound was accompanied by decreased translational efficiency (Bugaut *et al.*, 2010). *NRAS* is an important oncogene (Marshall *et al.*, 1982, Hall *et al.*, 1983, Shimizu *et al.*, 1983) and a member of the *RAS* superfamily of proteins. Permanent mutation leading to activation of *RAS* proteins are noted in 20 to 25% of all human cancers (Downward, 2003). RNA G-quadruplex structures can also promote translation. In the case of the *FGF2* (Bonnal *et al.*, 2003) and the *hVEGF* (Morris *et al.*, 2010) mRNA, the PQS in the internal ribosome entry site (IRES) in the 5'-untranslated region was found to be essential for IRES-mediated translation initiation. Mutations resulted in loss of G-quadruplex formation also resulted in loss of translation initiation. Both *FGF2* (House *et al.*, 2003) and *hVEGF* (Karkkainen and Petrova, 2000, Holmes *et al.*, 2007) are potent growth factors associated with angiogenesis. However, while the role of *FGF2* in cancer is unclear, overexpression of *hVEGF* is associated with poor prognosis, high

18

recurrence, and decreased survival in breast cancer (Gasparini, 2000, Guo *et al.*, 2010, Heer *et al.*, 2001). In addition to translational regulation, other proposed roles for RNA G-quadruplex structures include posttranscriptional processing of mRNA (Christiansen *et al.*, 1994), alternative splicing (Marcel *et al.*, 2011), and RNA turnover (Bashkirov *et al.*, 1997). No discussion of RNA G-quadruplex is complete without mentioning TERRA, the guanine-rich mRNA transcribed from cytosine-rich telomeric template strand. Several high-resolution structures have been reported which demonstrated that TERRA assumes parallel G-quadruplex structures (Xu *et al.*, 2008a, Xu *et al.*, 2008b, Martadinata and Phan, 2013). Although the exact cellular function of TERRA is unknown, a recent study reported that TERRA can inhibit telomerase activity in an *in vitro* assay (Schoeftner and Blasco, 2008). In addition, using RNA-FISH hybridization, the authors showed that TERRA localizes to the telomere. Together, these findings suggest that TERRA may function in regulation and protection of the telomere.

Visualization of G-Quadruplex Structures in Human Cells

Bioinformatics surveys have identified over 370,000 PQS in the human genome (Huppert and Balasubramanian, 2005, Huppert and Balasubramanian, 2007). For a select set of PQS, G-quadruplex formation has only been observed *in vitro* and purported biological functions have only been demonstrated through a reporter system (e.g. luciferase system). Thus, the true biological relevance of such sequences is unknown. Recently, Biffi *et al.* (2013) used an antibody against DNA G-quadruplex structures (anti-G4) to visualize G-quadruplex formation in the cell and provided the strongest evidence to date for the existence of G-quadruplex structures *in vivo*. When fixed cells were treated

with anti-G4 antibody and followed by incubation with a secondary detector antibody and a tertiary fluorochrome-containing reporter antibody, fluorescent anti-G4 foci can be detected by confocal microscopy. These foci disappeared when the anti-G4 antibody was pre-incubated with pre-folded G-quadruplex-forming oligonucleotides (QFO) but not when pre-incubated with single stranded oligonucleotides. In addition, the number of foci increased when the cells were transfected with pre-folded QFO but not when the cells were transfected with single stranded oligonucleotides. Together, these observations indicate that the anti-G4 antibody was able to bind to G-quadruplex structures in the cell. The number of foci decreased upon treatment of the cell with DNase but not upon treatment with RNase implying that the antibody was able to distinguish between DNA and RNA G-quadruplexes. Only about 25% of the anti-G4 foci were found to localize to the telomere. This observation indicates that formation of G-quadruplex structures in the cell occurs primarily in non-telomeric locations, such as oncogene promoters (Huppert and Balasubramanian, 2005, Huppert and Balasubramanian, 2007). Consistent with the hypothesis that G-quadruplex formation is more likely to occur during replication as the double stranded DNA is unwound by helicase into single strands that can more easily fold into the G-quadruplex form (Sarkies *et al.*, 2010), an increase in the number of anti-G4 foci was observed in the S-phase compared to the other phases in the cell cycle. In addition, the number of anti-G4 foci also increased upon treatment of the cell with a G-quadruplex-stabilizing small molecule. Taken together, these findings strongly suggest that G-quadruplex structure indeed form in human cells and that formation of G-quadruplexes can have significant implications in cellular functions.

Summary of Dissertation Works and Accomplishments

In the current dissertation work, we present three experimental approaches for investigating G-quadruplex structures. In the first experimental approach, we employed size exclusion chromatography (SEC) to separate the G-quadruplex structures formed from the G-quadruplex-forming Pu27 *c-Myc* promoter sequence. Examination by SEC revealed that Pu27 exists as a heterogeneous mixture of monomer and higher-order G-quadruplex structures. We investigated the effect of changing experimental conditions on Pu27 structural polymorphism as well as its interaction with the small molecule TMPyP4. Lastly, we compared our observation of the Pu27 sequence with four modified sequences reported in the literature to determine the effect of sequence modification on Pu27 structural polymorphism. The findings of this research project was published in *Organic and Biomolecular Chemistry* (Appendix) and is reproduced in Chapter II.

In our second experimental approach, we employed hydrodynamic bead modeling (HBM) to study the structural polymorphism surrounding G-quadruplex formation by the hTel22 human telomere sequence. We carried out sedimentation velocity experiments to obtained measurements for sedimentation coefficients, translational diffusion coefficients, and frictional ratios. To sample G-quadruplex structures for examination by HBM, we carried out microsecond timescale molecular dynamics simulations for ten different telomeric G-quadruplex forming oligonucleotides. Using our calibrated parameters for hydrodynamic bead modeling, we calculated the hydrodynamic properties for G-quadruplex structures sampled from MD and compared the calculated values to experimentally determined values. We performed cluster analysis on the sampled MD

21

structures using the calculated sedimentation coefficients as the clustering criteria to identify different hydrodynamic substates. To conclude our discussion of the telomeric G-quadruplex, we carried out grid mappings of water and ions surrounding the G-quadruplex structures to identify sites of water and ions binding. The findings of this research project was published in *The Journal of Physical Chemistry B* (Appendix) and is reproduced in Chapter III.

In our third experimental approach, we presented a screening platform that was successful in identifying a new G-quadruplex-interacting small molecule. In our approach, we presented an *in silico* generated model of an end-pasting site that was used for virtual screening. Compound 1 was identified as a G-quadruplex-interacting small molecule. Compound 1 binding to G-quadruplex structures was initially confirmed using a thermal denaturation analysis of FRET-labeled G-quadruplex DNA. We employed fluorescent and circular dichroism spectroscopy to demonstrate that Compound 1 binds to G-quadruplex structures by the same end-pasting mechanism identified in the virtual screen. Additionally, we submitted Compound 1 for testing with the NCI-60 DTP Human Tumor Cell Line Screen to assess the effect of Compound 1 on cancer cells. The findings of this. The findings of this research project is currently in submission (Appendix) and the manuscript is reproduced in Chapter IV.

CHAPTER II

G-quadruplexes, DNA tertiary structures highly localized to functionally important sites within the human genome, have emerged as important new drug targets. The putative G-quadruplex-forming sequence (Pu27) in the NHE-III1 promoter region of the *c-Myc* gene is of particular interest as stabilization of this G-quadruplex with TMPyP4 has been shown to repress *c-Myc* transcription. In this study, we examine the Pu27 G-quadruplex-forming sequence and its interaction with TMPyP4. We report that the Pu27 sequence exists as a heterogeneous mixture of monomeric and higher-order G-quadruplex species *in vitro* and that this mixture can be partially resolved by size exclusion chromatography (SEC) separation. Within this ensemble of configurations, the equilibrium can be altered by modifying the buffer composition, annealing procedure, and dialysis protocol thereby affecting the distribution of G-quadruplex species formed. TMPyP4 was found to bind preferentially to higher-order G-quadruplex species suggesting the possibility of stabilization of the junctions of the *c-Myc* G-quadruplex multimers by porphyrin end-stacking. We also examined four modified *c-Myc* sequences that have been previously reported and found a narrower distribution of G-quadruplex configurations compared to the parent Pu27 sequence. The findings reported here demonstrate that experimental conditions contribute significantly to G-quadruplex formation and should be carefully considered, controlled, and reported in detail.

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 *MESO*-TETRA(N-METHYL-4-PYRIDYL)PORPHINE.

Introduction

A G-quadruplex is a DNA tertiary structure formed by the unimolecular folding of a guanine-rich sequence bearing four or more runs containing at least two guanine bases (Williamson, 1994). Bi- or tetramolecular G-quadruplexes can form from strands containing fewer runs of guanine. A G-quadruplex is typically composed of two or three stacked G-tetrads. Each tetrad is composed of four guanines in a square planar arrangement stabilized by Hoogsteen hydrogen bonds. A cation, usually sodium or potassium, is associated with 1 or 2 stacked G-quartets by coordination with the O6 of the guanine molecules, stabilizing the tetrad arrangement and promoting G-quadruplex formation (Williamson, 1994, Huppert, 2008).

Among the first G-quadruplexes studied were those formed from the human telomere sequence(Neidle, 2010). Generally recognized as repeats of d(GGGTTA), telomeres are DNA sequences that cap the ends of chromosomes and are thought to contribute to genetic stability by preventing the ends of the chromosome from being eroded away during replication. The human telomere is 5-8 thousand base pairs in length with a single stranded 3' overhang of 100 to 200 bases (Wright *et al.*, 1997). The formation of G-quadruplexes in these 3' overhangs has been shown to decrease the

activity of telomerase, an enzyme which is responsible for maintaining the length of telomeric DNA. Since telomerase activation has been found to be involved in greater than 90% of all cancer (Shay and Bacchetti, 1997), G-quadruplex formation in the human telomere is an attractive anti-cancer drug target. Small molecules that stabilize G-quadruplexes formed from telomeric oligonucleotide sequences *in vitro* have been shown to inhibit telomerase activity leading to cellular senescence and cell death in cell-based experiments (Riou *et al.*, 2002, Cuesta *et al.*, 2003, De Cian *et al.*, 2008). Although most commonly associated with the human telomere, G-quadruplex forming sequences are also found throughout the genome (Huppert and Balasubramanian, 2005). A search of the human genome has revealed more than 370,000 potential G-quadruplex forming sequences (Huppert and Balasubramanian, 2005, Huppert and Balasubramanian, 2007). Recent research suggests that G-quadruplex-forming sequences are not randomly distributed but are concentrated in functionally important sites (Huppert and Balasubramanian, 2005). In particular, the occurrence of potential G-quadruplex-forming sequences is much higher in proto-oncogenes such as *c-Myc* (Ambrus *et al.*, 2005), *c-Kit* (Hsu *et al.*, 2009, Phan *et al.*, 2007a), *Bcl-2* (Dai *et al.*, 2006), *VEGF* (Sun *et al.*, 2005), and *HIF-1α* (De Armond *et al.*, 2005) than in other areas of the genome; whereas the occurrence of G-quadruplex-forming sequences in tumor suppressor genes tends to be much lower (Eddy and Maizels, 2006).

The human *c-Myc* gene is a particularly significant oncogene. Alteration of this gene or its expression has been associated with many types of cancer, including but not limited to breast, lung, prostate and hematological cancers (Nesbit *et al.*, 1999, Nilsson and Cleveland, 2003). A putative G-quadruplex-forming sequence has been discovered in

25

the promoter region of this oncogene. The NHE-III$_1$, a 27 base-pair guanine-rich sequence demonstrated to form G-quadruplexes *in vitro* (Simonsson *et al.*, 1998) and located upstream of the *c-Myc* P1 promoter, is believed to be the predominant regulator of *c-Myc* expression (Brooks and Hurley, 2010). In a study employing a cell-based luciferase reporter system, the formation of G-quadruplexes in NHE-III$_1$ was proposed as a mechanism for regulating *c-Myc* expression (Siddiqui-Jain *et al.*, 2002). Sequence mutations that destabilize G-quadruplex formation *in vitro* resulted in an increase of *c-Myc* expression in the cells, while actions that promote G-quadruplexes formation, e.g. treating the cell with the G-quadruplex stabilizing compound *meso*-Tetra (N-methyl-4-pyridyl)porphine (TMPyP4) (Grand *et al.*, 2002) (Figure 2A) among others (Brown *et al.*, 2011), led to a decrease in *c-Myc* expression. The binding mechanism of TMPyP4 to *c-Myc* and other G-quadruplexes remains an area of investigation (Figure 2B-C) (Anantha *et al.*, 1998, Haq *et al.*, 1999, Seenisamy *et al.*, 2004, Freyer *et al.*, 2007, Wei *et al.*, 2009). We believe that several factors are likely to have contributed to the discrepancies regarding the TMPyP4 binding mechanism, which include the lack of examining the inherent polymorphism and therefore a lack of definitive structures of the G-quadruplexes formed from the *c-Myc* promoter sequence without major perturbation of the sequence, general lack of understanding of the binding preferences, and variability and a lack of detail in reporting experimental conditions (Seenisamy *et al.*, 2004, Seenisamy *et al.*, 2005, Freyer *et al.*, 2007, González *et al.*, 2009, Ou *et al.*, 2011) leading to the formation of G-quadruplexes.

Figure 2. TMPyP4 binding to G-quadruplex. (A) The chemical structure of the cationic porphyrin, TMPyP4 and (B) NMR solution structure (PDB: 2A5R) of TMPyP4 binding a unimolecular G-quadruplex formed from the modified *c-Myc* promoter sequence, Pu24I, (C) X-ray crystal structure (PDB: 2HRI) of TMPyP4 binding to the bimolecular G-quadruplex formed from the human telomere sequence, $TAG_3T_2AG_3$.

A
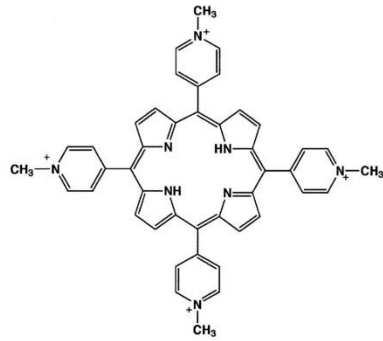
B

C
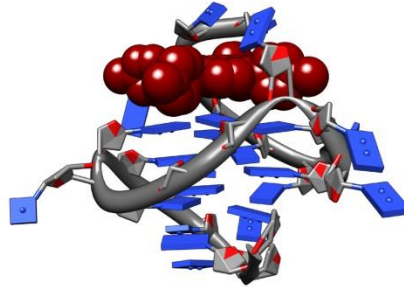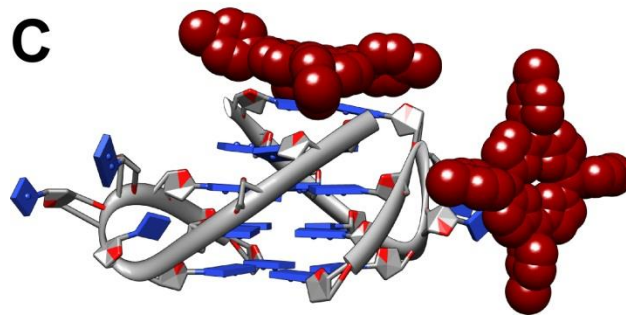
G-quadruplex formation can be highly polymorphic (Lane *et al.*, 2008, Dailey *et al.*, 2010). In the presence of sodium, it is generally accepted that the human telomere sequence, $G_3(T_2AG_3)_3$, folds into a single anti-parallel G-quadruplex topology, often termed a "basket" configuration, with a lateral, a diagonal, and a lateral connecting loop (Wang and Patel, 1993). In the presence of potassium, the predominant cation inside the cell and the nucleus, the same sequence can adopt a number of different topologies with variations in connecting loop types, strand/segment orientations, numbers of G-quartets, and glycosyl torsion angles depending on the conditions (e.g. buffer composition, the presence of organic solvents such as acetonitrile or ethanol, DNA concentration, ion concentration, annealing profile, and the presence of various biological molecules). While it is a relatively simple G-quadruplex-forming sequence, the human telomere sequence can potentially fold into more than 200 intramolecular conformations (Lane *et al.*, 2008). For more complex G-quadruplex-forming sequences, such as those commonly found in the promoter regions of many oncogenes, this issue can be greatly exacerbated. As an example of one such sequence, the *c-Myc* parent sequence (i.e. Pu27, Figure 3) has 5 runs of 3 or more guanines and can potentially display a much greater degree of polymorphism than in the telomere sequence. Consequently, when a G-quadruplex-forming sequence is studied, certain steps are usually taken to reduce this inherent polymorphism.

Figure 3. Sequence information for the *c-Myc* G-quadruplex-forming sequence and reported derivatives. Substituted bases are underlined. Complete NMR structures are available for sequences marked with asterisk.

```
Pu27               T GGGGAGGG TGGGGAGGGTGGGGAAGG
                        5    10   15   20   25

Pu18                      AGGGTGGGGAGGGTGGGG

Myc-1245           TGGGGAGGGTTTTTAGGGTGGGGAA
Myc-2345            TGAGGGTGGGGAGGGTGGGGAA

MYC22-G14T/G23T*      TGAGGGTGGGTAGGGTGGGTAA

Pu24                 TGAGGGTGGGGAGGGTGGGGAAGG
Pu24I*               TGAGGGTGGIGAGGGTGGGGAAGG

Pu19               TGGGGAGGGTGGGGAGGGT
Pu19_A2A11*        TAGGGAGGGTAGGGAGGGT
```

The most common approach to reducing the structural polymorphism of G-quadruplex-forming sequences is through sequence modification (Dai *et al.*, 2008, Yang and Okamoto, 2010). Initially, in the case of the parent *c-Myc* sequence (i.e. Pu27), the five runs of three or more guanines are truncated to four runs (Pu18 and Pu19) or truncated and then further modified by addition of a base (i.e. Myc-2345 and Pu24). Subsequent base substitutions and subtractions, generally reducing the number of guanines to exactly four runs of three guanines or forcing certain regions of the sequence to be loops, have also been used to yield several sequences with reduced polymorphism (i.e. Myc-2345 to MYC22-G14T/G23T, Pu19 to Pu19_A2A11, Myc-1245). Some sequences can be further modified by base substitutions with non-canonical bases such as inosine (i.e. Pu24 to Pu24I). Among these sequences, only a selected few are sufficiently enriched in one particular topology necessary for the determination of the complete structure by NMR (i.e. MYC22-G14T/G23T, Pu24I, Pu19_A2A11) (Seenisamy *et al.*, 2004, Phan *et al.*, 2004, Ambrus *et al.*, 2005, Phan *et al.*, 2005, Mathad *et al.*, 2011). Regardless of the approach taken, the results are often new sequences that are markedly different from the parent sequence (Figure 3).

Other sequence modifications include incorporation of 8-methylguanine or 8-bromoguanine which are known to produce G-quadruplex structures with a *syn* glycosidic configuration (Esposito *et al.*, 2004, Virgilio *et al.*, 2005a, Virgilio *et al.*, 2005b, Virgilio *et al.*, 2012), while use of O^6-methylguanine, inosine, or 6-thioguanine have been shown to destabilize G-quadruplex formation(Mekmaysy *et al.*, 2008, Petrovic and Polavarapu, 2008, Marathias *et al.*, 1999, Spackova *et al.*, 2004). Incorporation of 8-aminoguanine and 8-methylguanine promote formation of tetramolecular parallel G-

quadruplexes such as those formed by TG$_4$T (Gros *et al.*, 2008, Tran *et al.*, 2011). Modifications of the sugar-phosphate backbone by insertion of 5'-5' or 3'-3' polarity inversion have also been shown to have a dramatic effect on G-quadruplex formation and stability (Esposito *et al.*, 2005, Esposito *et al.*, 2009, Galeone *et al.*, 2008) and use of RNA or LNA force adoption of a *anti* glycosidic guanosine conformation (Bonifacio *et al.*, 2008, Kumar and Maiti, 2007, Tang and Shafer, 2006, Qi and Shafer, 2007). In addition to sequence modifications, another common approach to reduce polymorphism is by changing the solution conditions. The addition of biological molecules (Sannohe and Sugiyama, 2001) (e.g. sugar, proteins), presence of co-solvents (e.g. acetonitrile, PEG) (Xue *et al.*, 2007, Miller *et al.*, 2010), choice of divalent versus monovalent cations (Blume *et al.*, 1997, Miyoshi *et al.*, 2001), and cation concentration (Gray *et al.*, 2009a, Gray *et al.*, 2009b) all play a major role in directing G-quadruplex formation and determining stability.

Among the limitations of the methods described above for resolving the polymorphism of G-quadruplex structures is that such an approach can often result in drastic and unpredictable perturbation of the system. The untested assumption is that these means of reducing polymorphism enriches a member of the ensemble of species originally formed by the parent sequence. In fact, it is possible that such a perturbation can shift the equilibrium to favor species that might not actually form *in vitro* or *in vivo* (Lane *et al.*, 2008). G-quadruplex polymorphism has severely hindered investigation of G-quadruplex structure, biophysical properties, small molecule interaction, and the thermodynamics of G-quadruplex formation. We have reported that techniques commonly used to study G-quadruplex DNA are typically too low in resolution to

33

distinguish between species in a mixture, e.g. CD spectroscopy, UV-Vis spectroscopy, ultracentrifugation, and gel electrophoresis, while high resolution techniques, e.g. NMR spectroscopy, are of limited utility for mixtures containing multiple G-quadruplex topologies (Dailey *et al.*, 2010).

Recently, we reported results for ten different G-quadruplex-forming promoter sequences (Miller *et al.*, 2011) (including *c-Myc*) examined by SEC, a technique that was able to show their polymorphism *in vitro* without perturbing the system by sequence modifications (Miller and Trent, 2001). It was discovered that these ensembles of structures are more complex than previously thought, a factor masked by the determination of single supposedly representative structures. In addition to the diverse topologies from individual strands, we observed that G-quadruplexes can associate together to form dimers, trimers, and higher-order G-wire structures *in vitro* (Miller *et al.*, 2011). We employed SEC to resolve the polymorphism of the *c-Kit* promoter sequence, $(CG_3)_2(CG)_2(AG_3)_2T$ (PDB: 2KJ2), into three fractions that were revealed by AUC to be of a monomer species, a dimer species, and a higher-order G-wire species.

Here, we examine in detail the structural polymorphism of the *c-Myc* promoter G-quadruplex-forming sequence and describe the influence of experimental conditions (e.g. DNA concentration, ion concentration, buffer components, annealing conditions, dialysis procedure, etc.) on the polymorphism. We observe the interactions of TMPyP4 with the *c-Myc* G-quadruplex ensemble in solution and demonstrate that it binds preferentially to some G-quadruplex structures. As a consequence, TMPyP4 perturbs the distribution of species within the ensemble. In addition, we also examine several modified *c-Myc* sequences which showed reduced polymorphism compared to the parent Pu27 sequence.

34

However, we cannot definitely conclude whether such G-quadruplex structures existed within the original ensemble or were produced as a consequence of sequence modification.

Materials and Methods

Oligonucleotide Preparation, Annealing, and Small Molecule Acquisition.

The *c-Myc* G-quadruplex-forming sequence (Pu27) was purchased from Integrated DNA Technologies (Coralsville, IA) and consists of the sequence:

5' – TGGGGAGGGTGGGGAGGGTGGGGAAGG – 3'

The standard conditions for Pu27 G-quadruplex formation is described in this section and depicted in Figure 3. Deviations from standard conditions were noted accordingly in the results section and Figure 3. A stock solution of the Pu27 oligonucleotide was dissolved in KPEK 200 (200 mM K$^+$) buffer, which is composed of K_2HPO_4 (6mM), KH_2PO_4 (2mM), KCl (185mM), EDTA disodium salt dihydrate (1mM), pH 7.0. Dialysis was performed overnight using Spectrum Laboratories (Rockford, IL, USA) 0.1-0.5 kDa MWCO dialysis devices following manufacturer's instructions. The DNA was quantified using a NanoDrop 2000 instrument (Thermo Scientific, Wilmington, DE). Samples were made at 200 μM strand concentration (with $\varepsilon=279,900$ L·M^{-1}·cm^{-1} for the single-strand form) by dilution of the dialyzed Pu27 stock solution with KPEK 200. The oligonucleotide samples were annealed in a water bath by heating to 100°C, holding the samples at temperature for 10 minutes, and gradually cooling to room temperature overnight.

A 10mM concentrated stock solution of TMPyP4 (Frontier Scientific, Logan, UT) was produced by weighing out the compound and dissolving in DMSO. TMPyP4 concentration was quantified using a NanoDrop 2000 instrument. To prepare TMPyP4 for titration against Pu27, the stock solution of TMPyP4 was diluted by a factor of 4 in DMSO. 10μL of diluted TMPyP4 was mixed with 125μL of 200μM dialyzed and annealed Pu27 G-quadruplex (7.4% DMSO final volume) to form the 1:1 [TMPyP4]:[Pu27] mixture. TMPyP4 was further diluted for the 1:2, 1:4, 1:8, 1:16, 1:32 samples. A sample without TMPyP4 and 7.4% DMSO was examined by SEC and found that the presence of DMSO did not alter the distribution of G-quadruplex species in solution.

Size-Exclusion Chromatography (SEC)

SEC was performed over 600 minutes using a Waters system (Waters 2998 Photodiode Array Detector and Waters 600 Pump, Waters Corporation, Milford, MA) with monitoring of the *c-Myc* G-quadruplex at an absorbance of 260nm and of DNA-bound TMPyP4 at an absorbance of approximately 445nm. Two Superdex™ 75 10/300 columns connected serially (GE Healthcare, Piscataway, NJ) were employed for sample separation using a flow rate of 0.05 mL/min with 50μL injection volumes of the sample. The mobile phase was consisted of 100 mM KCl, 20 mM K_2HPO_4, pH 7.0. The absorbance of eluted DNA was adjusted by setting the minimum absorbance between 13.75 and 27.50 mL to 0. Absorbance was normalized by dividing the adjusted absorbance by the total area under the curve between 13.75 and 27.50 mL.

Circular Dichroism (CD)

CD experiments were performed on a Jasco J-710 spectropolarimeter (Jasco Inc., Easton, MD). CD scanning experiments were performed at an A_{260} of 0.50 from 340nm to 220nm with a data interval of 1nm, band width of 1nm, response of 1 second, scanning speed of 200 nm/minute and a total of four accumulated scans. The CD of the mixture samples were normalized with respect to strand concentration (ε=279,900 $L \cdot M^{-1} \cdot cm^{-1}$). For the SEC separated fractions, the CD were normalized to the number of G-quadruplexes based on the MW determined by AUC experiments.

Analytical Ultracentrifugation (AUC)

AUC was carried out in a Beckman Coulter ProteomeLab XL-A analytical ultracentrifuge (Beckman Coulter Inc., Brea, CA) at 20°C overnight at 50,000 rpm in standard 2 sector cells. Data were analyzed using the program Sedfit (free software: www.analyticalultracentrifugation.com). Buffer density was determined on a Mettler/Paar Calculating Density Meter DMA 55A at 20.0°C and buffer viscosity was calculated using Sednterp software (free software: www.jphilo.mailway.com). For the calculation of frictional ratio, 0.55 mL/g was used for partial specific volume and 0.3 g/g was assumed for the amount of water bound.

Differential Scanning Calorimetry (DSC)

DSC was conducted using a MicroCal VP-DSC (MicroCal Inc., Piscataway, NJ) for thermal denaturation of the *c-Myc* oligodeoxynucleotide in 25, 200 and 400 mM total potassium conditions. Prior to scanning DNA samples, 25, 200 or 400 mM potassium

37

buffer scans were collected for thermal equilibration of the instrument and collection of baseline scans for analysis of DSC thermograms. DNA samples were prepared at 200 μM strand concentration under the preparation conditions given above. DSC scan parameters included a temperature range of 20°C to 120°C to encompass the melting temperature even at high potassium concentrations. Data were collected for both heating and cooling scans with a temperature gradient of 30°C/hr to mimic slow annealing conditions. Additional scan parameters included zero hold times for pre and post-scan holds and low gain to minimize noise with slow scanning conditions. Data analysis was done using Origin 7.0 software. Blank buffer scans were subtracted from DNA sample scans and normalized for DNA concentration. Further baseline adjustments were made based on pre- and post-denaturation regions.

NMR Sample Preparation and Spectroscopy

Fractions of the *c-Myc* G-quadruplex were separated with a mobile phase consisting of 100 mM KCl and 20 mM K_2HPO_4 (pH=7.0). Fractions were collected at 0.1 ml intervals. The five to seven fractions corresponding to the maxima for the seven major peaks were collected and combined from five separate runs. HPLC fractions were kept frozen at -80°C between runs. This material was concentrated using Microcon spin columns (Millipore). The concentrated material was then diluted to 330 μL total volume with HPLC buffer. 30 μM DSS and 10% D_2O were then added. The un-separated *c-Myc* sample was prepared by adding 10% D_2O and 30 μM DSS to a 1 mM sample. Samples were loaded into 5 mm Shigemi NMR tubes. NMR spectra were recorded using a 5 mm

inverse triple resonance (HCN) probe on Varian Inova spectrometer at 14.1 T using a cold probe.

Results and Discussion

Examination of the Polymorphism of the *c-Myc* Promoter G-Quadruplex Forming Sequence (Pu27).

G-quadruplexes formed from a single G-quadruplex-forming DNA sequence can sometimes adopt multiple topologies in solution that exist in equilibrium with one another (Lane *et al.*, 2008). The complexity of such a G-quadruplex containing system *in vitro* is poorly understood and often underrepresented. The Pu27 sequence has been previously demonstrated by non-denaturing gel electrophoresis to form higher-order species (Siddiqui-Jain *et al.*, 2002). This observation and its possible implications for data analysis are frequently ignored and rarely discussed in many studies, especially those examining interactions between small molecules and G-quadruplexes formed from the Pu27 sequence (Siddiqui-Jain *et al.*, 2002, Ou *et al.*, 2011). In the current work, we prepared the Pu27 sequence by initially rehydrating commercially obtained Pu27 sequence in KPEK 200 buffer (6mM K_2HPO_4, 2 mM KH_2PO_4, 186 mM KCl, 1 mM EDTA disodium salt dehydrate, pH 7.0) to a stock solution with a concentration of approximately 1 mM. The stock solution was dialyzed overnight (~16 hours) in buffer. After dialysis, the sample was diluted to a concentration of 200 μM, annealed in a 100°C water bath for 10 minutes, and allowed to cool to room temperature overnight. The annealed samples were then analyzed by SEC and NMR at 200 μM or further diluted to

39

0.50 OD (~2 μM) for CD and AUC analysis. These conditions are outlined in detail in the Experimental Section and will be referred to throughout the text as the standard conditions. Pu27 G-quadruplexes prepared using these conditions were highly heterogeneous and shown to contain at least seven species by SEC analysis (Figure 4A). Dilution of the sample for AUC did not alter the distribution of species observed (Figure 5); however, AUC analysis of the same sequence revealed a mixture consisting of at least 3 species of different sizes (Figure 4B).

40

Figure 4. The structural polymorphism of the *c-Myc* G-quadruplex-forming sequence (Pu27) prepared under standard conditions was examined by (A) SEC, (B) AUC, and (C) CD. (A) Absorbance of DNA was monitored at 260 nm and normalized to the area under the curve. (B), c(s) is the concentration distribution of sedimenting species based on absorbance at 260 nm and normalized to the area under the curve. (C), circular dichroism, Δε, was normalized to strand concentration.
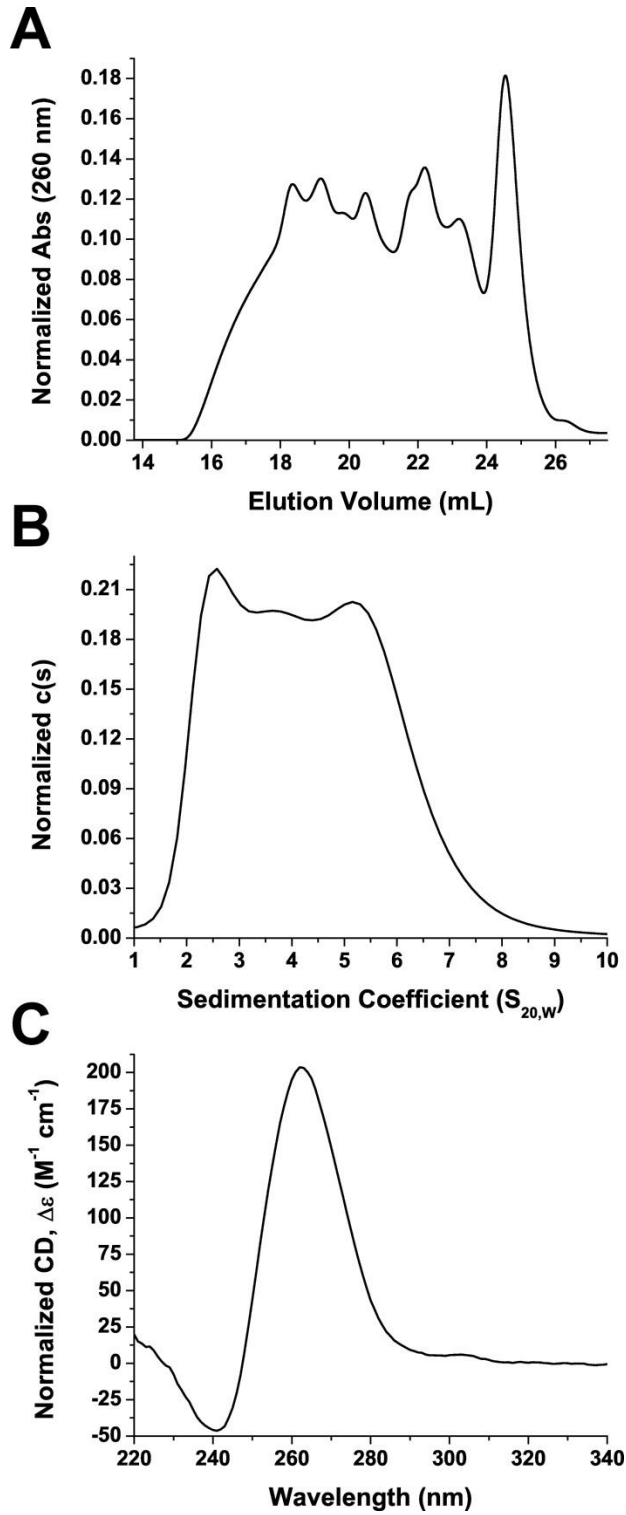
Figure 5. The distribution of Pu27 G-quadruplex species in sample before and after 100 fold dilution was examined by SEC. Absorbance of DNA was monitored at 260 nm and normalized to the area under the curve.
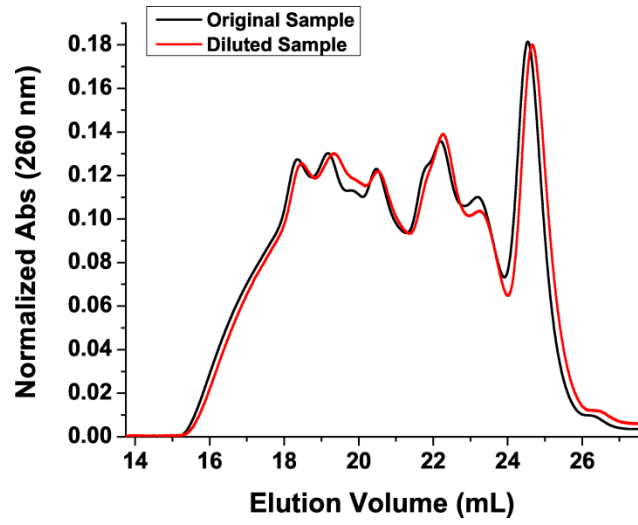
The heterogeneity of the G-quadruplex species was undetectable by CD spectroscopy (Figure 4C), a method commonly used to characterize G-quadruplex structures. The CD spectrum of the *c-Myc* G-quadruplex mixtures appeared as one species of what had been attributed to an all-parallel G-quadruplex with a positive peak at 260 nm and a negative peak at 240 nm (Đapić *et al.*, 2003, Karsisiotis *et al.*, 2011). 1D ^{1}H-NMR analysis was conducted on the mixture and showed a broad spectrum of overlapping GN1H resonances in the imino/amino proton region (Figure 6), characteristic of the G-quadruplex associated Hoogsteen hydrogen bonds. The overlapping resonances indicated that the solution contained either a complex mixture of G-quadruplexes or a structure of multiple DNA strands (i.e. G-wire) or both. Lastly, in 5 subsequent Pu27 samples that were also prepared under the same conditions, we observed the same general distribution of the 7 G-quadruplex species with only slight variations within experimental reproducibility (Figure 7).

Figure 6. 1D ¹H-NMR spectrum of the Pu27 G-quadruplex mixtures prepared under standard conditions. The imino region from the 1D ¹H-NMR spectrum of Pu27 G-quadruplex-forming sequence prepared under standard conditions demonstrates the formation of G-quadruplexes *in vitro* and the existence of a complex mixture of monomers and higher-order species.

46
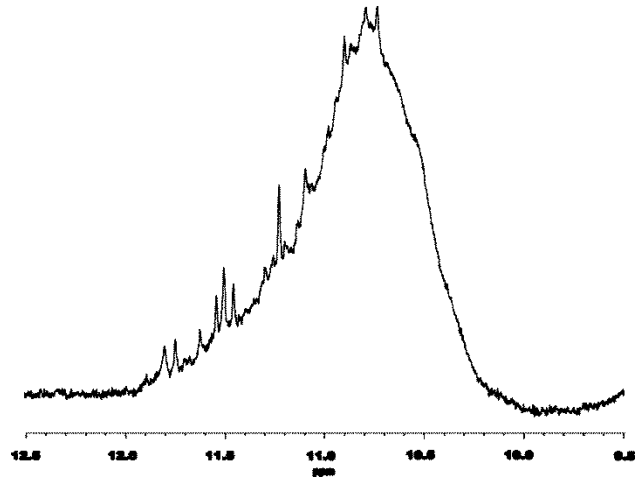
Figure 7. The distribution of G-quadruplex species from six different commercially obtained batches of Pu27 oligonucleotides was examined by SEC. G-quadruplex-forming sequences were prepared under standard conditions. Absorbance of DNA was monitored at 260 nm and normalized to the area under the curve.

48
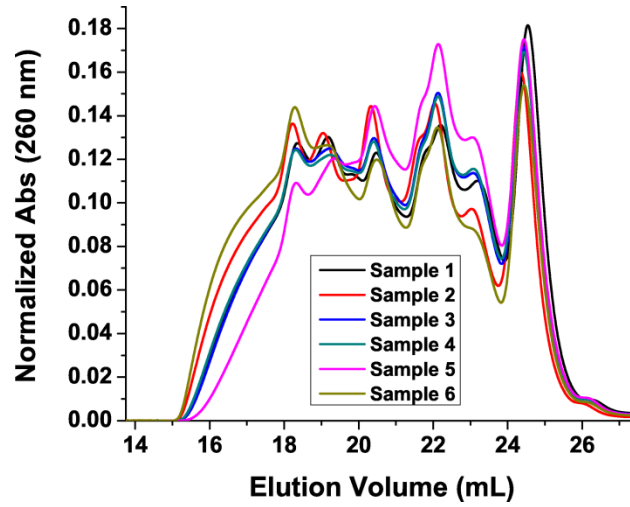
Examination of Fractions Obtained from SEC Separation of *c-Myc* G-quadruplex-Forming Sequence.

Chromatographic separation of the Pu27 sequence yielded 7 fractions. 1D ^{1}H-NMR spectroscopy on the 7 fractions demonstrated GN1H resonances in the imino/amino proton region for all 7 fractions indicating that all fractions contained G-quadruplex structures (Figure 8). Similar to the NMR spectrum of the mixture sample (Figure 6), NMR spectra for fractions 1 to 6 all displayed broad overlapping resonances. Thus, these fractions could also contain G-wire structures in addition to G-quadruplexes. Initial analysis of the fractions was accomplished by individually re-injecting each fraction into SEC to examine whether there was re-equilibration between the fractions (Figure 9A). Fractions 1 and 7 appeared predominantly as one species with very little re-equilibration observed for these samples. For fractions 2-6, while also appearing predominantly as one species, some minor re-equilibration was observed compared to fractions 1 and 7. Furthermore, when analyzed individually by AUC, each fraction sedimented essentially as a single species, which indicated that each component was thermodynamically stable or that re-equilibration and multimerization occurred very slowly at room temperature (Figure 9B). The AUC samples were saved and stored at room temperature. Even after one week, the samples remained unchanged (Figure 10) with only minor variations within experiment reproducibility. In addition, remixing of species separated by SEC did not change the distribution of species upon re-analysis by both SEC and AUC (Figure 11).

Figure 8. 1D ^{1}H-NMR spectra of the 7 fractions from SEC separation of the Pu27 G-quadruplex mixtures. The imino region from the 1D ^{1}H-NMR spectra of the 7 fractions demonstrated that all 7 fractions contained G-quadruplexes.
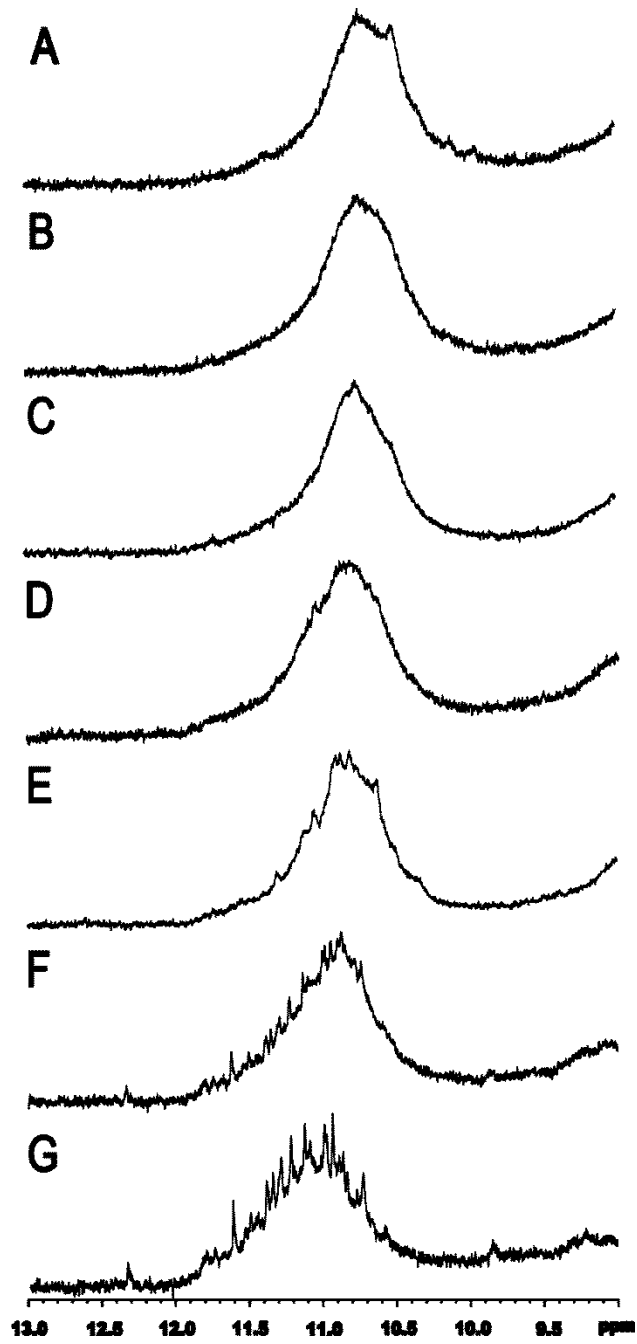
Figure 9. Seven fractions collected from SEC separation of the Pu27 sequence prepared under standard conditions were examined by (A) SEC, (B) AUC, and (C) CD. (A) Absorbance of DNA was monitored at 260 nm and normalized to the area under the curve. (B), c(s) is the concentration distribution of sedimenting species based on absorbance at 260 nm and normalized to the area under the curve. (C), Δε was calculated and normalized to G-quadruplex concentration.
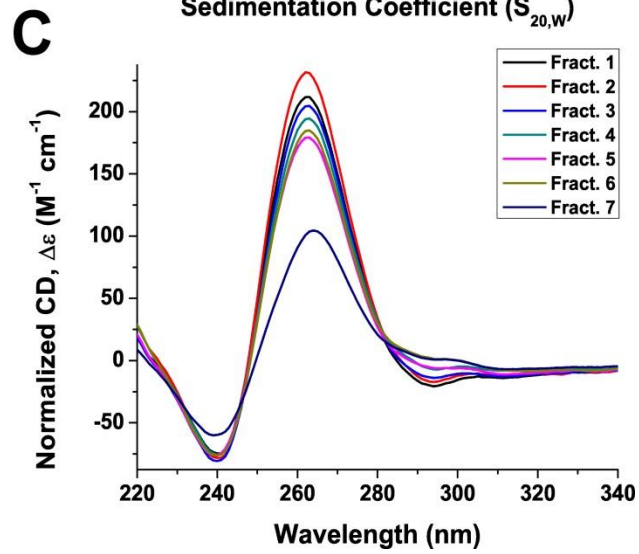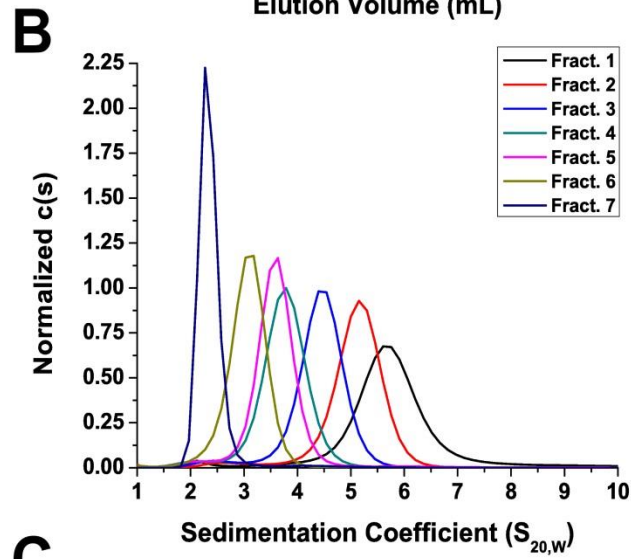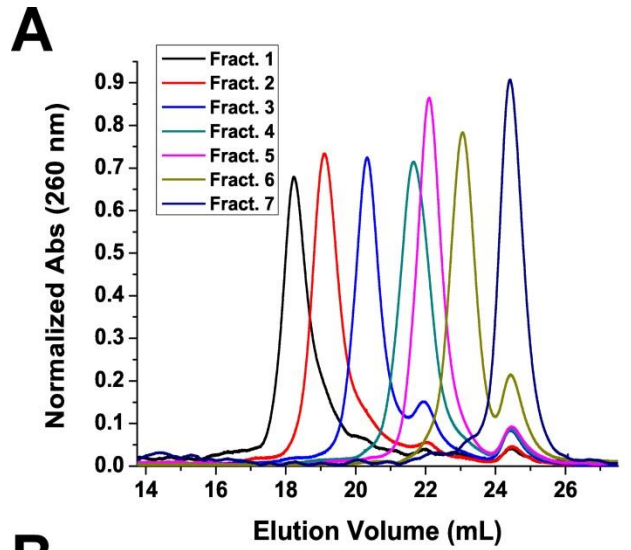
Figure 10. Fractions from SEC separation of Pu27 G-quadruplex mixtures before and one week following AUC analysis were examined by SEC. Absorbance of DNA was monitored at 260 nm and normalized to the area under the curve.

55
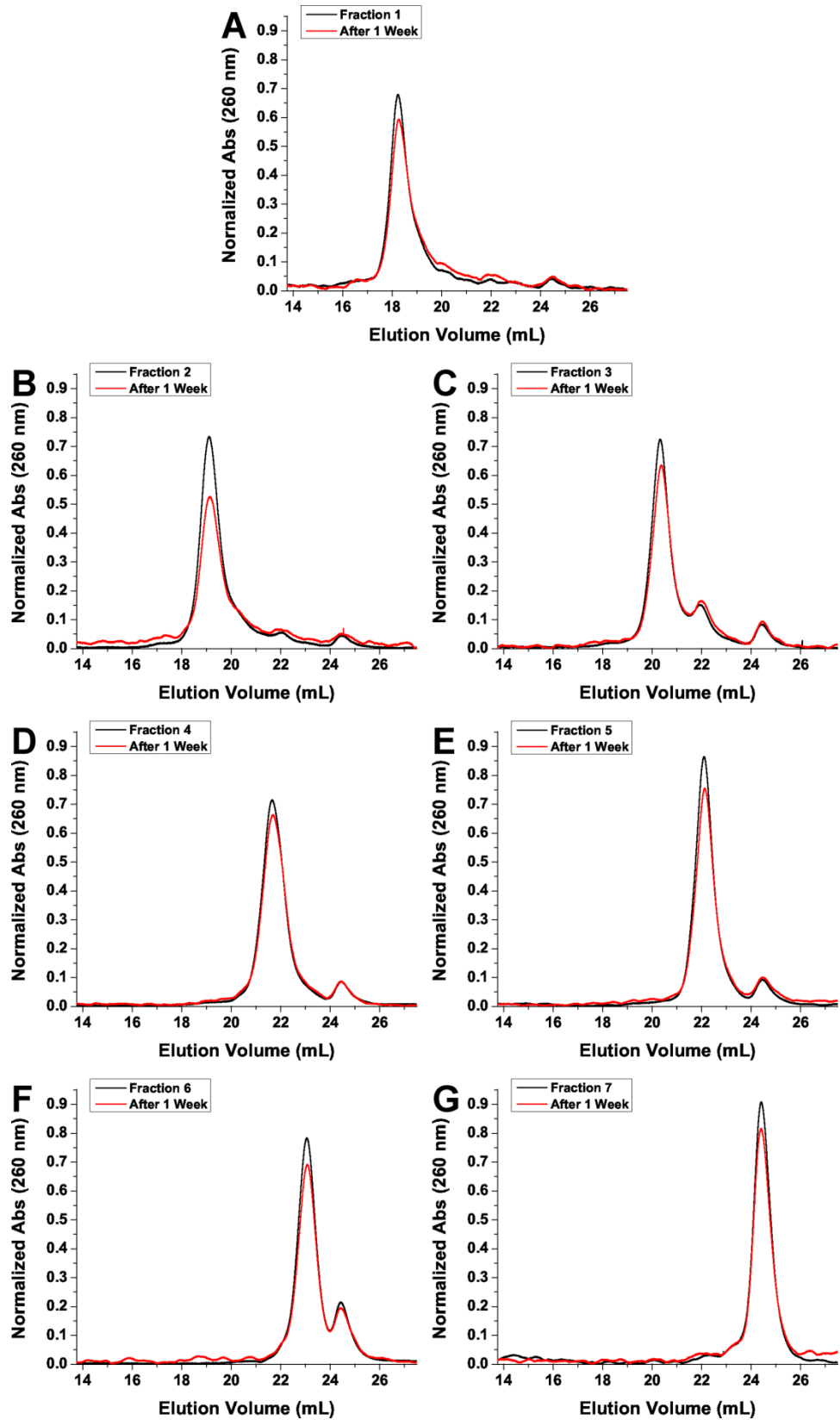
Figure 11. The distribution of Pu27 G-quadruplex species in a sample obtained by remixing SEC separated fractions were examined by SEC (A) and AUC (B). (A) Absorbance of DNA was monitored at 260 nm and normalized to the area under the curve. (B), c(s) is the concentration distribution of sedimenting species based on absorbance at 260 nm and normalized to the area under the curve.

A

B
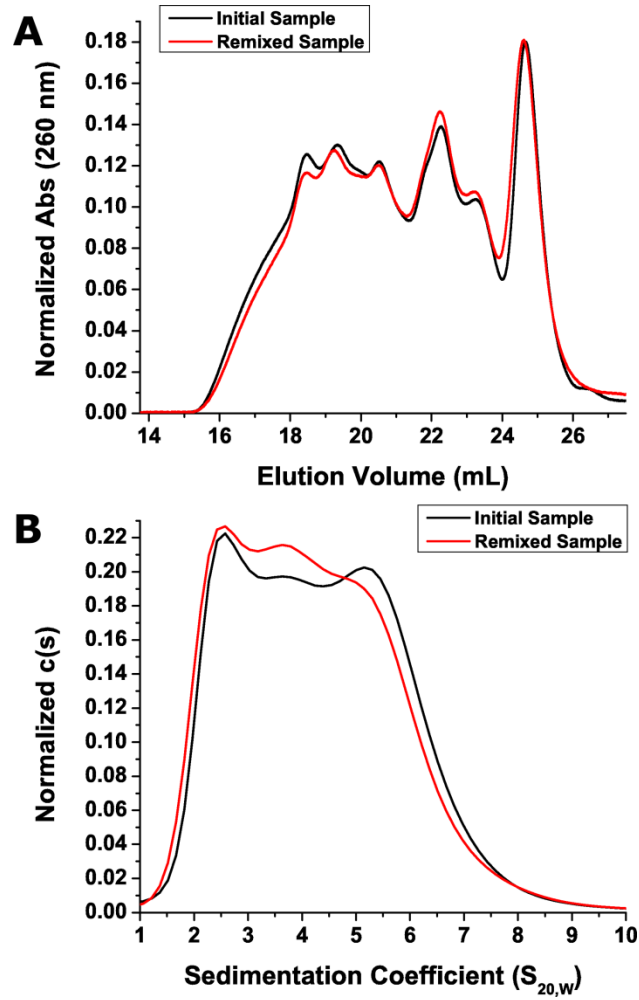
AUC analysis of fraction 7 yielded an experimentally determined molecular weight of 8,900 Da, which corresponds closely with the calculated molecular weight for one Pu27 strand (8,687.6 Da) indicating that this fraction is likely one or more unimolecular G-quadruplexes (Table 2). Fraction 6 (14,600 Da) consisted of G-quadruplex species that appeared to be intermediates between monomer and dimer. Fractions 5 (18,100 Da) and 4 (21,200 Da) each appeared to contain dimer G-quadruplexes while fractions 3 (30,000 Da), 2 (35,700 Da), and 1 (47,600 Da) were higher-order trimer, tetramer, and pentamer G-quadruplexes, respectively. This confirmed the previous AUC observation (Figure 4B) that the Pu27 sequence existed as a mixture between monomers (fraction 7) and higher-order G-quadruplex species (fractions 1-6). In addition, going from fraction 7 to fraction 1, the frictional ratio increased, which indicated that the higher-order Pu27 G-quadruplexes exhibited a more elongated solution structure compared to the monomer G-quadruplexes (Table 2). Together, these findings suggest the aggregation of lower-order G-quadruplex species to form higher-order G-quadruplex species *in vitro*.

Table 2. Hydrodynamic Properties of SEC Separated Pu27 *c-Myc* G-Quadruplex Fractions

Fraction	Elution Volume (mL)	Sedimentation Coefficient ($S_{20,W}$)	Molecular Weight (D)	Frictional Ratio
1	18.2	5.89	47,600	1.468
2	19.1	5.22	35,700	1.369
3	20.3	4.52	30,000	1.408
4	21.6	3.83	21,200	1.318
5	22.1	3.61	18,100	1.259
6	23.1	3.11	14,600	1.264
7	24.3	2.38	8,900	1.188

We employed CD spectroscopy to further characterize the chromatographically separated G-quadruplex species. Similar to the mixture, all 7 fractions appeared by CD as one species of what has been attributed to an all-parallel G-quadruplex with a positive peak at 260 nm and a negative peak at 240 nm (Figure 9C) (Đapić *et al.*, 2003, Karsisiotis *et al.*, 2011). The spectrum for the monomer G-quadruplex (fraction 7) had the smallest magnitude while the magnitudes of the spectra for the higher-order species (fraction 1-6) were greater. Increase in normalized circular dichroism ($\Delta\varepsilon$) is observed for higher-order G-quadruplex structures (Petraccone *et al.*, 2011). In the current work, we normalized the circular dichroism values with respect to the number of strands using the extinction coefficient calculated by the nearest neighbor method. This value corresponded to the absorbance of the single-stranded form of the Pu27 sequence. However, the absorbance of the G-quadruplex form of a sequence is known to be less than the single-stranded absorbance (Mergny *et al.*, 2005) which leads to an underestimation of the G-quadruplex concentration and an overestimation of the $\Delta\varepsilon$. This unknown was a limitation to our normalization method and prevents us from making any definitive conclusions regarding the CD data of the Pu27 G-quadruplex structures from the 7 fractions.

Lastly, it should be noted that the 7 fractions collected might not represent the complete picture of the polymorphism exhibited by the Pu27 sequence. These fractions did not account for any G-quadruplex species that might have eluted at less than 18.2 ml, as indicated by the shoulder of the SEC chromatogram (Figure 4A). In addition, it should be noted that techniques such as SEC and AUC are low resolution techniques that might not be able to discern between species which have similar hydrodynamic properties. The tailing of the SEC peaks (Figure 9A) and broadening of AUC distribution curve (Figure

9B) suggested that several, if not all, fractions may contain more than one G-quadruplex species. In fact, fraction 7 was shown by 1D ^{1}H-NMR to be a mixture of at least two species (Figure 8). The enrichment by SEC separation allowed for individual GN1H resonances to be observed rather than just the broad spectrum that was observed for the mixture (Figure 6). However, there were more resonances than expected for one topology implying that fraction 7 remained a mixture of several G-quadruplex monomers. The NMR spectra of the other fractions revealed that all displayed overlapping GN1H resonances indicative of higher order species and/or possibly mixtures of several species.

Effect of Annealing on G-Quadruplex Distribution *in vitro*.

The annealing protocol used can play an important role in G-quadruplex formation. We conducted several experiments to determine whether different aspects of the annealing process may affect the distribution of Pu27 G-quadruplex species in solution. Our protocol to prepare samples for AUC and CD analysis called for the samples to be diluted from the annealed concentration of 200 μM (~50 A_{260}) to 2 μM (~0.50 A_{260}). We asked the question, "Will the distribution of G-quadruplex species change if the samples are annealed at 2 μM instead of 200 μM?" When annealed at the lower concentration (2 μM), we observed by both SEC and AUC a dramatic increase in the amount of monomer G-quadruplexes with a corresponding decrease in the amount of higher-G-quadruplexes (Figure 12). To further explore the concentration dependency of G-quadruplex formation we examined the distribution of Pu27 G-quadruplexes at three different concentrations of 100 μM, 200 μM, and 400 μM (Figure 13). SEC analysis revealed a concentration dependent effect between 400 μM and 200 μM with the higher

concentration of DNA shifting the equilibrium toward higher-order G-quadruplexes. This concentration-dependent effect was not observed between 200 μM and 100 μM.

Figure 12. The concentration dependency of G-quadruplex formation was examined by SEC (A) and AUC (B). (A) Absorbance of DNA was monitored at 260 nm and normalized to the area under the curve. (B), c(s) is the concentration distribution of sedimenting species based on absorbance at 260 nm and normalized to the area under the curve.

A

B

Figure 13. The DNA concentration dependency of G-quadruplex formation was examined by SEC. Absorbance of DNA was monitored at 260 nm and normalized to the area under the curve.
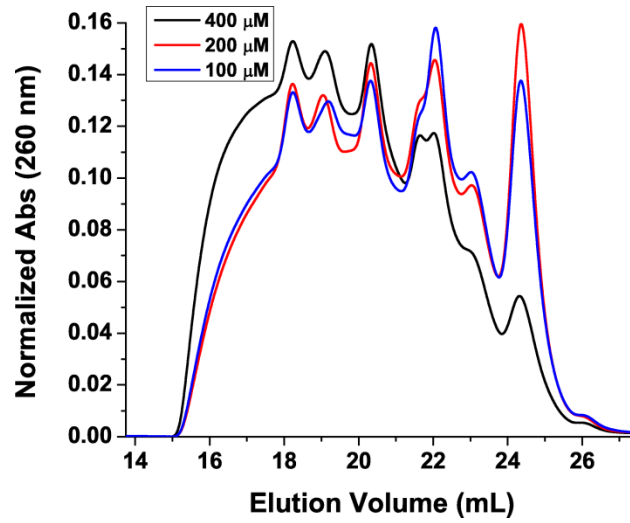
In addition to the concentration, we also investigated how the distribution of G-quadruplex species can be affected by annealing temperature and quenching process. In the first experiment, Pu27 G-quadruplex solutions were prepared at 200 μM. One was annealed at 100°C in a water bath, one at 120°C under pressure in a silicon oil bath, and one was not annealed (room temperature). All were examined by SEC (Figure 14A). There were only minor differences between the chromatograms of the sample annealed at 100°C and the sample that was not annealed. However, with the sample annealed at 120°C, a dramatic shift occurred and a decreased amount of higher-order G-quadruplex species was observed with the complementary increased amount of lower-order and monomer species. It is possible that at 120°C, the structures of some of the higher melting G-quadruplexes were disrupted. When these G-quadruplexes refolded while cooling, a new equilibrium was established resulting in an altered composition of G-quadruplex species in solution. In addition to the annealing temperature dependency, we also observed a change in distribution of G-quadruplex species with a higher amount of monomer species when the annealed sample was quenched on ice for 10 minutes rather than allowed to slowly cool to room temperature overnight (Figure 14B). The findings from these three experiments highlight the importance of experimental protocol and annealing in determining the distribution of G-quadruplexes in solution.

Figure 14. The (A) annealing temperature dependency and (B) quenching protocol dependency of G-quadruplex formation were examined by SEC. Absorbance of DNA was monitored at 260 nm and normalized to the area under the curve.
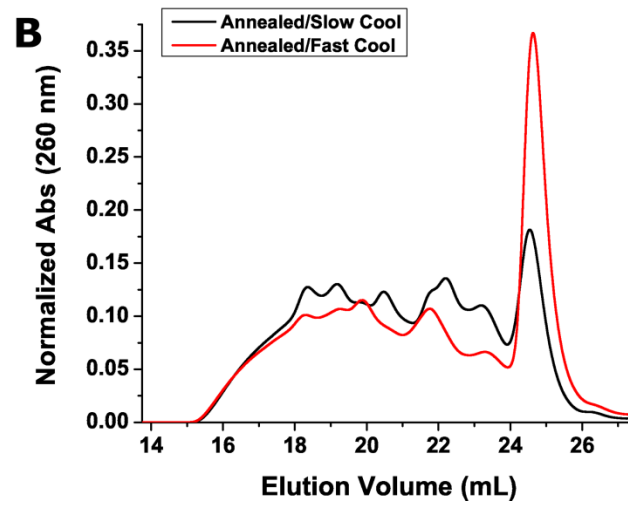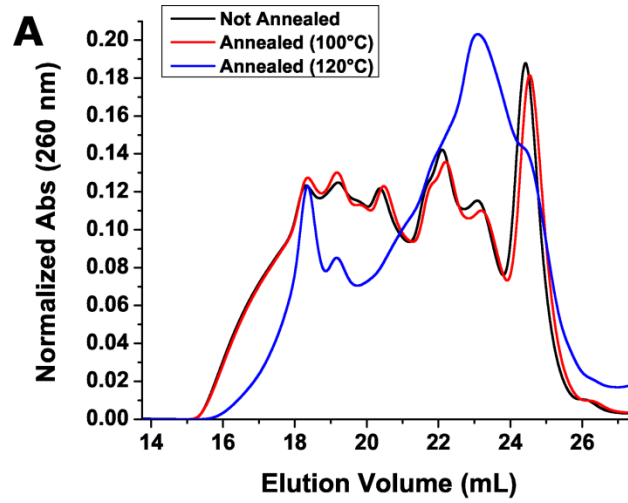
**Effect of Sample Preparation on G-Quadruplex Distribution *in vitro.***

In addition to annealing, preparation of G-quadruplex samples often involves dialysis to remove small fragments remaining from synthesis of the oligonucleotide sequence and to facilitate the folding of the oligonucleotides into G-quadruplexes. An experiment was conducted to determine the effect of dialysis conditions on Pu27 G-quadruplex species distribution. Four different batches of Pu27 oligonucleotide were prepared by rehydrating the sample in KPEK 200 to a concentration of 1000 μM. One sample was kept without dialysis while three samples were dialyzed overnight in KPEK 200, one using a 0.1-.0.5 kDa MWCO membrane, one using a 0.5-1.0 kDa MWCO membrane, and one using a 3.5-5.0 kDa MWCO membrane. All four samples were diluted to a concentration of 200 μM, annealed at 100°C for 10 minutes, slowly cooled overnight to room temperature, and analyzed using SEC. A greater amount of higher-order species was observed in dialyzed samples compared to samples not dialyzed (Figure 15). In addition, as the MWCO of the dialysis membrane was increased, an increase in higher-order species was observed. The A_{260} of the sample from before and after dialysis differed by only about 5-10%, which is within the standard for typical recovery from dialysis, indicating that the increased amount of larger G-quadruplex species observed was not due to the smaller species being dialyzed out of solution. Rather, it suggests some type of interactions between G-quadruplex structures and the dialysis membrane leading to the formation of higher-order structures. These interactions could be specific to the particular dialysis membrane being used in this study and it is possible that a different set of results could be obtained if a different type of dialysis membrane was utilized.

72

Figure 15. The dialysis dependency of G-quadruplex formation was examined by SEC. Absorbance of DNA was monitored at 260 nm and normalized to the area under the curve.

74
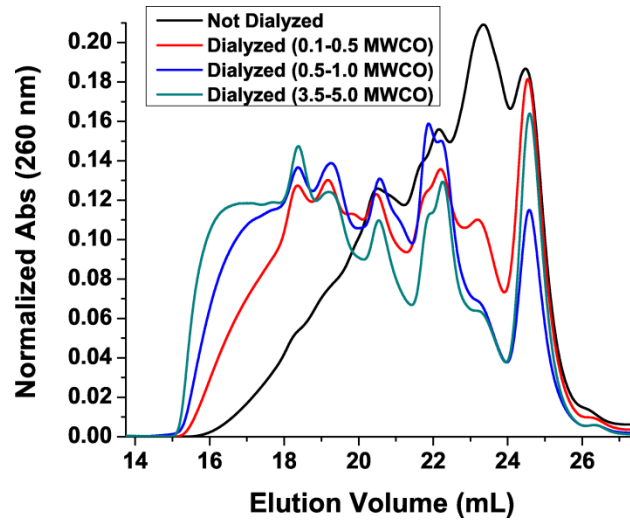
We also observed that the distribution of G-quadruplexes species formed from the Pu27 sequence was sensitive to other minor, seemingly innocuous, changes in sample preparation. For example, DNA obtained commercially is usually shipped lyophilized. We observed more monomeric G-quadruplexes when the lyophilized DNA was initially dissolved in deionized water instead of KPEK (Figure 16). In addition to changes in sample preparation processes, minor differences in buffer composition can have significant effect on G-quadruplex formation (Figure 17). When the acid form of EDTA was substituted for the disodium salt dihydrate form, more monomeric species were observed. These findings further supported that G-quadruplex formation can be affected at any step of the sample preparation process and the procedure employed should be carefully reported.

Figure 16. The rehydration of oligonucleotide protocol dependency of G-quadruplex formation was examined by SEC. A greater proportion of monomeric and dimeric G-quadruplex species were produced when the lyophilized DNA was initially dissolved in deionized water instead of KPEK. Absorbance of DNA was monitored at 260 nm and normalized to the area under the curve.
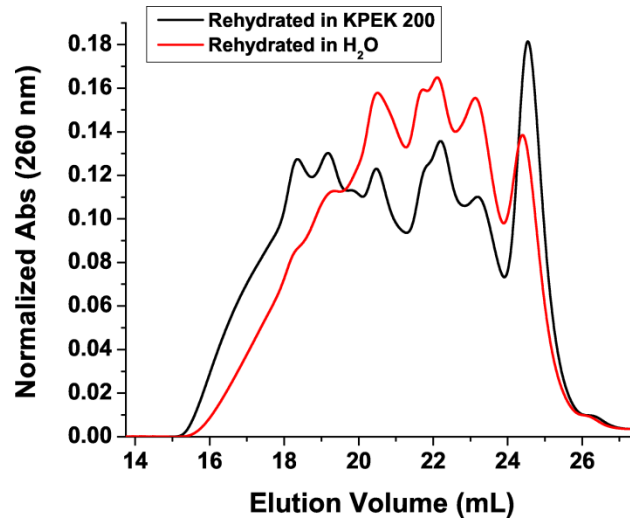
Figure 17. The EDTA dependency of G-quadruplex formation was examined by SEC. When the acid form of EDTA was substituted for the disodium salt dihydrate form, more monomeric species were observed. Absorbance of DNA was monitored at 260 nm and normalized to the area under the curve.
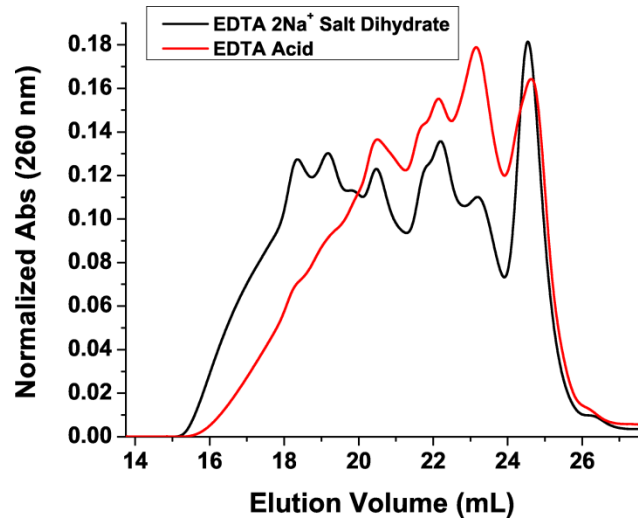
Effect of Potassium Concentration on G-Quadruplex Distribution *in vitro*.

The concentration of potassium in solution has been known to play a large role in G-quadruplex stability and formation (Gray *et al.*, 2009b). We examined the effect of altered potassium concentration on the distribution of Pu27 G-quadruplexes in solution. The Pu27 sequence was prepared in buffers with varying concentrations of potassium 25 mM, 200 mM, and 400 mM. The potassium concentration of 25 mM was chosen because it has been previously demonstrated that this is the minimum concentration of potassium required for G-quadruplex formation (Gray *et al.*, 2009b). The Pu27 samples were initially rehydrated and dialyzed in the corresponding buffer, diluted to 200 μM, annealed at 100°C for 10 minutes with slow cooling, and analyzed by SEC or further diluted to 2 μM for AUC analysis. In high potassium conditions (400 mM), the distribution was predominantly higher-order species while in low potassium conditions (25 mM) it was predominantly monomers (Figure 18A-B). Additionally, ^{1}H-NMR analysis of the sample in 25 mM potassium revealed a GN1H resonance pattern similar to that of fraction 7 (Figure 19). These findings suggest that, in addition to stabilization of the G-quartet stem, potassium ions also interacts with G-quadruplexes externally and help drive the formation of higher-order G-quadruplex species.

Figure 18. The potassium concentration dependency of G-quadruplex formation was examined by SEC (A), AUC (B), and DSC (C). (A) Absorbance of DNA was monitored at 260 nm and normalized to the area under the curve. (B), c(s) is the concentration distribution of sedimenting species based on abso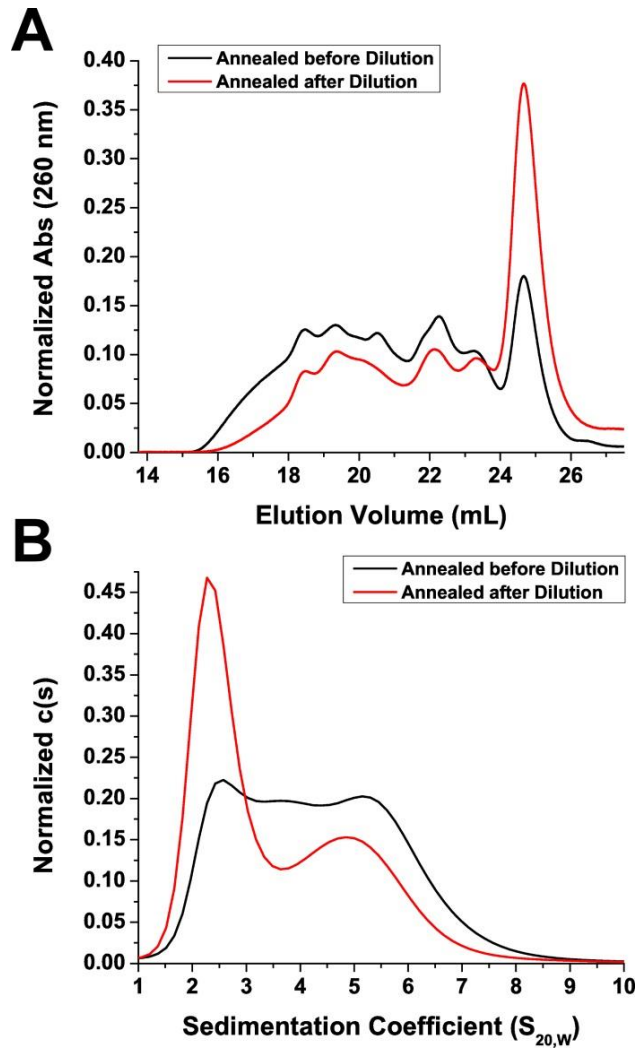rbance at 260 nm and normalized to the area under the curve. (C) Excess heat capacity measurement for DSC was normalized with respect to strand concentration (200 μM).
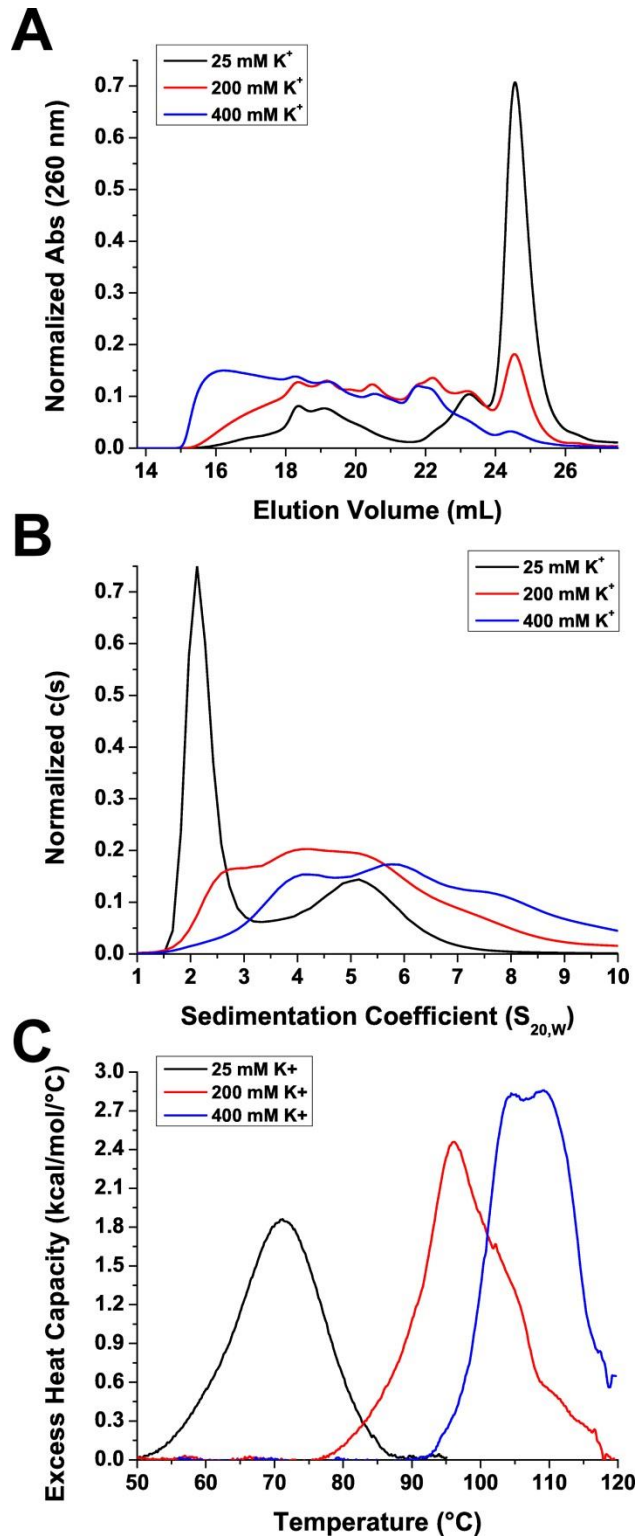
Figure 19. 1D ^{1}H-NMR spectrum of the Pu27 G-quadruplex mixtures prepared under low potassium conditions (25 mM K$^+$). The imino region from the 1D ^{1}H-NMR spectrum of the Pu27 G-quadruplex mixtures in low potassium conditions (25mM) displayed individual resonances indicating a mixture of monomers G-quadruplexes. This NMR spectrum highly resembles the 1D ^{1}H-NMR spectrum of SEC separated fraction 7 (Figure 8G).
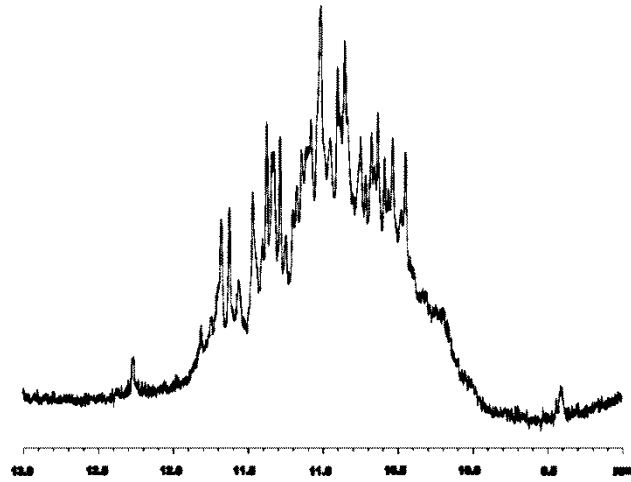
The sample of G-quadruplexes in 400 mM K^+ was diluted in a low potassium buffer until the potassium concentration was reduced to 25 mM. In order to account for any effect the dilution process might have on G-quadruplex equilibrium, another sample was diluted while keeping the potassium concentration at 400 mM. Similarly, the 25 mM K^+ sample was diluted in high potassium buffer to bring the final potassium concentration to 400 mM for the test sample or kept at 25 mM K^+ as a control. The process of dilution alone did not alter the distribution of G-quadruplex species (Figure 20). Going from high potassium conditions (400 mM) to low potassium conditions (25 mM) resulted in less of the higher-order species and more of the lower-order species such as monomers while a change to higher potassium concentration led to lesser amount of monomers and higher amount of higher-order G-quadruplexes (Figure 21). It should be noted that this is a very slow process. These changes were only observable after nearly two weeks of incubation as one day of incubation revealed no changes.

Figure 20. The distribution of Pu27 G-quadruplex species in (A) low potassium conditions (25 mM K$^+$) and (B) high potassium conditions (400 mM K$^+$) before and after dilution was examined by SEC. Dilution of the *c-Myc* G-quadruplex-forming sequence did not alter the distribution of G-quadruplex species. Absorbance of DNA was monitored at 260 nm and normalized to the area under the curve. The findings here agree with previous observations that dilution after annealing does not alter the distribution of G-quadruplex species even after more than two weeks at room temperature.
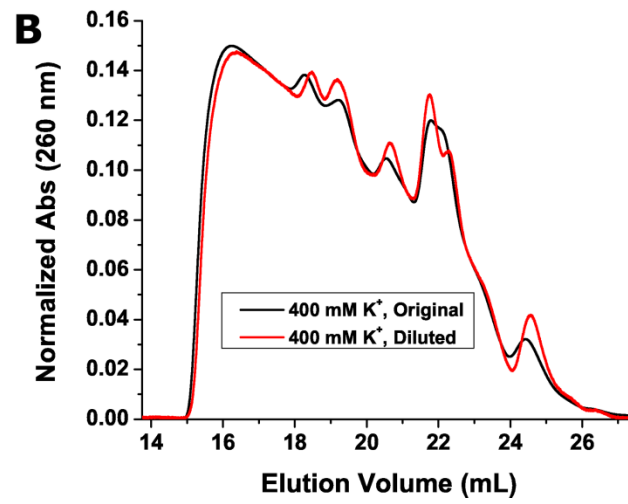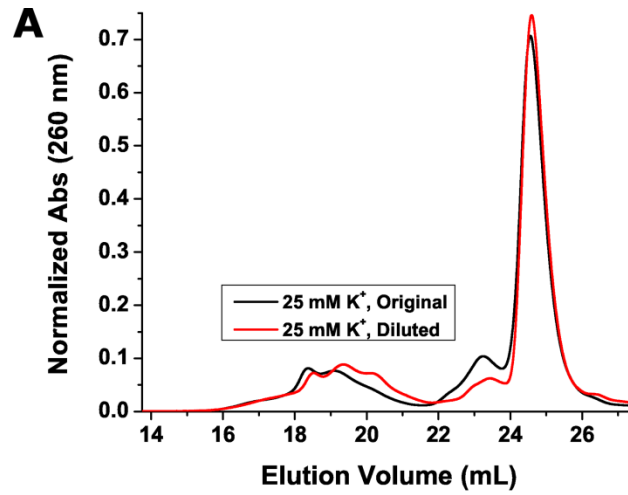
Figure 21. Changes in distribution of Pu27 G-quadruplex species when (A) potassium concentrations were increased from 25 mM to 400 mM and (B) potassium concentrations were decreased from 400 mM to 25 mM were examined by SEC. Absorbance of DNA was monitored at 260 nm and normalized to the area under the curve. It should be noted that significant changes in distribution of Pu27 G-quadruplex species were only observable more than two weeks after potassium concentrations were altered. In addition, it should also be noted that these samples were annealed prior to the changes in potassium concentration. There was no annealing after the potassium concentrations were altered.
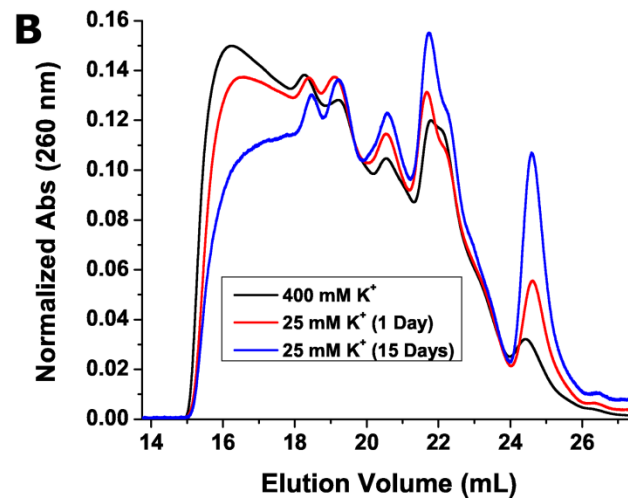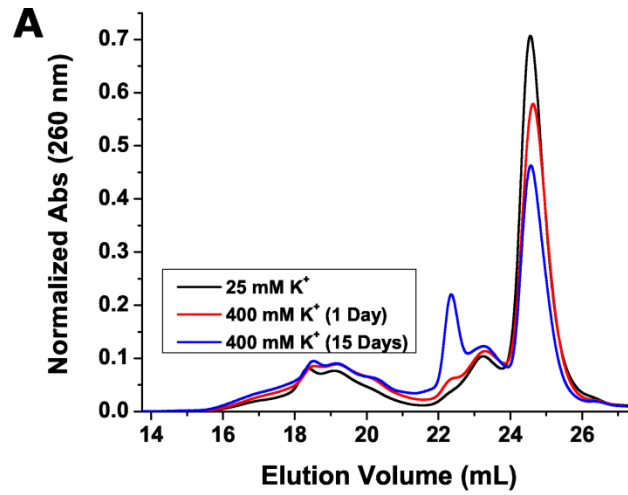
The effect of potassium concentration on G-quadruplex formation was also examined using DSC melting (Figure 18C). Under standard conditions in 200 mM K^+ buffer, the Pu27 G-quadruplexes were extremely stable, requiring superheating past the boiling point of water to melt. The high melting temperature was an indicator of the stability of the G-quadruplex structures. In addition, the large change in excess heat capacity of the system between 80°C to 120°C further indicated the stability of these G-quadruplex structures. The DSC data agreed with previous observations by SEC that annealing at 100°C is not at a sufficiently high enough temperature to unfold some of the higher-order G-quadruplex species (Figure 14A). Compared to the thermogram at 200 mM K^+, the thermogram under high potassium conditions (400 mM) shifted to the right to indicate increased stability while the thermogram under low potassium conditions (25 mM) shifted to the left to indicate decreased stability. This observation correlates with that seen by SEC and AUC that in high potassium (400 mM) there is an increase in higher-order, presumably more stable, G-quadruplex species, while in low potassium (25 mM), there are more of lower-order, presumably less stable G-quadruplexes. It should be noted that the presence of additional cations alone could result in an increase in melting temperature. Therefore, from the data we were not able to determine how much of the increase in melting temperature was due to the increase in higher-order structures and how much was due to the stabilizing effects of higher potassium concentration. However, these findings indicate that, in additional to sample preparation protocols (e.g. annealing, dialysis, etc.), experimental conditions are also important and that detailed reporting of methodology is necessary when studying with G-quadruplex structures.

Interaction of TMPyP4 with the Pu27 G-Quadruplex-Forming Sequence.

As the discovery of small-molecule G-quadruplex-based inhibitors is one goal of G-quadruplex research, we examined the effect of the polymorphism of a G-quadruplex-forming sequence on *in vitro* small molecule binding experiments by looking at the interactions of TMPyP4 with the Pu27 sequence. TMPyP4 displayed a preference for the higher-order subpopulations of the Pu27 G-quadruplex distribution (Figure 22). The species that eluted at 20.3 and 21.6 mL, corresponding to fractions 4 and 5 (i.e. dimer G-quadruplex species) were the most preferred and their association with TMPyP4 can be seen at as low as a 1:32 [TMPyP4]:[Pu27] molar ratio. The next preferred subset was the monomer species that eluted at 24.3 mL, corresponding to fraction 7, followed by the higher-order species that eluted at less than 20.0 mL, corresponding to fraction 1, 2, and 3. Lastly, it is interesting to note a lack of observed TMPyP4 interaction with the species that eluted at 23.1 mL, corresponding to fraction 6 (i.e. the monomer/dimer intermediate G-quadruplex species), but it does not rule out possible TMPyP4 binding to these species at higher concentrations. These results indicate that small molecules may not bind the different G-quadruplex forms within the same sequence distribution equally, a significant finding when contemplating using structure-based drug design.

Figure 22. The interactions of TMPyP4 with Pu27 G-quadruplexes at different molar ratios were examined by SEC. The DNA bound form of TMPyP4 were monitored at 445 nm and normalized to the area under the curve.
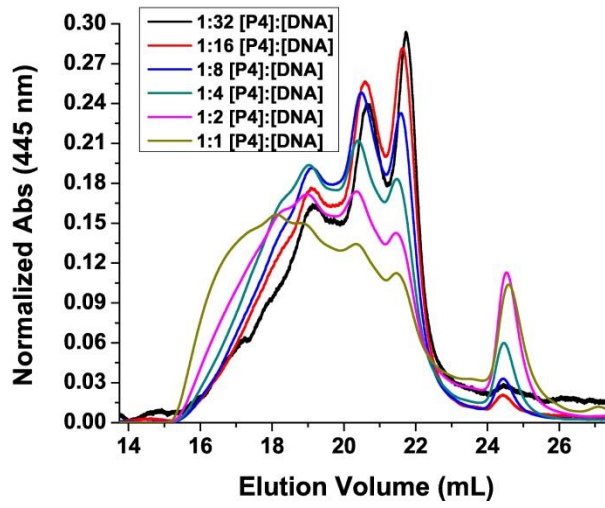
Another important observation is that TMPyP4 not only prefers to bind to higher-order species but it also directly alters the polymorphic distribution of G-quadruplex towards formation of these higher-order species (Figure 23). The addition of TMPyP4 to the Pu27 system resulted in lower amounts of the species that eluted 24.3 mL, 23.1 mL, 22.1 mL, 21.6 mL and 20.3 mL and greater amounts of the species that eluted at 19.1 mL and 18.2 mL as well as other higher-order species that elute before 18.2 mL. This observation is important as it suggests that there is induced change in the original distribution thus altering the baseline and potentially altering the response in any given experiment. On the other hand, the findings from this and previous experiments were able to provide some insight into the binding mode of TMPyP4. These findings suggest that TMPyP4 binds to Pu27 *c-Myc*, and possibly other G-quadruplexes, through end-pasting which is consistent with previous reports (Wei *et al.*, 2006, Freyer *et al.*, 2007). This binding mode facilitates TMPyP4 stabilization of G-quadruplex aggregation by the TMPyP4 interactions between G-quadruplexes.

Figure 23. The effect of TMPyP4 interactions with DNA on the distribution of Pu27 G-quadruplexes in solution were examined by SEC. Absorbance of DNA was monitored at 260 nm and normalized to the area under the curve.
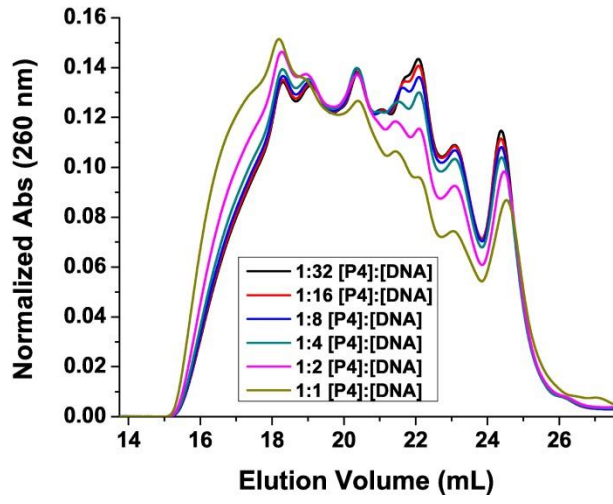
Polymorphism of Modified *c-Myc* G-Quadruplex-Forming Sequences.

We end our discussion by examining four modified *c-Myc* G-quadruplex forming sequences reported in the literature (Myc-1245, Myc-2345, MYC22-G14T/G23T, and Pu24). These sequences were prepared under the same standard conditions that we used for the Pu27 sequence and analyzed using NMR (Figure 24), SEC, AUC and CD (Figure 25). The 1D ^{1}H-NMR spectra for the four sequences were comparable to previously reported spectra. Individual resonances were observed for all four sequences compared to the broad overlapping resonances observed for the Pu27 sequence. The NMR data were in agreement with SEC and AUC findings which demonstrated the enrichment of monomer G-quadruplexes (Figure 25A-B). In fact, of the four sequences, MYC22-G14T/G23T was enriched the most in single monomeric species and yielded the cleanest NMR spectrum (Figure 24C). In addition, the G-quadruplex structures from the modified sequences did not aggregate to the same extent as the G-quadruplex structures from the Pu27 sequence, suggesting some fundamental differences between the two systems. This observation was further supported by CD spectroscopy, which showed a Δε spectrum of greater magnitude for the Pu27 sequence compared to the modified sequences (Figure 25C).

Figure 24. 1D ^{1}H-NMR spectra of G-quadruplexes formed from derivatives of the Pu27 G-quadruplex-forming sequence. The imino region from the 1D ^{1}H-NMR spectra of the modified *c-Myc* G-quadruplex-forming sequence Myc-1245 (A), Myc-2345 (B), MYC22-G14T/G23T (C), and Pu24 (D) shows marked reduction in polymorphism compared to the parent Pu27 sequence (Figure 5). The spectra here agreed with previously reported spectra for these sequences.
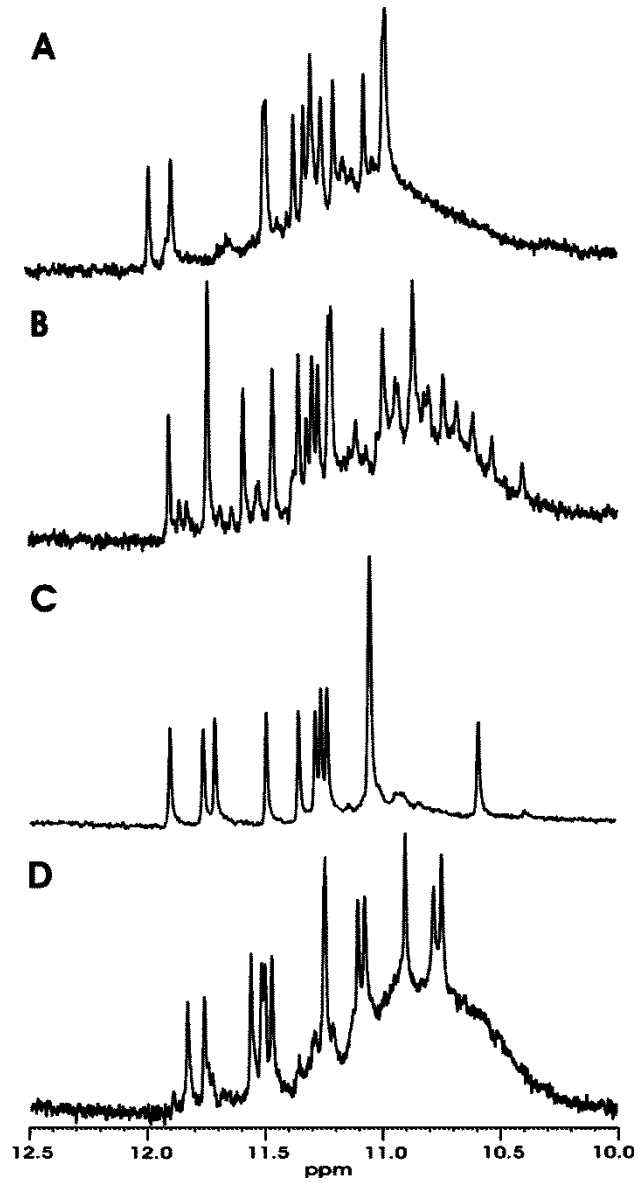
Figure 25. The polymorphism of *c-Myc* G-quadruplexes formed from the Pu27 sequence and four derivative sequences prepared under standard conditions was examined by (A) SEC, (B) AUC, and (C) CD. (A) Absorbance of DNA was monitored at 260 nm and normalized to the area under the curve. (B), c(s) is the concentration distribution of sedimenting species based on absorbance at 260 nm and normalized to the area under the curve. (C), circular dichroism, Δε, was normalized to strand concentration.
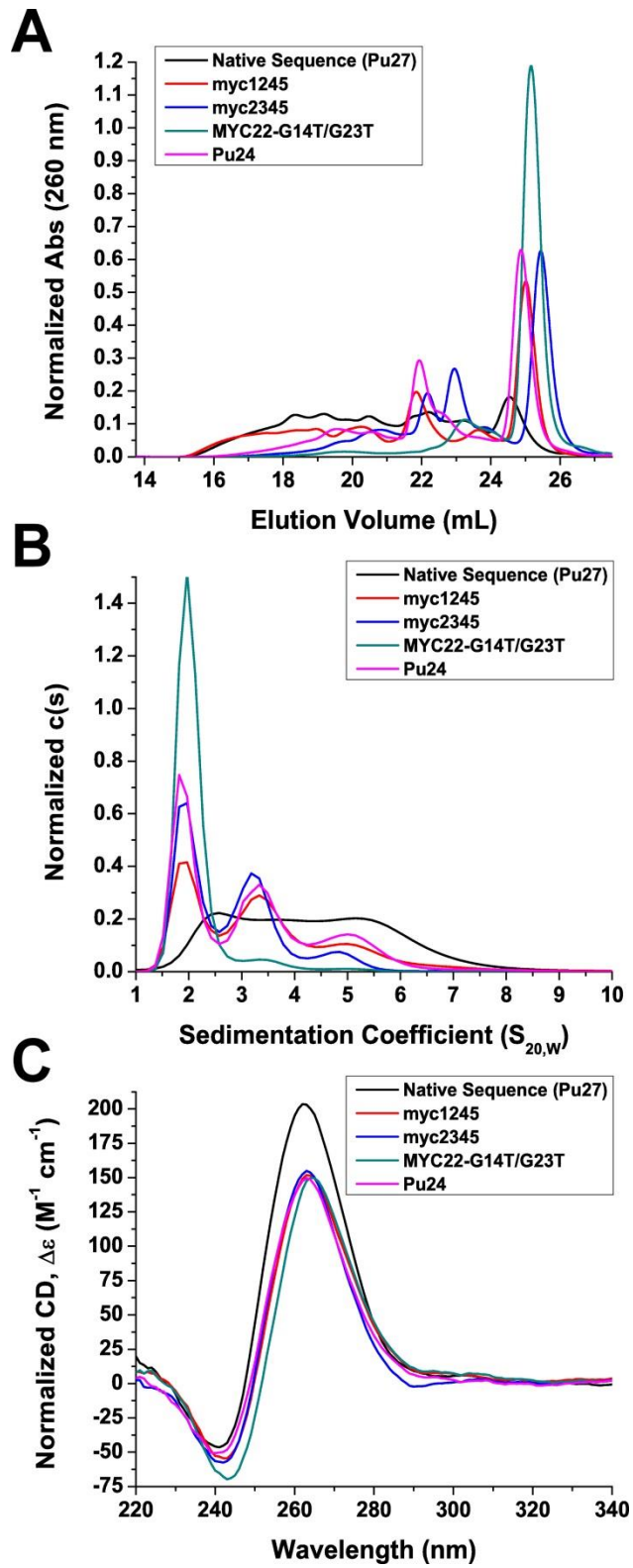
Furthermore, it should be noted that the monomer G-quadruplexes formed from the modified sequences eluted and sedimented more slowly than the monomer G-quadruplexes formed from the Pu27 sequence. One possible explanation is that the modified sequences are smaller (22-25 bp) than the parent sequence (27 bp). However, another possible explanation is that sequence modifications did not select for a G-quadruplex topology from the original ensemble but induced a new topology altogether. While we were not able to make a definitive comparison between the G-quadruplex structures formed by the modified sequences and the G-quadruplex structures formed by the parent sequence, these findings further emphasized the limitation of sequence modification in that a new topology can potentially be induced that is not part of the original distribution, or, at the very least, a topology is artificially selected that might not be preferred *in vivo* (Lane *et al.*, 2008).

Conclusion

The *c-Myc* sequence in solution is a highly polymorphic ensemble of structures containing a mixture of monomers and higher-order G-quadruplexes. We demonstrated that the higher-order species always exist *in vitro*. Even in the low potassium conditions (25 mM) where the monomers are predominant, a small fraction of the G-quadruplex species still existed in the higher-order form. The existence of higher order *c-Myc* G-quadruplex structures *in vitro* is a critical detail often ignored since it is believed that only the monomers are relevant *in vivo*. It also leads to the question "What is the relevance of an affinity constant measured for a compound against the *c-Myc* G-quadruplexes if the

compound prefers the higher order species *in vitro*?" With careful controls and taking into account the polymorphism a more realistic measurement can be reported.

We have shown that the distribution of G-quadruplex species is highly sensitive to variability in sample preparation. Small differences in experimental conditions can lead to dramatic differences in the system being studied. The concentration of ions and DNA played major roles in determining the distribution between higher-order and monomer species of G-quadruplexes. Preparation of G-quadruplex samples often involves dialysis to remove small fragments left over from synthesis, and annealing at high temperature to unfold the DNA allowing for G-quadruplex formation as the sample cools. We show here that the choice and method of dialysis and annealing protocol can greatly influence the distribution of *c-Myc* G-quadruplex species. EDTA is often added to the DNA buffer as a chelating agent to deplete metal ions, i.e. Ca^{2+} and Fe^{3+}, as these divalent cations could affect G-quadruplex stabilities by directly interacting with the DNA(Blume *et al.*, 1997, Miyoshi *et al.*, 2001) or by acting as cofactors for metal-dependent DNA hydrolytic enzymes. However, the addition of EDTA to the sample buffer can result in significantly different G-quadruplex populations. The polymorphism observed becomes even more complicated when ligands are introduced. It was demonstrated that not only does TMPyP4 prefer one species of the G-quadruplex population over another but that the ligand itself actively plays a role in changing the equilibrium between the species. With so many different factors to consider and multiple opportunities along the way to introduce new variability, it is not difficult to see why there are differences in reports regarding TMPyP4 interactions with G-quadruplexes.

The biological relevance of such *in vitro* G-quadruplex systems is currently unknown. However, these findings indicate that sample preparation should be carefully considered, controlled, and reported when working with G-quadruplexes. G-quadruplexes represent an exciting area in the future of anticancer therapy. An understanding and appreciation of the complexity associated with G-quadruplex would result in higher reproducibility of data and more efficient targeted drug discovery.

CHAPTER III

Guanine-rich oligonucleotides can adopt non-canonical tertiary structures known as G-quadruplexes, which can exist in different forms depending on experimental conditions. High-resolution structural methods, such as X-ray crystallography and NMR spectroscopy, have been of limited usefulness in resolving the inherent structural polymorphism associated with G-quadruplex formation. The lack of, or the ambiguous nature of, currently available high-resolution structural data, in turn, have severely hindered investigations into the nature of these structures and their interactions with small-molecule inhibitors. We have used molecular dynamics in conjunction with hydrodynamic bead modeling to study the structures of the human telomeric G-quadruplex-forming sequences at the atomic level. We demonstrated that molecular dynamics can reproduce experimental hydrodynamic measurements and thus can be a powerful tool in the structural study of existing G-quadruplex sequences or in the prediction of new G-quadruplex structures.

105

AN INVESTIGATION OF G-QUADRUPLEX STRUCTURAL POLYMORPHISM IN THE HUMAN TELOMERE USING A COMBINED APPROACH OF HYDRODYNAMIC BEAD MODELING AND MOLECULAR DYNAMICS SIMULATION

Introduction

In solutions with physiological Na^+ and K^+ concentration, single-stranded guanine-rich oligonucleotide sequences can self-assemble and fold into unimolecular G-quadruplexes, noncanonical DNA tertiary structures comprised of a four-stranded helical stem and three interconnecting loops (Burge *et al.*, 2006). Within the human genome, over 370,000 putative G-quadruplex-forming sequences have been identified and most of these are observed to localize to genomic regions with important cellular functions, such as the telomere, immunoglobulin switch regions, proto-oncogene promoters, and mRNA untranslated regions (Huppert and Balasubramanian, 2005, Huppert and Balasubramanian, 2007). Many of these sequences are found to be evolutionarily conserved between humans, mice, and rats (Verma *et al.*, 2008) suggesting that G-quadruplex structures play important regulatory roles within the cell. The formation of G-quadruplex at the distal 3' end of the human telomere region (Wright *et al.*, 1997), which contains a single-stranded guanine-rich overhang of approximately 100 to 200 bases, has been investigated as a potential target for novel small molecule-based anti-cancer therapy. Small molecules that stabilize telomeric G-quadruplex structures have been shown to decrease the activity of telomerase *in vitro* (Riou *et al.*, 2002, Cuesta *et al.*,

2003, De Cian *et al.*, 2008, Lopes *et al.*, 2011, Rodriguez *et al.*, 2012). Since telomerase is activated in more than 90% of all cancers (Shay and Bacchetti, 1997), G-quadruplex-based anti-telomerase therapy could be an attractive strategy for the development of anti-cancer therapeutics.

Despite considerable research being devoted to targeting telomeric G-quadruplexes (Balasubramanian and Neidle, 2009, De Cian *et al.*, 2008, Han and Hurley, 2000, Neidle and Read, 2000, Ou *et al.*, 2008, Saretzki, 2003, White *et al.*, 2001), the development of small-molecule G-quadruplex-based inhibitors has progressed slowly with only one candidate drug making it to clinical trials (Drygin *et al.*, 2009). A challenge to the rational design of small molecules that bind specifically to G-quadruplexes is the lack of, or the ambiguous nature of, high-resolution structural data for many putative G-quadruplex-forming sequences. In fact, the hTel22 sequence, AGGGTTAGGGTTAGGGTTAGGG, which is often used as an *in vitro* model for G-quadruplex formation in the human telomere (Wang and Patel, 1993, Parkinson *et al.*, 2002), has been found to exist in many forms depending on experimental conditions and sequence composition (Table 3). In the presence of sodium, it is widely accepted that this sequence folds into an antiparallel "basket" topology (Wang and Patel, 1993) which consists of three stacked G-tetrads with one diagonal and two lateral loops. In the presence of potassium, it exists as an ensemble of structures, which includes two mixed "hybrid" topologies, i.e. hybrid-1 (Luu *et al.*, 2006, Dai *et al.*, 2007b, Phan *et al.*, 2007b) and hybrid-2 (Dai *et al.*, 2007a, Phan *et al.*, 2007b), a parallel "propeller" topology (Parkinson *et al.*, 2002), and a new antiparallel "basket" topology (Lim *et al.*, 2009, Zhang *et al.*, 2010b). Hybrid-1 consists of three stacked G-tetrads with a double chain-

107

reversal loop followed by two lateral loops. Hybrid-2 also consists of three stacked G-tetrads but with reversed loop order, two lateral loops followed by a double chain-reversal loop. The parallel "propeller" topology consists of three stacked G-tetrads and three double chain-reversal loops. Lastly, the K^+ antiparallel "basket" topology consists of two stacked G-tetrads with a diagonal and two lateral loops.

Table 3. G-Quadruplex-Forming Sequences for HYDROPRO Calculations

Sequence[1]	ID[2]	Fold[3]	Type	Cations
*A*GGGTTA**GGG**TTA**GGG**TTA**GGG**	143D	Basket	NMR	Na$^+$
*A*GGGTTA**GGG**TTA**GGG**TTA**GGG**	1KF1	Prop	X-Ray	K$^+$
*TT***GGG**TTA**GGG**TTA**GGG**TTA**GGG***A*	2GKU	Hyb-1	NMR	K$^+$
*AAA***GGG**TTA**GGG**TTA**GGG**TTA**GGG***AA*	2HY9	Hyb-1	NMR	K$^+$
*TA***GGG**TTA**GGG**TTA**GGG**TTA**GGG**	2JSM	Hyb-1	NMR	K$^+$
*TTA***GGG**TTA**GGG**TTA**GGG**TTA**GGG***TT*	2JPZ	Hyb-2	NMR	K$^+$
*TA***GGG**TTA**GGG**TTA**GGG**TTA**GGG***TT*	2JSL	Hyb-2	NMR	K$^+$
GGGTTA**GG**GTTA**GG**TTA**GG***GT*	2KF8	Basket	NMR	K$^+$
*A***GG**GTTA**GGG**TTA**GGG**TTA**GG***GT*	2KKA-G	Basket	NMR	K$^+$
*A*GGGTTA**GGG**TTA<u>I</u>**GG**TTA**GG***GT*	2KKA-I	Basket	NMR	K$^+$

[1]G-quartet stem residues (**bold**) and flanking residues (*italic*). Noncanonical residues (<u>underline</u>).[2]Protein Data Bank code (www.pdb.org). [3]Hyb-1 is hybrid-1, Hyb-2 is hybrid-2, and Prop is propeller

The inherent structural polymorphism associated with G-quadruplex formation has severely hindered investigations of G-quadruplex structures and their formation. As a consequence, steps are usually taken to artificially reduce the structural polymorphism with the goal of enrichment of one species for NMR structure elucidation (Dai *et al.*, 2008, Yang and Okamoto, 2010). Sequence modification, as seen with human telomere sequence, is one common approach. The reported sequence variants for the human telomere sequence (Table 3) demonstrate how small changes with respect to the flanking bases can result in dramatically different dominant topologies. While each sequence contains the identical G-quadruplex-forming core GGGTTAGGGTTAGGGTTAGGG, the flanking bases differ from the hTel22 sequence (5'-A-core-3'), which contains a mixture of G-quadruplex structures, compared to the hybrid-1 dominant sequences (2GKU: 5'-TT-core-A-3', 2HY9: 5'-AAA-core-AA-3', and 2JSM: 5'-TA-core-3') (Luu *et al.*, 2006, Dai *et al.*, 2007b, Phan *et al.*, 2007b), the hybrid-2 dominant sequences (2JPZ: 5'-TTA–core-TT-3', 2JSL: 5'-TA-core-TT-3') (Dai *et al.*, 2007a, Phan *et al.*, 2007b), and the antiparallel dominant sequences (2KF8: 5'-core-T-3', and 2KKA: 5'-A-core-T-3') (Lim *et al.*, 2009, Zhang *et al.*, 2010b). In addition to changes in the flanking bases, G-quadruplex-forming sequences can also be truncated or elongated. Often, sequence modifications also involve the incorporation of non-canonical bases, as is the case with the 2KKA sequence which contains an inosine substitution for guanine. A list of G-quadruplex modifying constituents and their effects on G-quadruplex formation can be found in a recent review (Sagi, 2013).

Aside from sequence modification, another common approach to reduce the structural polymorphism is by changing the solution conditions. The presence of

111

biological molecules (Sannohe and Sugiyama, 2001) (e.g. sugar, proteins), co-solvents (e.g. acetonitrile, PEG) (Xue *et al.*, 2007, Miller *et al.*, 2010, Buscaglia *et al.*, 2013), the use of divalent versus monovalent cations (Blume *et al.*, 1997, Miyoshi *et al.*, 2001), and cation concentration (Gray *et al.*, 2009a, Gray *et al.*, 2009b, Le *et al.*, 2012), all play a major role in determining G-quadruplex stability. The parallel topology of hTel22 clearly illustrates the effect of experimental conditions on G-quadruplex formation. This topology was first reported in potassium conditions as a crystal structure (Parkinson *et al.*, 2002). It was later determined that this is not the predominant topology in solution (Li *et al.*, 2005, Buscaglia *et al.*, 2013, Hänsel *et al.*, 2011) and accounts for only about 14% of the total G-quadruplex structures (Buscaglia *et al.*, 2013). However, under the effect of dehydration (Miller *et al.*, 2010) or in the presence of PEG (Buscaglia *et al.*, 2013) (both factors present in the crystallization conditions), the parallel topology is enriched to become the predominant form. In fact, as proof of this principle, a recently reported NMR solution structure (Heddi and Phan, 2011) for hTel22 in 40% polyethylene glycol (PEG 200) was similar to the previously reported crystal structure.

The unintended consequence of sequence modification or alteration of experimental conditions is that such an approach can result in an unpredictable perturbation of the system and the selection of a topology which may or may not be representative of the original ensemble of topologies (Lane *et al.*, 2008). However, these approaches for the artificial reduction of the structural polymorphism are used because of the limitations of traditional biophysical methods with regard to elucidating G-quadruplex structure. Low-resolution spectroscopy methods, such as circular dichroism or UV-Vis spectroscopy, usually cannot distinguish between the different topologies

112

within the ensemble (Dailey *et al.*, 2010, Le *et al.*, 2012), while high-resolution structural methods, such as NMR spectroscopy and X-ray crystallography, are often of limited utility when it comes to resolving the structural polymorphism of G-quadruplexes. As observed with hTel22, a definitive structure cannot be obtained by NMR spectroscopy because this sequence exists as a mixture of multiple G-quadruplex species in solution (Burge *et al.*, 2006). The alternative method, X-ray crystallography, may not be appropriate either, as under dehydrating crystallization conditions, this sequence adopts a topology that may not be representative of the ensemble in solution (Buscaglia *et al.*, 2013) or *in vivo* (Hänsel *et al.*, 2011). This has a significant effect on what structure or structures can be claimed as "biologically relevant". Thus, there is a need for new experimental approaches that can explore the conformational space surrounding the G-quadruplex topologies without significantly disrupting or perturbing the system.

We propose an alternative approach for the unperturbed investigation of G-quadruplex structures, hydrodynamic bead modeling (HBM). HBM has emerged as a useful tool for studying biological macromolecules and their complexes for which high-resolution structural data are either unavailable or ambiguous (Byron, 2008). HBM has been used, in a limited scope, to study G-quadruplex structures. Niermann *et al.* (1999) used HBM to calculate the rotational and translational diffusion coefficients for the Watson-Crick double helical B-DNA, the single-stranded duplex "hairpin," and the tetramolecular G-quadruplex structures. Li *et al.* (2005) employed HBM to demonstrate that the crystal structure of the hTel22 telomere sequence in potassium is not the predominant topology in solution. More recently Petraccone *et al.* used HBM to study higher-order G-quadruplex formation by the human telomeric sequence $(T_2AG_3)_nT_2$ (n =

113

4, 8, 12) (Petraccone *et al.*, 2010, Petraccone *et al.*, 2011, Petraccone *et al.*, 2008). The purpose of the current work is to use hydrodynamic bead modeling in tandem with molecular dynamics (MD) simulations to explore the structural polymorphism of the human telomere G-quadruplex sequence. In particular, we exploited recent advances in computing hardware, which makes it feasible for routine microsecond-timescale simulations through traditional MD methods. Compared to shorter nanosecond simulations, longer simulations are better at sampling conformations while avoiding the bias of the starting structures (Islam *et al.*, 2013). Using MD, we explored the conformational space surrounding the five different folding topologies associated with the human telomere sequence. HBM was used to calculate sedimentation coefficients ($s_{20,w}$) and other hydrodynamic properties for "snapshot" structures obtained by MD simulations. Overall, the calculated hydrodynamic values agreed with experimental values obtained via analytical ultracentrifugation (AUC) studies. Clustering of the MD trajectories using $s_{20,w}$ values revealed the existence of different hydrodynamic substates within the ensemble of structures. Using principal component analysis (PCA) of the MD trajectory, the key motions of the G-quadruplex structures that are responsible for variations observed by hydrodynamic measurements were identified. Grid mapping of water and cations around the DNA also provides valuable insights into the role of hydration and ion binding in hydrodynamic measurements. Lastly, a novel use for HBM is proposed to estimate the number of counterions bound to a particular G-quadruplex structure when accurate information about the size and shape of the DNA is known. This work demonstrates that hydrodynamic bead modeling in conjunction with MD simulations is a powerful technique to study G-quadruplex structures.

114

Molecular Dynamics (MD) Simulation

Molecular models of G-quadruplex structures were obtained from the Protein Data Bank using the PDB IDs in Table 3. For structures containing multiple models, the first model in the file was selected for AMBER MD simulations. Appropriate coordinating ions were added to the stacked G-tetrads of each model and additional ions were added to neutralize the G-quadruplex structures. The system was solvated in a truncated octahedral box of TIP3P water molecules with 10Å buffer. The system was heated and equilibrated using the following protocol: (i) minimize water and ions (1000 steps - 500 steepest descents) holding the DNA fixed (50 kcal/mol/Å), (ii) 50ps MD (heating to 300 K) holding the DNA fixed, (iii) repeat step (i), (iv) minimize all atoms (2500 steps - 1000 steepest descents), (v) repeat step (ii), (vi) 50ns MD (T = 300K) equilibration holding the DNA fixed (50 kcal/mol/Å), and (vii) 50 ns MD to finish the equilibrium period. Production runs of 1μs after the final equilibration step were carried out to obtain snapshots at 100-ps interval for a total of 10,000 snapshots. Simulations were performed in the isothermal isobaric ensemble (P = 1atm, T = 300K) using sander and GPU version of pmemd. Periodic boundary conditions and Particle-Mesh-Ewald algorithms were used. A 2.0 fs time step was used with bonds involving hydrogen atoms frozen using SHAKE. Analysis of the trajectory was performed using the *cpptraj* module of the AmberTools 13 Package. Calculations of hydrodynamic properties were done using the program HYDROPRO. All AMBER and HYDROPRO calculations were

conducted in part using the resources of the University of Louisville's research computing group and the Cardinal Research Cluster.

Oligonucleotide Preparation and Annealing

The human telomere G-quadruplex-forming oligonucleotide, $dAG_3(T_2AG_3)_3$, and its derivatives (Table 3) were purchased from Integrated DNA Technologies (Coralsville, IA). A stock solution (1mM) of each oligonucleotide was prepared by dissolving the lyophilized DNA in TBAP buffer (10mM tetrabutylammoniumphosphate monobasic, 1mM EDTA, pH 7.0). The DNA was quantified using a Nanodrop 2000 instrument (Thermo Scientific, Wilmington, DE). Molar extinction coefficient (ε) for each oligonucleotide was calculated via the nearest-neighbor method. Prior to sedimentation velocity experiments, the DNA samples were diluted in TBAP buffer to an A_{260} of 0.5 and salt was added to the solution to bring the final concentration of NaCl or KCl to 400mM. The oligonucleotide samples were annealed in a water bath by heating to 100°C, holding the sample at temperature for 10 minutes, and gradually cooling to room temperature overnight.

Analytical Ultracentrifugation (AUC)

AUC was carried out in a Beckman Coulter ProteomeLab XL-A analytical ultracentrifuge (Beckman Coulter Inc., Brea, CA) at 20.0 °C overnight at 50,000 rpm in standard 2 sector cells. Data were analyzed using the program Sedfit (www.sefit.com). The discrete noninteracting species model in Sedfit were used to determine experimental values for $s_{20,w}$ and $D_{t20,w}$ for comparison with HYDROPRO. The concentration-

116

dependent distributions of sedimenting species were calculated using the c(s) continuous distribution model using measured values for buffer density and viscosity. Buffer density was determined on a Mettler/Paar Calculating Density Meter DMA 55A at 20.0 °C and viscosity was measured using an Anton Parr AMVn Automated Microviscometer. For the calculation of frictional ratio, 0.55 mL/g was used for partial specific volume and 0.3 g/g was assumed for the amount of water bound.

Calibration of HYDROPRO Parameters

The atomic element radius (AER), or bead size, of the primary HYDROPRO hydrodynamic model was calibrated to the Stokes radius of the DNA obtained by AUC. In HYDROPRO, the Stokes radius is reported as the equivalent translational radius, defined as the radius of the sphere with equivalent translational diffusion coefficient value. The calculated values were taken as an average over the number of poses available in the Protein Data Bank. For 143D, 6 poses were deposited in the PDB record. For 2GKU, 12 poses were deposited. For 2HY9, 2JPZ, 2JSL, 2JSM, and 2KF8, 10 poses were deposited. The experimental values were determined as previously described. The calculated values were fitted to the experimental values using a global-fit approach (Equation 1) described in the latest HYDROPRO calibration report(Ortega *et al.*, 2011).

$$\Delta^2 = \frac{1}{N_{\text{G-Quadruplex}}} \sum^{N_{\text{G-Quadruplex}}} \left[\frac{\text{calculated} - \text{experimental}}{\text{experimental}} \right]^2 \qquad \text{Equation 1}$$

The Δ value is the root mean-square relative difference between the calculated values and the experimental values with 100Δ representing the percent difference typically used to characterize the goodness of prediction.

117

Hierarchical Agglomerative Cluster Analysis

Cluster analysis was performed using the *cpptraj* module of the AmberTools 13 Package. In hierarchical agglomerative clustering, each data point began in its own cluster and the two closest clusters were merged into a new cluster following after one run of the clustering iteration. The clustering process stopped when a certain number of clusters remained. In order to determine the optimum number of clusters, cluster analyses were performed until one to nine clusters remained. For each cluster analysis, ANOVA was employed to calculate the sums of squares. The percent of variance explained by the clustering was determined by dividing the regression sum of squares by the total sum of squares. The cluster number corresponding to the greatest increase in the percent of variance explained was taken as the optimum number of clusters. For instance, if the greatest increase in percent variance occurred when the number of clusters increased from two to three, three was designated the optimum cluster number. Following cluster analysis, the "snapshot" structures from each cluster were extracted for further investigation.

Free Energy Calculation

Free energy calculations were performed using the MMPBSA module of the AmberTools 13 Package. Solvation free energies were estimated with a nonlinear Poisson-Boltzmann electrostatic continuum method with a hydrophobic component from a surface area-dependent term. The solute dielectric constant was $\square = 1$. A cubic lattice with linear dimensions ~50% larger than the longest dimension was applied with 0.25 Å grid spacing; potentials at the boundaries of the finite-difference lattice were set to a sum

118

of Debye-Huckel potentials. Salt effects were set to be 0.400 M, which corresponded to the salt concentration of sodium and potassium used in the sedimentation velocity experiments. To estimate the nonpolar contributions to solvation, $\Delta G_{nonpolar}$, the solvent accessible surface area (SASA) algorithm of Sanner was used in a parameterization where $\Delta G_{nonpolar} = \gamma(SASA) + \beta$, where $\gamma = 0.00542$ kcal/Å^2 and $\beta = 0.92$ kcal/mol.

Principal Component Analysis (PCA)

PCA was performed using the *cpptraj* module of the AmberTools 13 Package. The eigenvectors and eigenvalues were calculated from the diagonalization of the covariance matrix which contained the atomic positional fluctuation, about the average structures, in Cartesian coordinate spaces for all three coordinate axes. The covariance matrix is a $3N \times 3N$ matrix with $3N - 6$ eigenvectors possible, where N is the number of atoms in the system(Hayward and Groot, 2008). The eigenvectors describe the nature of the fluctuation while the eigenvalues describe the contribution of each eigenvector to the overall atomic fluctuation. The results were the partition of the atomic positional fluctuations reported in the previous section into individual components in a way that all components are orthogonal to each other and that the first component accounts for the most variance possible and that each subsequent component, in turn, accounts for the highest variance possible while remaining orthogonal to the preceding components.

Radial Distribution Functions (RDF) and Grid Mapping

RDF and grid mapping were performed using the *cpptraj* module of the AmberTools 13 Package. The number of water molecule bound to the G-quadruplex

119

structure was calculated by integrating the RDF function up to the first minimum, which represents the boundary of the primary hydration shell. Prior to grid mapping, the trajectory was prepared using the autoimage command. The structures were RMS fit to the bases of the G-tetrads. Grid mappings of water and cation distributions were calculated by binning atom positions at 100-ps intervals into $0.5 \times 0.5 \times 0.5$ Å^3 grids over 1-μs duration of the trajectories. In other words, the value of each grid element represents the number of times the coordinates of the center of a particular atom of interest (i.e. water oxygen) were within the $0.5 \times 0.5 \times 0.5$ Å^3 represented by that particular grid element. For 10000 frames, the expected number of waters per grid element, assuming bulk water density (55.5 M), was 42. The reference density was 6 M for sodium and 4 M for potassium, which represents the solubility of NaCl and KCl in water, respectively. This corresponded to an expected 4.5 sodium atoms and 3.0 potassium atoms per grid element.

Results and Discussion

Molecular Dynamics Simulations of Telomeric G-quadruplex Structures

The aim of this research was the detailed evaluation and parameterization of HBM in the study of G-quadruplex structural polymorphism. To that end, MD simulations were performed to sample the conformational spaces around the human telomere G-quadruplex structures and to produce "snapshot" structures for subsequent hydrodynamic calculations. Simulations were performed on ten different G-quadruplex structures representing the five folding topologies (Table 3): the antiparallel "basket" in

sodium, the parallel "propeller", the mixed "hybrid-1", the mixed "hybrid-2", and the antiparallel "basket" in potassium. For the parallel "propeller" topology, the crystal structure originally reported by Parkinson *et al.*(Parkinson *et al.*, 2002) was used. In addition, simulations were performed on two separate models of the 2KKA structure(Zhang *et al.*, 2010b). The 2KKA-I model, which contains an inosine substitution for guanine at position 14, is the model deposited in the PDB database. 2KKA-G is a model created by changing the inosine residue back to guanine in order to study the effect of inosine substitution on G-quadruplex formation.

The best-fit root mean-square deviations (RMSD) over the full 1μs of the MD trajectories for the G-quadruplexes indicated that the stem structures are rigid and the loop and flanking structures are more flexible (Figure 26). The stem of two or three stacked G-tetrads, held together by Hoogsteen hydrogen bonds and supported by the π-π stacking interaction between adjacent G-tetrads and the electrostatic interaction between the centrally coordinated cations, was the more rigid structural feature (average RMSD of 1.0 Å). The lower RMSD values that were observed for just the guanine bases alone (average RMSD of 0.5 Å) compared to the complete stem can be attributed to the more flexible phosphate backbone. Compared to just the stem alone, the RMSD for the G-quadruplex as a whole was much higher, with RMSD values as high as 4.5 Å. To highlight the mobility of different structural components in the G-quadruplex structures, the atomic positional fluctuations were calculated on a per-atom and per-residue basis (Figures 27 and 28) and mapped to representative structures from the MD trajectories (Figure 29). The fluctuation calculation is a better indicator of mobility, whereas the RMSD calculation, which measures the deviation from a reference set of coordinates, is a

better indicator of individual substates. The stacked G-tetrad stems were observed to remain remarkably rigid during the course of the simulation (Figure 29, colored red and orange) with an average positional fluctuation of about 0.5 Å (Figures 27 and 28). This observation agreed with previously reported MD results (Haider *et al.*, 2008) and helps explain the stability of these structures *in vitro* (Lane *et al.*, 2008, Chaires, 2010, Lane, 2012). In contrast to the rigid stems, the loop and flanking bases were found to be more flexible (Figure 29, colored yellow and green). The thymine loop residues were observed to be more mobile than the adenine loop residues, which have been previously shown to have possible stacking interactions with the G-tetrad stem (Haider *et al.*, 2008, Haider and Neidle, 2010). In addition, different degrees of fluctuation were observed for different loop types. The diagonal loops, which transverse the G-tetrad and are more likely to stack with the G-tetrad bases, appeared more rigid (Figure 29A, H, I) compared to the chain-reversal loops (Figure 29B-G). These findings regarding loop dynamics have significant implications for ligand recognition as it is theorized that interaction with loop and flanking bases contribute to a ligand's G-quadruplex selectivity.

Figure 26. Root mean square deviations with fitting, compared to the first frame, for 143D (A), 1KF1 (B), 2GKU (C), 2HY9 (D), 2JSM (E), 2JPZ (F), 2JSL (G), 2KF8 (H), 2KKA-G (I), and 2KKA-I (J). RMSD calculations were carried out for all non-hydrogen atoms in the G-quadruplex structures (black), in the stacked G-tetrads only (red), and only for G-tetrad guanine bases excluding sugars and phosphate groups (green).
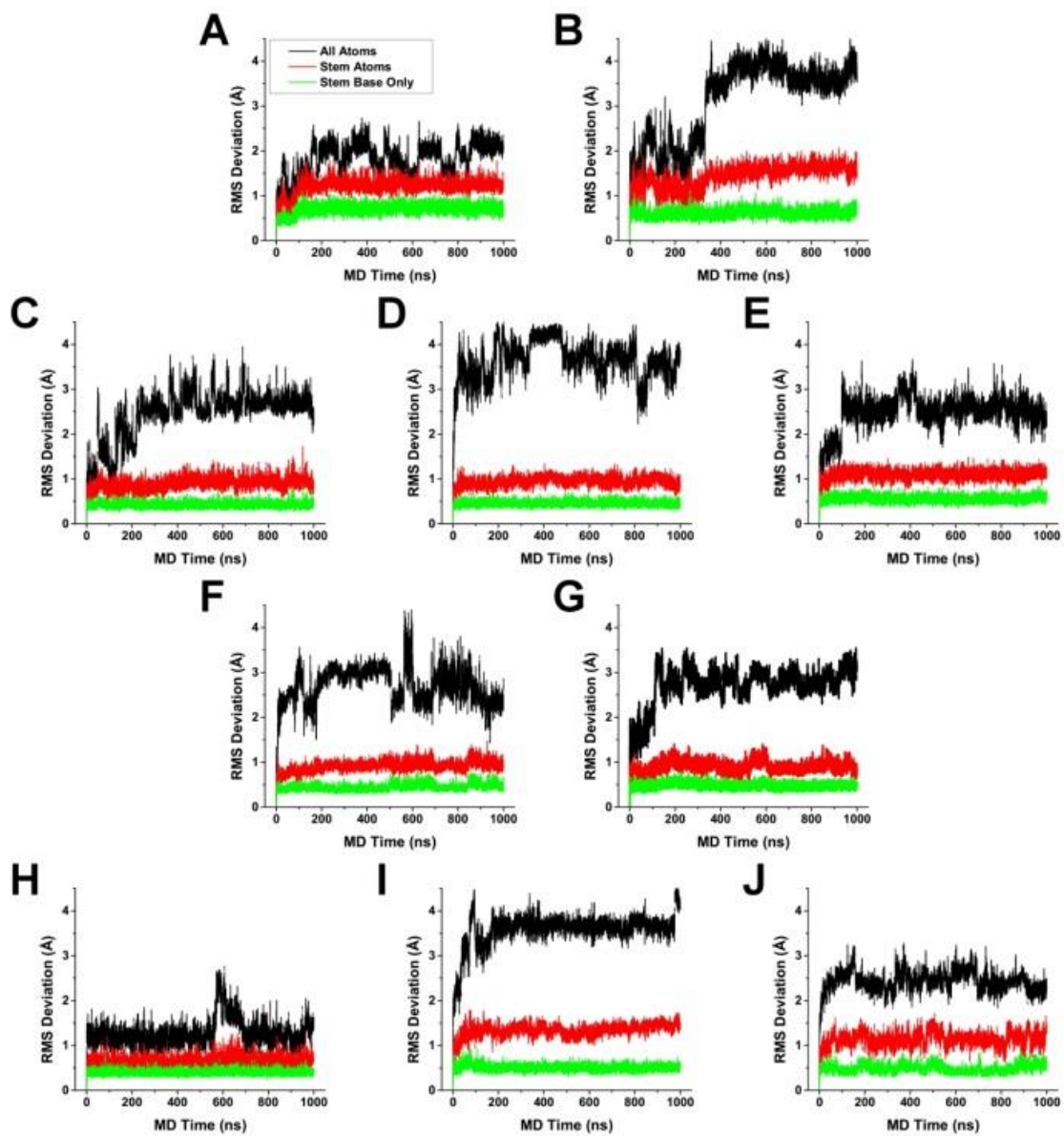
Figure 27. Per-atom atomic positional fluctuations over 10,000 frames for 143D (A), 1KF1 (B), 2GKU (C), 2HY9 (D), 2JSM (E), 2JPZ (F), 2JSL (G), 2KF8 (H), 2KKA-G (I), and 2KKA-I (J).
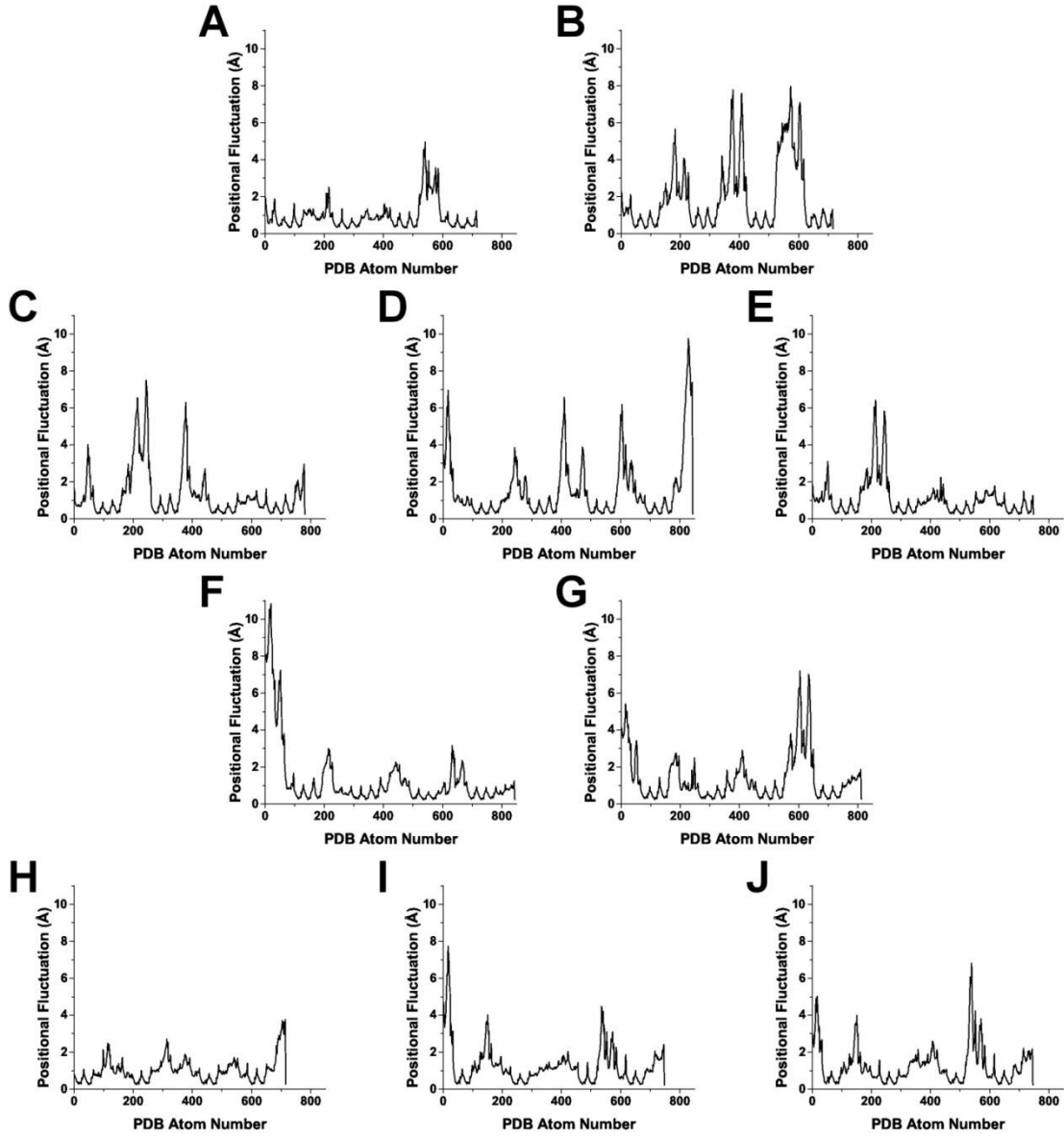
Figure 28. Per-residue atomic positional fluctuations over 10,000 frames for 143D (A), 1KF1 (B), 2GKU (C), 2HY9 (D), 2JSM (E), 2JPZ (F), 2JSL (G), 2KF8 (H), 2KKA-G (I), and 2KKA-I (J).
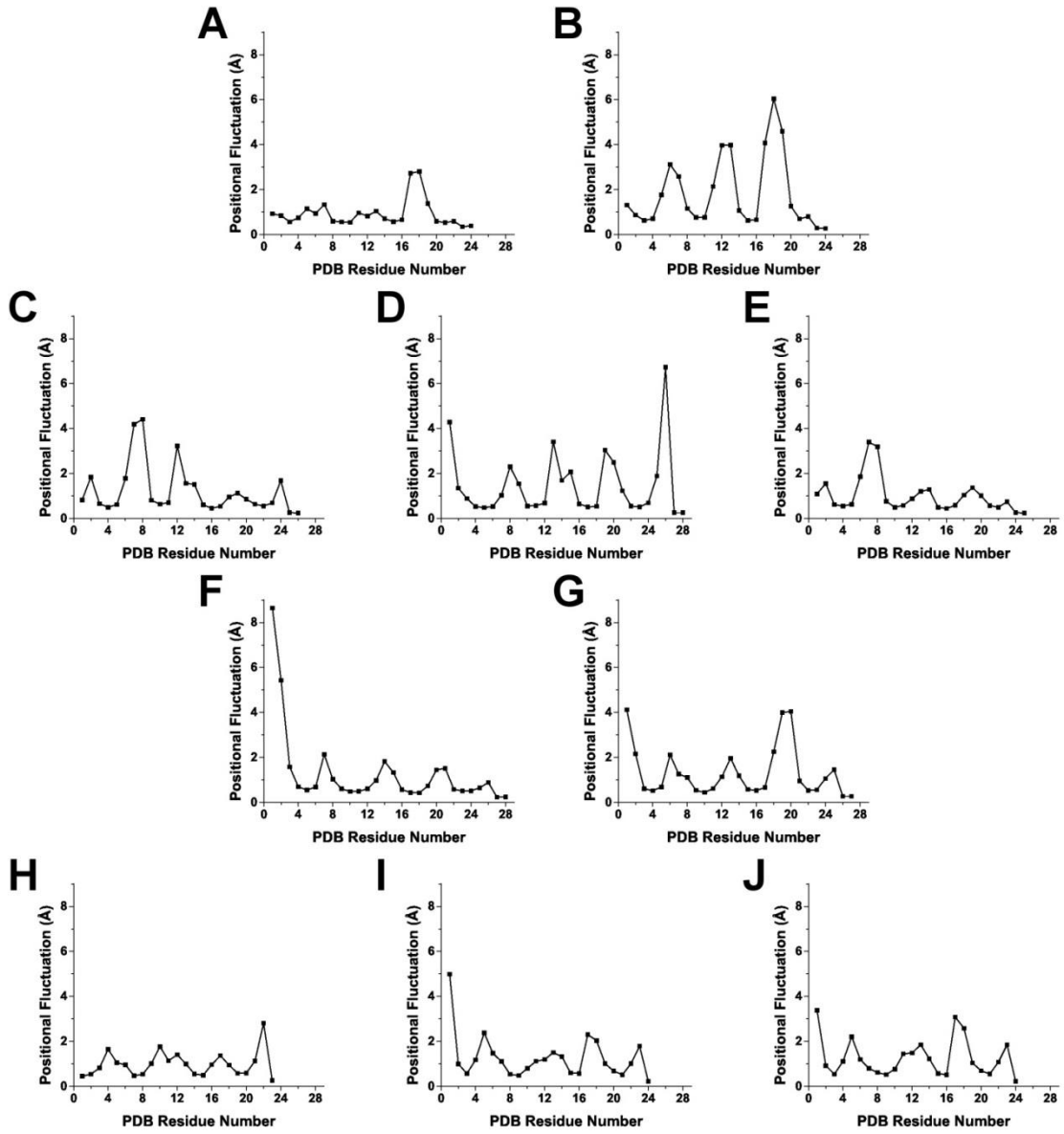
Figure 29. Representative structures from MD trajectories for 143D (A), 1KF1 (B), 2GKU (C), 2HY9 (D), 2JSM (E), 2JPZ (F), 2JSL (G), 2KF8 (H), 2KKA-G (I), and 2KKA-I (J). Residues and atoms are colored by atomic fluctuation units calculated by AMBER *cpptraj* module with the color gradient progressing from low fluctuation values to high fluctuation values (red-green, respectively).
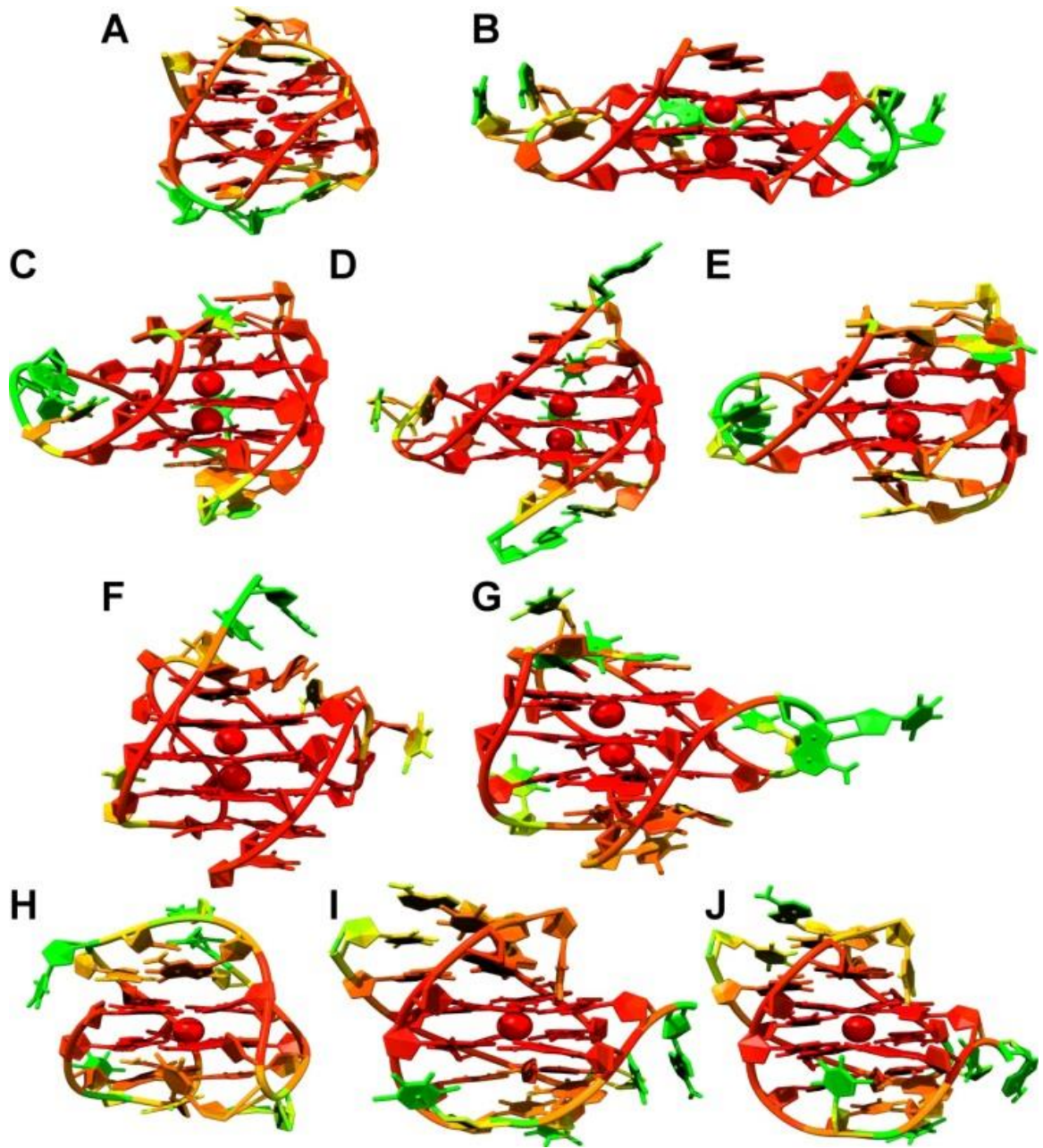
Determination of Hydrodynamic Values by Sedimentation Velocity Experiments

MD simulations provided a detailed picture of the G-quadruplex structures *in silico*. Sedimentation velocity experiments were carried out by AUC to obtain experimental information about the G-quadruplex structures. The 2KKA-I sequence was excluded because of the inosine substitution in order to limit experimental measurements to sequences containing only canonical bases. AUC experiments were performed for the unsubstituted 2KKA-G sequence, which served as an experimental reference for both the 2KKA-G and 2KKA-I MD models. As the data will show, the calculated hydrodynamic values for the 2KKA-I model agreed with the experimental values for the 2KKA-G sequence suggesting that the inosine substitution did not substantially alter the hydrodynamic behavior of the 2KKA-G sequence in solution. Sedimenting mixtures of polyelectrolytes often exhibit non-ideal behaviors due to the electrostatic interaction between different charged components resulting in smaller measured molecular weights and sedimentation coefficients (Yphantis and Roark, 1971). To reduce the effect of non-ideality, AUC experiments were carried out in high salt (400 mM NaCl/KCl) buffers to "swamp out" the electrostatic interactions. The distributions of sedimenting species are shown in Figure 30 (black lines). The Stokes radius (a_T), molecular weight (MW), and frictional ratios ($^f/_{f0}$) were calculated from the diffusion ($D_{t20,w}$) and sedimentation coefficients ($s_{20,w}$) (Table 4). The frictional ratio is a dimensionless value comparing the observed translational diffusion coefficient of a macromolecule with the translational diffusion coefficient of an equivalent sphere of the same molecular weight with the higher frictional ratios being indicative of a more asymmetric and less spherical shape

(Smith, 1988, Harding, 2002). The value for f is determined from the experimental measurement of $D_{t20,w}$ using Equation 2:

$$f = \frac{RT}{N_A D_{t20,W}}$$

Equation 2

where R is the gas constant and N_A is Avogrado's number. The value for f_0 is determined from the molecular weight using Equation 3:

$$f_0 = 6\pi\eta_{20,W}\left(\frac{3M(\overline{v}+\delta/\rho_{20,W})}{4\pi N_A}\right)^{1/3}$$

Equation 3

where $\eta_{20,w}$ is the viscosity of water at 20°C, $\overline{v}$ is the partial specific volume, δ is a uniform expansion factor to account for hydration, $\rho_{20,w}$ is the density of water at 20°C, and M is the molecular weight determined by the Svedberg equation (Equation 4):

$$M = \frac{s_{20,W}RT}{D_{20,W}(1-\overline{v}\rho_{20,W})}$$

Equation 4

From these equations, it becomes apparent that the frictional ratio of a macromolecule is dependent on its shape, flexibility, and the amount of hydration associated. In the current work, the amount of associated hydration or δ was assumed to be 0.3 $^{g\ of\ water}/_{g\ of\ G\text{-quadruplex}}$. For nucleic acids, a value of 0.3-0.35 $^{g}/_{g}$ is typically used although this parameter can be quite difficult to determine with accuracy (Fernandes *et al.*, 2002).

Figure 30. Comparison of experimentally determined and HYDROPRO calculated sedimentation coefficient distributions for 143D (A), 1KF1 (B), 2GKU (C), 2HY9 (D), 2JSM (E), 2JPZ (F), 2JSL (G), 2KF8 (H), and 2KKA [2KKA-G red, 2KKA-I blue] (I). For each G-quadruplex structure, sedimentation coefficients ($s_{20,w}$) were determined experimentally by AUC (black) and calculated from MD "snapshots" using HYDROPRO (red and blue).
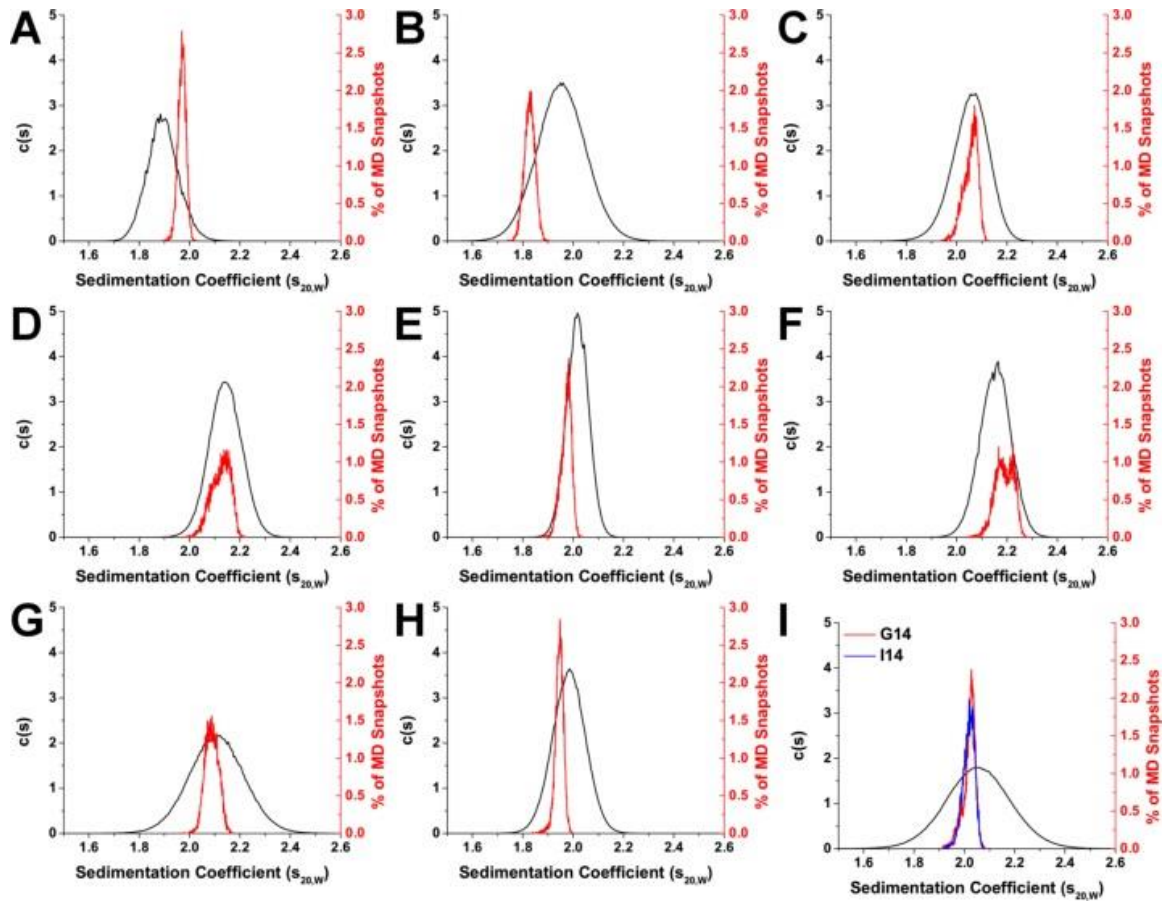
Table 4. Hydrodynamic Values Determined by Sedimentation Velocity Experiments

G-quadruplex	$s_{20,w}$[1,2]	$D_{t20,w}$[1,2]	a_T[2,3]	MW[4]	f/f_0[5]
143D	1.90	1.42	1.50	7199.97	1.10
1KF1	1.97	1.45	1.48	7330.24	1.07
2GKU	2.05	1.45	1.48	7644.45	1.06
2HY9	2.14	1.44	1.49	8039.62	1.05
2JSM	2.02	1.61	1.33	6779.45	1.00
2JPZ	2.15	1.45	1.48	8049.02	1.05
2JSL	2.11	1.41	1.52	8113.66	1.07
2KF8	1.98	1.54	1.39	6975.38	1.03
2KKA-G	2.06	1.58	1.35	7021.38	1.02

[1]$s_{20,w}$ and $D_{t20,w}$ were calculated from sedimentation velocity data using discrete non-interacting species model of Sedfit; [2]$s_{20,w}$ expressed in units of 10^{-13} s, $D_{t20,w}$ expressed in units of 10^{-6} cm^2/s, a_T (Stokes radius) expressed in units of 10^{-7} cm; [3]a_T calculated from $D_{t20,w}$, solvent viscosity, and temperature; [4]MW is calculated from $D_{t20,w}$, solvent density, temperature, and 0.55 $^{mL}/_g$ value for partial specific gravity; [5]f/f_0 calculated using 0.3 $^g/_g$ water bound and 0.55 $^{mL}/_g$ value for partial specific gravity.

For each of these sequences, the G-quadruplex structures sedimented as one species (Figure 30), including the hTel22 sequence in potassium and the 2KKA-G sequence, which are known to exist as a mixture of multiple G-quadruplex species(Dai *et al.*, 2008, Zhang *et al.*, 2010b). The inability of AUC to resolve the existence of different G-quadruplex species in solution can be attributed to the low-resolution nature of the technique. Overall, the G-quadruplex structures formed from different telomeric sequences all assumed similar spherical shapes ($^f/_{f0}$ = 1.00-1.10). The G-quadruplex structures formed by the hTel22 sequence in sodium (143D) sedimented at a lower rate ($s_{20,w}$ = 1.90) than G-quadruplex structures in potassium (1KF1) ($s_{20,w}$ = 1.97), which indicated the sodium form is less compact than the potassium form. Since the sequences are identical in both cases, the difference in sedimentation might be attributed to a change in shape. In fact, the sodium form appeared more elongated ($^f/_{f0}$ = 1.10) when compared to the potassium form ($^f/_{f0}$ = 1.07). It is important to note, however, that the difference between the two sequences might not be distinguishable as it falls within the precision limit of 5% for the experimental technique. For the 2JSM, 2KF8, and 2KKA-G sequences, the changes in sedimentation were also attributed to differences in shape. The sequence has a lower mass compared to the hTel22 sequence but are more compact in shape ($^f/_{f0}$ = 1.00-1.03) and thus sedimented at a higher rate ($s_{20,w}$ = 1.98-2.06). The differences in frictional ratios of these sequences compared to the 1KF1 sequence were distinguishable as this difference is greater than the 5% precision limit of the experimental technique. In contrast, the changes in sedimentation for the 2GKU, 2HY9, 2JPZ, and 2JSL sequences compared to the hTel22 sequence can be attributed mainly to difference in sequence sizes. These sequences appeared to be of similar shape or more

137

compact (f/f_0 = 1.05-1.07) yet sedimented at a higher rate ($s_{20,w}$ = 2.02-2.15) due to the increased sequence molecular weight. Taken together, these findings indicated that AUC can be informative when used to study G-quadruplex structures.

Calculation of Hydrodynamic Properties using the Program HYDROPRO

To determine if the *in silico* models can reproduce the hydrodynamic values observed experimentally, HYDROPRO (Ortega *et al.*, 2011) was used to calculate the hydrodynamic values (i.e. $s_{20,w}$, D_{t20w}) for each "snapshot" structure obtained from MD simulations. Prior to calculating the hydrodynamic properties, the optimum size of the beads in the primary hydrodynamic model was determined using the procedure previously described(Ortega *et al.*, 2011). In brief, the size of the beads (termed the atomic element radius or AER in the HYDROPRO input file) was varied and the AER that yielded the smallest difference between HYDROPRO calculated Stokes radius (a_T) and the Stokes radius determined by AUC experiments was accepted as the optimum bead size (Figure 31 and Table 5). For the atomic-level model, where each non-hydrogen atom is replaced with the beads, the physical meaning of AER can be thought of as the radius of a typical non-hydrogen atom plus a uniform expansion to account for hydration. For the residue-level calculations, where each residue (or nucleotide) is replaced with a bead, the physical meaning of AER is the size of the nucleotide plus a uniform expansion to account for hydration. The calibration was done using previously reported NMR structures (PDB ID: 143D, 2GKU, 2HY9, 2JPZ, 2JSL, 2JSM and 2KF8). All structural poses deposited in the PDB database were used for HYDROPRO calibration. For the first calculation mode where the non-hydrogen atoms were replaced by a collection of

overlapping spheres and the hydrodynamic values were calculate using a shell-model methodology(Bloomfield *et al.*, 1967), the best-fit AER was determined to be 2.19 Å (Table 5) with a difference of 2.35 % between the predicted hydrodynamic values and the experimental values. For a non-hydrogen atom, the atomic radius is approximately 1.8 Å (Ortega *et al.*, 2011). The determined best fit AER suggests a hydration sphere of about 0.4 Å which is less than the 1.1 Å typically used in hydrodynamics calculations to estimate hydration (García De La Torre, 2001). When the default AER (2.84 Å) was used instead of the best fit AER, the difference increased from 2.35 to 6.32 % (Table 5). For the second calculation mode where each residue is replaced by a sphere instead of each atom and the hydrodynamic values are also calculated using a shell-model methodology, the best fit AER was 3.98 Å with an difference of 2.50% compared to the 4.84 Å for the standard AER (7.78% difference). For the third calculation mode, which calculates the hydrodynamic values directly on the primary bead models from the second calculation mode, the best fit AER was 5.04 Å with a difference of 2.60% compared to 6.11 Å for the standard AER (8.10% difference). Overall, the calibrated AERs were able to reproduce the experimental hydrodynamic values better than the standard default AERs. While the three modes to calculate hydrodynamic properties provide the same general information about the structure being modeled, the residue-level models were much faster in the amount of time required for calculation and could be useful for preliminary analysis of large MD simulations where thousands of structures are sampled.

Figure 31. HYDROPRO parameters calibrated by global fitting of translational equivalent radii with AUC data. Translational equivalent radius is also known as the Stokes radius. The values for $100\Delta X$ and 100Δ as a function of the atomic element radius (AER) for the primary hydrodynamic model as calculated using the seven G-quadruplex structures formed from the human telomere sequence. Hydrodynamic properties of G-quadruplexes were calculated using the atomic-level hydrodynamic shell-model calculation (A), residue-level shell-model calculation (B), and residue-level bead-model calculation (C).

140

A

B

C

Average	143D
2GKU	2HY9
2JPZ	2JSL
2JSM	2KF8
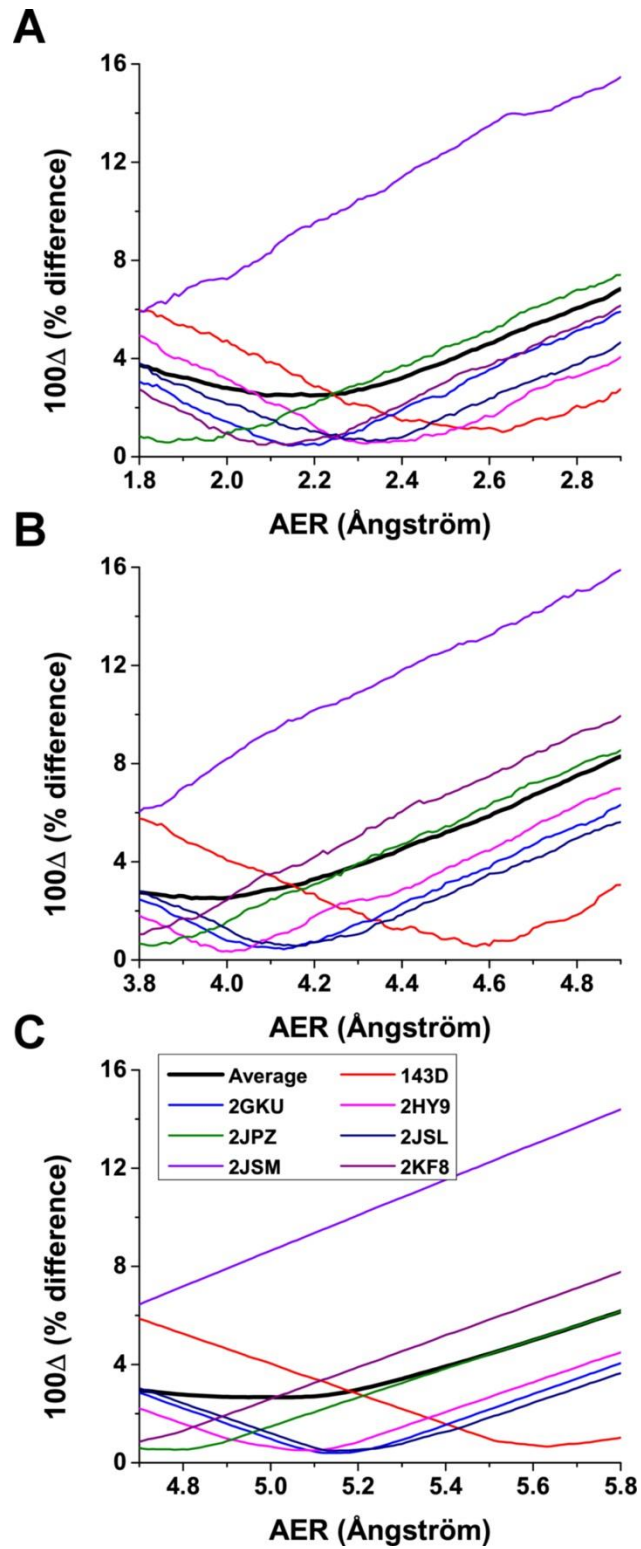
Table 5. Results of HYDROPRO global fit analysis

Calculation Mode	AER (Å)	% Difference a_T	% Difference $s_{20,w}$	% Difference $D_{t20,w}$
Atom, shell model	2.19	2.35	1.19	2.24
Atom, shell model	2.84	6.32	4.27	5.83
Residue, shell model	3.98	2.50	1.39	2.42
Residue, shell model	4.84	7.78	5.59	7.12
Residue, bead model	5.04	2.60	1.41	2.50
Residue, bead model	6.11	8.10	5.86	7.38

143

Ten thousand "snapshot" structures were sampled from each MD trajectory and hydrodynamic properties were calculated for each structure using the calibrated AER values. The average calculated hydrodynamic values for each of the MD simulations are shown in Table 6. The hydrodynamic values calculated using MD snapshots agreed with previously calculated values using reported NMR structures with a difference of 1.85 % for the atomic-level/shell model mode, 1.89 % for residue-level/shell model mode, and 1.88 % for the residue-level/bead model mode (Table 7). In general, all three modes of calculation by HYDROPRO were able to accurately predict the hydrodynamic values of the G-quadruplex structures (Table 8). The distributions of the $s_{20,w}$ values calculated using the HYDROPRO atomic-level model shell-model calculation mode (Figure 30, colored lines) are shown with the concentration-dependent $c(s)$ experimental $s_{20,w}$ distributions (Figure 30, black line). For the $c(s)$ experimental $s_{20,w}$ distributions, the width of the distribution curve is related to the homogeneity of species in solution (Schuck *et al.*, 2002, Dam and Schuck, 2004). Samples containing multiple species that sediment at similar rates will have broad $s_{20,w}$ distributions while highly homogenous samples with a single predominant species (accounting for more than 90% of total ensemble) will have narrow $s_{20,w}$ distributions. In general, it was observed that sequences (143D, 2GKU, 2HY9, 2JPZ, 2JSL, and 2KF8) for which there was enough enrichment of a major species for NMR structure elucidation have narrower distributions compared to sequences for which NMR structure elucidation was not possible (1KF1 and 2KKA). The only exception was the 2JSM sequence, which was highly enriched (>70% of the total ensemble) (Phan *et al.*, 2007b) in a major species yet has a broad $s_{20,w}$ distribution. Electrostatic interaction, which contributed to the non-ideal behaviors of the DNA at

lower salt concentrations, does not affect the width of the distributions but will shift the distribution to the left (lower apparent $s_{20,w}$). At higher concentrations of ions, the hydrodynamic behavior became more ideal, the distribution shifts to the right, and the apparent $s_{20,w}$ approaches the true $s_{20,w}$. Compared to the experimental distributions, the calculated $s_{20,w}$ distributions were narrower and in certain cases were asymmetric. The shape of the calculated distributions is determined by the sampling capability of MD simulations which is on a much shorter timescale (1 µs) compared to the sampling capability of AUC experiments (~6 hours). The time-course graph of $s_{20,w}$ values for the 2JPZ model highlight the effect of sampling time on the distributions peak shape (Figure 32F). The distributions of calculated $s_{20,w}$ values across the full 1 µs of MD simulation was bimodal (Figure 32F). However, had the simulation completed within the first 500 ns, the distribution would have been narrower and unimodal.

145

Table 6. Summary of HYDROPRO Calculated Hydrodynamic Values

G-quadruplex	Atom/shell	Residue/shell	Residue/bead
143D	$s_{20,w}[1] = 1.97 \pm 0.02$ $D_{t20,w}[1] = 1.51 \pm 0.01$	1.94 ± 0.01 1.49 ± 0.01	1.92 ± 0.01 1.48 ± 0.01
1KF1	$s_{20,w} = 1.83 \pm 0.02$ $D_{t20,w} = 1.40 \pm 0.02$	1.89 ± 0.01 1.45 ± 0.03	1.87 ± 0.03 1.44 ± 0.03
2GKU	$s_{20,w} = 2.06 \pm 0.03$ $D_{t20,w} = 1.45 \pm 0.02$	2.03 ± 0.01 1.44 ± 0.01	2.02 ± 0.01 1.43 ± 0.01
2HY9	$s_{20,w} = 2.12 \pm 0.04$ $D_{t20,w} = 1.38 \pm 0.02$	2.15 ± 0.02 1.40 ± 0.01	2.14 ± 0.02 1.39 ± 0.01
2JSM	$s_{20,w} = 1.97 \pm 0.02$ $D_{t20,w} = 1.45 \pm 0.02$	1.95 ± 0.01 1.44 ± 0.01	1.94 ± 0.01 1.43 ± 0.01
2JPZ	$s_{20,w} = 2.19 \pm 0.04$ $D_{t20,w} = 1.43 \pm 0.02$	2.17 ± 0.02 1.42 ± 0.01	2.17 ± 0.02 1.42 ± 0.01
2JSL	$s_{20,w} = 2.09 \pm 0.03$ $D_{t20,w} = 1.42 \pm 0.02$	2.08 ± 0.01 1.41 ± 0.01	2.08 ± 0.01 1.41 ± 0.01
2KF8	$s_{20,w} = 1.95 \pm 0.02$ $D_{t20,w} = 1.49 \pm 0.01$	1.91 ± 0.01 1.46 ± 0.01	1.89 ± 0.01 1.45 ± 0.01
2KKA-G[2]	$s_{20,w} = 2.02 \pm 0.03$ $D_{t20,w} = 1.49 \pm 0.02$	2.00 ± 0.01 1.47 ± 0.01	1.99 ± 0.01 1.47 ± 0.01
2KKA-I[2]	$s_{20,w} = 2.02 \pm 0.03$ $D_{t20,w} = 1.49 \pm 0.02$	2.00 ± 0.01 1.47 ± 0.01	1.99 ± 0.01 1.46 ± 0.01

[1]$D_{t20,w}$ expressed in units of $10^{-6} cm^2/s$ and $s_{20,w}$ expressed in units of $10^{-13}s$;

[2]Experimental values for 2KKA-G and 2KKA-I were obtained with the oligonucleotide sequence $dAG_3(T_2AG_3)_3T$.

Table 7. Comparisons of $s_{20,w}$ Calculated from MD Simulations and NMR Structures

Structure	% Diff. Atom/shell	% Diff. Residue/shell	% Diff. Residue/bead
143D	3.05 %	0.06 %	0.34 %
2GKU	0.32 %	1.87 %	2.03 %
2HY9	5.30 %	3.04 %	3.38 %
2JPZ	1.17 %	0.31 %	0.28 %
2JSL	0.09 %	0.98 %	0.71 %
2JSM	1.35 %	3.61 %	3.11 %
2KF8	2.44 %	2.64 %	2.69 %
Weighted average[1]	1.85 %	1.89 %	1.88 %

[1]Average weighted by number of poses deposited in PDB Database (143D – 6 poses, 2GKU – 12 poses, 2HY9 – 10 poses, 2JPZ – 10 poses, 2JSL – 10 poses, 2JSM – 10 poses, 2KF8 – 10 poses)

Table 8. Comparison of HYDROPRO-Calculated and Experimental $s_{20,w}$

Structure	% Diff. Atom/shell	% Diff. Residue/shell	% Diff. Residue/bead
143D	3.76	1.96	1.19
2GKU	0.30	0.83	1.23
2HY9	0.90	0.35	0.16
2JPZ	1.70	0.95	0.61
2JSL	1.15	1.46	1.68
2JSM	2.23	3.23	3.60
2KF8	1.95	3.95	4.59
1KF1	7.23	3.80	4.91
2KKA-G[1]	1.88	2.67	3.11
2KKA-I[1]	1.98	2.98	3.45

[1]Experimental values for 2KKA-G and 2KKA-I obtained with the oligonucleotide sequence $dAG_3(T_2AG_3)_3T$.

With the exception of the hTel22 sequence in potassium (1KF1, the HYDROPRO calculated hydrodynamic values agreed with the experimental values (Table 8). As hydrodynamic experiments are essentially low-resolution measurements, a difference between calculated and experimental values by 5% (the precision limit for hydrodynamic measurements (Fernandes *et al.*, 2002)) can be considered acceptable. The % difference in Table 8 was determined by comparing the HYDROPRO calculated values from Table 6 with the experimental values in Table 4 using Equation 1. For the seven models in used in the HYDROPRO calibration procedure, the calculated distributions agreed with the experimental distributions (0.30-3.76% difference) as shown in Table 8. With the 143D model, it appeared initially that the model does not agree with the experimental data (Figure 28A), however, the difference of 3.76% (Table 8) was still within the experimental precision limit. Of the three models which were excluded from the calibration procedure, the calculated distribution of 1KF1 did not agree with the experimental distribution with a difference of 7.23% (Table 8). The current findings agreed with previously report data, which demonstrated that the parallel form of 1KF1 have different hydrodynamic behaviors compared to the ensemble of structures in solution (Li *et al.*, 2005). The conclusion that the 1KF1 structure was distinct from the 1KF1 sequence in solution was made on the basis of the calculated $s_{20,w}$ value alone and not $D_{t20,w}$ value, which appeared to agree better the experimental value than the $s_{20,w}$. The reason that $D_{t20,w}$ was not used for that purpose was that in solution contain multiple species, the differential migrations of different species result in a broadening of the sedimentation boundary leading and an incorrect apparent $D_{t20,w}$ (Schuck *et al.*, 2002, Dam and Schuck, 2004). For the 1KF1 structures in solution, there is a mixture at least

three to four G-quadruplex species in solution (Buscaglia *et al.*, 2013). For that reason, the apparent $D_{t20,w}$ value determined by experimental measurement might not be representative of the structures in solution and was reported only to complete the data even though it might not be appropriate to use it to rule in or rule out a certain structure in solution. For the other two models (2KKA-G and 2KKA-I), the calculated hydrodynamic values differed from the experimental values only slightly (1.88-1.98 % difference). It should be noted that the unsubstituted 2KKA-G (1.88 %) model agreed better with experimental results than the inosine-substituted 2KKA-I (1.98 %) model. However, this difference is too small to draw any definitive conclusion regarding possible difference between the two models. These findings highlighted a major limitation of HBM, which is that hydrodynamic measurements are inherently low resolution. The model that accurately predicts the hydrodynamic parameters measured in solution is not necessarily representative of the definitive structure for that molecule but rather one of the possible conformations amongst many others (Byron, 2008). Thus it can be concluded that the 2KKA-G and 2KKA-I models represent a possible set of conformations for the G-quadruplex structures in solution. However, this prediction should be confirmed by non-hydrodynamic experiments (e.g. 2-aminopurine fluorescence spectroscopy to probe the solvent accessibility of adenine bases or DMS footprinting to probe guanine base-pairing interactions), such as was done by Li *et al.* (2005) Overall, the present findings demonstrated that HYDROPRO calculations can accurately predict the correct hydrodynamic properties and can be used as a screening tool to initially rule out "incorrect" structures associated with other G-quadruplex-forming sequences.

Clustering of Molecular Dynamics Trajectories into Hydrodynamic Substates

The results of HYDROPRO calculations revealed that sedimentation rates can differ dramatically between different "snapshot" structures within the same MD trajectory (Figure 32). To identify different hydrodynamic substates within the MD trajectories, hierarchical cluster analysis were performed using HYDROPRO calculated $s_{20,w}$ values as the clustering criterion. The results of the cluster analysis indicated that heterogenous mixtures of G-quadruplex structures were present in simulations (Figures 33-42). Cluster analysis was able to identify key differences between the 2KKA-G model and the 2KKA-I model which were not apparent by $s_{20,w}$ calculation alone (Figure 30I). The 2KKA G-quadruplex structures consist of a two-stack G-tetrad stem that is capped by a triple-base cap and a double-base cap on either end. In the 2KKA-I model, the G-G-I triple-base cap was slightly offset from the G-tetrad bases (Figure 42) while in the 2KKA-G model, the G-G-G triple-base cap was directly on top of the G-tetrads (Figure 41). It was unclear whether the direct triple conformation is simply a new stable conformation or if it is an indication of the G-quadruplex structure transitioning to a new folding topology as it is beyond the current capacity of MD to model such phenomenon. The current findings clearly demonstrated that the substitution of inosine for guanine played a critical role in stabilizing the selected G-quadruplex structures. In addition, these findings raised the need for additional experimental investigation into the role of these noncanonical bases in selecting for a G-quadruplex topology from an ensemble of structures.

154

Figure 32. Hydrodynamics substates identified by clustering of HYDROPRO calculated sedimentation coefficients (S20,W) for 143D (A), 1KF1 (B), 2GKU (C), 2HY9 (D), 2JSM (E), 2JPZ (F), 2JSL (G), 2KF8 (H), and 2KKA (I). The most populated cluster for each model was colored black and the second most populated cluster was colored red. Additional clusters were colored accordingly (green, blue, cyan, and magenta) in the order of decreasing population density.
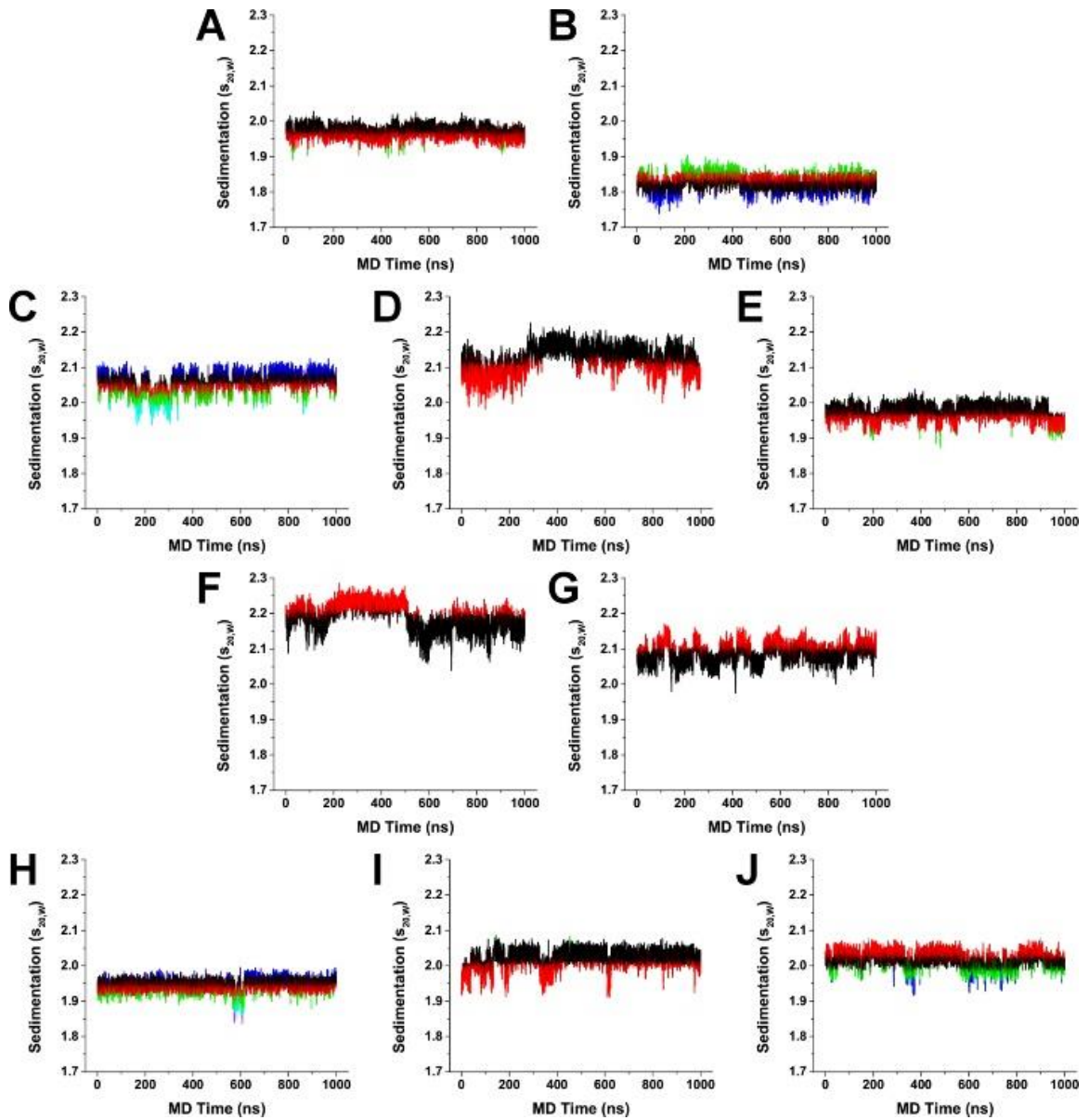
Figure 33. Cartoon depictions of representative structures from cluster analysis of the 143D MD trajectory. The clusters are shown from the most populated cluster to the least populated cluster. Nucleotides are colored according to NDB formats: guanine is green, adenine is red, and thymine is blue. Sodium ions are colored purple.

157
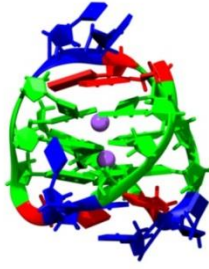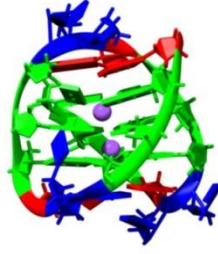
A   B   C
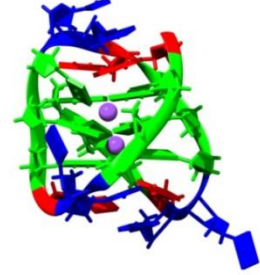
Figure 34. Cartoon depictions of representative structures from cluster analysis of the 1KF1 MD trajectory. The clusters are shown from the most populated cluster to the least populated cluster. Nucleotides are colored according to NDB formats: guanine is green, adenine is red, and thymine is blue. Potassium ions are colored purple.
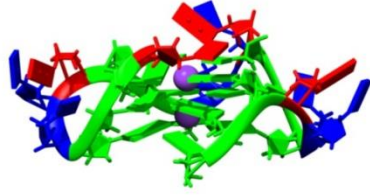
A   B
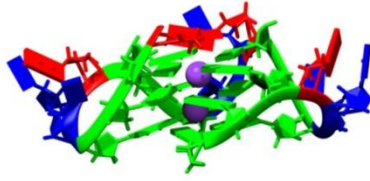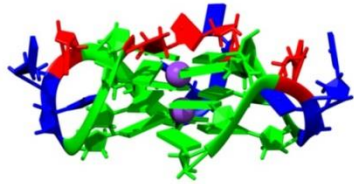
C   D
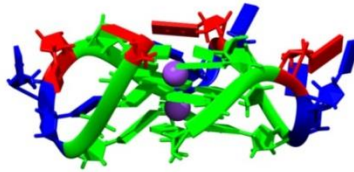
Figure 35. Cartoon depictions of representative structures from cluster analysis of the 2GKU MD trajectory. The clusters are shown from the most populated cluster to the least populated cluster. Nucleotides are colored according to NDB formats: guanine is green, adenine is red, and thymine is blue. Potassium ions are colored purple.

A   B

C   D   E
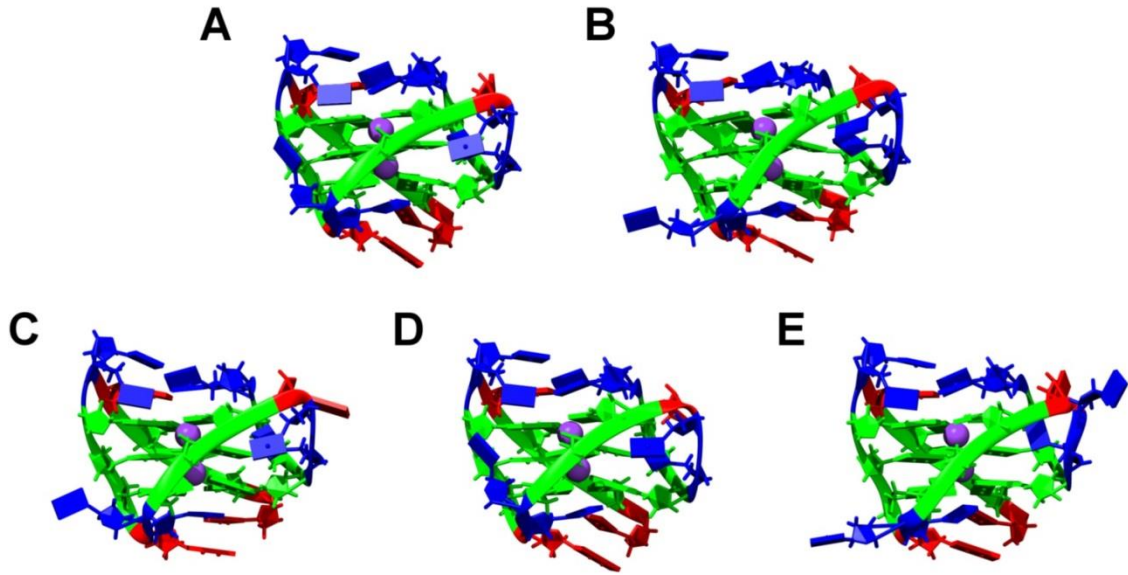
Figure 36. Cartoon depictions of representative structures from cluster analysis of the 2HY9 MD trajectory. The clusters are shown from the most populated cluster to the least populated cluster. Nucleotides are colored according to NDB formats: guanine is green, adenine is red, and thymine is blue. Potassium ions are colored purple.
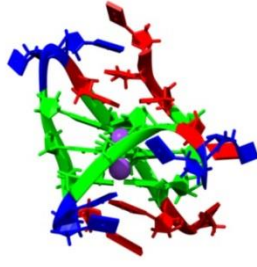
A   B
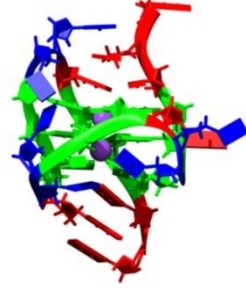
Figure 37. Cartoon depictions of representative structures from cluster analysis of the 2JSM MD trajectory. The clusters are shown from the most populated cluster to the least populated cluster. Nucleotides are colored according to NDB formats: guanine is green, adenine is red, and thymine is blue. Potassium ions are colored purple.
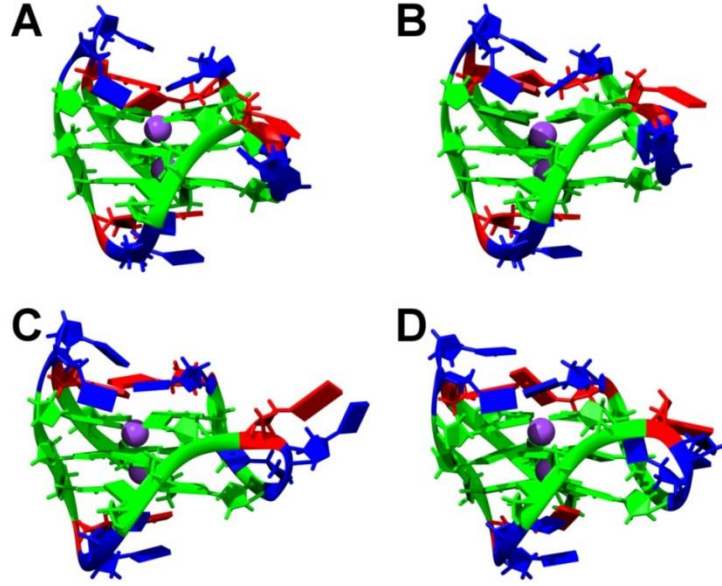
Figure 38. Cartoon depictions of representative structures from cluster analysis of the 2JPZ MD trajectory. The clusters are shown from the most populated cluster to the least populated cluster. Nucleotides are colored according to NDB formats: guanine is green, adenine is red, and thymine is blue. Potassium ions are colored purple.

168
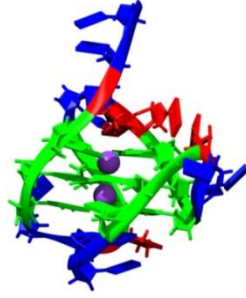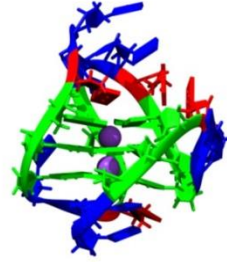
Figure 39. Cartoon depictions of representative structures from cluster analysis of the 2JSL MD trajectory. The clusters are shown from the most populated cluster to the least populated cluster. Nucleotides are colored according to NDB formats: guanine is green, adenine is red, and thymine is blue. Potassium ions are colored purple.
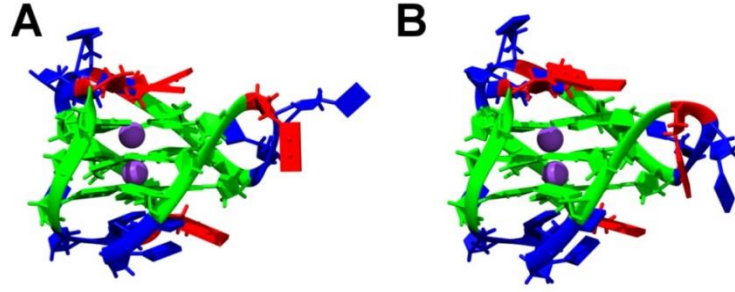
Figure 40. Cartoon depictions of representative structures from cluster analysis of the 2KF8 MD trajectory. The clusters are shown from the most populated cluster to the least populated cluster. Nucleotides are colored according to NDB formats: guanine is green, adenine is red, and thymine is blue. Potassium ions are colored purple.
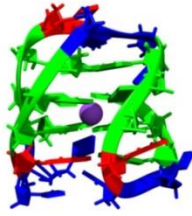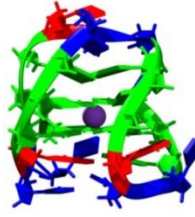
A   B   C
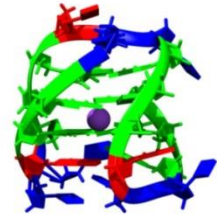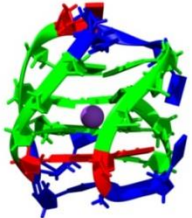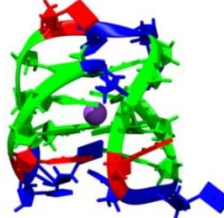
D   E   F
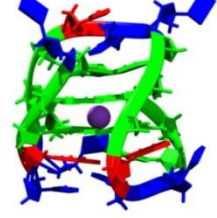
Figure 41. Cartoon depictions of representative structures from cluster analysis of the 2KKA-G MD trajectory. The clusters are shown from the most populated cluster to the least populated cluster. Nucleotides are colored according to NDB formats: guanine is green, adenine is red, and thymine is blue. Potassium ions are colored purple.
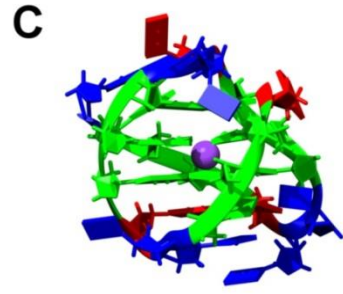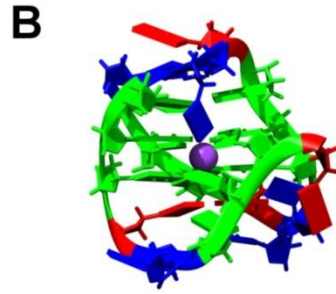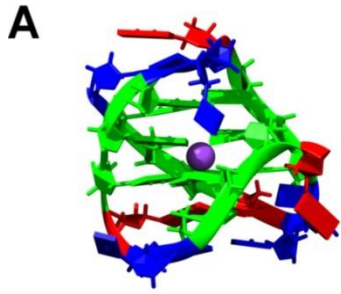
Figure 42. Cartoon depictions of representative structures from cluster analysis of the 2KKA-I MD trajectory. The clusters are shown from the most populated cluster to the least populated cluster. Nucleotides are colored according to NDB formats: guanine is green, adenine is red, and thymine is blue. Potassium ions are colored purple.

175

A   B

C   D   E
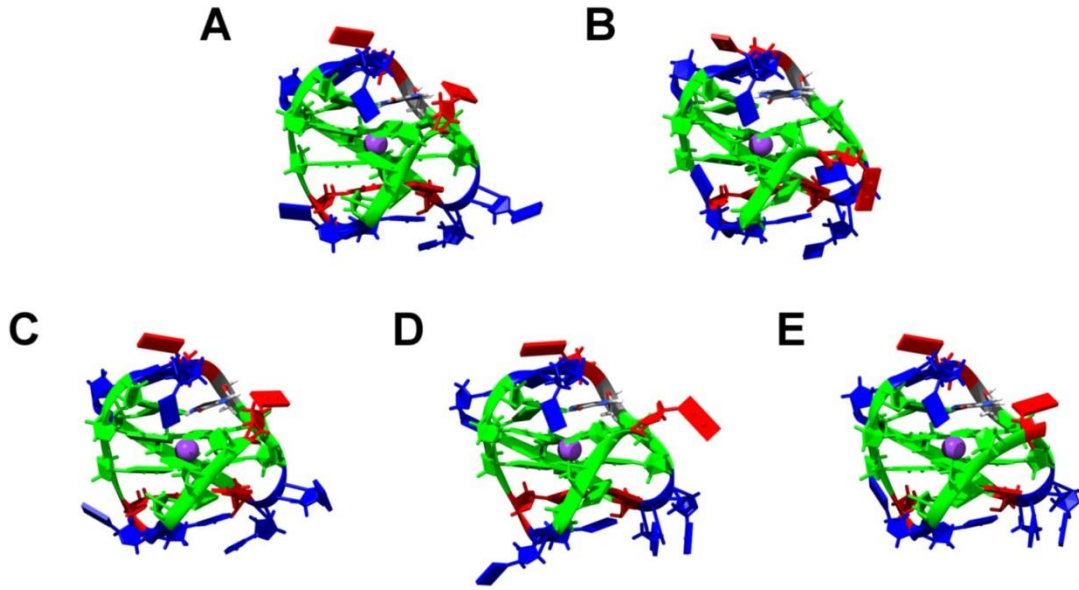
In addition to the HYDROPRO calculated hydrodynamic values, free energy values (ΔG) and the number of water bound were also determined for each cluster (Table 9). Free energy was calculated using the MMPBSA method and is the sum of all the molecular mechanics energies (bond, angle, torsion, van der Waals, and electrostatic) and the solvation energy (calculated using a numerical solution of the Poisson-Boltzmann equation) (Kollman *et al.*, 2000). The free energy values did not include an estimate of solute entropy, which is the least reliable component of MMPBSA calculations (Štefl *et al.*, 2003, Fadrná *et al.*, 2004, Cang *et al.*, 2011). The results of the cluster analysis demonstrated that G-quadruplex structures remained highly polymorphic even with the limited sampling of MD simulations. In general, it was observed that the more populated substates were more energetically favorable (i.e. lower ΔG). The G-quadruplex structures in these substates were more spherical (i.e. lower f/f_0) and compact (i.e. higher $s_{20,w}$), and were characterized by favorable interactions between the loops and flanking bases with the G-tetrad stems (Figures 33-42). For example, in the case of the 2HY9 sequence, which contained a 3-residue long 5' flanking sequence and a 2-residue long 3' flanking sequence, the more compact structures featured stacking interactions between the flanking bases and the G-tetrads while the less compact structures did not. Previous studies have reported that the folding of a single-stranded DNA structure into the G-quadruplex form was associated with the release of water molecules from the DNA into the surrounding environment (Olsen *et al.*, 2006, Miyoshi *et al.*, 2007). It was observed that the more compact structures were also associated with a reduced number of bound water. As hydration plays an important role in ligand binding and recognition (Poornima and Dean, 1995a, Poornima and Dean, 1995b, Poornima and Dean, 1995c, García-Sosa,

2013), for instance, the binding of groove-binding small molecules to duplex DNA is driven by water being displaced from the minor grooves (Chaires, 2006). Therefore, clustering by $s_{20,w}$ values can be used in the process of drug design to identify substates that could interact more favorably with small-molecule inhibitors.

178

Table 9. Results of $s_{20,w}$ Cluster Analysis on MD Trajectories

Structure	Cluster #	N	$s_{20,w}$[1]	f/f_0	ΔG[1]	# Water
143D	1	6170	1.98	1.12	-4421.89	165.26
	2	3802	1.95	1.14	-4421.62	166.92
	3	28	1.91	1.16	-4416.97	169.89
1KF1	1	5244	1.82	1.22	-4360.05	174.63
	2	3503	1.84	1.20	-4363.84	173.62
	3	768	1.87	1.19	-4368.20	173.43
	4	485	1.78	1.24	-4357.99	176.97
2GKU	1	4694	2.07	1.14	-4664.48	174.66
	2	2482	2.04	1.15	-4664.43	176.61
	3	1413	2.01	1.17	-4664.29	178.65
	4	1167	2.09	1.13	-4663.50	172.86
	5	244	1.98	1.19	-4664.01	181.28
2HY9	1	6374	2.15	1.16	-5094.00	190.93
	2	3626	2.08	1.19	-5091.86	197.55
2JSM	1	6852	1.98	1.15	-4536.80	175.92
	2	3115	1.95	1.17	-4533.68	177.01
	3	32	1.90	1.19	-4525.48	178.38
	4	1	2.04	1.13	-4545.97	172.00
2JPZ	1	6009	2.16	1.14	-4933.18	188.59
	2	3991	2.23	1.12	-4930.52	180.50
2JSL	1	6681	2.07	1.16	-4811.53	186.55
	2	3319	2.12	1.14	-4817.04	184.74

2KF8	1	5471	1.96	1.14	-4241.88	166.86
	2	3892	1.93	1.15	-4239.20	168.79
	3	411	1.90	1.16	-4224.43	170.41
	4	186	1.98	1.13	-4247.06	164.19
	5	36	1.87	1.18	-4215.53	172.75
	6	4	1.84	1.20	-4205.31	171.50
2KKA-G	1	7863	2.03	1.13	-4434.13	171.78
	2	2135	1.98	1.15	-4427.76	175.58
	3	2	2.09	1.11	-4409.13	179.00
2KKA-I	1	4673	2.01	1.14	-4347.30	172.24
	2	4146	2.04	1.12	-4347.99	170.08
	3	1068	1.98	1.15	-4343.95	173.96
	4	111	1.94	1.17	-4340.41	176.68
	5	2	2.08	1.11	-4331.61	162.00

[1]$s_{20,w}$ is expressed in units of 10^{-13}s; ΔG is expressed in units of kcal/mol

Principal Component Analysis of Molecular Dynamics Trajectories

PCA was employed to identify the major patterns of motions in the MD models and investigate the underlying mechanism for the formation of the different substates identified by clustering analysis. Motions from MD can appear random and chaotic, however, most of the fluctuations associated with the MD models can usually be reduced to several low-frequency eigenvectors with large eigenvalues (Teeter and Case, 1990). The first, second, and third eigenvectors are depicted using "porcupine" plots (Tai *et al.*, 2001, Tai *et al.*, 2002) with the arrows representing the magnitude and direction for the eigenvector for each atom (Figures 43-45). In general, it was observed that the overall dynamics of the G-quadruplex structures were composed of only one to five major movements as indicated by the large eigenvalues associated with the first several eigenvectors compared to subsequent eigenvectors (Figure 46). In fact, the first three eigenvectors accounted for 30-80 % of the variance in the models (Figures 47). The porcupine plots indicated that motions related to the more rigid green colored stems (Figures 43-45) tend to be small and localize while the motions related to the more flexible red loops (Figures 43-45) and blue flanking bases (Figure 43-45) were more prominent. Each structural component was defined by characteristic movements. Most of the movements associated with the stem structures were the twisting motions of the phosphate backbone around the quadruple helix stack. The major movements associated with the loop structures were rotation of the phosphate backbone which moved the bases either toward the G-tetrad stem or away from it. In agreement with previous atomic positional fluctuations results, higher magnitude movement was seen with the chain-reversal loops compared to the lateral and diagonal loops. Lastly, the transition between

182

flanking bases stacking on the G-tetrad bases to the unstacking of the flanking bases was observed as the major movement associated with the flanking bases. The movements associated with structures containing longer flanking sequences (i.e. 2HY9 and 2JPZ) were much higher in magnitude. The movements identified by PCA correlated with the structural differences between the different substates previously identified by cluster analysis.

Figure 43. Porcupine plots of the first eigenvectors for 143D (A), 1KF1 (B), 2GKU (C), 2HY9 (D), 2JSM (E), 2JPZ (F), 2JSL (G), 2KF8 (H), 2KKA-G (I), and 2KKA-I(J). Principal component analysis was carried out on MD trajectories in order to determine the major patterns of motions. Motions associated with stem residues are colored green, loop residues are red, flanking residues are blue, and central ions are yellow.
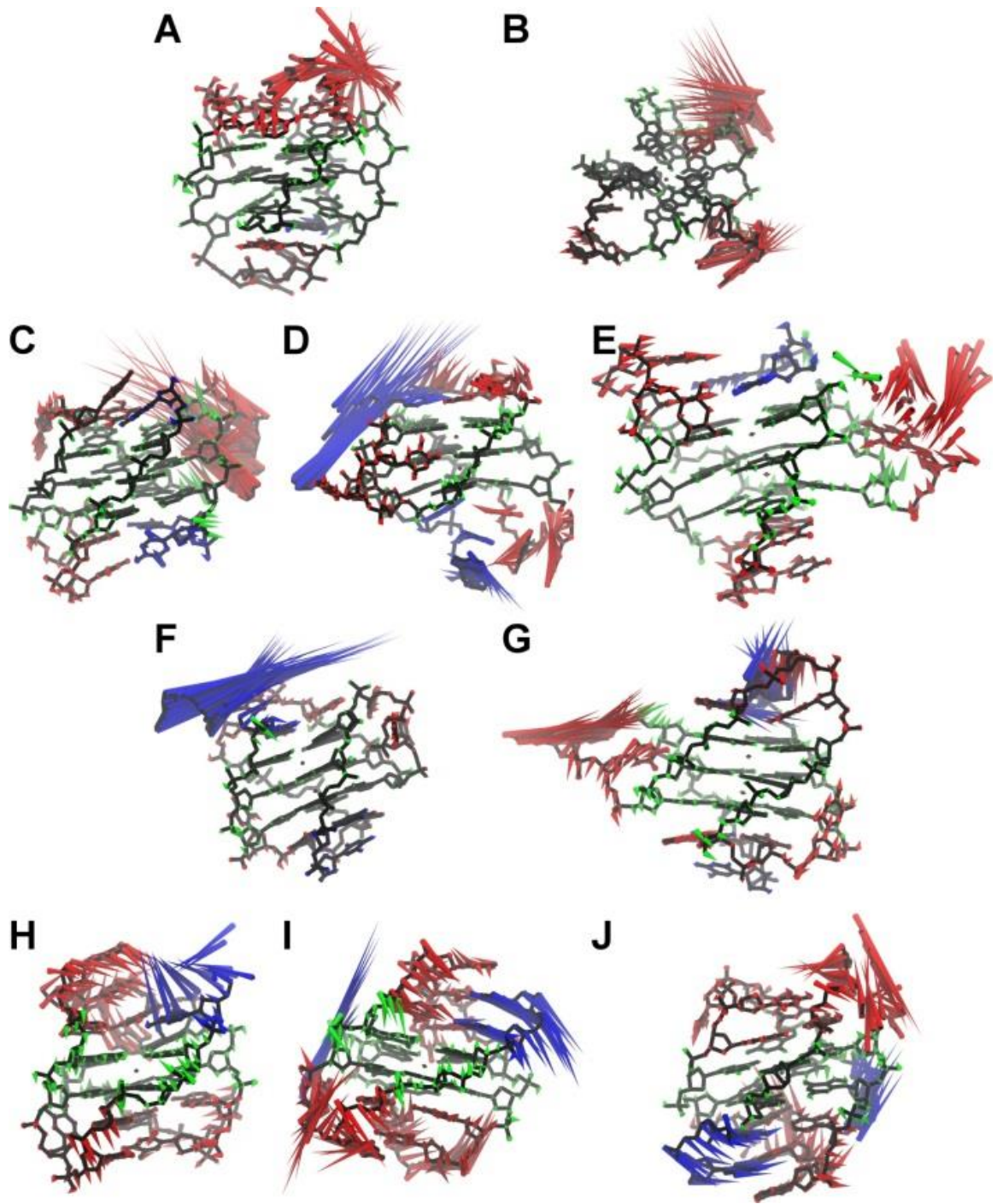
Figure 44. Porcupine plots of the second eigenvectors for 143D (A), 1KF1 (B), 2GKU (C), 2HY9 (D), 2JSM (E), 2JPZ (F), 2JSL (G), 2KF8 (H), 2KKA-G (I), and 2KKA-I(J). Principal component analysis was carried out on MD trajectories in order to determine the major patterns of motions. Motions associated with stem residues are colored green, loop residues are red, flanking residues are blue, and central ions are yellow.
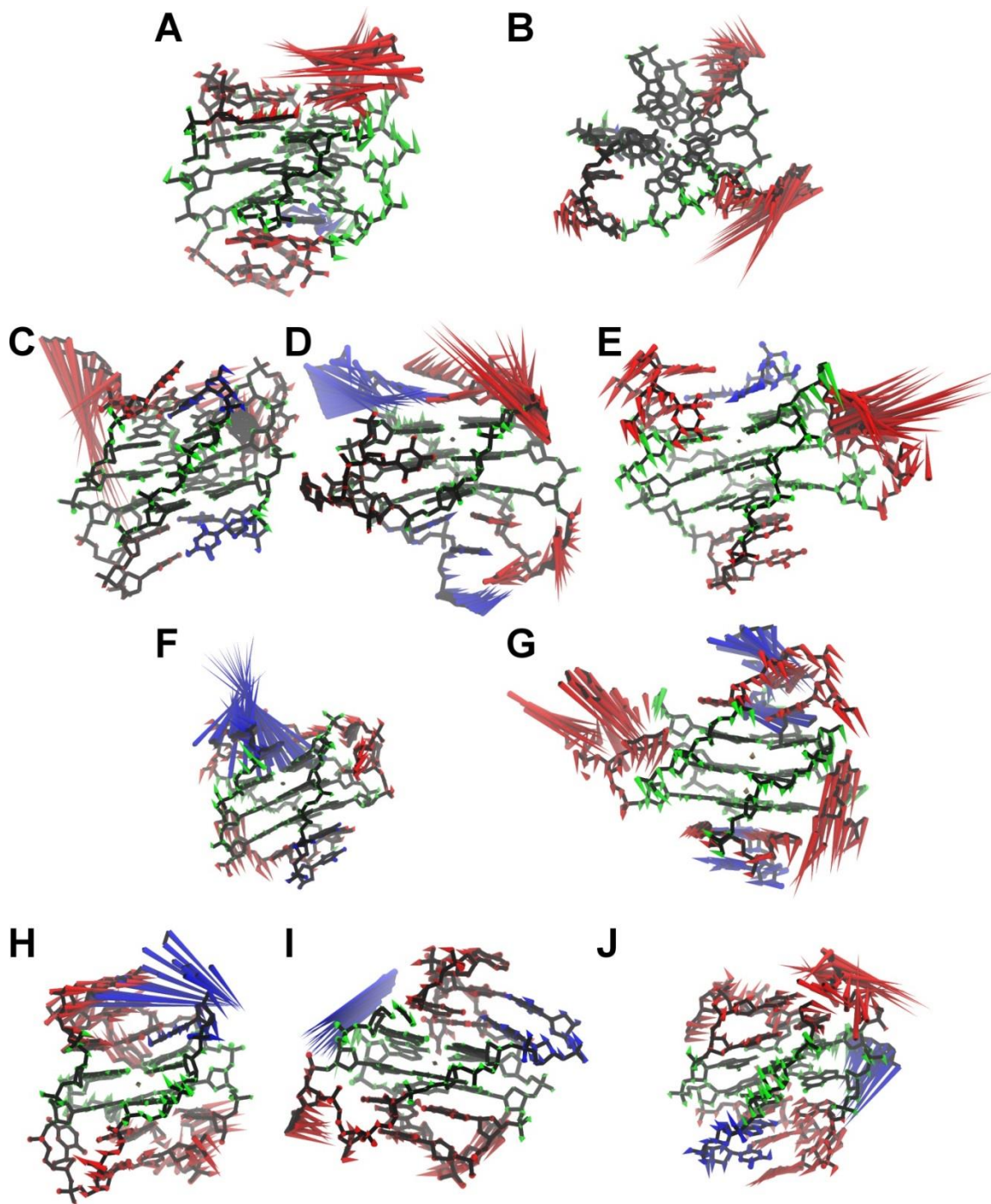
Figure 45. Porcupine plots of the third eigenvectors for 143D (A), 1KF1 (B), 2GKU (C), 2HY9 (D), 2JSM (E), 2JPZ (F), 2JSL (G), 2KF8 (H), 2KKA-G (I), and 2KKA-I(J). Principal component analysis was carried out on MD trajectories in order to determine the major patterns of motions. Motions associated with stem residues are colored green, loop residues are red, flanking residues are blue, and central ions are yellow.

A   B

C   D   E

F   G

H   I   J
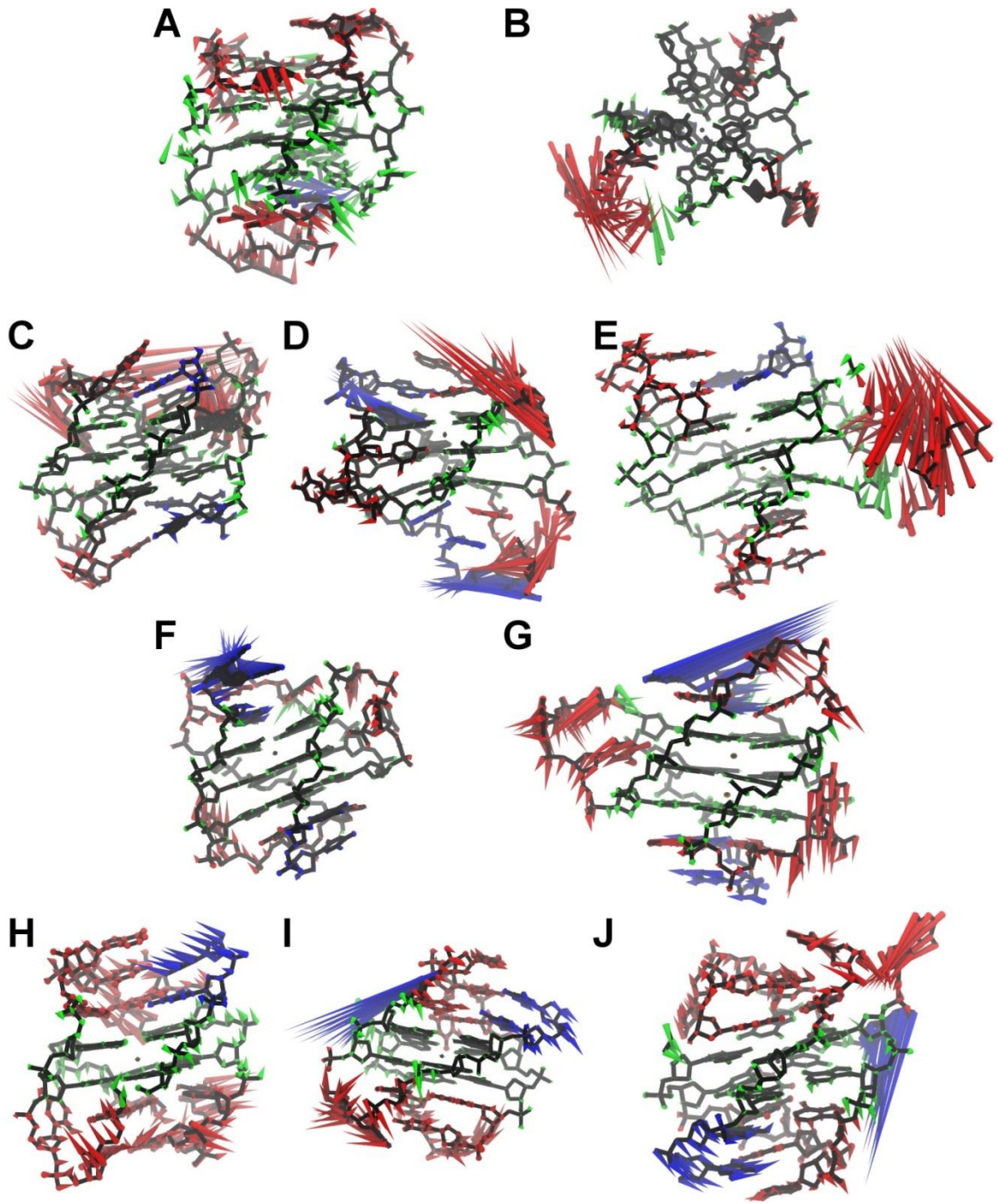
Figure 46. The first ten eigenvalues for 143D (A), 1KF1 (B), 2GKU (C), 2HY9 (D), 2JSM (E), 2JPZ (F), 2JSL (G), 2KF8 (H), 2KKA-G (I), and 2KKA-I(J).
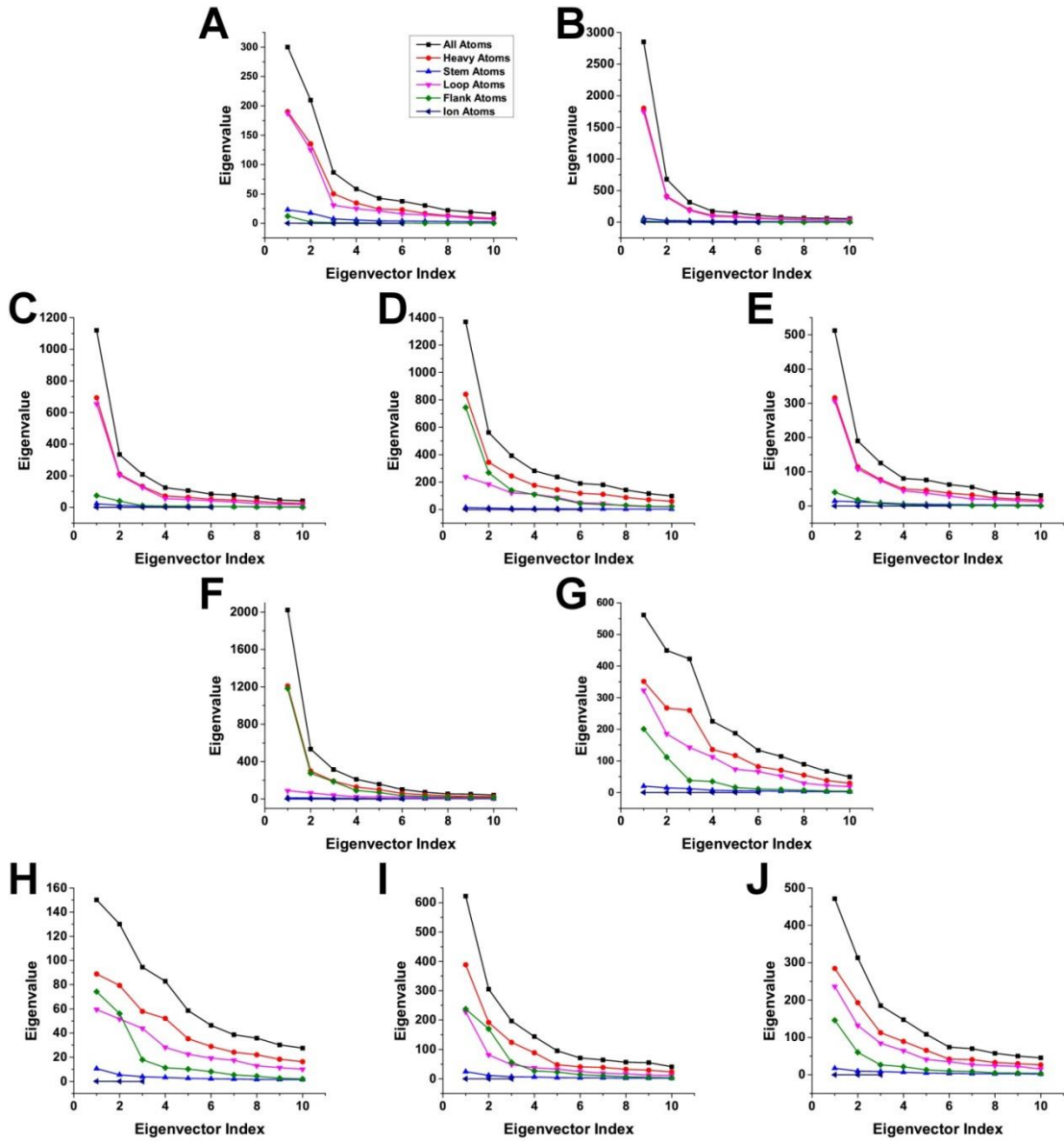
Figure 47. The percent of total variance explained by principal component analysis for 143D (A), 1KF1 (B), 2GKU (C), 2HY9 (D), 2JSM (E), 2JPZ (F), 2JSL (G), 2KF8 (H), 2KKA-G (I), and 2KKA-I(J).
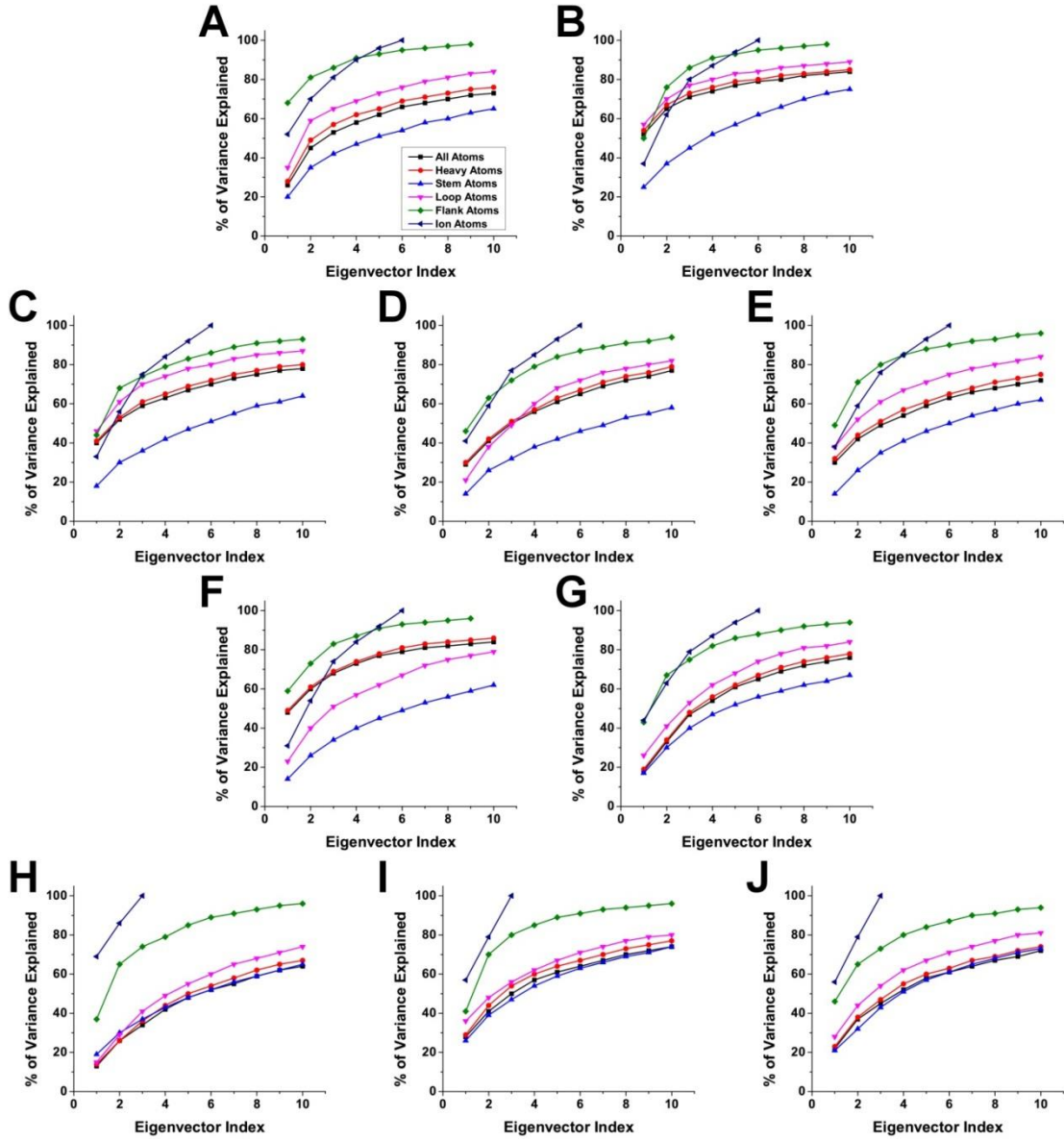
The first ten eigenvalues are reported (Figure 46) as calculated for the G-quadruplex structures as a whole and for various structural components (i.e. stem, loops, flanking bases). The eigenvalues were slightly reduced when the more mobile hydrogen atoms were removed from the analysis. Because PCA is sensitive to the scaling of the original data, the eigenvectors and eigenvalues of similar motions should be similar to one another. An overlap of the eigenvalues of loop and flanking bases with the eigenvalues of the overall G-quadruplex structures indicated that overall dynamic motions of G-quadruplex structures are mainly due to motions of loop and flanking bases. To demonstrate this correlation between the dynamics motions of individual structural components and the overall dynamic motions of the G-quadruplex structures, the first principal components were plotted as a function of the second principal components for all ten MD models for overall G-quadruplex structure and the individual stem, loop, and flanking structures (Figures 48-57). From the PCA results, each model can be classified as having loop-dominated dynamics (i.e. 143D, 1KF1, 2GKU, 2JSM), flanking bases-dominated dynamics (i.e. 2HY9 and 2JPZ), and mixed dynamics with contributions from both loop and flanking bases (i.e. 2JSL, 2KF8, 2KKA-G, and 2KKA-I). Overall, the PCA findings help explain the differences in hydrodynamic values for individual substates identified by previous cluster analysis. In addition, the differences observed between the dynamic motions of different loop and flanking bases could have significant implications in drug interaction and the choices of sequence chosen for *in vitro* drug binding studies.

Figure 48. The second principal component as a function of the first principal component for 143D MD trajectory. Principal components are plotted for all non-hydrogen atoms (A), non-hydrogen G-tetrad stem atoms (B), non-hydrogen loop atoms (C), and non-hydrogen flanking bases atoms (D). For the 143D MD trajectory, dynamic motions of loop atoms have a dominant effect on the overall dynamic motions of the G-quadruplex structures.
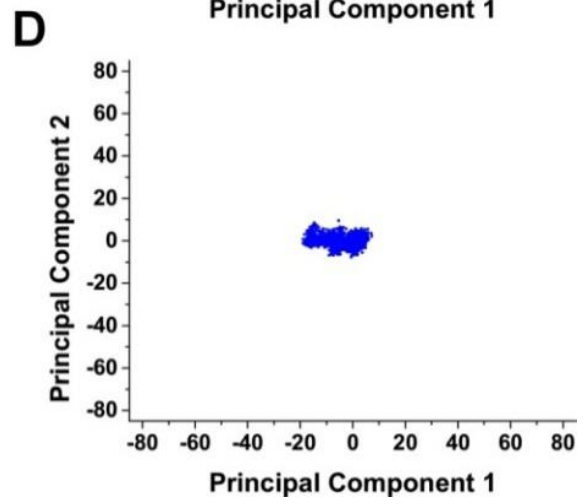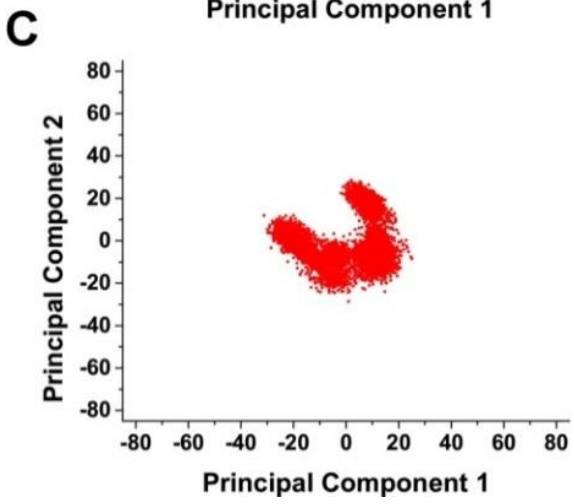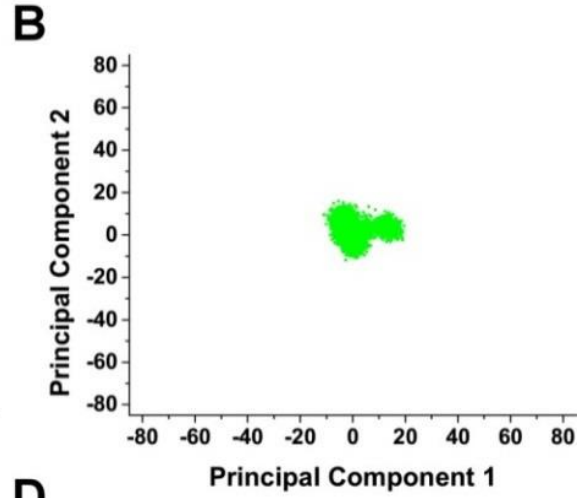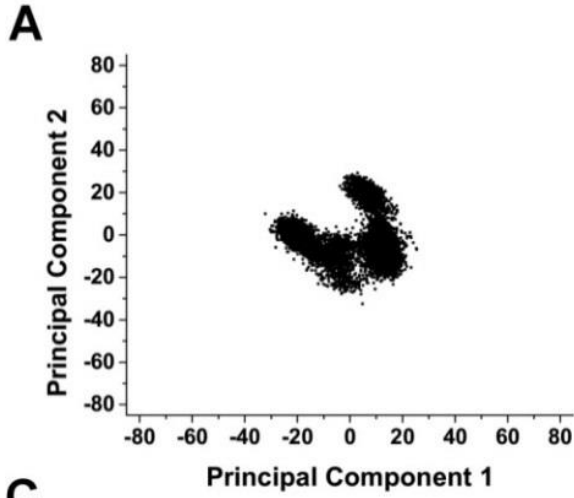
Figure 49. The second principal component as a function of the first principal component for 1KF1 MD trajectory. Principal components are plotted for all non-hydrogen atoms (A), non-hydrogen G-tetrad stem atoms (B), non-hydrogen loop atoms (C), and non-hydrogen flanking bases atoms (D). For the 1KF1 MD trajectory, dynamic motions of loop atoms have a dominant effect on the overall dynamic motions of the G-quadruplex structures.
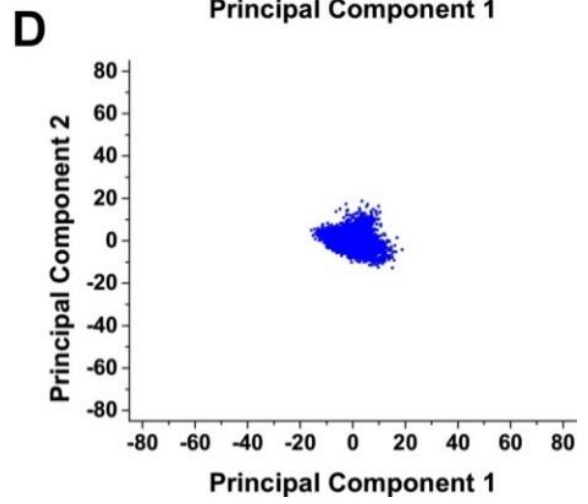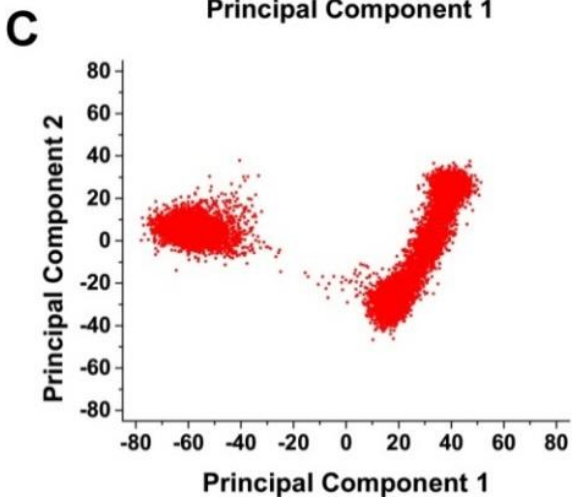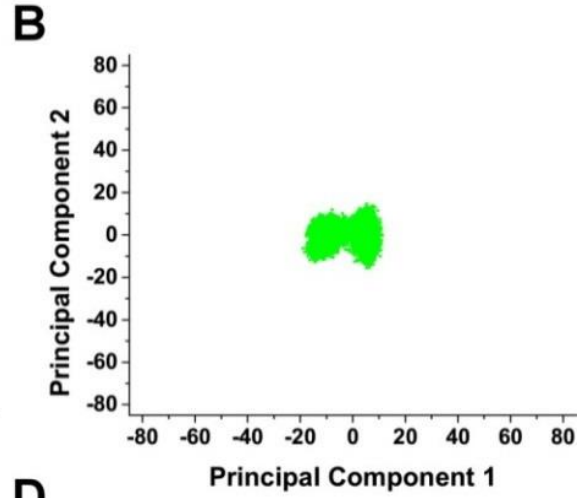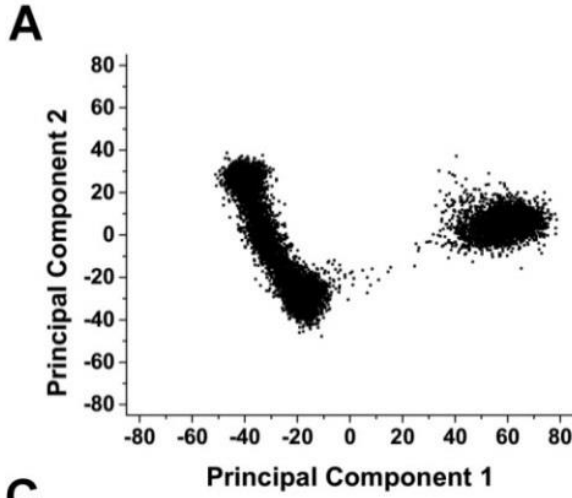
Figure 50. The second principal component as a function of the first principal component for 2GKU MD trajectory. Principal components are plotted for all non-hydrogen atoms (A), non-hydrogen G-tetrad stem atoms (B), non-hydrogen loop atoms (C), and non-hydrogen flanking bases atoms (D). For the 2GKU MD trajectory, dynamic motions of loop atoms have a dominant effect on the overall dynamic motions of the G-quadruplex structures.
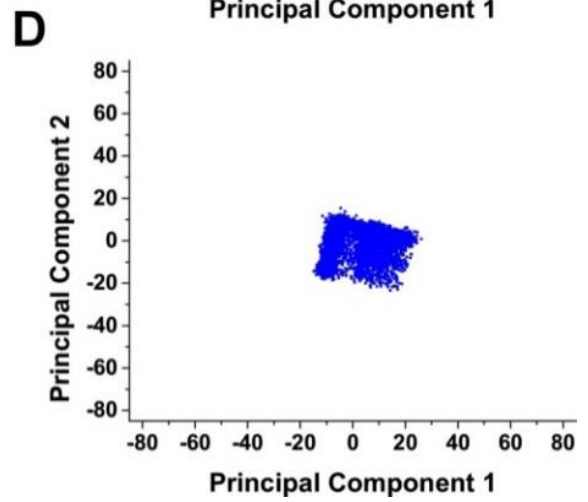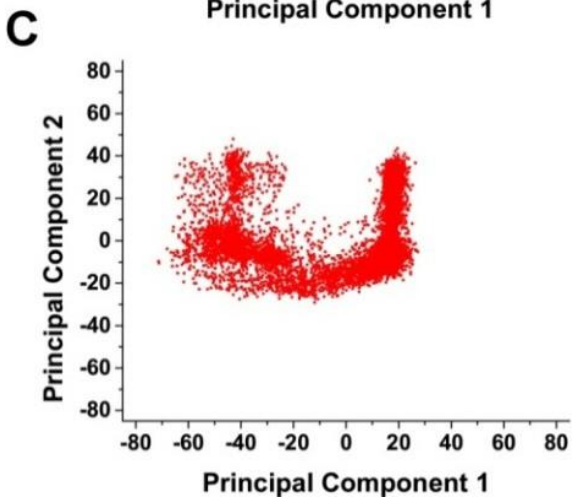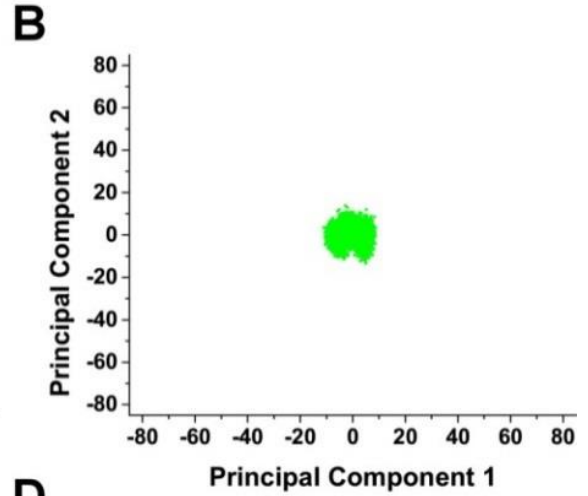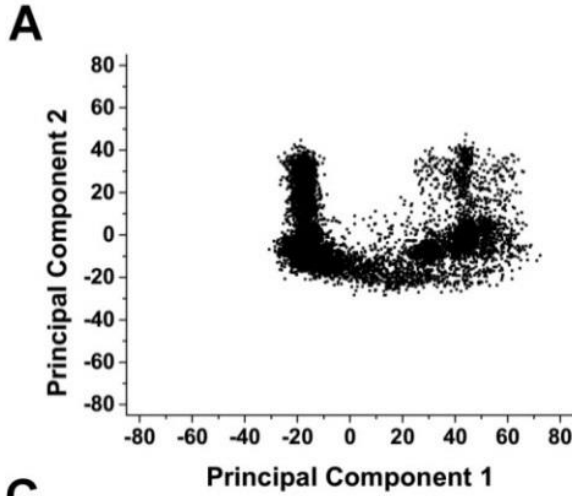
Figure 51. The second principal component as a function of the first principal component for 2HY9 MD trajectory. Principal components are plotted for all non-hydrogen atoms (A), non-hydrogen G-tetrad stem atoms (B), non-hydrogen loop atoms (C), and non-hydrogen flanking bases atoms (D). For the 2HY9 MD trajectory, dynamic motions of flanking atoms have a dominant effect on the overall dynamic motions of the G-quadruplex structures with only small contributions from loop atoms.
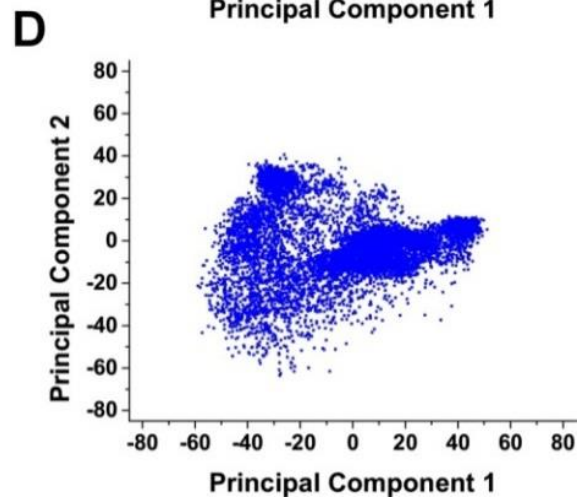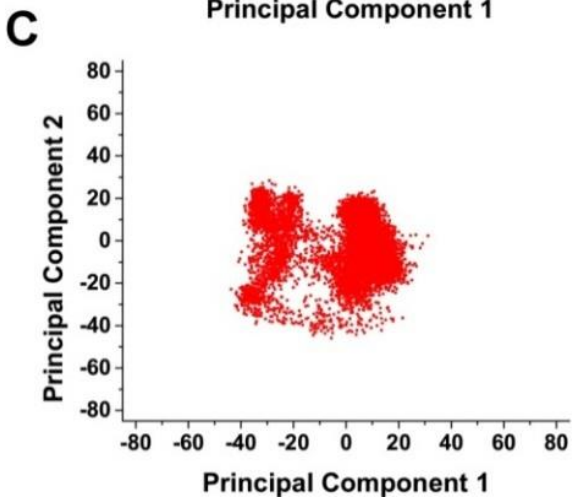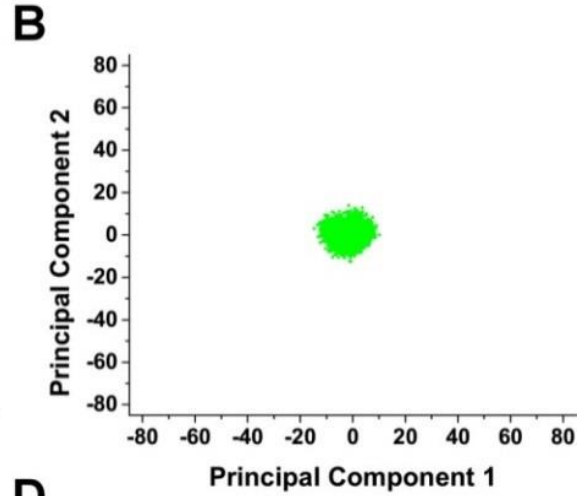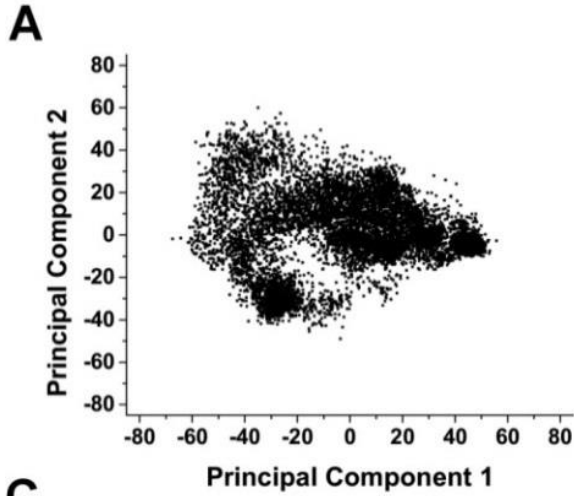
Figure 52. The second principal component as a function of the first principal component for 2JSM MD trajectory. Principal components are plotted for all non-hydrogen atoms (A), non-hydrogen G-tetrad stem atoms (B), non-hydrogen loop atoms (C), and non-hydrogen flanking bases atoms (D). For the 2JSM MD trajectory, dynamic motions of loop atoms have a dominant effect on the overall dynamic motions of the G-quadruplex structures.
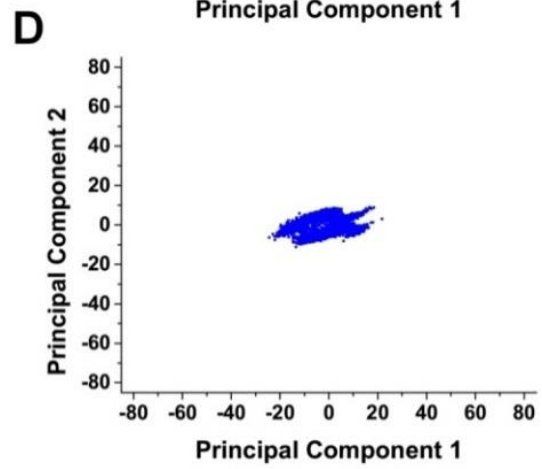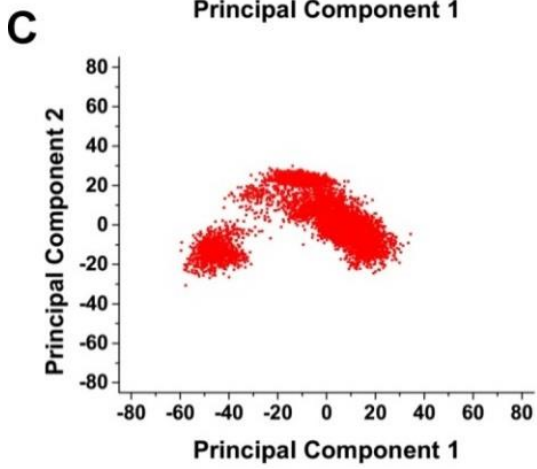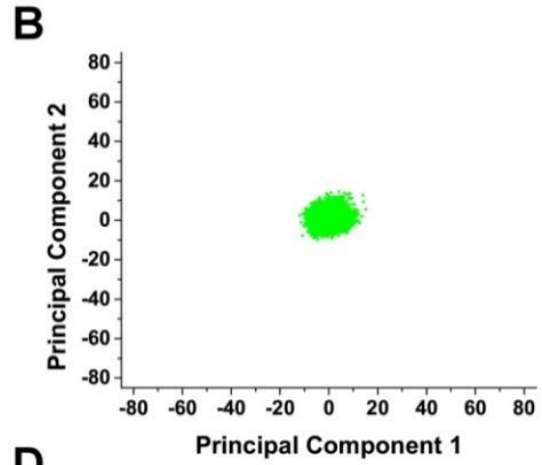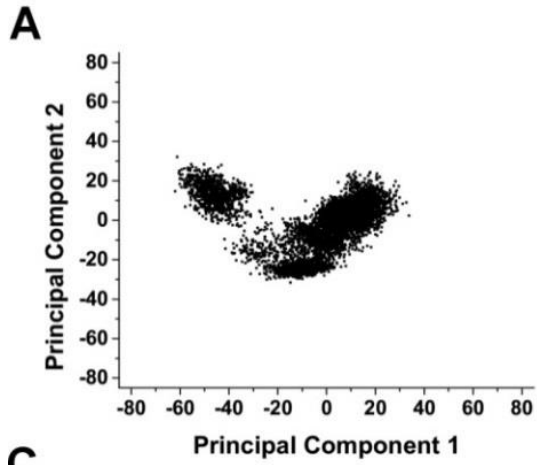
Figure 53. The second principal component as a function of the first principal component for 2JPZ MD trajectory. Principal components are plotted for all non-hydrogen atoms (A), non-hydrogen G-tetrad stem atoms (B), non-hydrogen loop atoms (C), and non-hydrogen flanking bases atoms (D). For the 2JPZ MD trajectory, dynamic motions of flanking atoms have a dominant effect on the overall dynamic motions of the G-quadruplex structures.

205
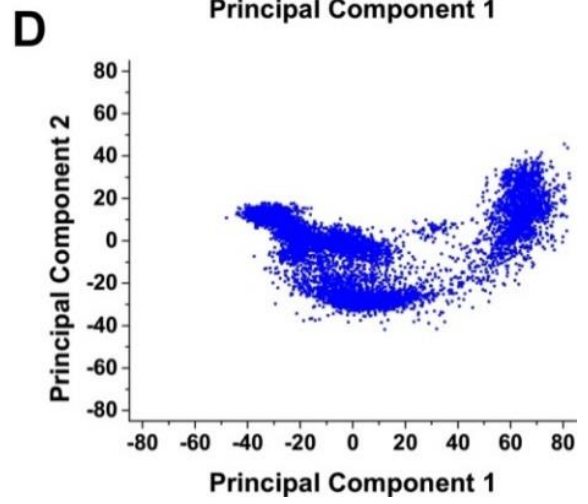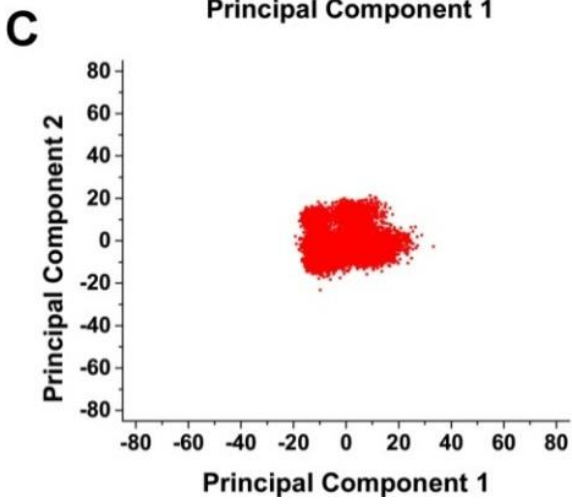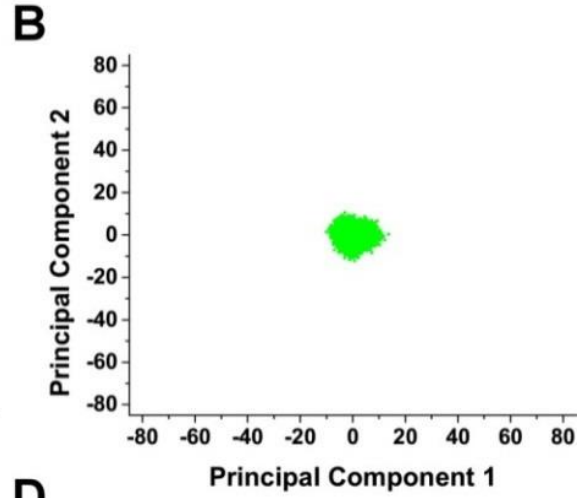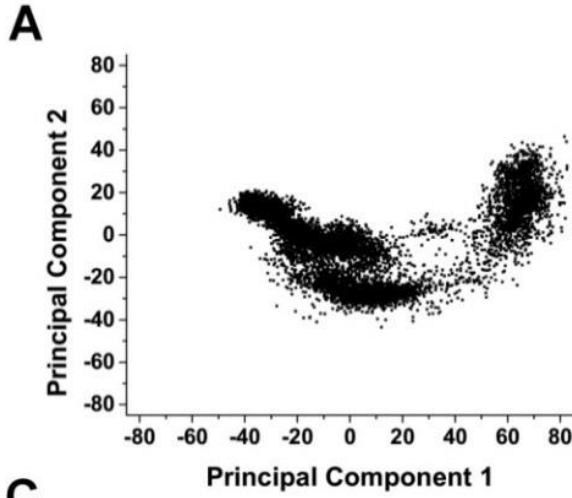
Figure 54. The second principal component as a function of the first principal component for 2JSL MD trajectory. Principal components are plotted for all non-hydrogen atoms (A), non-hydrogen G-tetrad stem atoms (B), non-hydrogen loop atoms (C), and non-hydrogen flanking bases atoms (D). For the 2JSL MD trajectory, dynamic motions of both loop and flanking atoms were observed to influence the overall dynamic motions of the G-quadruplex structures.
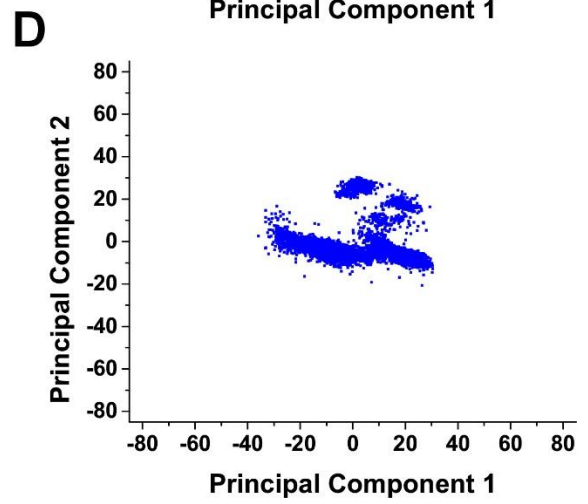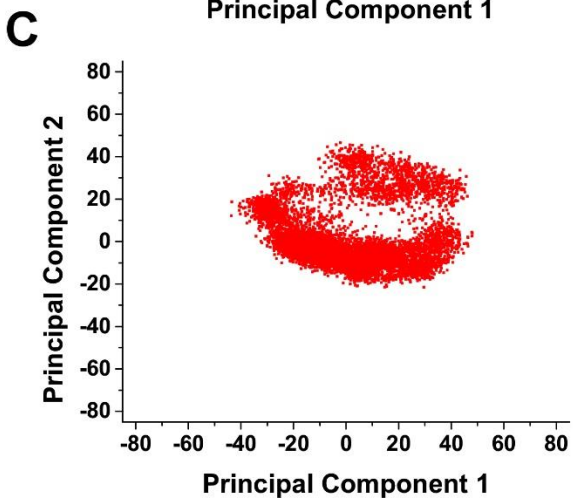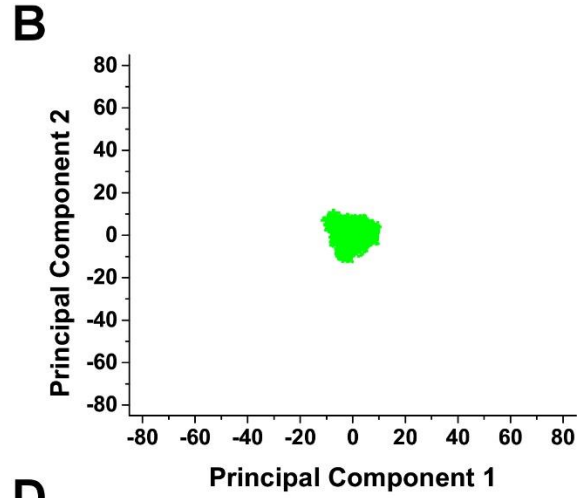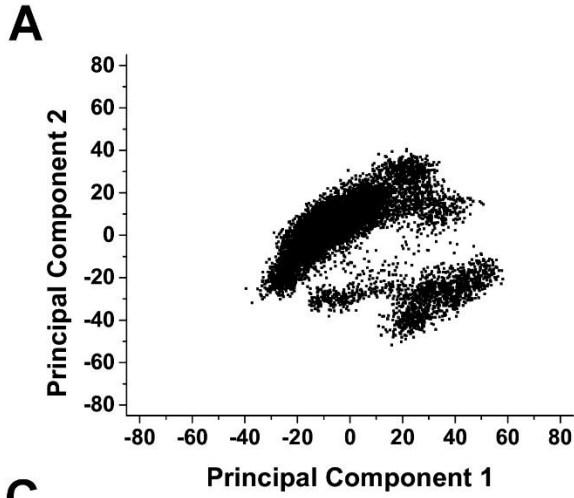
Figure 55. The second principal component as a function of the first principal component for 2KF8 MD trajectory. Principal components are plotted for all non-hydrogen atoms (A), non-hydrogen G-tetrad stem atoms (B), non-hydrogen loop atoms (C), and non-hydrogen flanking bases atoms (D). For the 2KF8 MD trajectory, dynamic motions of both loop and flanking atoms were observed to influence the overall dynamic motions of the G-quadruplex structures.
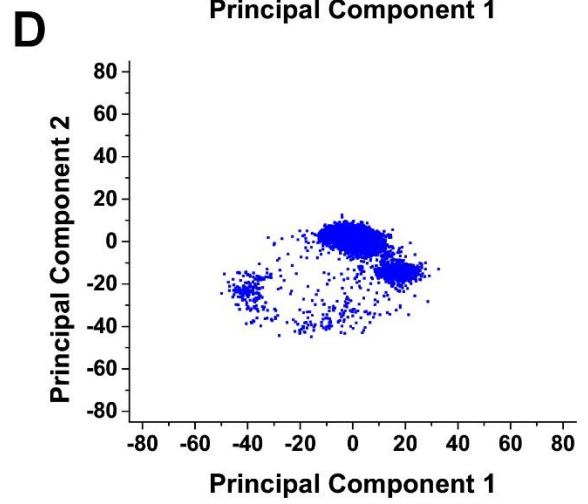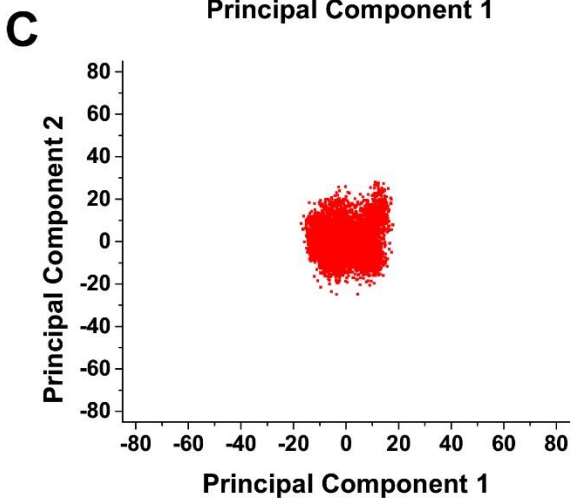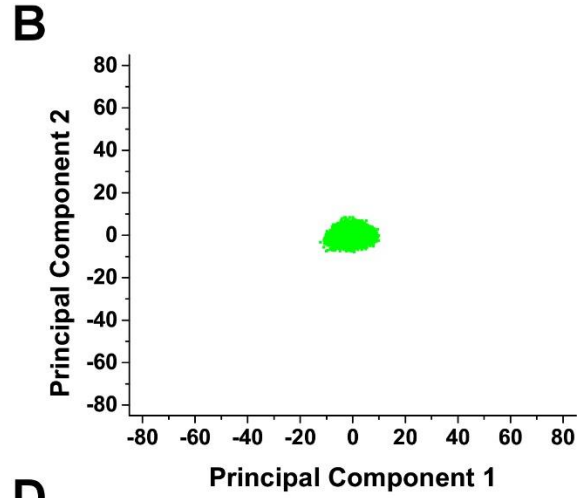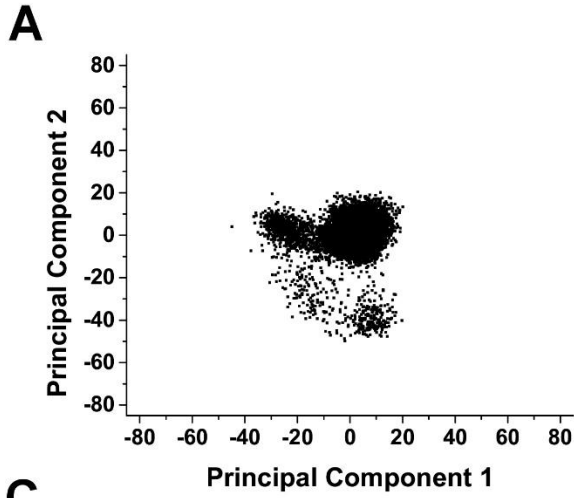
Figure 56. The second principal component as a function of the first principal component for 2KKA-G MD trajectory. Principal components are plotted for all non-hydrogen atoms (A), non-hydrogen G-tetrad stem atoms (B), non-hydrogen loop atoms (C), and non-hydrogen flanking bases atoms (D). For the 2KKA-G MD trajectory, dynamic motions of both loop and flanking atoms were observed to influence the overall dynamic motions of the G-quadruplex structures.

211
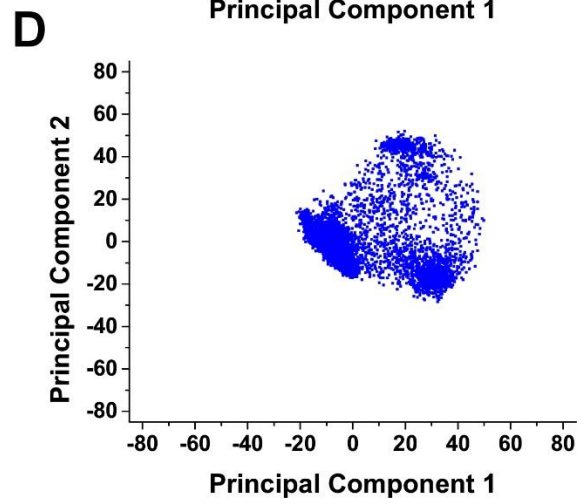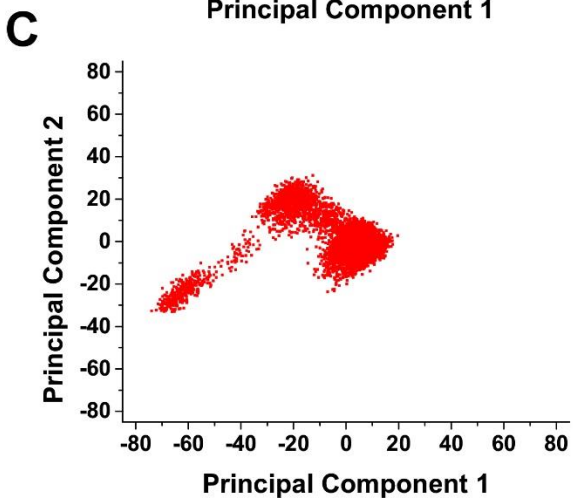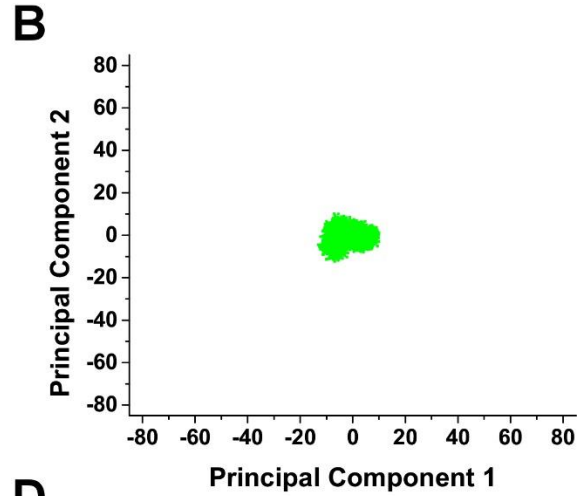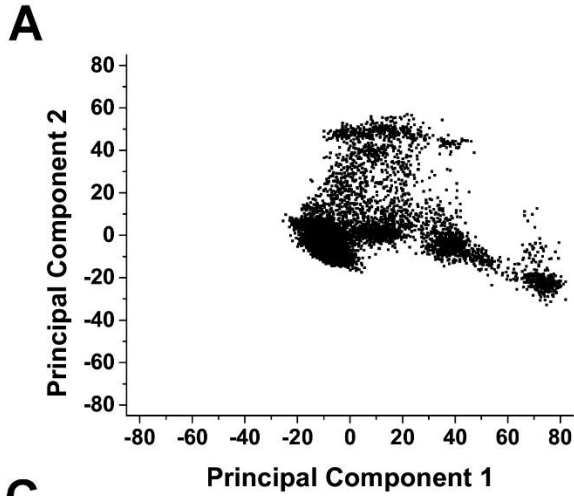
Figure 57. The second principal component as a function of the first principal component for 2KKA-I MD trajectory. Principal components are plotted for all non-hydrogen atoms (A), non-hydrogen G-tetrad stem atoms (B), non-hydrogen loop atoms (C), and non-hydrogen flanking bases atoms (D). For the 2KKA-I MD trajectory, dynamic motions of both loop and flanking atoms were observed to influence the overall dynamic motions of the G-quadruplex structures.

213
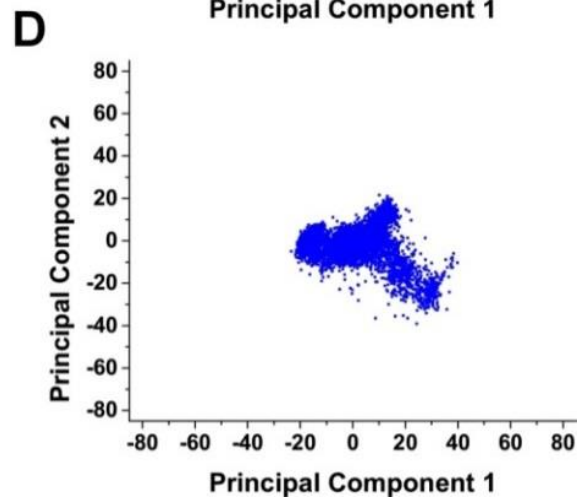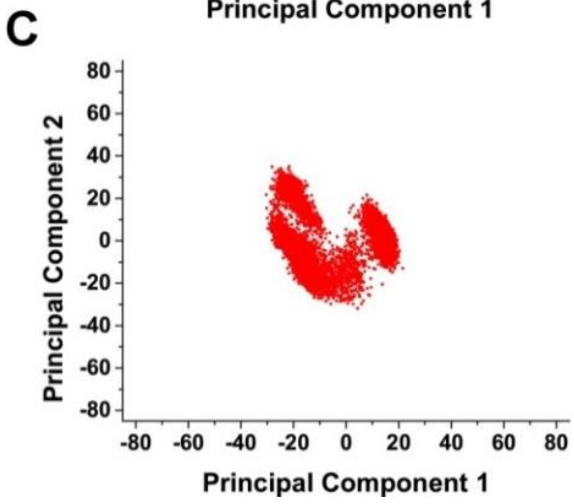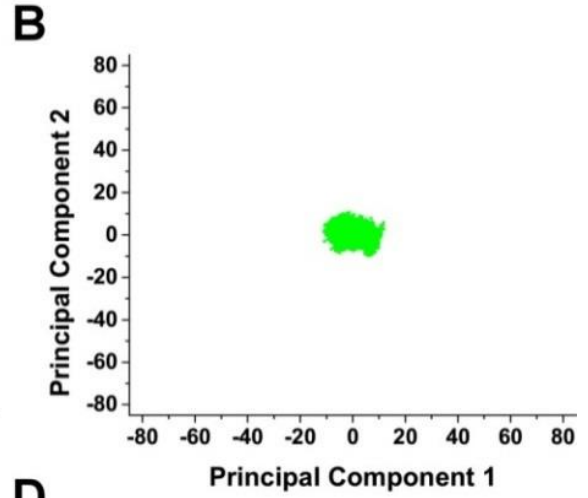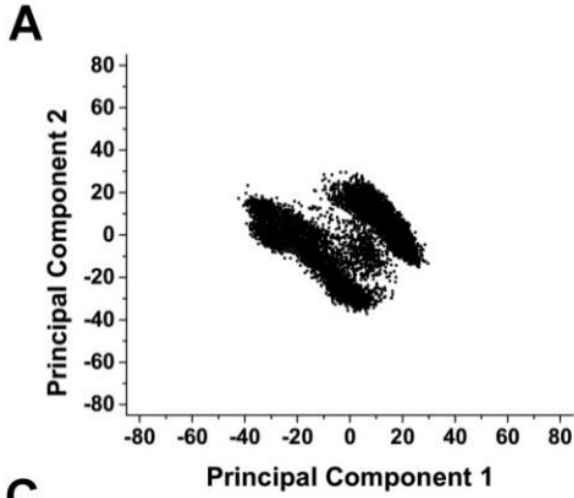
Investigation of Water and Cation Distribution around G-quadruplex Structures

Given the importance of hydration and cations in influencing G-quadruplex formation, as well as the role of water and ions in determining hydrodynamic properties of a macromolecule, density mapping and radial distribution functions (RDF) were used to visualize the distribution of these molecules around the G-quadruplex structures and provide a general overview of hydration and cation binding. RDF of the distance between oxygen atoms of the water molecules and non-hydrogen atoms on the surface of the G-quadruplex structures indicated the formation of a well-defined first hydration shell between 2.4 to 3.2 Å around the G-quadruplex structures (Figure 58). The second and third hydration shells can also be observed, although these peaks were not as sharp as the peaks for the first hydration shell. To visualize these hydration shells, density mapping of the water molecules over the duration of the trajectory was performed. Water density is shown contoured at equivalent levels (about two times the expected bulk water density) around the average structures from the MD trajectories (Figure 59). The condensation of the water around the G-quadruplex structure was evident. While the water appeared to be evenly distributed around the G-quadruplex structures, higher density of water was observed around the rigid stem structures and lower density around the more mobile loops and flanking structures. This observation agreed with a previous study, which reported that rigid structures are associated with more well-defined water positions and higher apparent density (Cheatham and Kollman, 1997). When contoured at a higher level (about three times the expected bulk water density), preferential binding of water to the grooves of the G-quadruplex structures became more apparent. At this higher contour level, a "spine of hydration" which traverses the grooves of the G-quadruplex structures

215

was observed (Figure 60). These findings demonstrated the interaction between water molecules DNA and suggested a role for hydration in maintaining the G-quadruplex structures.

216

Figure 58. Radial distribution functions (RDF) between G-quadruplex surface heavy atoms (no hydrogen) and water oxygen atoms. RDFs were calculated for distances from 1.5-6.5 Å (A) and for distances from G-quadruplex surface to edge of the periodic box.

A

B
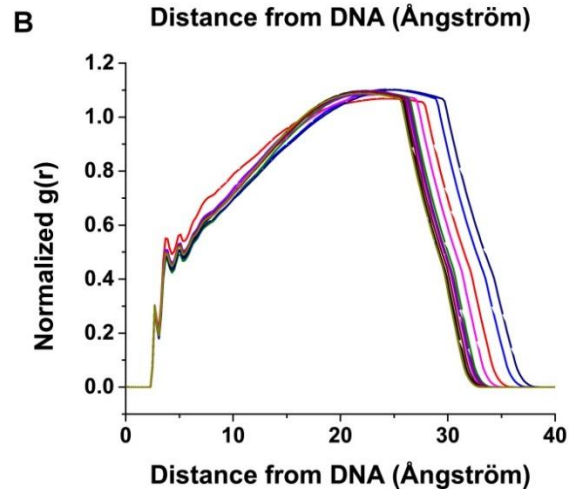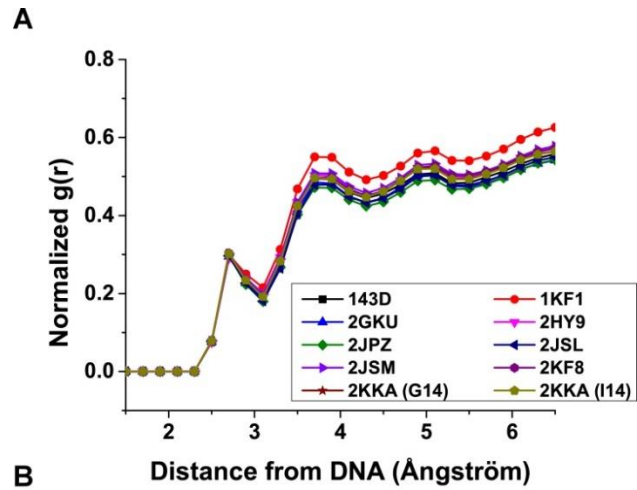
Figure 59. Pseudo-density grid maps of water oxygen atoms for 143D (A), 1KF1 (B), 2GKU (C), 2HY9 (D), 2JSM (E), 2JPZ (F), 2JSL (G), 2KF8 (H), 2KKA-G (I), and 2KKA-I (J). Water density (blue) was contoured at 2X the density (55.5 M) of bulk water. The average structure of each G-quadruplex over the trajectory is shown.

A B C D E F G H I J
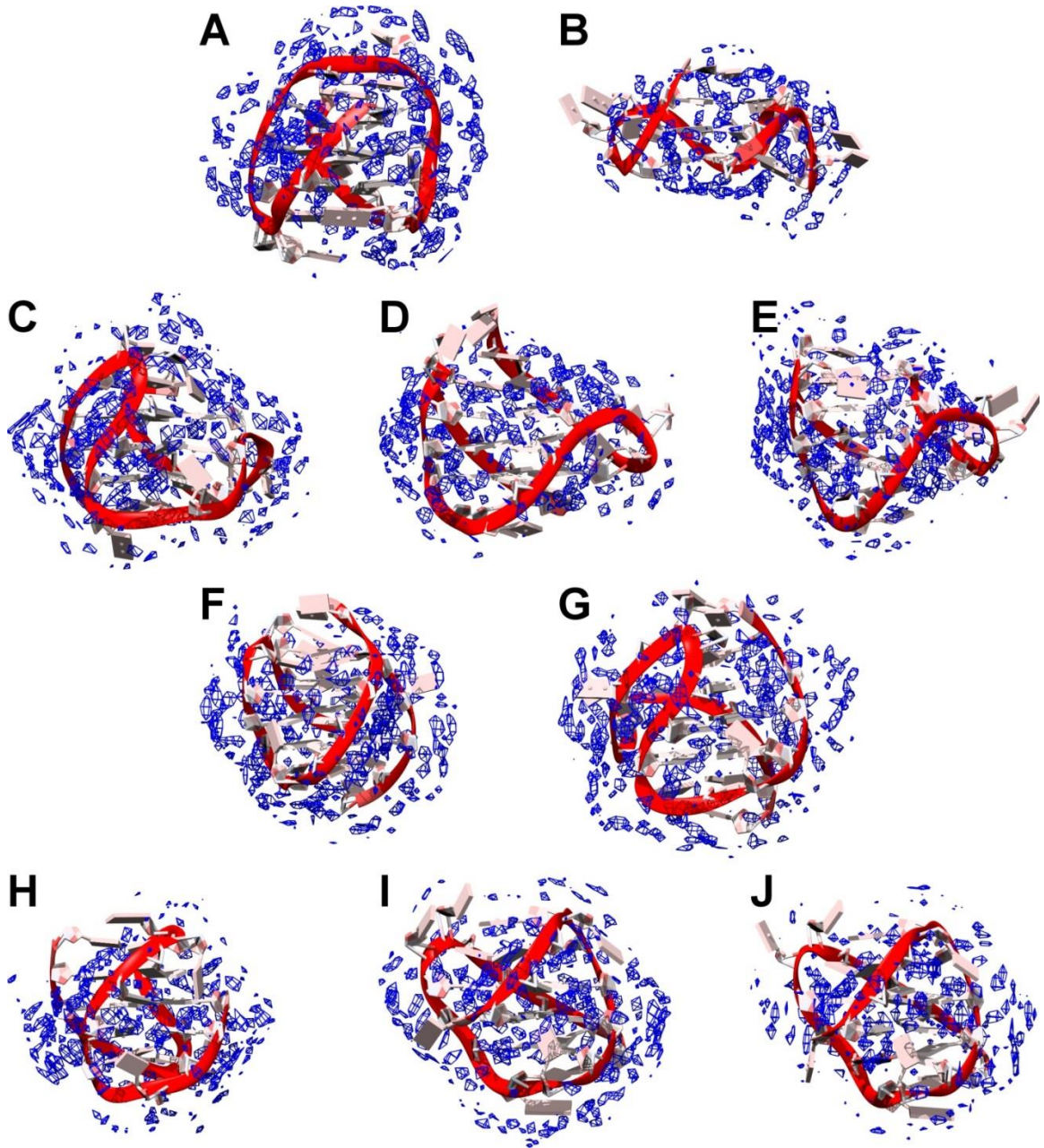
Figure 60. Pseudo-density grid maps of water oxygen atoms for 143D (A), 1KF1 (B), 2GKU (C), 2HY9 (D), 2JSM (E), 2JPZ (F), 2JSL (G), 2KF8 (H), 2KKA-G (I), and 2KKA-I (J). Water density (blue) was contoured at 3X the density (55.5 M) of bulk water. The average structure of each G-quadruplex over the trajectory is shown.
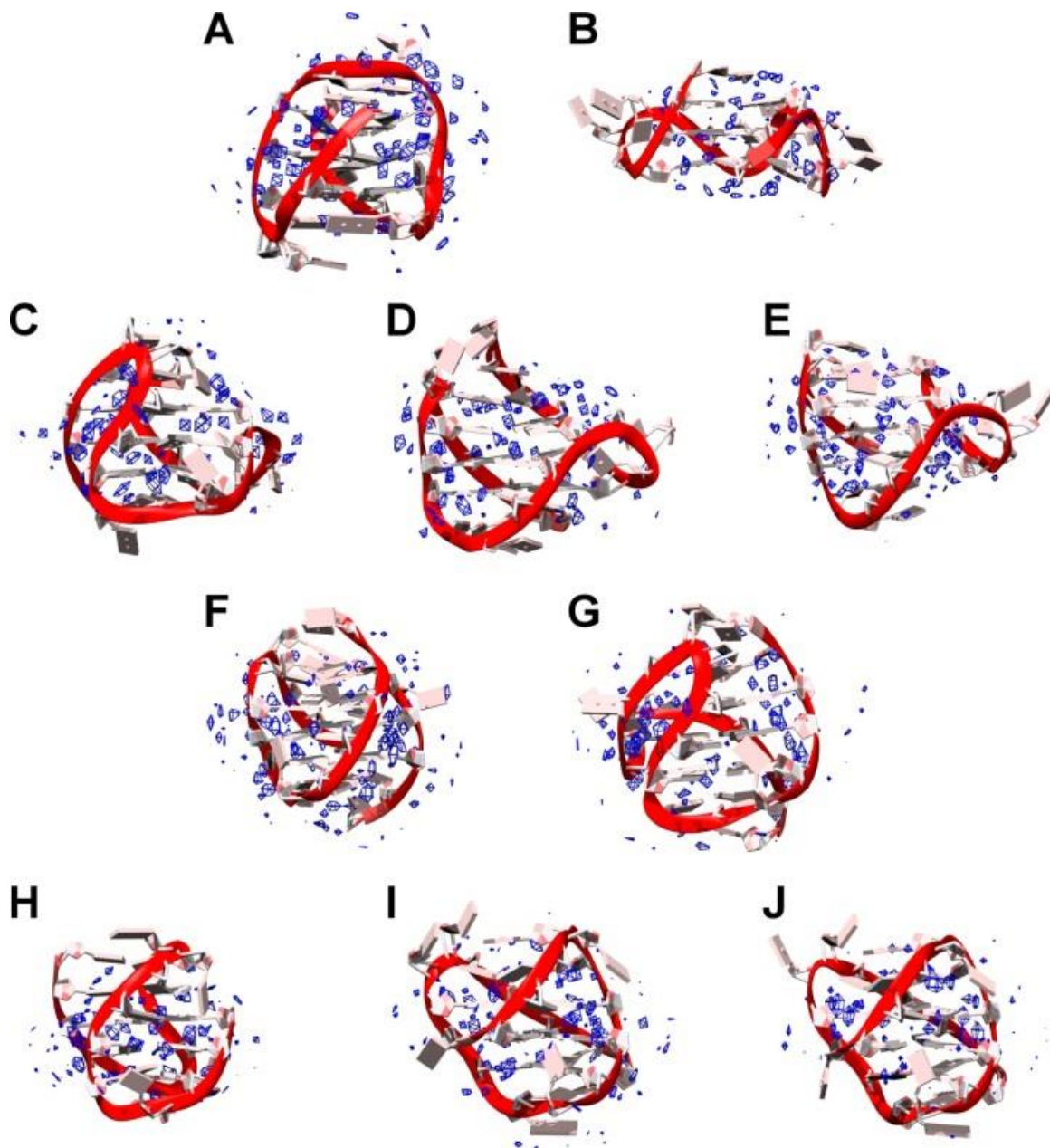
The RDF plots of the distance between cations and G-quadruplex structures indicated cations interacted with G-quadruplex structures differently from water (Figure 61). The RDF plots showed a clear attraction of positively charged cations to the negatively charged nucleic acids with a high density of cations closer to the DNA and decreased density further out. The interaction shells for the sodium ions were closer to the DNA than those for the potassium ions. This observation is expected given that within the force field the sodium ions are smaller (radius = 1.369 Å vs. 1.705 Å) (Joung and Cheatham, 2008) and thus can move closer to the G-quadruplex structures. Four cation interaction shells were observed for each structure. The first two interaction shells were well-defined while the second two interaction shells were more diffused. Density mapping analyses were performed to visualize these cation shells. The density was contoured at the reference density (Figure 62). At this contour level, two different modes of interaction were observed. The first mode was the coordination of cations within the central G-tetrads. It was observed that the cations remained positioned between the stacked G-tetrads throughout the simulation with no exchange between the G-quadruplex stem and the bulk solvent as observed in other simulations (Reshetnikov *et al.*, 2011, Akhshi *et al.*, 2012, Pagano *et al.*, 2008). The 143D model of the hTel22 sequence in sodium containing solution is noteworthy, as the starting structure contained three sodium ions placed within the plane of the G-tetrads. However, when the system was equilibrated, one of the sodium ions were ejected from the G-quadruplex stem into the surrounding solvent and the remaining two sodiums assumed positions between the G-tetrads and remained so for the course of the MD trajectory. In addition to the central coordination, diffused interaction of the cations along the grooves of the G-quadruplex

223

Footer has watermark and www.manaraa.com

structures was also observed. These "spines of cations" were not deep within the grooves as were the "spines of hydration". Instead, they were closer to the backbones and interacted with the backbone phosphate groups. When the cations were contoured at a higher level, these electrostatic interactions began to disappear and a third mode of interaction, external coordination with loops and flanking bases, was observed (Figure 63). These externally coordinated cations were critical in maintaining loop and flanking base positions. For instance, potassium binding was responsible for stabilizing the capping structures in the 2KF8 model. The binding site at the top of the G-quadruplex structure remained 100 % occupied with the potassium ion either coordinated within the plane of the triple-base cap or stacked between the planes of the inner triple-base cap and the outer double-base cap. Cation stabilization of loop structures has also been observed in previous simulations (Gray *et al.*, 2009b) as well as experimentally in the crystal structure of the human c-Kit DNA promoter sequence (Wei *et al.*, 2012). In that study, one Mg^{2+} and two K^+ ions were observed in the G-quadruplex loops and grooves in addition to the K^+ ions within the canonical central ion channel. In the crystal structure, all three external cations are believed to play a role in maintaining the c-Kit G-quadruplex structure. It should be noted that one of the K^+ ions appeared to be transient and capable of adopting one of several distinct positions while the other K^+ ion appeared to be more static suggesting the existence of a high affinity binding site. While the K^+ ions are believed to have a primary role in stabilizing the G-quadruplex structure by direct electrostatic interaction with the DNA, the Mg^{2+} ion is thought to assume a secondary role of shielding the anionic charge of the phosphate groups. In fact, it is well-known that polyanions, like DNA, can attract a shell of cations to help partially

224

neutralizes the negatively-charged backbone phosphates (Manning, 1978, Record *et al.*, 1978). Together, these findings suggest that in addition to stabilizing the G-tetrads within the G-quadruplex core, cations can also play other roles in promoting G-quadruplex formation and stability.

Figure 61. Radial distribution functions (RDF) between G-quadruplex surface heavy atoms (no hydrogen) and cations. RDFs were calculated for distances from 1.5-6.5 Å (A) and for distances from G-quadruplex surface to edge of the periodic box.

226

A

B
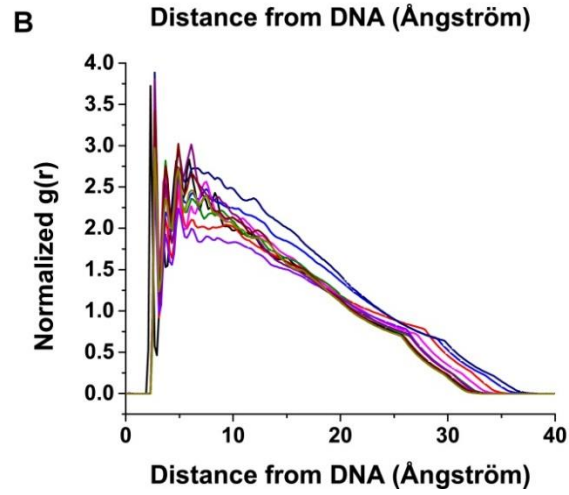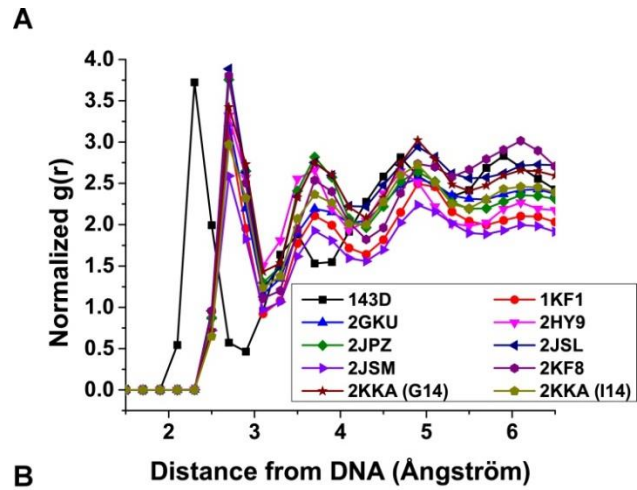
Figure 62. Pseudo-density grid maps of cations for 143D (A), 1KF1 (B), 2GKU (C), 2HY9 (D), 2JSM (E), 2JPZ (F), 2JSL (G), 2KF8 (H), 2KKA-G (I), and 2KKA-I (J). Cations density (purple) was contoured at the solubility density of sodium (6 M) for 143D and at the solubility density of potassium (4 M) for the other G-quadruplex structures. The average structure of each G-quadruplex over the trajectory is shown.

A B

C D E

F G

H I J
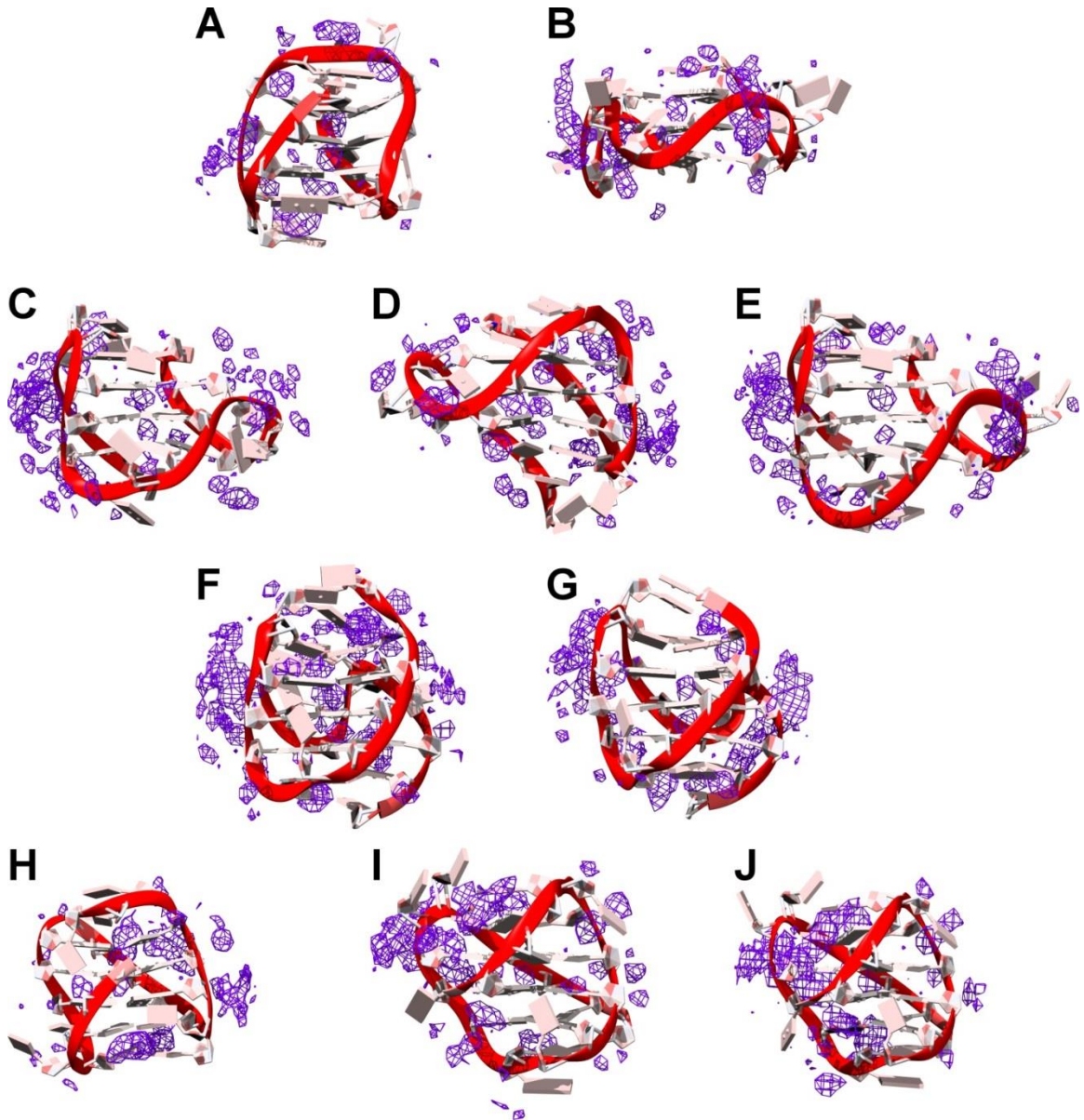
Figure 63. Pseudo-density grid maps of cations for 143D (A), 1KF1 (B), 2GKU (C), 2HY9 (D), 2JSM (E), 2JPZ (F), 2JSL (G), 2KF8 (H), 2KKA-G (I), and 2KKA-I (J). Cations density (purple) was contoured at 3X the density of sodium (6 M) for 143D and at 3X the density of potassium (4 M) for the other G-quadruplex structures. The average structure of each G-quadruplex over the trajectory is shown.
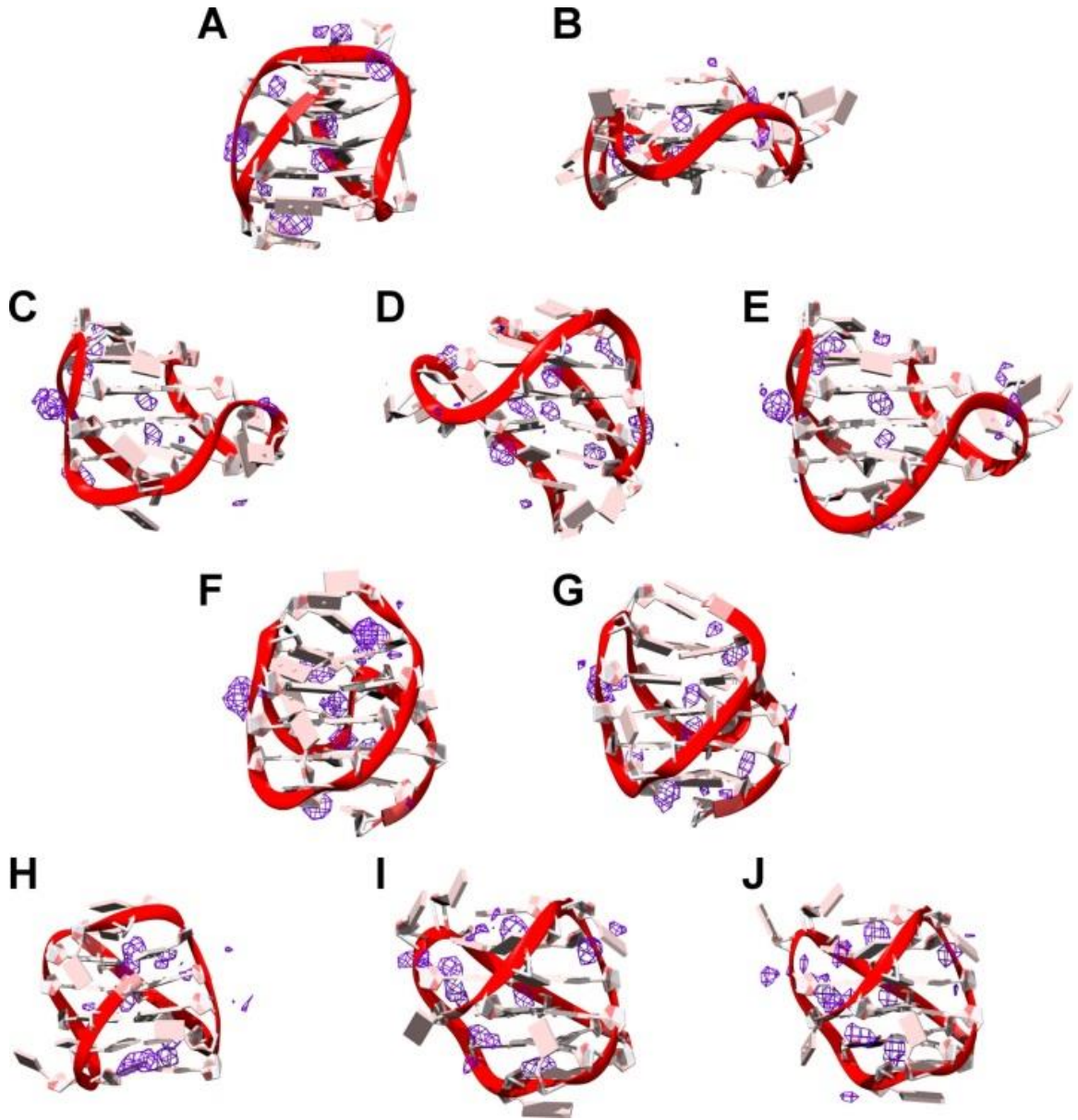
Effects of Cations Binding on Hydrodynamic Calculations

HBM was used to provide information about the number cations bound to the G-quadruplex structures. This use of HBM is based on the premise that the hydrodynamic properties are dictated by the size (volume and mass) and shape of a macromolecule in solution. When accurate experimental information is known about the volume and shape of the G-quadruplex structure, HBM can be used to estimate the mass of the G-quadruplex and by extension, the number of cations bound to it. To illustrate this, a series of HYDROPRO calculations were performed. The additions of potassium or sodium to the G-quadruplex structure were accounted for by making the appropriate increase in the molecular weight value in the HYDROPRO parameter file. The calculations were done using all three modes of HYDROPRO calculation with the default AER instead of the calibrated AER. In order to display the data for the sequences on the same scale, the number of ions bound was normalized using Equation 5:

$$\% \text{ charge saturation} = \frac{\text{\# potassium bound}}{\text{sequence length (\# bases)-1}} \qquad \text{Equation 5}$$

The results of the HYDROPRO calculations are shown in Figure 64. For the atomic-level hydrodynamic bead models, the lowest errors were observed at about 50 % charge neutralization (i.e. about 10 to 13 cations); while for the coarser residue-level models the lowest errors were observed at about 60 % charge saturation (i.e. about 12 to 15 cations). The findings agreed well with previous studies which reported 10 to 11 potassium ions bound to the hTel22 sequence in 30 mM KCl buffer (Gray and Chaires, 2011). This demonstrates that HBM may have application beyond structural prediction.

Figure 64. $s_{20,w}$ values as a function of the number of bound cations. The values for 100Δ as a function of percent charge saturation for the primary hydrodynamic model calculated using the seven G-quadruplex structures formed from the human telomere sequence. Hydrodynamic properties of G-quadruplexes were calculated using atomic-level shell-model calculation (A), residue-level shell-model calculation (B), and residue-level bead-model calculation (C).
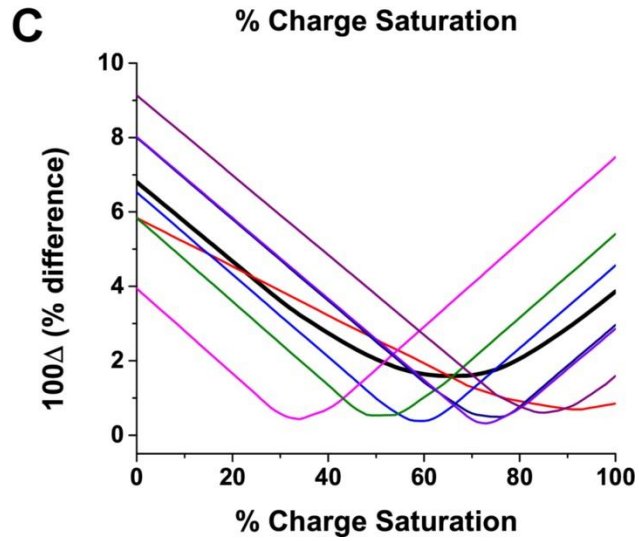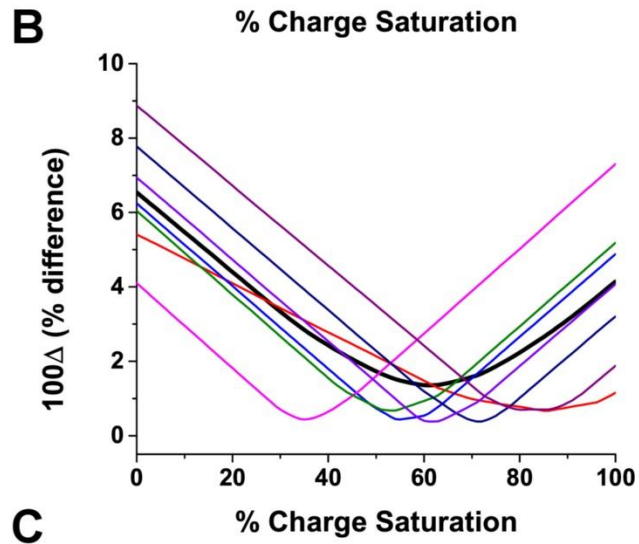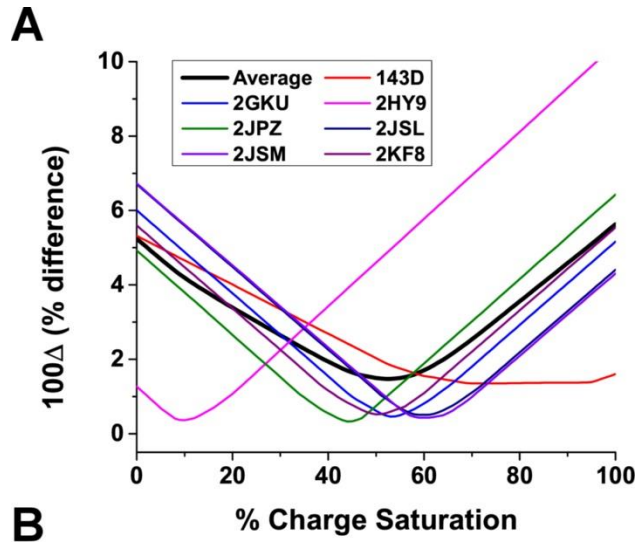
Conclusion

HBM was demonstrated to be a powerful tool for studying G-quadruplex structures when atomic-level structural representations are unavailable, ambiguous, or can't be determined experimentally. In such cases, the use of low-resolution techniques such as hydrodynamics, combined with readily accessible biophysical measurements (e.g. CD spectroscopy, fluorescent spectroscopy), can be used to obtain general information regarding the G-quadruplex structure, size, and shape. HBM can bridge low-resolution hydrodynamic measurements and high-resolution molecular modeling to provide further information regarding these structures. For example, HBM can be used to estimate the number of cations bound to the G-quadruplex structure. When applied to the study of novel G-quadruplex-forming sequences, HBM can be used to rule out structures that are not representative of the ensemble as was the case with 1KF1 whose calculated values differed greatly from the experimental values. However, hydrodynamics remains a low-resolution technique and molecular models for which calculated hydrodynamic values agreed (or differed only slightly) with experimental values will need to be confirmed with additional hydrodynamic and biophysical measurements, as was the case with the 2KKA-G and 2KKA-I models.

Another limitation of hydrodynamics and HBM is that the calculation is typically performed on one structure giving a static look at an otherwise dynamic system. HYDROPRO can only calculate hydrodynamic values for a single structure at any given time. However, as demonstrated, HBM can be used in tandem with MD simulations to provide a more dynamic representation of the macromolecule. In addition, the high-

www.manaraa.com

resolution nature of molecular dynamics can help complement the low-resolution nature of hydrodynamic measurements. As was observed with the 2KKA-G and 2KKA-I models, hydrodynamic was not able to distinguish between the two sequences as both gave rise to identical sedimentation and diffusion values. However, cluster analysis of the MD trajectories revealed that 2KKA-G model was more heterogeneous than the 2KKA-I model and demonstrated the effect of inosine substitution on selecting for a particular G-quadruplex topology. The findings demonstrated that molecular dynamics can supplement HBM in the study of G-quadruplex structures.

For G-quadruplexes it is recommended to use either the atomic-level model with shell-model calculation mode (AER = 2.19 Å) or the residue-level model with bead-model calculation mode (AER = 5.04 Å) for HYDROPRO calculations, as both modes can predict the hydrodynamic properties accurately with a reasonable estimate of the size of the macromolecule. The atomic-level model with shell-model calculation is considered the gold standard and should be used whenever the most rigorous calculation is required. The residue-level model with bead-model calculation mode has the advantage of being significantly faster (by several orders of magnitude) but with a slightly higher error in predicting hydrodynamic properties. For molecular weight, the recommended value is the molecular weight of the DNA and internal coordinating cations. For Na^+, because these ions are smaller and can fit within the G-quartet plane, the number of internal coordinating ions is equal to the number of G-quartets. For K^+, which is bigger and thus can only fit in between the planes of the G-quartets, the number of internal coordinating ions is one less than the number of G-quartets. The parameters for HYDROPRO presented can be used for hydrodynamic calculation of G-quadruplexes or can be further

optimized by fitting with additional hydrodynamic values (i.e. rotational diffusion coefficients, NMR relaxation time, intrinsic viscosity, etc.). Although this work was conducted on the human telomere sequence, the experimental approach outlined can be easily adapted for other G-quadruplex-forming sequences to propose relevant G-quadruplex structures that can be used as a basis for drug design.

CHAPTER IV

There is current interest in biomedical research to identify small molecules that can bind to and stabilize G-quadruplex structures as this represents a new strategy in anti-cancer therapeutics. We report a screening platform with the combined approach of virtual screening and biophysical measurements that was successful identifying a new G-quadruplex-interacting small molecule. Compound 1 was initially identified through molecular docking screening of the ZINC Drug-Like Databse against an *in silico* generated model of the human telomeric G-quadruplex structure. The binding of Compound 1 to G-quadruplex DNA was confirmed and characterized using fluorescent and circular dichroism spectroscopy. The screening platform was successful in identifying a new G-quadruplex-interacting small molecule and can be used to identify other G-quadruplex-interacting agents.

238

FROM CYBERSPACE TO CANCER DRUG: BIOPHYSICAL CHARACTERIZATION OF A G-QUADRUPLEX-INTERACTING SMALL MOLECULE IDENTIFIED BY VIRTUAL SCREENING

Introduction

G-quadruplexes, associated with guanine-rich nucleic acids, are non-canonical four-stranded tertiary structures comprised of stacked G-quartets (four guanine bases associating by Hoogsteen hydrogen bonding in a square planar arrangement) (Burge *et al.*, 2006, Chaires, 2010, Lane *et al.*, 2008, Lane, 2012). The most predominant type of G-quadruplex structures in human cells is thought to be unimolecular G-quadruplexes, formed from the intramolecular folding of a nucleic acid sequence containing four runs of two or more guanines each. The guanine runs make up the stacked G-quartets stem and the bases, in between and surrounding the guanine runs, form the loop and flanking structures, both of which can play critical roles in G-quadruplex stability. Within human cells, one of the major sites where G-quadruplex formation occurs at the distal 3' end of the telomere, which contains a 100-200 bases long guanine-rich single-stranded overhang consisting of the hexanucleotide repeat d[TTAGGG] (Wright *et al.*, 1997). The stabilization of these G-quadruplex structures has been proposed as a novel strategy for inhibiting telomerase (Balasubramanian and Neidle, 2009, De Cian *et al.*, 2008, Han and Hurley, 2000, Neidle and Read, 2000, Ou *et al.*, 2008, White *et al.*, 2001), the main enzyme responsible for maintaining the length of the telomere and is also activated in over 90% of cancers allowing cancer cells to achieve immortality (Shay and Bacchetti,

1997). Small molecules that stabilize telomeric G-quadruplexes *in vitro* have also been shown to inhibit telomerase activity leading to senescence and cell death in cell-based experiments (Riou *et al.*, 2002, Cuesta *et al.*, 2003) suggesting that such therapeutics might be effective in the treatment of cancer (Cuesta *et al.*, 2003, Shay and Wright, 2006, Saretzki, 2003).

In addition to the telomere, G-quadruplex formation also occurs in other areas of the human genome. In fact, over 370,000 putative G-quadruplex-forming sequences (PQSs) have been identified through bioinformatics surveys (Huppert and Balasubramanian, 2005, Huppert and Balasubramanian, 2007). Many of these sequences are conserved between human, mouse, and rat, suggesting a critical role for G-quadruplex structures in cellular functions (Verma *et al.*, 2008). PQSs are not evenly distributed throughout the genomes but, instead, are found to localize to functionally important areas, such as the promoters of several important proto-oncogenes including *c-Myc* (Ambrus *et al.*, 2005), *c-Kit* (Hsu *et al.*, 2009, Phan *et al.*, 2007a), *Bcl-2* (Dai *et al.*, 2006), *VEGF* (Sun *et al.*, 2005), and *HIF-1α* (De Armond *et al.*, 2005). In contrast, the occurrences of PQSs in association with tumor suppressor genes tend to be much lower (Eddy and Maizels, 2006). G-quadruplex formation appears to play a critical role in the regulation of oncogene transcription (Balasubramanian *et al.*, 2011), for example *c-Myc*, an oncogene whose overexpression is strongly associated with the development of several types of cancer including but not limited to breast, lung, prostate and hematological cancers (Nesbit *et al.*, 1999, Nilsson and Cleveland, 2003). Stabilization of the G-quadruplex structures in the nuclease hypersensitivity element III$_1$ (NHE-III$_1$) region of the *c-Myc* promoter by the small molecule TMPyP$_4$ inhibited transcription in a

luciferase assay (Siddiqui-Jain *et al.*, 2002). As the NHE-III region is the primary control of *c-Myc* transcription (Patel *et al.*, 2007), G-quadruplex formation wass proposed as a mechanism of regulating *c-Myc* expression (Siddiqui-Jain *et al.*, 2002). In addition to controlling transcription, G-quadruplex formation can also affect gene expression at the translation level. For example, the stabilization of G-quadruplex structures located in the 5'-untranslated region of the *NRAS* mRNA by small molecule inhibitors was accompanied by decreased translational efficiency (Bugaut *et al.*, 2010). Taken together, these findings suggest that G-quadruplex-based small-molecule inhibitors might be effective as anti-cancer therapies.

The potential of G-quadruplex-based anti-cancer therapies has led to studies aimed at discovering small molecules that selectively bind these structures. Small molecules can interact with G-quadruplexes by one of three possible binding modes: 1) groove binding, 2) intercalation between adjacent G-quartets, and 3) end-pasting, where the ligand is bounded on one side by a G-quartet and on the other side by the loop structures (Pan and Zhang, 2009). There is interest in the discovery of small molecules that bind G-quadruplexes by end-pasting as this binding mode is believed to confer selectivity for G-quadruplexes over other nucleic acid structures (e.g. G-quadruplex vs. duplex), as well as discriminates between G-quadruplex structures formed from different sequences (e.g. telomeric G-quadruplex vs. promoter G-quadruplex) by taking into account both G-quartet and loop interactions. Virtual screening using molecular docking methods has been proposed as an approach to identify such small molecules (Dailey *et al.*, 2009). Originally developed, parameterized, and optimized for protein targets, molecular docking software are also suitable for probing the interaction of small

241

molecules with nucleic acids (Dailey *et al.*, 2009, Holt *et al.*, 2008, Holt *et al.*, 2009, Holt *et al.*, 2011). Two docking programs, Surflex (Tripos International, St. Louis, MO) and Autodock (The Scripps Research Institute, La Jolla, CA), were able to rationalize known small molecules/DNA interaction by accurately reproducing the crystallographic poses of intercalating and groove binding drugs bound to duplex DNA (Holt *et al.*, 2008). More recently, Surflex was utilized in a virtual screening approach to discover new triple-helical DNA intercalating agents (Holt *et al.*, 2009). In that study, the *in silico* library of chemicals was prescreened using molecular similarity search and Surflex was employed to dock the most similar compounds into the DNA target and identify the compound with the highest binding affinity. Prescreening using molecular similarity search or other pharmarcophore-based methods is a valid strategy to identify new lead compounds for drug development and such methodologies have been used previously with G-quadruplex-interacting small molecules (Castillo-Gonzalez *et al.*, 2013, Cosconati *et al.*, 2012). A limitation to this strategy is that a potential hit could be improperly eliminated from consideration if its structure is highly dissimilar from the known agent. Therefore, it remains unknown if molecular docking software can be used without prescreening to identify new G-quadruplex-interacting small molecules with novel chemical scaffolds.

Accordingly, the success of such receptor-based drug discovery approaches depends on a critical understanding and appreciation of the structural complexity associated with the molecular target of interest. The folding of a PQS from its single-stranded structures into the unimolecular G-quadruplex form can be highly polymorphic with hundreds of possible G-quadruplex topologies varying in the number of G-quartets, strand directions (i.e. parallel, antiparallel, mixed), loop combinations (i.e. lateral,

242

diagonal, double chain-reversal), and glycosidic bond angles (i.e. *syn* or *anti*) (Lane *et al.*, 2008). Often, a PQS exists in solution as a mixture of multiple G-quadruplex species and the equilibrium between species in solution can be influenced by experimental conditions (e.g. choice of annealing protocol) (Le *et al.*, 2012), by the choice of cation and its concentration in buffer (Gray *et al.*, 2009a, Gray *et al.*, 2009b, Le *et al.*, 2012), by the presence of biological molecules (e.g. sugar, proteins) (Sannohe and Sugiyama, 2001) or co-solvents (e.g. acetonitrile, PEG) (Xue *et al.*, 2007, Miller *et al.*, 2010, Buscaglia *et al.*, 2013), or by the use of divalent versus monovalent cations (Blume *et al.*, 1997, Miyoshi *et al.*, 2001). An example highlighting the structural polymorphism associated with G-quadruplex structures is the hTel22 sequence, AGGGTTAGGGTTAGGGTTAGGG, which is often used as an *in vitro* model to examine small molecules/G-quadruplex interaction (Wang and Patel, 1993, Parkinson *et al.*, 2002). In the presence of sodium, the hTel22 G-quadruplex structures exist predominantly as a single species, an antiparallel "basket" topology (Wang and Patel, 1993) consisting of three stacked G-quartets and a lateral-diagonal-lateral loop combination. Whereas, in the presence of potassium (the more relevant intracellular cation), hTel22 exists as a mixture of G-quadruplex species. The first high-resolution structure of hTel22 in potassium was a crystal structure, which revealed a parallel "propeller" topology consisting of three stacked G-quartets and three double chain-reversal loops (Parkinson *et al.*, 2002). Subsequent studies determined that the parallel topology is not the major species in solution (Li *et al.*, 2005, Buscaglia *et al.*, 2013, Hänsel *et al.*, 2011) and accounts for only about 14% of the total ensemble (Buscaglia *et al.*, 2013). However, under the effect of dehydration (Miller *et al.*, 2010) or in the presence of polyethylene glycol (PEG) (Buscaglia *et al.*, 2013) (both factors

243

contributing to crystallizing environment), the parallel topology can be enriched to become the major species. In addition to changing the experimental conditions, sequence modification is another commonly accepted approach to influence the equilibrium of G-quadruplex species in solution with the goal of enrichment of a single species for NMR structure elucidation (Dai *et al.*, 2008, Yang and Okamoto, 2010). In sequence modification, a PQS can be truncated or elongated by adding flanking bases, and non-canonical bases (Sagi, 2013) can be incorporated to select for a particular topology. In the case of hTel22, addition of flanking bases have resulted in sequences that are predominantly of the mixed "hybrid-1" topology (three stacked G-quartets with a double chain-reversal loop followed by two lateral loops) (Luu *et al.*, 2006, Dai *et al.*, 2007b, Phan *et al.*, 2007b), sequences that are predominantly of the mixed "hybrid-2" topology (three stacked G-quartets with two lateral loops followed by a double chain-reversal loop) (Dai *et al.*, 2007a, Phan *et al.*, 2007b), and sequences that are predominantly of an antiparallel "basket" topology (similar to the sodium form but with two stacked G-quartets) (Lim *et al.*, 2009, Zhang *et al.*, 2010b).

These strategies for reducing G-quadruplex structural polymorphism are utilized with the assumption that these means enrich for a member of the ensemble of species originally formed by the parent sequence. The unintended consequence, however, might be an unpredictable perturbation of the system and the selection of a topology that may not be representative of the original ensemble of topologies and can have significant implications on which structures are suitable for drug discovery and can be claimed as "biologically relevant" (Lane *et al.*, 2008). The crystal structure of hTel22 is often used to rationalize G-quadruplex/small molecules interaction (Luedtke, 2009, Neidle, 2009)

244

but it is the "hybrid" forms, rather than the "propeller" form, that are thought to be predominant in solution and are hypothesized to be the more "relevant" structures *in vivo* (Dai *et al.*, 2007a, Dai *et al.*, 2007b, Phan, 2010, Dai *et al.*, 2008, Hänsel *et al.*, 2011, Hansel *et al.*, 2013a, Hansel *et al.*, 2013b). In a recently published NMR structure of the G-quadruplex structure in the "hybrid-1" form bound to the small-molecule telomerase inhibitor telomestatin (Chung *et al.*, 2013), it was revealed that the end-pasting site for the "hybrid" form differs significantly from the end-pasting sites of the "propeller" form. In the "hybrid" form, the binding of small molecule to the G-quadruplex structure was characterized by π-π interaction between the G-quartet, the aromatic core of the compound, and the loop structure, whereas in the "propeller" form π-π interaction was observed only between the compound and the G-quartet not the loops. Furthermore, in the "propeller" form, the side chain of the compound was observed to interact with the grooves formed by adjacent strand of the G-quartet stem, whereas in the "hybrid" form, the side chain interacted more with the loop structures. As a result, the choice of representative structure of the PQS is critical in virtual screening studies using receptor-based approaches such as molecular docking.

In the current work, the successful development of a high throughput *in silico* molecular docking platform that identified a small molecule with anti-cancer properties and binds to G-quadruplex structures is reported. This compound possesses a novel DNA binding chemical scaffold, naptholphenoxazine (Figure 65), that have not been examined in the literature. As the representative structure for the G-quadruplex, a "hybrid-1" structure which contains the hTel22 sequence (PDB: 2HY9) was utilized. An end-pasting binding site was introduced between the terminal G-quartet and the loop and flanking

structures at the 5' end of the G-quadruplex structure. Compound 1 was identified from the virtual screening as a potential G-quadruplex-interacting small molecule. Using spectroscopic methods, the binding affinity, binding stoichiometry, and binding mode were determined for Compound 1 interaction with G-quadruplex structures. In addition, *in silico* molecular models of all possible end-pasting sites of the "hybrid-1" and "hybrid-2" forms of hTel22 PQS were generated and the newly discovered Compound 1 was docked into the sites to assess interaction between nucleic acids and small molecule. Finally, Compound 1 was subjected to the NCI-60 DTP Human Tumor Cell Line Screen and discovered to possess good inhibitory properties against cancer cells. Overall, the virtual screening platform reported was successful and computationally plausible for screening millions of small molecules against a nucleic acid target and can be used to identify new small molecules which interact with G-quadruplex structures in a predicted mechanism.

Materials and Methods

Preparation of Nucleic Acids Target and in Silico Libraries for Molecular Docking

For virtual screening, the G-quadruplex structure (PDB ID: 2HY9) was downloaded from the Protein Data Bank in PDB file format. The end-pasting site was introduced between terminal G-quartet and the loop structures near the 5' end of the oligonucleotide as previously described (Read *et al.*, 2001). Briefly, the phosphate backbones were broken and the loop and flanking structures were separated from the terminal G-quartet structure. To expose the end-pasting site, a virtual ligand consisted of

246

a planar, aromatic, small molecule was constructed that would stack well upon the terminal G-quartet (Figure 66). Using Macromodel (Schrödinger, Portland, OR), the virtual ligand was initially positioned between the terminal G-quartet and loop region. After the phosphate backbones were reconnected, the G-quadruplex/ligand complex was minimized using a Steepest Descent algorithm for 1000 iterations while holding the ligand fixed. A second round of minimization was performed using the Polak Ribier Conjugate Gradient algorithm for 500 steps with the nucleotides comprising the end-pasting site including the terminal G-quartet and loop nucleotides designated as flexible and the remaining nucleic acid bases designated fixed. For the hTel22 molecular docking experiments, the representative hybrid-1 and hybrid-2 G-quadruplex structures (PDB ID: 2HY9 and 2JPZ, respectively) were downloaded from the Protein Data Bank in PDB file format. For each structure, the first two residues and the last two residues were removed from the PDB files to create models of the hTel22 sequence, respectively. Each of the four possible end-pasting binding sites was created using the procedure previously described in this paragraph. *In silico* small molecules were downloaded from the ZINC (Irwin and Shoichet, 2005) database 2008 "Drug-Like" dataset, "Reference" subset, which has been classified based on adherence to adherence to Lipinski's "rule of fives" for increased oral bioavailability (Lipinski *et al.*, 2001). For the 2008 version of the ZINC database, 6.6 million compounds fell under this classification. The small molecules were downloaded and used without any further modification.

Figure 65. (A) A G-quadruplex structure (PDB ID: 2HY9) that contains the hTel22 sequence with a representative end-pasting site that was initially used for virtual screening with Surflex and (B) the newly discovered small molecule, Compound 1, from the *in silico* virtual screening experiments.

A

B
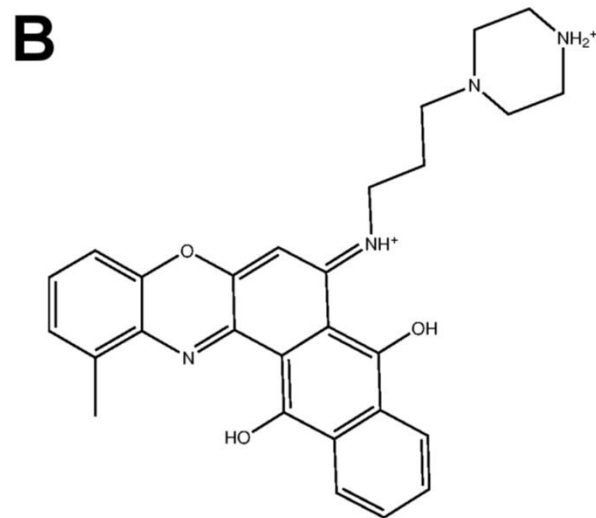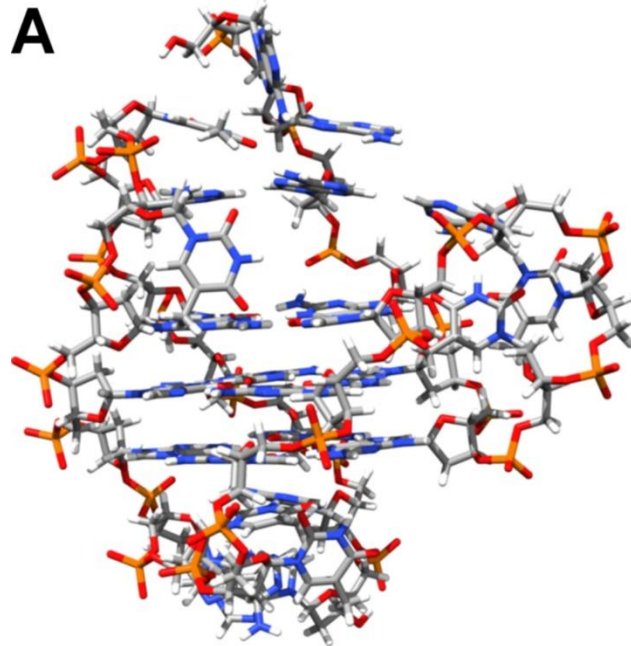
Figure 66. The structure of the quaterpurine used for *in silico* construction of G-quadruplex end-pasting sites.
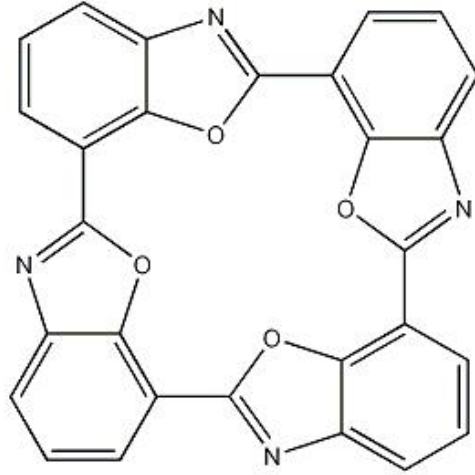
Molecular Docking Experiments

All molecular docking experiments were performed on the Brown Cancer Center Modeling Facility 440 core server consisting of 2.66 GHz Intel(R) Xeon(R) E5430 processors using Surflex and Autodock docking parameters previously described (Holt *et al.*, 2008). For molecular docking with Surflex, the end-pasting cavity was specified with a ligand-based approach by using the bound virtual ligand (Figure 66) to generate a Surflex "protomol" which guides the molecular docking of the *in silico* ligands to the end-pasting site. The "protomol" was constructed by altering the "proto_bloat" and "proto_thresh" functions and visualized in Sybyl (Tripos International, St. Louis, MO) to ensure reasonable interactions in the end-pasting site. The significance of the protomol and the Surflex docking and scoring functions have been described in detail previously (Jain, 2003). Briefly, the "protomol" consists of a series of small chemical fragments that model important forces in the nucleic acid pocket: steric effects and hydrogen bond acceptors and donors. For each ligand being docked, the molecule is fragmented, aligned against the protomol, and subsequently scored based on the interactions in the binding site. For molecular docking with Autodock, the ligand and G-quadruplex structures were converted from the MOL2 format to PDBQT format using the Python scripts included with Autodock. For each G-quadruplex structure, a grid map with a grid spacing of 0.375 Å was generated centered on the end-pasting site. The parameter for grid generation is shown (Table 10). A genetic algorithm, the Lamarkian Genetic Algorithm, is used to assess the interactions of the ligand with the pre-calculated energy grids until typically the specified number of energy evaluation was reached. The final top "pose" returned by Autodock is the computed lowest energy docked structure of the ligand.

Table 10. Autodock Parameters for Hybrid-1 and Hybrid-2 Docking

Autodock Target	Dimensions of Grid (X, Y, Z)	Grid Center
Hybrid-1 End-Paste Site 1	$66 \times 64 \times 40$	$-1.644 \times 6.950 \times -0.460$
Hybrid-1 End-Paste Site 2	$66 \times 64 \times 40$	$-0.491 \times 5.057 \times -0.484$
Hybrid-2 End-Paste Site 1	$66 \times 64 \times 40$	$0.642 \times 7.237 \times -0.239$
Hybrid-2 End-Paste Site 2	$66 \times 64 \times 40$	$0.298 \times 4.392 \times 0.048$

Preparation of Oligonucleotides and Reagents

The DNA sequences used in this study are listed in Table 11. Unlabeled oligonucleotides were purchased from Integrated DNA Technologies (Coralsville, IA). FRET-labeled oligonucleotides containing fluorescein (FRET donor) at the 5' end and tetramethylrhodamine (FRET acceptor) at the 3' end were purchased from Sigma-Aldrich (St. Louis, MO). Lyophilized DNA was rehydrated in ddH$_2$O at a stock concentration of 1 mM for the unlabeled oligonucleotides and 250 μM for the FRET-labeled oligonucleotides. The DNA was quantified using a Nanodrop 2000 instrument (Thermo Scientific, Wilmington, DE) using molar extinction coefficients (ε) calculated via the nearest-neighbor method (Table 11). Prior to experiments, the stock oligonucleotides were diluted to the desired concentration using the appropriate buffer. The diluted oligonucleotides were annealed by heating in a boiling water bath for 10 minutes and slowly cooled overnight to room temperature. Compound 1 with ZINC ID 20263704 were purchased from Vitas-M Laboratory (Stock #: STK554084, Moscow, Russia). Thiazole orange (TO) and 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide (MTT) were purchased from Sigma-Alrich. Stock solutions were made in DMSO to concentrations of 1 mM (Compound 1) and 10 mM (TO). MTT stock solution was made at 5 $^{mg}/_{mL}$ in ddH2O, syringe filtered through 0.2 μm filters, and stored at 4°C away from light.

255

Table 11. Oligonucleotide Sequences Used in this Study and Their Molar Extinction Coefficients

Name	Sequence	ε (M^{-1}·cm^{-1})
AT duplex	5'-ATA TAT ATC CCC ATA TAT AT-3'	205,700
GC duplex	5'-GCG CGC GCT TTT GCG CGC GC-3'	166,300
ALT duplex	5'-AAA AAA AAC CCC TTT TTT TT-3'	191,300
hTel duplex	5'-GGG TTA GGG TTT TCC CTA ACC C-3'	202,200
AT triplex	5'-AAA AAA AAC CCC TTT TTT TTC CCC TTT TTT TT-3'	284,900
GC triplex	5'-CCC CCC CCT TTT GGG GGG GGT TTT CCC CCC CC-3'	261,600
hTel22 G-quadruplex	5'-AGG GTT AGG GTT AGG GTT AGG G-3'	228,500
cMyc22 G-quadruplex	5'-TGA GGG TGG GTA GGG TGG GTA A-3'	228,700
2GKU G-quadruplex	5'-TTG GGT TAG GGT TAG GGT TAG GGA-3'	244,300
2HY9 G-quadruplex	5'-AAA GGG TTA GGG TTA GGG TTA GGG AA-3'	278,200
2JSM G-quadruplex	5'-TAG GGT TAG GGT TAG GGT TAG GG-3'	236,500
2JPZ G-quadruplex	5'-TTA GGG TTA GGG TTA GGG TTA GGG TT-3'	261,200
2JSL G-quadruplex	5'-TAG GGT TAG GGT TAG GGT TAG GGT T-3'	253,100
2KF8 G-quadruplex	5'-GGG TTA GGG TTA GGG TTA GGG T-3'	223,500
2KKA G-quadruplex	5'-AGG GTT AGG GTT AGG GTT AGG GT-3'	237,000
hTel-dimer	5'-TTA GGG TAA GGG TTA GGG TTA GGG	509,500

| G-quadruplex | TTA GGG TTA GGG TTA GGG TTA GGG TT-3' | |

Differential Scanning Fluorometry (DSF)

DSF experiments were performed on a StepOnePlus Real-Time PCR System (Applied Biosystems, Carlsbad, CA) adapted for use in thermal melting experiments. In the first melting experiment to assess Compound 1 binding to FRET-labeled DNA, the final concentration of FRET-labeled oligonucleotides was fixed at 0.25 µM. Compound 1 was added to the FRET-labeled DNA at different final concentrations ([Compound 1] = 0.25, 0.50, 1, 2.5, 5, and 12.5 µM). In the second melting experiment to measure the ability of unlabeled DNA to compete with FRET-labeled DNA for Compound 1 binding, the final concentrations of FRET-labeled DNA and Compound 1 were held fixed at 0.1 and at 10µM, respectively. Unlabeled oligonucleotides were added to the FRET-labeled DNA and Compound 1 mixture at different final concentrations ([oligonucleotide] = 5, 10, 25, 50, 100 µM). All DSF experiments were performed in buffer (10 mM tetrabutylammonium phosphate, 1 mM EDTA acid form, 30 mM KCl, 10% DMSO, pH 7.0). The temperature range for melting was 20-98°C with 0.2°C steps. Fluorescence was monitored using a fluorescence filter that quantifies emission at 520 nm. Data were exported and analyzed using Origin software (OriginLab, Northampton, MA). The melting temperature (T_M) was determined by the location of the peak in first derivative plot of the melting curve.

Fluorescent Emission Spectroscopy (FE)

FE experiments were performed in quadruplicate on a Safire2 96-well microplate reader (Tecan US, Morrisville, NC) using the following instrument parameters: excitation wavelength 650 nm, emission scanning range 670-850 nm, emission step size 1 nm,

excitation bandwidth 9 nm, emission bandwidth 9 nm, integration time 200 μs, gain 200, and number of scans 16. For the Job plot experiments, Compound 1 and DNA samples were diluted from stock concentration to 20 μM in low salt buffer (10 mM tetrabutylammonium phosphate, 1 mM EDTA acid form, 30 mM KCl, 10% DMSO, pH 7.0). After the DNA has been annealed, Compound 1 and DNA were mixed together in different ratios to obtain 24 samples with varying molar fractions of Compound 1 (molar fraction = 0.04 to 0.96). For titration experiments, the final concentration of Compound 1 was held fixed at 1 μM and DNA were added at different concentrations (log[DNA] = -9 to -3.5 with 0.25 difference). Titration experiments were conducted either in same low salt buffer as the Job plot experiments or in high salt buffer (10 mM tetrabutylammonium phosphate, 1 mM EDTA acid form, 200 mM KCl, 10% DMSO, pH 7.0). Data were exported and analyzed using Origin. The affinity constant was determined by plotting fluorescent reading versus DNA concentration at the max wavelength of the fluorescent binding differential spectra. The data were fitted using non-linear regression to the following set of equations:

$B = x + L_T + (1/K_A)$;

$C = (B - sqrt(B^2 - 4 \cdot x \cdot L_T))/2$;

$y = (1/L_T) \cdot ((F_I \cdot (L_T - C)) + (F_F \cdot C))$;

where x is the DNA concentration, y is fluorescent reading, L_T is total ligand concentration, K_A is the affinity constant of ligand to the DNA sequence, F_I is fluorescent reading of Compound 1 in the unbound form, and F_F is the fluorescent reading of Compound 1 in the bound form.

260

Induced Circular Dichroism Spectroscopy (ICD)

ICD experiments were performed on a J-810 spectropolarimeter (Jasco, Easton, MD) using the following instrument parameters: scanning range 450-725 nm, data pitch 1 nm, bandwidth 1 nm, Digital Integration Time 8 seconds, scanning speed of 50 $^{nm}/_{min}$, and number of scans 4. ICD experiments were carried out for Compound 1 (50 μM) alone, Compound 1 (50 μM) with hTel22 (444 μM), and Compound 1 (50 μM) with cMyc22 (452 μM). All experiments were carried out in buffer (10 mM tetrabutylammonium phosphate, 1 mM EDTA acid form, 30 mM KCl, 10% DMSO, pH 7.0). Data were exported and analyzed using Origin.

Thiazole Orange Fluorescent Intercalator Displacement Assay (TO-FID)

TO-FID experiments were performed in duplicate on a Safire2 96-well microplate reader using the following instrument parameters: excitation wavelength 500 nm, emission scanning range 510-750 nm, emission step size 1 nm, excitation bandwidth 5 nm, emission bandwidth 5 nm, integration time 200 μs, gain 125, and number of scans 16. FID experiments were carried out at the following concentrations: TO (1 μM), DNA (2 μM), and Compound 1 (5 μM). The percentage of TO displacement (%FID) was calculated from the fluorescent emission intensity at 527 nm using Equation 6:

$$\%FID = \left(1 - \frac{F_{corrected}}{F0_{corrected}}\right) \times 100 \qquad \text{Equation 6}$$

where $F_{corrrected}$ is the corrected fluorescence of TO bound to the oligonucleotide in the presence of compound and $F0_{corrected}$ is the corrected fluorescence of TO bound to the oligonucleotide in the absence of compound. To determine $F_{corrrected}$, the fluorescent readings of the well containing TO alone and the well containing Compound 1 with DNA

261

only were subtracted from the fluorescent reading of the well containing TO, DNA, and Compound 1. To determine F0$_{corrected}$, the fluorescent reading of the well contain TO alone was subtracted from the sample containing TO and DNA. TO-FID experiments involving AT triplex DNA were performed in Buffer A (6 mM K$_2$HPO$_4$, 4 mM KH$_2$PO$_4$, 15 mM KCl, 1 mM MgCl$_2$, 1% DMSO, pH 7.2). TO-FID experiments involving GC triplex DNA were performed in Buffer B (1 mM K$_2$HPO$_4$, 9 mM KH$_2$PO$_4$, 15 mM KCl, 1% DMSO, pH 6.2). All other TO-FID experiments were performed in Buffer C (6 mM K$_2$HPO$_4$, 4 mM KH$_2$PO$_4$, 15 mM KCl, pH 7.2). Data were exported and analyzed using Origin.

Cell Culture and MTT Assay

Cancer cell cultures were grown in Dulbecco's modified Eagle's media (DMEM) supplemented with 10% heat-inactivated fetal bovine serum (FBS, 15 minutes at 65°C) and 1% penicillin/streptomycin (10,000 units/ml). Upon passage, cells were plated in media in 96-well plates (6,000 cells per well) and incubated overnight before treatment with Compound 1. After 72 hours, or when untreated cells reach confluence, MTT was added to each well and incubated at 37°C for 4 hours. The cells were lysed with lysis buffer (10% sodium lauryl sulfate, 0.01 N HCl) and MTT was solubilized overnight. Absorbance at 570 nm (A$_{570}$) was determined using a Safire2 96-well microplate reader. Data were exported and analyzed using Origin. The percentage of cell growth (%Growth) was calculated using Equation 7:

$$\%Growth = \left(\frac{A_C - A_0}{A_U - A_0}\right) \times 100 \qquad \text{Equation 7}$$

for positive growth and Equation 8:

262

$$\%Growth = \left(\frac{A_C - A_0}{A_0}\right) \times 100 \qquad \text{Equation 8}$$

for negative growth. A_0 is the A_{570} value obtained from a separate experiment where the cells are treated with MTT and lysed after the initial overnight incubation. A_C is the A_{570} values of wells containing cells treated with Compound 1. A_U is the A_{570} value of wells containing untreated cells.

Results and Discussion

Creation of a Representative Binding Site for Virtual Screening of G-Quadruplex-Interacting Small Molecules

The identification of new G-quadruplex-interacting small molecules is of great interest as these molecules may serve as lead compounds in the development of novel anti-cancer drugs (Cuesta *et al.*, 2003, Shay and Wright, 2006, Saretzki, 2003). In addition to the traditional binding modes of intercalation and groove binding, small molecules can also interact with G-quadruplex structures by end-pasting. In end-pasting, the small molecule stacks onto the surface of the terminal G-quartet and interact with both the terminal G-quartet and the G-quadruplex loops and flanking structures. Because different PQS can adopt different folding topologies and therefore different loops and flanking structures, small molecules binding by end-pasting can potentially be sequence selective and can discern between G-quadruplex structures from different areas of the genome. With few published NMR solution structures of G-quadruplexes with small molecule bound by the end-pasting mechanism, the alternate approach utilized in the current work involved an *in silco* end-pasting site introduced between the terminal G-

263

quartet and the loop and flanking structures near the 5' end of a human telomeric PQS (Figure 65). While this approach has been used previously to rationalize binding of known small molecules to G-quadruplex structures (Read *et al.*, 2001), this is the first instance where such of such approach is for the identification of new G-quadruplex interacting small molecules. After the completion of all works presented in this study, an NMR structure of a human telomeric PQS with telomestatin bound by end-pasting was published by Chung *et al.* (2013). The binding pocket of the published NMR structure agreed with the *in silico* generated structure (Figure 67). The findings suggested that the *in silico* model was suitable for the purpose of screening millions of compound to identify new structures. Importantly, the utilized approach demonstrated a strategy for circumventing receptor-based virtual screening problems when the target site is not entirely available from the known X-ray crystal or NMR solution structures.

264

Figure 67. Overlay of *in silico* generated end-pasting site (green) with end-pasting site reported by NMR (yellow)
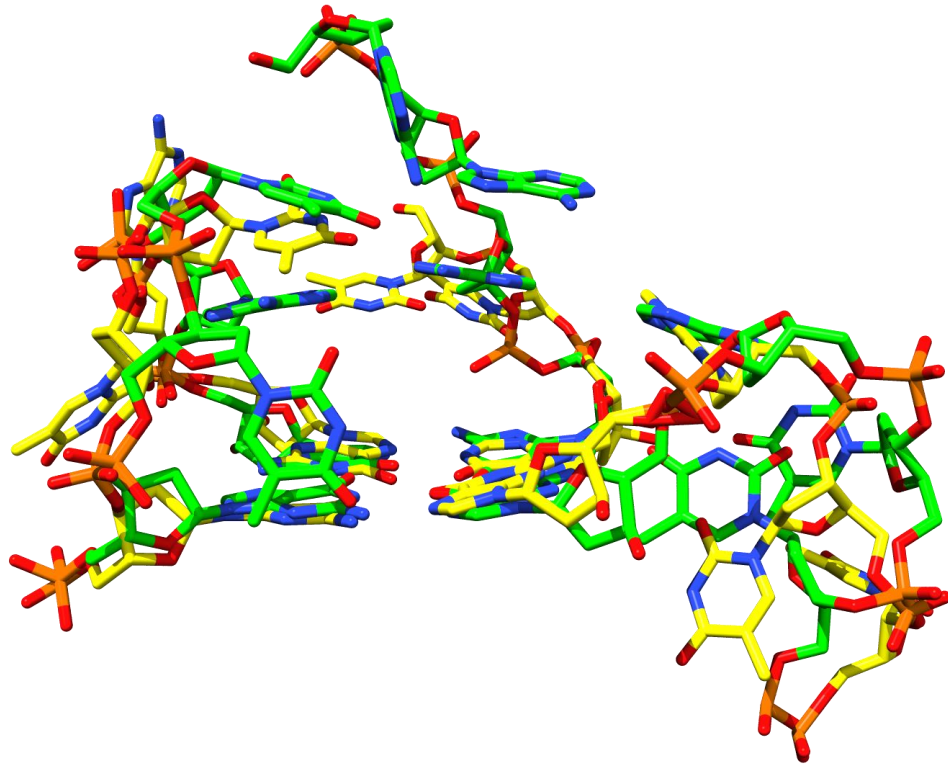
Identification of Compound 1 as a G-Quadruplex-Interacting Small Molecule

A high-throughput virtual screening effort using Surflex to dock the "Drug-Like" subset of the 2008 ZINC Database into the representative G-quadruplex structure identified Compound 1 as a potential G-quadruplex-interacting small molecule (Figure 65). For an initial assessment of whether Compound 1 interact with G-quadruplexes, melting experiments were performed with nine different FRET-labeled PQS from the human telomere to determine whether Compound 1 can thermally stabilize known G-quadruplex structures. Melting experiments were carried out for two PQS from the *c-Myc* promoter region (cMyc22 and Pu27), however, the melting temperatures (T_M) for these sequences were greater than 90°C, too close to the experimental limit of 100°C to be analyzed. The T_M of the DNA were observed to increase upon the addition of Compound 1 (Figure 68). The changes in melting temperature (ΔT_M) were calculated by subtracting the T_M of DNA alone from the T_M of the DNA in the presence of Compound 1. At low relative concentrations of Compound 1 compared to FRET-labeled DNA ($1\times$ to $5\times$), all nine sequences had similar ΔT_M. There was a dose-dependent increase in the ΔT_M, which reached a maximum of 5 to 20°C at a relative concentration of $50\times$ Compound 1. The dimer sequence (5°C), which contains 8 runs of guanines and is thought to form two contiguous G-quadruplex structures, had a much lower max ΔT_M compared to the other eight sequences (10-20°C), which only contain 4 runs of guanines each and form monomer G-quadruplex structures. The results suggested that the linking of G-quadruplex structures together in series resulted in a loss of binding sites and could have implications for targeting these structures in the telomere which contains a 100-200 bases single-stranded segment where multiple contiguous G-quadruplex structures are thought

267

to form. In addition, it should be noted that the salt concentration used in the melting experiments (30 mM KCl) is much lower than the physiological salt concentration inside the cells (about 140-200 mM). The salt concentration was chosen for practical purpose as at higher salt concentrations the T_M of the G-quadruplex structures (about 80-85°C) would be too close to the experimental limit of 100°C for any meaningful analysis. Overall, the data from the melting experiments demonstrated that Compound 1 interacts with G-quadruplex structures and the sequence-dependent max ΔT_M indicated that Compound 1 might have different preferences for different G-quadruplex topologies and might be able to distinguish between G-quadruplex structures formed from different PQS.

268

Figure 68. Effect of Compound 1 on the melting temperature of the FRET-labeled human telomeric G-quadruplex structures. FRET-labeled DNA sequences used were hTel22 $d[AG_3(T_2AG_3)_3]$, 2GKU $d[T_2G_3(T_2AG_3)_3A]$, 2HY9 $d[A_3G_3(T_2AG_3)_3A_2]$, 2JSM $d[TAG_3(T_2AG_3)_3]$, 2JPZ $d[(T_2AG_3)_4T_2]$, 2JSL $d[TAG_3(T_2AG_3)_3T_2]$, 2KF8 $d[G_3(T_2AG_3)_3T]$, 2KKA $d[AG_3(T_2AG_3)_3T]$, and hTel-dimer $d[(T_2AG_3)_8T_2]$. Data plotted are the average of experiments performed in triplicate. Error bars represent plus/minus one standard deviation.
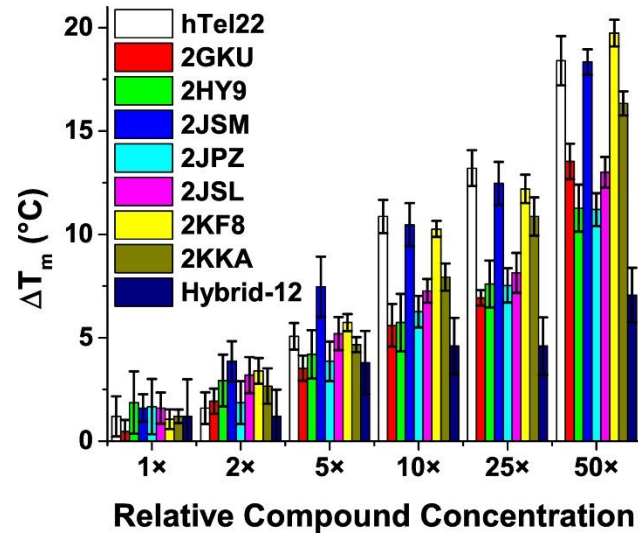
Compound 1 Selectively Binds G-quadruplex Structures over Duplex Structures

To determine if Compound 1 prefers G-quadruplex structures over other DNA tertiary structures, melting experiments were performed to assess whether unlabeled DNA can compete with the FRET-labeled DNA for binding with Compound 1. The addition of unlabeled hTel22 to mixture of FRET-labeled hTel22 and Compound 1 showed a concentration dependent decrease in ΔT_M (Figure 69). At high excess of unlabeled DNA compared to Compound 1 ($5\times$ and $10\times$), the unlabeled DNA was able to bind all Compound 1 in solution and the T_M for the FRET-labeled DNA was similar to T_M of FRET-labeled DNA in the absence of Compound 1. The findings demonstrated that the increase in T_M observed with the FRET-labeled PQS in the previous melting experiments were from Compound 1 stabilization of the G-quadruplex structure and not from Compound 1 stabilization of the FRET-pair formation. Compared with the unlabeled hTel22, the addition of unlabeled AT-rich duplex DNA did not affect the T_M while the addition of unlabeled GC-rich duplex DNA showed only a moderate decrease in T_M. This observation indicated that Compound 1 preferred G-quadruplex structures over duplex structures. In addition, it appears that Compound 1 slightly preferred GC-rich DNA over AT-rich DNA, providing a clue that Compound 1 might bind to G-quadruplex structures by either intercalating or end-pasting as previous investigations of known intercalating agents have demonstrated similar preferences (Müller and Crothers, 1975, Müller *et al.*, 1975, Müller and Gautier, 1975).

271

Figure 69. Effect of unlabeled competitor DNA on the ability of Compound 1 to thermally stabilize the FRET-labeled hTel22 G-quadruplex structures. Unlabeled DNA sequences used were hTel22 d[AG3(T2AG3)3], AT duplex d[(AT)5C4(AT)5], and GC duplex d[(GC)5T4(GC)5]. Data plotted are the average of experiments performed in triplicate. Error bars represent plus/minus one standard deviation

Relative Competitor DNA Concentration
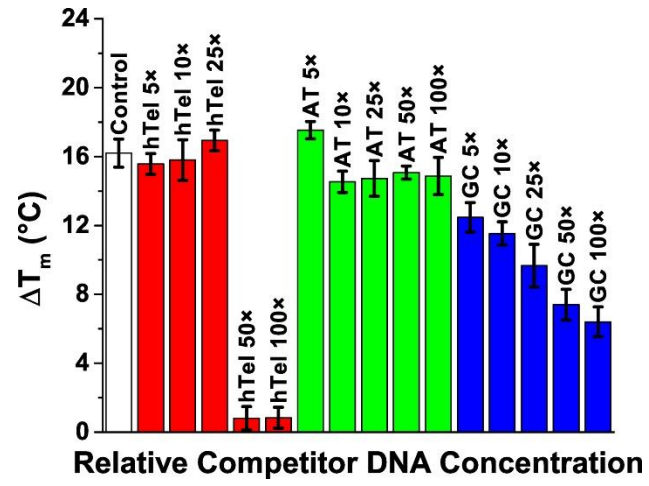
Determination of Binding Stoichiometry by Job Plot

The melting experiments with FRET-labeled PQS answered the question "Does Compound 1 interact with G-quadruplex structures?" and confirmed that the proposed virtual screening approach was successful in identifying a G-quadruplex-interacting small molecule. The questions that remained unanswered are "How does Compound 1 interact with G-quadruplex structures? How many? How tightly? How so?" The construction of a Job plot using the continuous variation method is a way to answer the question of "How many?" For Compound 1 interaction with hTel22 and cMyc22 G-quadruplex structures, both Job plots showed a single inflection point indicating that Compound 1 binds to these G-quadruplex structures with a single binding mode (Figure 70). For the interaction with hTel22 G-quadruplex structures, the inflection point was at 0.53 mol fraction Compound 1 indicating a binding stoichiometry of approximately one molecule of Compound 1 per one G-quadruplex structure. For the interaction with cMyc22 G-quadruplex structures, the inflection point was at 0.65 mol fraction Compound 1 indicated a binding stoichiometry approximately two molecules of Compound 1 per one G-quadruplex structure. The differences observed in stoichiometry for Compound 1 binding to the two PQS are related to the specific G-quadruplex topology adopted: "hybrid" form for hTel22 and "propeller" form for cMyc22. In a recently published NMR solution structure of one telomestatin molecule bound by end-pasting to a telomeric G-quadruplex, which contains the hTel22 sequence and adopts the "hybrid-1" topology, it was noted that one end of the G-quadruplex structure was more open than the other to molecule binding. In the case of cMyc22 G-quadruplex, an NMR solution structure with two quinacrine molecules bound by end-pasting showed two identical binding sites on either ends of the two terminal G-

quartets. For this reason, cMyc22 might be able to bind two small molecules compared to hTel22, which can only bind one small molecule. Overall, the results of the Job plots agreed with previously published high-resolution structures of G-quadruplex bound ligands suggesting that Compound 1 might interact with G-quadruplex structures with similar mechanism.

Figure 70. Job plots of fluorescent emission spectroscopy at different molar fractions of Compound 1 showing the interaction of Compound 1 with (A) hTel22 d[AG$_3$(T$_2$AG$_3$)$_3$] and (B) cMyc22 d[TGA(G$_3$TG$_3$TA)$_2$A] G-quadruplex structures.
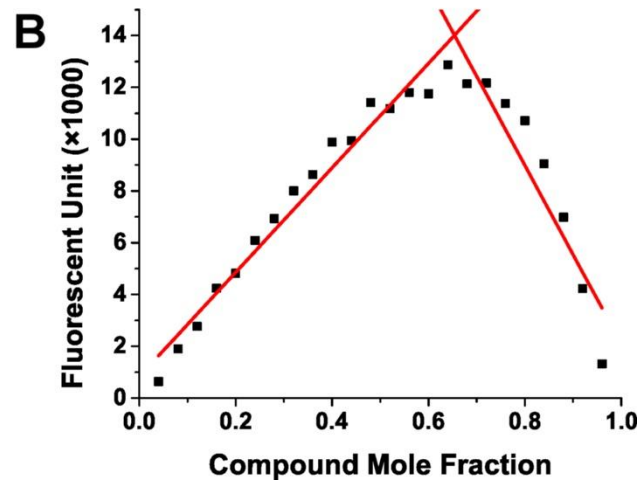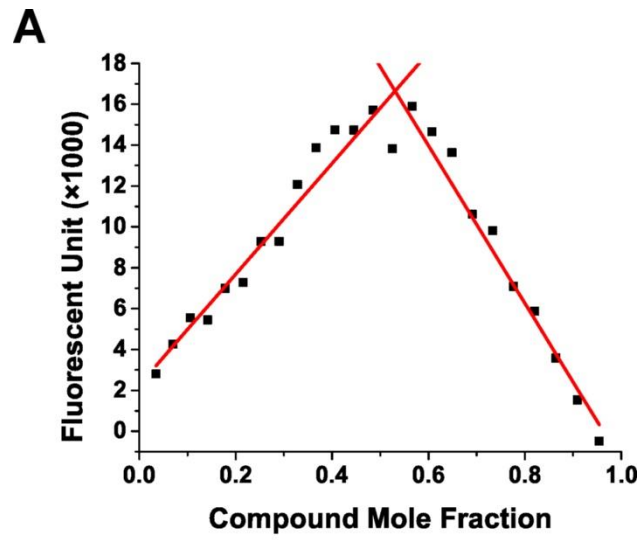
Determination of Binding Affinity by Fluorescent Emission Spectroscopy

With the Job Plot provided the answer to "How many," titration experiments using fluorescent emission spectroscopy were carried out to answer the question "How tightly?" One advantage of fluorescent method is the ability to work at lower sample concentrations, useful as Compound 1 has low solubility in water. However, a limitation is that the compound itself must be fluorescent (Figure 71). From titration experiments, whereby DNA is added to a fixed concentration of Compound 1 and the changes in fluorescent is measured, the binding affinity, K_A, and free energy of binding (by the Gibb's equation, $\Delta G = -RT\ln K_A$) can be determined. Examination of the salt dependency on the observed K_A allows for the decomposition of the binding free energy into the nonelectrostatic ($\Delta G_{nonelectrostatic}$) and the electrostatic ($\Delta G_{electrostatic}$) components. The titration experiments in low salt (30 mM KCL) showed that fluorescent emission of Compound 1 was enhanced in the presence of increasing amounts of DNA (Figure 72, left column). The fluorescent differential spectra of Compound 1 (Figure 72, middle column), calculated by finding the difference between the fluorescent spectra of Compound 1 in the presence of maximum concentration of DNA and the fluorescent spectra of Compound 1 in the absence of DNA indicated a maximum increase of fluorescent emission at about 690 nm upon binding with DNA. Interaction of Compound 1 with the hTel22 and the cMyc22 G-quadruplex structures also resulted in a second plateau in the 730-760 nm range for the fluorescent differential spectra. The existence of the second plateau was more subtle for interaction with the G-rich duplex structures and was not present for the interaction with the AT-rich duplex structures. The binding affinities of Compound 1 for different nucleic acids structures (Table 12) were

278

determined by non-linear regression fitting of fluorescent emission data of Compound 1 at 690 nm at increasing amounts of DNA (Figures 72, right column). For the cMyc22 G-quadruplex structure, which was previously shown to bind two molecules of Compound 1, the existence of only one transition on the binding isotherms suggested that Compound 1 has similar affinities for both binding sites. The relative affinities of Compound 1 for different nucleic acid structures were, in decreasing affinity: cMyc22 G-quadruplex >> AT-rich duplex > hTel22 G-quadruplex >> GC-rich duplex. While the measured affinities of Compound 1 appeared to contradict the observation made by melting experiments that AT-rich duplex do not bind Compound 1, one possible explanation for the inconsistency between the methods is that that since AT-rich duplex melts at very low temperature (40°C) compared to hTel22 (65°C), this could have resulted in re-equilibration of Compound 1 with the FRET-label hTel22 G-quadruplex structures after the AT-rich duplex structures have already denatured. Of the four DNA sequences examined, the experimental error associated with the cMyc22 sequence was dramatically higher compared to the other three sequences. This sequence is known to aggregate at higher concentrations (Le *et al.*, 2012), which may affect the binding and fluorescence of Compound 1. This resulted in a disagreement between the measured data and the line of best fit at higher DNA concentrations and the higher observed experimental error for the sequence (Figure 72, bottom right).

Figure 71. (A) Absorption spectrum and (B) fluorescent emission and excitation of Compound 1 in DMSO.

A

B
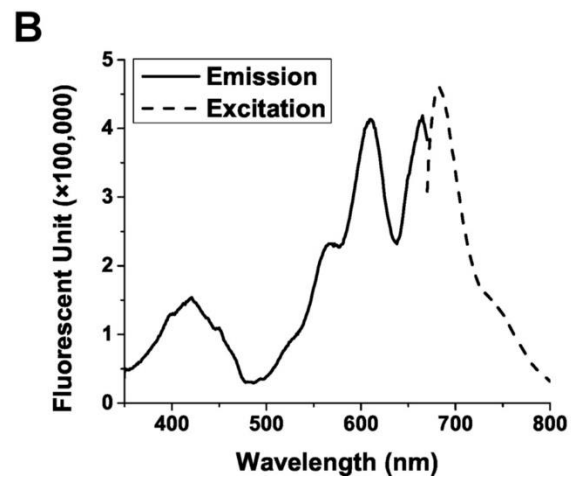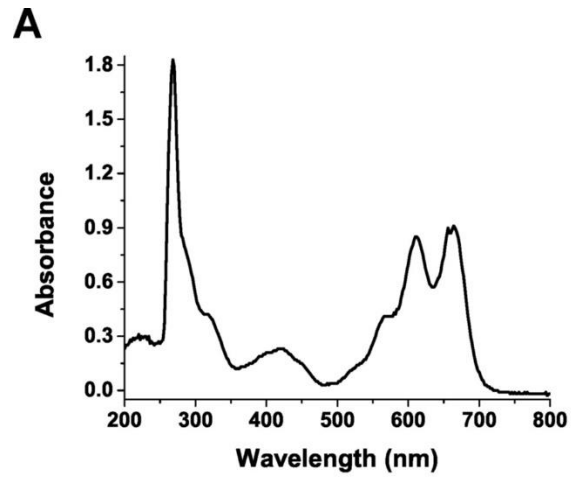
Figure 72. Fluorescent emission titrations at increasing DNA:Compound 1 ratios demonstrating the interaction of Compound 1 with DNA structures in low salt (30mM KCl) buffer. Left panels are fluorescent emission spectra of Compound 1, middle panels are fluorescent emission differential spectra, and right panels are titration curve with best fit lines by non-linear regression (red). DNA sequences used were (A) AT duplex d[$(AT)_5C_4(AT)_5$], (B) GC duplex d[$(GC)_5T_4(GC)_5$], (C) hTel22 d[$AG_3(T_2AG_3)_3$], and (D) cMyc22 d[$TGA(G_3TG_3TA)_2A$].
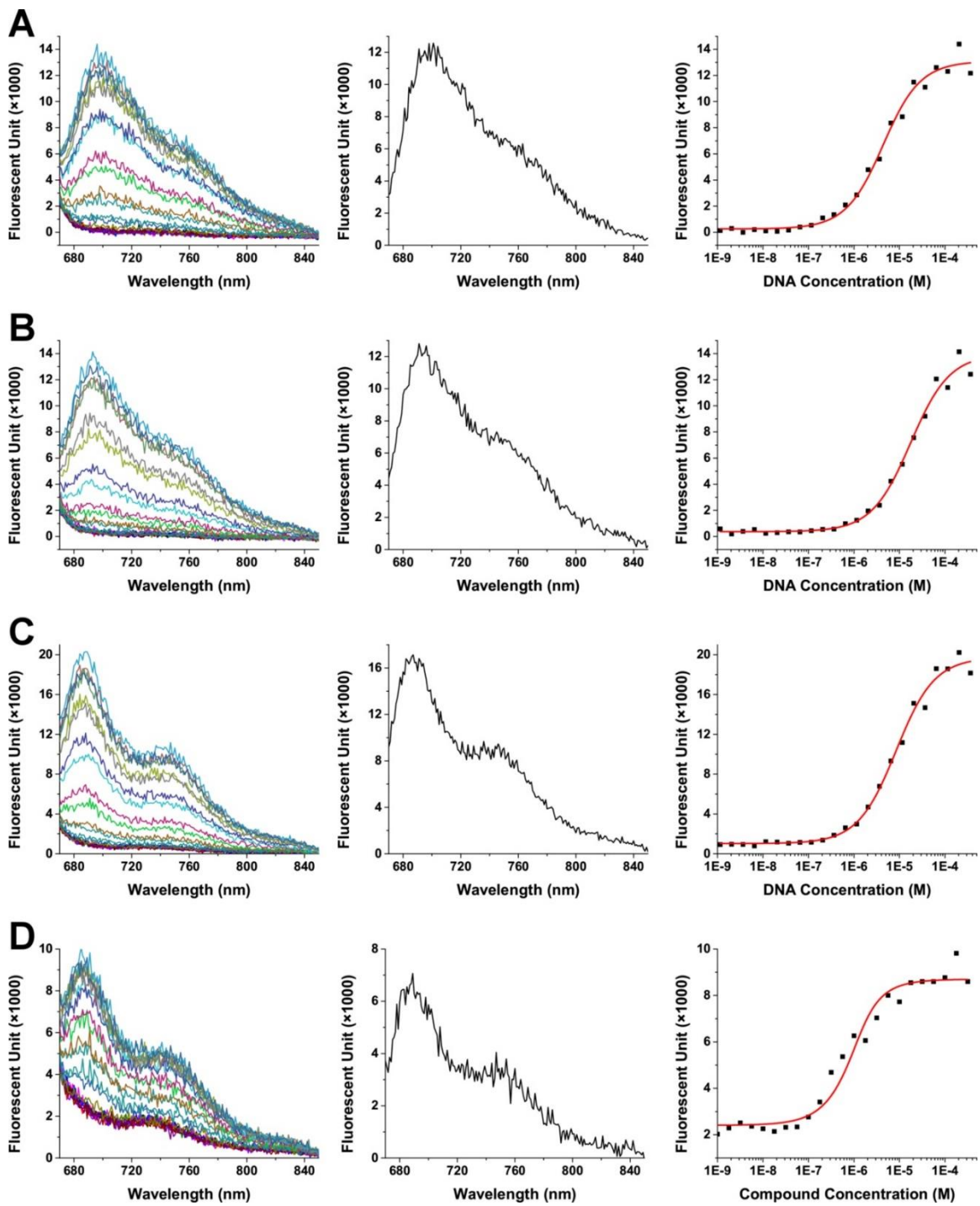
Table 12. Association Constants (K_A) Determined from Fluorescent Titration Experiments

Sequence	Low Salt Buffer (25 mM KCl)			High Salt Buffer (200mM KCl)			$-Z\varphi$
	K_{obs} (M^{-1})	ΔG_{bind} (kJ)	ΔG_{elec} (kJ)	K_{obs} (M^{-1})	ΔG_{bind} (kJ)	ΔG_{elec} (kJ)	
AT duplex $d[(AT)_5C_4(AT)_5]$	258481 ± 30345	-32.97	-13.82	15294 ± 1810	-25.49	-6.34	-1.49
GC duplex $d[(GC)_5T_4(GC)_5]$	58537 ± 6078	-29.04	-0.86	49107 ± 5819	-28.57	-0.39	-0.09
hTel22 $d[AG_3(T_2AG_3)_3]$	125471 ± 11775	-31.05	-3.17	65650 ± 5860	-29.34	-1.45	-0.34
cMyc22 $d[TGA(G_3TG_3TA)_2A]$	2242692 ± 726385	-38.68	N/A	N/A	N/A	N/A	N/A

The results for titration experiments in high salt buffer (200 mM KCl) also showed an enhancement of Compound 1 fluorescent emission in the presence of increasing amounts of DNA (Figure 73). Titration experiments were not performed with the cMyc22 sequence because of significant DNA aggregation at higher salt. For Compound 1 interaction with the hTel22 G-quadruplex structures, a second plateau was observed in the 730-760 nm similar to the low salt experiments. Similarly, the second plateau was more subtle for the interaction with G-rich duplex structures and was not seen for interaction with AT-rich duplex structures. Compared to the results at lower salt, the relative affinities of Compound 1 at high salt were different and displayed a clear preference for G-quadruplex over duplex structures: hTel22 G-quadruplex > G-rich duplex >> AT-rich duplex. Overall, the measured affinities supported the conclusion that Compound 1 is G-quadruplex selective. While the results at low salt indicated some preference for AT-rich duplex, the results at higher salt (which is closer to the physiological salt concentration of 140 mM KCl) are more relevant.

Figure 73. Fluorescent emission titrations at increasing DNA:Compound 1 ratios demonstrating the interaction of Compound 1 with DNA structures in high salt (200mM KCl) buffer. Left panels are fluorescent emission spectra of Compound 1, middle panels are fluorescent emission differential spectra, and right panels are titration curve with best fit lines by non-linear regression (red). DNA sequences used were (A) polyAT duplex d[$(AT)_5C_4(AT)_5$], (B) polyGC duplex d[$(GC)_5T_4(GC)_5$], and (C) hTel22 d[$AG_3(T_2AG_3)_3$].
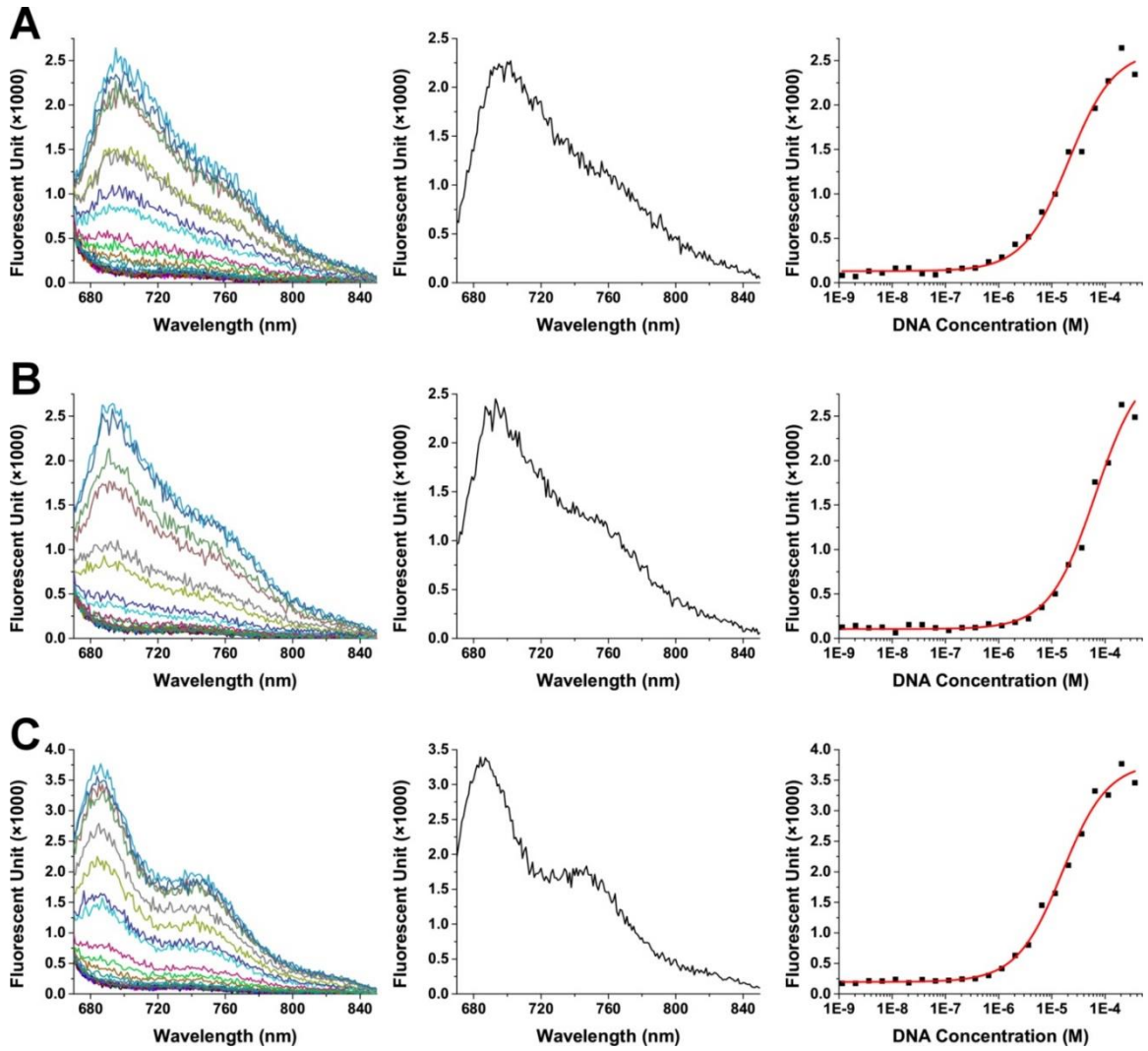
The number of counterions released upon drug binding can be determined from the salt-dependency of the binding constant using Equation 9:

$$-Z\varphi = \frac{\delta lnK}{\delta ln[K^+]} \qquad \text{Equation 9}$$

where δlnK is the change in the observed binding constant, $\delta ln[K^+]$ is the change in the salt concentration, Z is the charge of compound 1 (+2), and φ is the fraction of K^+ associated with each DNA phosphate group. From the number of counterions released, the electrostatic component of the binding free energy can be calculated using Equation 10:

$$\Delta G_{electrostatic} = Z\psi RTln[K^+] \qquad \text{Equation 10}$$

where R is the gas constant and T is the absolute temperature. With the AT-duplex structures, electrostatic interaction contributed significantly to Compound 1 binding (Table 12), accounted for 41% ($^{-13.82}/_{-32.97}$) of the free energy of binding (ΔG) at low salt and 25% ($^{-6.34}/_{-25.49}$) of ΔG at high salt. With the hTel22 G-quadruplex structures, electrostatic interaction played a lesser role in Compound 1 binding, 10% ($^{-3.17}/_{-31.05}$) at low salt and 5% ($^{-1.45}/_{-29.34}$) at high salt. With the GC-duplex structures, electrostatic interaction was not a major factor in Compound 1 binding, 3% ($^{-0.86}/_{-29.04}$) at low salt and 1% ($^{-0.39}/_{-28.57}$) at high salt. The different levels of contribution of electrostatic interaction to the overall binding free energy suggested that Compound 1 might bind to different DNA structures with different binding modes. For example, Compound 1 might bind to AT-duplex structures by groove binding where electrostatic interaction might be more predominant (i.e. between the negatively charged DNA phosphate backbone and the positively charged piperazine side chain). For binding with the GC-duplex structures, Compound 1 might be binding by intercalation where nonelectrostatic interaction is more

289

of a factor (i.e. π-π stacking interaction between the GC base pair and the aromatic naphthophenoxazine core). For binding with hTel22 G-quadruplex structures, electrostatic interaction played a bigger role than with GC-duplex structures but not as much of a factor as the AT-duplex structures. The binding mode of end-pasting is consistent with the observed free energy decomposition as it is the binding mode that allows for Compound 1 to interact with the G-quadruplex structures with both nonelectrostatic (π-π stacking between the aromatic core and the terminal G-tetrad) and electrostatic interaction (between the positively charged side chain and the negatively charged phosphate groups of the loop structures).

Determination of Binding Mechanism by Induced CD Spectroscopy

Titration experiments began to answer the question "How so? By what binding mode does Compound 1 interact with nucleic acids?" Circular dichroism (CD) spectroscopy experiments were carried out to provide further evidence that Compound 1 interact with G-quadruplex structures by the end-pasting mechanism. For small molecules, which typically lack a CD signal, binding to nucleic acids can result in an induced CD (ICD) effect whose magnitude and sign would allow for the classification of the binding mode as either groove binding (a positive signal) or intercalation/end-pasting (a negative signal) (Garbett *et al.*, 2007). Two negative ICD peaks were observed at 625 nm and 675 nm for Compound 1 interaction with hTel22 and cMyc22 G-quadruplex structures (Figure 74). The locations of the ICD peaks were consistent with the two absorbance peaks of Compound 1 in the area (Figure 71). Combined with previous

290

observations, the induced CD from Compound 1 interaction with hTel22 and cMyc22 G-quadruplex structures showed that Compound 1 binds by the end-pasting mechanism.

Figure 74. Induced CD spectroscopy at increasing DNA:Compound 1 ratios showing the interaction of Compound 1 with hTel22 $d[AG_3(T_2AG_3)_3]$ and cMyc22 $d[TGA(G_3TG_3TA)_2A]$ G-quadruplex structures.
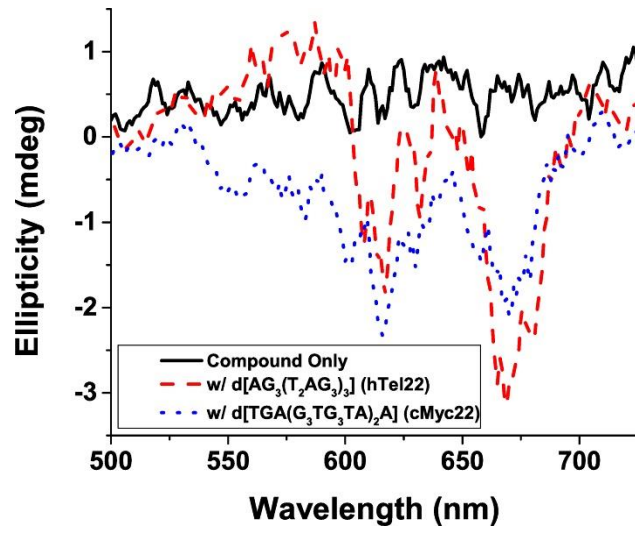
Displacement of Thiazole Orange Binding to Nucleic Acids by Compound 1

TO-FID experiments were performed to provide additional evidence that Compound 1 interacts with G-quadruplex structures by the end-pasting mechanism. Thiazole orange (TO) is believed to bind to the hTel22 by the end-pasting mechanism (Monchaud *et al.*, 2008). When bound to DNA, the quantum yield of TO increases significantly resulting in an observed increase in fluorescence at 570 nm. A quenching of TO fluorescence upon addition of Compound 1 indicates the successful displacement of TO binding by Compound 1 and confirms both the binding of Compound 1 to the G-quadruplex structures and the probable binding site. Compound 1 was able to successfully displace TO binding from the hTel22 G-quadruplex structures as well as other nucleic acids structures examined (Figure 75). The maximum displacement was observed for Compound 1 interaction with cMyc22 G-quadruplex structures (60%) while the minimum displacement was observed for Compound 1 interaction with the GC-duplex structures (20%). The relative activities of Compound 1 by TO-FID were in agreement with previous titration experiments performed in low salt buffer: cMyc22 G-quadruplex > AT-triplex > GC triplex > AT-duplex > hTel22 G-quadruplex > hTel duplex > ALT duplex > GC Duplex. For the hTel22 G-quadruplex structures, previous studies have suggested that TO bind by the end-pasting mechanism (Monchaud *et al.*, 2006, Monchaud *et al.*, 2008). The displacement of TO by Compound 1 suggested that Compound 1 also binds by the same end-pasting mechanism. For the other nucleic acid structures, the binding mechanism of TO is unknown, however, the displacement of TO binding to these structures by Compound 1 indicated that Compound 1 interacts with nucleic acid structures at the same binding site as TO. Overall, the TO-FID data were

consistent with previous observations and supported the hypothesis that Compound 1 likely interacted with G-quadruplex structures by end-pasting. In addition, it is important to note the high activity observed with AT-triplex and GC-triplex structures, which suggest another possible application for Compound 1 as small molecules that stabilize triplex DNA structures might have therapeutic values (Holt *et al.*, 2009).

Figure 75. TO-FID results showing the displacement of thiazole orange (TO) from different DNA structures by Compound 1. DNA sequences used were (A) AT duplex d[(AT)$_4$C$_4$(AT)$_4$], (B) GC duplex d[(GC)$_4$T$_4$(GC)$_4$], (C) ALT duplex d[A$_8$C$_4$T$_8$], (D) H-Tel duplex d[G$_3$T$_2$AG$_3$T$_4$C$_3$TA$_2$C$_3$], (E) AT triplex d[A$_8$C$_4$T$_8$C$_4$T$_8$], (F) GC triplex d[C$_8$T$_4$G$_8$T$_4$C$_8$], (G) hTel22 d[AG$_3$(T$_2$AG$_3$)$_3$], and (H) cMyc22 d[TGA(G$_3$TG$_3$TA)$_2$A]. Data plotted are the average of experiments in duplicate. Error bars represent the spread of the data.
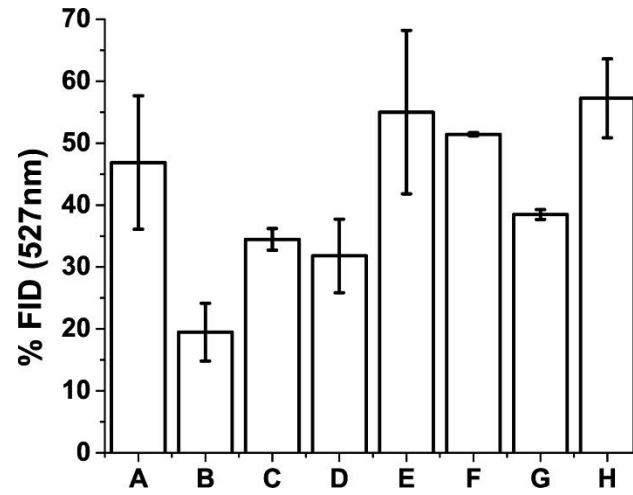
Molecular Docking of Compound 1 to Hybrid-1 and Hybrid-2 Structures

The biophysical findings demonstrated that Compound 1 interacts with G-quadruplex structures by the end-pasting mechanism. As such, several models of Compound 1 bound to the end-pasting sites of the hTel22 G-quadruplex structures were generated to visualize the interaction. In solution, there are two predominant topologies for the hTel22 G-quadruplex structures, the "hybrid-1" and the "hybrid-2 forms (Dai *et al.*, 2007a, Dai *et al.*, 2007b). These two topologies each have two external G-quartet end-pasting sites for a total of four possible sites. The top poses of Compound 1 as determined by two docking software, Autodock and Surflex, are shown (Figure 76) with the resulting scores shown in Table 13. The Autodock and Surflex scores observed for Compound 1 were comparable. A closer view of the top poses without the nucleic acid present (Figure 77) revealed that the top ranked poses for Surflex and Autodock appear to have a higher amount of overlap for the two Hybrid 1 end-pasting sites compared to Surflex and Autodock top ranked poses for the end-pasting sites of Hybrid-2. In all four models, there were general agreement between the two docking software with regard to the stacking of the aromatic core of Compound 1 onto the terminal G-quartets. In contrast, with regard to the interaction of the side chain with the loops and groove structures, the two docking software did not agree. The models suggested that Compound 1 can interact with the various loops and grooves to different extents. It is important to note, however, that while the two software disagree on the specific loops and grooves interaction, the models generated from both software indicated that loops, grooves, and terminal G-quartets interactions are necessary to stabilize Compound 1 binding to G-quadruplex structures. These findings are supported by previous biophysical

measurements which demonstrated both the contribution of electrostatic and nonelectrostatic interactions to Compound 1 binding.

Figure 76. The top ranked poses for Compound 1 docked using Surflex (Red) and Autodock (Blue) to the following G-quadruplex nucleic acid structures: (A) hybrid-1 end-paste site 1, (B) hybrid-1 end-paste site 2, (C) hybrid-2 end-paste site 1, and (D) hybrid-2 end-paste site 2. The nucleic acid structures are shown in dark grey except for the small molecule interacting terminal G-quartets, which are shown in green.

A   B

C   D
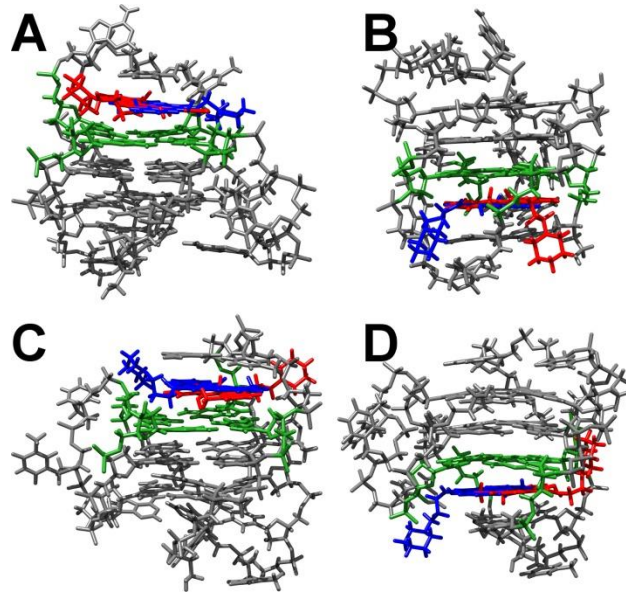
Table 13. The Scores for the Top-Ranked Poses from Autodock or Surflex

G-Quadruplex Structure	Surflex Score ($-\log(K_d)$)	Autodock Score (kcal/mol)
Hybrid-1 End-Paste Site 1	11.36	-13.39
Hybrid-1 End-Paste Site 2	12.93	-14.67
Hybrid-2 End Paste Site 1	14.12	-14.29
Hybrid-2 End Paste Site 2	15.13	-14.98

Figure 77. The top ranked poses for Compound 1 docked using Surflex (Red) and Autodock (Blue) to the following G-quadruplex nucleic acid structures: (A) hybrid-1 end-paste site 1, (B) hybrid-1 end-paste site 2, (C) hybrid-2 end-paste site 1, and (D) hybrid-2 end-paste site 2.
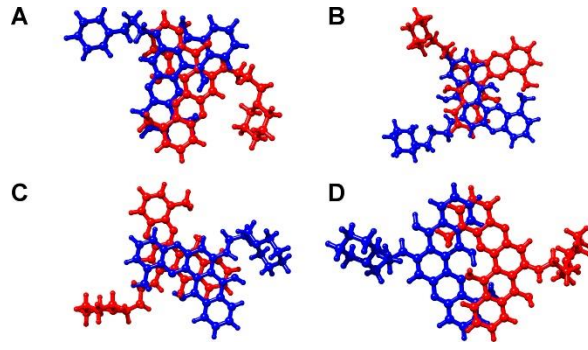
Inhibition of Cancer Cells Growth by Compound 1 Treatment

With the initial discovery of Compound 1 and subsequent biophysical testing which demonstrated that Compound 1 binds to G-quadruplex structures by the end-pasting mechanism, the discussion on Compound 1 is concluded with an investigation of the effect of Compound 1 on cancer cells. The results of MTT assays showed that Compound 1 is cytotoxic to cancer cells (Figure 78). The MTT results were in agreement with the results of the NCI-60 DTP Human Tumor Cell Line Screen (Figures 79 and 80). In general, Compound 1 was able to inhibit the growth of all 60 cancer cell lines at the concentrations tested. The activity of Compound 1 in inhibiting growth was less active in renal cancer cell lines and more in leukemia, colon, melanoma, ovarian, and prostate cancer cell lines. Taken together with the previous biophysical experiments, the results of the cell-based experiments suggest that Compound 1 might be able to serve as a lead compound in the development of a novel anti-cancer therapeutics.

Figure 78. Results of MTT assay showing the effect of Compound 1 on cancer cell proliferation.

Figure 79. Summary of dose-response experiments showing the effect of Compound 1 on human cancer cells. The experiments were conducted as part of the NCI-60 DTP Human Tumor Cell Line Screen.

National Cancer Institute Developmental Therapeutics Program
In-Vitro Testing Results

NSC : D - 753649 / 1	Experiment ID : 1008NS09	Test Type : 08	Units : Molar
Report Date : February 09, 2012	Test Date : August 23, 2010	QNS :	MC :
COMI : cmpd7_zinc20263704 (96374)	Stain Reagent : SRB Dual-Pass Related	SSPL : 0TBH	

Log10 Concentration

Panel/Cell Line	Time Zero	Ctrl	Mean Optical Densities -8.0	-7.0	-6.0	-5.0	-4.0	Percent Growth -8.0	-7.0	-6.0	-5.0	-4.0	GI50	TGI	LC50
Leukemia															
CCRF-CEM	0.186	0.956	0.931	1.008	0.644	0.099	0.280	97	107	59	-47	12	1.23E-6	.	> 1.00E-4
K-562	0.192	1.495	1.440	1.475	1.119	0.104	0.166	96	98	71	-46	-14	1.52E-6	4.05E-6	> 1.00E-4
MOLT-4	0.375	1.502	1.594	1.701	1.397	0.199	0.325	108	118	91	-47	-13	1.97E-6	4.55E-6	> 1.00E-4
RPMI-8226	0.785	2.141	2.309	2.346	2.285	0.738	0.662	112	115	111	-6	-16	3.31E-6	8.87E-6	> 1.00E-4
SR	0.243	1.554	1.504	1.412	0.844	0.160	0.415	96	89	46	-34	13	8.00E-7	.	> 1.00E-4
Non-Small Cell Lung Cancer															
A549/ATCC	0.216	1.292	1.302	1.323	1.218	0.097	0.442	101	103	93	-55	21	1.95E-6	.	.
EKVX	0.667	1.561	1.507	1.430	1.430	0.739	0.761	94	85	85	8	10	2.87E-6	> 1.00E-4	> 1.00E-4
HOP-62	0.431	0.955	0.960	0.978	0.971	0.149	0.668	101	104	103	-66	45	2.06E-6	.	.
HOP-92	0.910	1.254	1.231	1.225	1.178	1.125	0.609	93	91	78	62	-33	1.35E-5	4.51E-5	> 1.00E-4
NCI-H226	0.903	2.135	2.010	2.110	1.992	2.027	1.040	90	98	88	91	11	3.27E-5	> 1.00E-4	> 1.00E-4
NCI-H23	0.567	1.561	1.549	1.484	1.466	0.663	0.425	99	92	90	10	-25	3.17E-6	1.90E-5	> 1.00E-4
NCI-H322M	0.671	1.243	1.319	1.300	1.272	0.403	0.461	113	110	105	-40	-31	2.40E-6	5.30E-6	> 1.00E-4
NCI-H460	0.171	1.312	1.424	1.422	1.285	0.019	0.390	110	110	98	-89	19	1.80E-6	.	.
NCI-H522	0.649	1.577	1.516	1.510	1.502	0.370	0.815	93	93	92	-43	18	2.05E-6	.	> 1.00E-4
Colon Cancer															
COLO 205	0.250	1.049	1.152	1.192	1.163	0.110	0.464	113	118	114	-56	27	2.38E-6	.	.
HCC-2998	0.525	1.736	1.690	1.664	1.597	0.298	0.430	96	94	89	-43	-18	1.96E-6	4.70E-6	> 1.00E-4
HCT-116	0.237	2.121	2.078	2.241	2.053	0.040	0.498	98	106	96	-83	14	1.81E-6	.	.
HCT-15	0.306	2.076	2.034	2.000	2.014	0.293	0.389	98	96	96	-4	5	2.89E-6	.	> 1.00E-4
HT29	0.214	1.252	1.240	1.262	1.124	0.051	0.583	99	101	88	-76	36	1.70E-6	.	.
KM12	0.277	1.249	1.324	1.274	1.271	0.019	-0.005	108	103	102	-93	-100	1.85E-6	3.34E-6	6.01E-6
SW-620	0.222	1.198	1.219	1.202	1.188	0.038	0.403	102	100	99	-83	19	1.86E-6	.	.
CNS Cancer															
SF-268	0.420	1.304	1.281	1.299	1.277	0.075	0.390	97	99	97	-82	-7	1.83E-6	3.48E-6	.
SF-295	0.470	1.434	1.357	1.339	1.416	0.485	0.534	92	90	98	2	7	3.15E-6	> 1.00E-4	> 1.00E-4
SF-539	0.564	1.698	1.609	1.605	1.587	0.098	0.385	92	92	90	-83	-32	1.71E-6	3.32E-6	.
SNB-19	0.481	1.589	1.602	1.586	1.623	0.572	0.454	101	100	103	8	-6	3.62E-6	3.92E-5	> 1.00E-4
SNB-75	0.797	1.416	1.273	1.237	1.219	1.317	0.661	77	71	68	84	-17	2.17E-5	6.78E-5	> 1.00E-4
U251	0.243	1.177	1.206	1.225	1.200	0.025	0.451	103	105	103	-90	22	1.87E-6	.	.
Melanoma															
LOX IMVI	0.256	1.721	1.711	1.610	1.602	0.100	0.701	99	92	92	-61	30	1.88E-6	.	.
M14	0.353	1.383	1.403	1.431	1.499	0.181	0.454	102	105	111	-49	10	2.41E-6	.	> 1.00E-4
MDA-MB-435	0.347	1.349	1.301	1.271	1.298	0.087	0.642	95	92	95	-75	29	1.84E-6	.	.
SK-MEL-2	1.010	2.646	2.586	2.607	2.552	2.190	0.443	96	98	94	72	-56	1.49E-5	3.65E-5	8.96E-5
SK-MEL-28	0.546	1.318	1.370	1.351	1.295	0.290	0.719	107	104	97	-47	22	2.12E-6	.	> 1.00E-4
SK-MEL-5	0.357	1.861	1.656	1.852	1.627	0.701	0.429	86	99	84	23	5	3.62E-6	> 1.00E-4	> 1.00E-4
UACC-257	0.490	1.177	1.246	1.238	1.245	0.190	0.392	110	109	110	-61	-20	2.24E-6	4.39E-6	.
UACC-62	0.690	2.051	2.074	2.139	2.127	1.272	0.660	102	106	106	43	-4	7.68E-6	8.06E-5	> 1.00E-4
Ovarian Cancer															
IGROV1	0.345	1.109	1.257	1.214	1.158	0.187	0.548	119	114	106	-46	27	2.35E-6	.	> 1.00E-4
OVCAR-3	0.345	0.857	0.890	0.941	0.922	0.022	0.292	106	116	113	-94	-16	2.01E-6	3.52E-6	.
OVCAR-4	0.523	1.471	1.462	1.429	1.430	0.282	0.707	99	96	96	-46	19	2.10E-6	.	> 1.00E-4
OVCAR-5	0.505	1.286	1.241	1.244	1.226	0.397	0.106	94	95	92	-21	-79	2.35E-6	6.47E-6	3.13E-5
OVCAR-8	0.255	1.230	1.279	1.296	1.270	0.078	0.072	105	107	104	-69	-72	2.05E-6	3.98E-6	7.73E-6
NCI/ADR-RES	0.422	1.432	1.437	1.404	1.397	0.131	0.537	100	97	97	-69	11	1.91E-6	.	.
SK-OV-3	0.624	1.301	1.365	1.342	1.378	1.364	0.606	109	106	111	109	-3	3.37E-5	9.41E-5	> 1.00E-4
Renal Cancer															
786-0	0.717	2.233	2.210	2.289	2.123	1.660	0.760	98	104	93	62	3	1.61E-5	> 1.00E-4	> 1.00E-4
A498	0.977	1.398	1.386	1.325	1.374	1.311	0.457	97	83	94	79	-53	1.66E-5	3.96E-5	9.45E-5
ACHN	0.353	1.557	1.616	1.553	1.570	1.152	0.160	105	100	101	66	-55	1.36E-5	3.53E-5	9.15E-5
CAKI-1	0.446	1.673	1.674	1.606	1.572	1.349	0.504	100	95	92	74	5	2.20E-5	> 1.00E-4	> 1.00E-4
RXF 393	0.491	0.973	0.927	0.974	0.897	0.063	0.517	90	100	84	-87	5	1.58E-6	.	.
SN12C	0.482	1.791	1.808	1.814	1.762	1.387	0.514	101	102	98	69	2	1.94E-5	> 1.00E-4	> 1.00E-4
TK-10	0.515	1.189	1.144	1.118	1.111	0.405	0.566	93	89	88	-21	8	2.24E-6	.	> 1.00E-4
UO-31	0.599	1.206	1.211	1.227	1.158	0.889	0.414	101	103	92	48	-31	8.91E-6	4.05E-5	> 1.00E-4
Prostate Cancer															
PC-3	0.271	0.720	0.703	0.711	0.674	0.055	0.486	96	98	90	-80	48	1.72E-6	.	.
DU-145	0.356	1.184	1.163	1.251	1.231	0.005	0.023	97	108	106	-99	-94	1.87E-6	3.29E-6	5.77E-6
Breast Cancer															
MCF7	0.351	1.584	1.522	1.506	1.441	0.106	0.594	95	94	88	-70	20	1.75E-6	.	.
MDA-MB-231/ATCC	0.520	1.261	1.227	1.308	1.234	0.342	0.432	95	106	96	-34	-17	2.26E-6	5.46E-6	> 1.00E-4
HS 578T	0.621	1.062	1.073	1.085	1.071	0.238	0.606	103	105	102	-62	-2	2.08E-6	4.20E-6	.
BT-549	0.960	1.958	2.025	1.994	2.008	1.940	0.678	107	104	105	98	-29	2.39E-5	5.88E-5	> 1.00E-4
T-47D	0.640	1.336	1.353	1.316	1.365	1.082	0.792	103	97	104	63	22	2.11E-5	> 1.00E-4	> 1.00E-4
MDA-MB-468	0.488	1.138	1.075	1.077	1.051	0.277	0.546	90	91	87	-43	9	1.91E-6	.	> 1.00E-4

Figure 80. Mean graphs comparing the activity of Compound 1 in various human cancer cells. The experiments were conducted as part of the NCI-60 DTP Human Tumor Cell Line Screen.

311

National Cancer Institute Developmental Therapeutics Program
Mean Graphs

NSC : D - 753649/1	Units :Molar	SSPL :0TBH	EXP. ID :1008NS09
Report Date :February 09, 2012		Test Date :August 23, 2010	

Panel/Cell Line	Log_{10} GI50	Log_{10} TGI	Log_{10} LC50
Leukemia			
CCRF-CEM	-5.91		> -4.00
K-562	-5.82	-5.39	> -4.00
MOLT-4	-5.70	-5.34	> -4.00
RPMI-8226	-5.48	-5.05	> -4.00
SR	-6.10		> -4.00
Non-Small Cell Lung Cancer			
A549/ATCC	-5.71		
EKVX	-5.54		> -4.00
HOP-62	-5.69		> -4.00
HOP-92	-4.87	-4.35	> -4.00
NCI-H226	-4.49	> -4.00	> -4.00
NCI-H23	-5.50	-4.72	> -4.00
NCI-H322M	-5.62	-5.28	> -4.00
NCI-H460	-5.75		
NCI-H522	-5.69		> -4.00
Colon Cancer			
COLO 205	-5.62		
HCC-2998	-5.71	-5.33	
HCT-116	-5.74		> -4.00
HCT-15	-5.54		
HT29	-5.77		
KM12	-5.73	-5.48	-5.22
SW-620	-5.73		
CNS Cancer			
SF-268	-5.74	-5.46	
SF-295	-5.50	> -4.00	> -4.00
SF-539	-5.77	-5.48	
SNB-19	-5.44	-4.41	> -4.00
SNB-75	-4.66	-4.17	
U251	-5.73		
Melanoma			
LOX IMVI	-5.73		
M14	-5.62		> -4.00
MDA-MB-435	-5.74	-4.44	-4.05
SK-MEL-2	-4.83		> -4.00
SK-MEL-28	-5.67		> -4.00
SK-MEL-5	-5.44	> -4.00	> -4.00
UACC-257	-5.65	-5.36	> -4.00
UACC-62	-5.11	-4.09	
Ovarian Cancer			
IGROV1	-5.63		> -4.00
OVCAR-3	-5.70	-5.45	
OVCAR-4	-5.68		> -4.00
OVCAR-5	-5.63	-5.19	-4.50
OVCAR-8	-5.69	-5.40	-5.11
NCI/ADR-RES	-5.72	-4.03	
SK-OV-3	-4.47		> -4.00
Renal Cancer			
786-0	-4.79	> -4.00	> -4.00
A498	-4.78	-4.40	-4.02
ACHN	-4.87	-4.45	-4.04
CAKI-1	-4.86	> -4.00	> -4.00
RXF 393	-5.80		> -4.00
SN12C	-4.71	> -4.00	> -4.00
TK-10	-5.65	-4.39	
UO-31	-5.05		> -4.00
Prostate Cancer			
PC-3	-5.77		
DU-145	-5.73	-5.48	-5.24
Breast Cancer			
MCF7	-5.76		
MDA-MB-231/ATCC	-5.65	-5.26	
HS 578T	-5.68	-5.38	> -4.00
BT-549	-4.62	-4.23	
T-47D	-4.68	> -4.00	> -4.00
MDA-MB-468	-5.72		> -4.00
MID	-5.46	-4.71	-4.11
Delta	0.64	0.77	1.13
Range	1.63	1.48	1.24

Conclusion

This study reported a virtual screening approach that was successful in identifying Compound 1, a new G-quadruplex-interacting small molecule. Compared to traditional drug discovery methods where chemical libraries, containing only thousands of compounds, are screened in an expensive and labor-intensive process, *in silico* virtual screening allows for the examination of larger libraries consisted of millions of compounds capable of a more extensive exploration of theoretical chemical space in a rapid and inexpensive manner. Previous virtual screening studies with libraries containing only 10^3-10^5 small molecules (Ma *et al.*, 2008, Cosconati *et al.*, 2009, Dailey *et al.*, 2009, Li *et al.*, 2009, Wu *et al.*, 2009, Lee *et al.*, 2010, Chen *et al.*, 2011, Holt *et al.*, 2011, Trotta *et al.*, 2011, Alcaro *et al.*, 2013, Castillo-Gonzalez *et al.*, 2013, Kar *et al.*, 2012, Ma *et al.*, 2012a, Ma *et al.*, 2012b) have not even begin to reach the potential of the technique. In the current work, 6.6 million compounds were investigated. This was one of the largest virtual screening effort to date. The selective binding of Compound 1 to G-quadruplex structures was validated by extensive biophysical measurements using fluorescent and CD spectroscopy. Four *in silico* molecular models were developed that demonstrated the interaction of the newly discovered Compound 1 with the G-quartet and loop regions of the end-pasting sites present in the hTel22 G-quadruplex structures,. Furthermore, Compound 1 was tested in cancer cells and showed good growth inhibitory properties. These results supported that an integrated platform of *in silico* virtual screening and *in vitro* biophysical and biological validation can be used to identify new lead compounds that have G-quadruplex binding activity.

CHAPTER V

CONCLUSION

In the current dissertation work, we presented three experimental approaches for investigating G-quadruplex structures. Our experimental approaches differ from approaches commonly used by others in that our methods do not significantly perturb the equilibrium of G-quadruplex structures in solution. Current approaches to study G-quadruplex structures often employ sequence modifications or changes to experimental conditions to select for a single conformation for structural study. Such strategies for resolving the polymorphism of G-quadruplex structures can result in drastic and unpredictable perturbation of the system. While the untested assumption is that these methods enrich for a member of the ensemble of species originally formed by the parent sequence. Such a perturbation can shift the equilibrium to favor species that might not actually form *in vitro* or *in vivo* and can have significant implications on what structure can be claimed as "biologically relevant."

In the first experimental approach, we employed size exclusion chromatography (SEC) to separate the G-quadruplex structures formed from the Pu27 *c-Myc* promoter sequence. The significance of *c-Myc* is its overexpression in many types of cancer (Nesbit *et al.*, 1999, Nilsson and Cleveland, 2003). Modulation of *c-Myc* expression through small molecule interaction with G-quadruplex structures might be an effective strategy for anti-cancer therapy. Examination by SEC revealed that the Pu27

314

sequence exists as a heterogeneous mixture of monomer and higher-order G-quadruplex structures. We observed that the distribution of G-quadruplex species in solution is highly sensitive to changes in experimental conditions including variations in buffer composition, salt concentration, DNA concentration, annealing procedure and temperature, and dialysis protocol. Our findings emphasized the need for accurate and detailed reporting of experimental procedures in studies investigating G-quadruplex structures. We investigated the interaction of the porphyrin TMPyP4 with Pu27 sequence. TMPyP4 binding has been shown previously to repress *c-Myc* expression in a luciferase assay. The binding mechanism of TMPyP4 to *c-Myc* and other G-quadruplexes remains controversial (Anantha *et al.*, 1998, Haq *et al.*, 1999, Seenisamy *et al.*, 2004, Freyer *et al.*, 2007, Wei *et al.*, 2009). We observed that in the mixture of G-quadruplex structures, TMPyp4 bound preferentially to the higher-order G-quadruplex species. Our findings emphasized that a proper analysis of small molecule binding to a G-quadruplex-forming oligonucleotide must also include an investigation of the effect of the G-quadruplex-forming oligonucleotides (QFOs) structural polymorphism on its interaction with small molecule inhibitors. Lastly, we compared our observation of the Pu27 sequence with four modified sequences reported in the literature. We observed that while the parent Pu27 sequence formed aggregates in solution, the modified sequences did not. The difference in behaviors between the parent and modified sequences suggested a fundamental difference in structures and can have implication on the relevance of the modified sequence and findings of such a study. Overall, our investigation added to the current knowledge on the *c-Myc* G-quadruplex structure as well as the knowledge on G-quadruplex structural polymorphism.

In our second experimental approach, we employed hydrodynamic bead modeling (HBM) to study the structural polymorphism associated with G-quadruplex formation by the hTel22 human telomere sequence. G-quadruplex formation in the telomere is of significance as stabilization of these structures by small molecules have been shown to inhibit telomerase (Riou *et al.*, 2002, Cuesta *et al.*, 2003, De Cian *et al.*, 2008, Lopes *et al.*, 2011, Rodriguez *et al.*, 2012). With the activation of telomerase in more than 90% of all cancers (Shay and Bacchetti, 1997), G-quadruplex-based anti-telomerase therapy is an attractive strategy for the development of anti-cancer therapeutics. To sample G-quadruplex structures for examination by HBM, we carried out microsecond timescale molecular dynamics simulations for ten different telomeric G-quadruplex forming oligonucleotides and sampled 100,000 total G-quadruplex structures. This represents one of the most extensive sampling of G-quadruplex structures by MD. We reported the successful parameterization of HYDROPRO (a program to calculate hydrodynamic properties using HBM) with G-quadruplex specific parameters. Using our calibrated parameters, we calculated the hydrodynamic properties for ten different variants of the hTel22 sequence. Consistent with a previous report (Li *et al.*, 2005), we showed that the hydrodynamic property of the crystal structure of hTel22 did not agree with the solution properties. We clustered the sample structures by calculated sedimentation coefficients and showed that even on the limited timescale of sampling (1 μs), the QFO still exists as a heterogeneous mixture of G-quadruplex structures. We noted a correlation between sedimentation, molecular shapes, water binding, and energetic stability. G-quadruplex structures that sediment at higher rates are more compact in shape, bound less water, and are lower in energy. As hydration plays an important role in ligand binding and

recognition we postulated that hydrodynamic properties can be used in the process of drug design to identify conformations that could interact more favorably with small-molecule inhibitors. To conclude our discussion of the telomeric G-quadruplex, we carried out grid mappings of water and ions surrounding the G-quadruplex structures and identified sites of ordered water binding similar to the classic "spine of hydration" observed for some duplex structures and also sites of external ions binding that might play a role in supporting the G-quadruplex structures. The findings, thus, demonstrated that molecular dynamics and hydrodynamic bead modeling can be used to predict the structures for a novel putative G-quadruplex-forming sequence. Overall, our investigation added to the current knowledge regarding the polymorphism of telomeric G-quadruplex and the applications of hydrodynamic and hydrodynamic bead modeling.

In our third experimental approach, we presented a screening platform that was successful in identify a new G-quadruplex-interacting small molecule. As we have previously discussed, there is interest in identifying new G-quadruplex-interacting small molecules as these might serve as lead compounds in the development of a novel class of anti-cancer therapeutics. The approach of using receptor-based drug design to develop new G-quadruplex-interacting small molecules is hindered by the lack of high-resolution structures of relevant QFOs with small molecules bound to it. In our approach, we presented an *in silico* generated model of an end-pasting site that was used for virtual screening. We identified Compound 1 as a potential G-quadruplex-interacting small molecule. The Compound 1 binding was initially confirmed using a thermal denaturation analysis of FRET-labeled G-quadruplex DNA. We employed fluorescent and circular dichroism spectroscopy to demonstrate that Compound 1 binds to G-quadruplex

317

structures by the same end-pasting mechanism identified in the virtual screen. Additionally, we submitted Compound 1 for testing with the NCI-60 DTP Human Tumor Cell Line Screen and reported that Compound 1 is cytotoxic to cancer cells. Overall, our findings suggested that Compound 1 might be suitable as a lead compound for development of G-quadruplex-based small molecule inhibitor.

In conclusion, the experimental approaches highlighted in this dissertation work represent a new paradigm for drug design. Current approach to rational drug design involves firstly the determination of a G-quadruplex structure for a given putative G-quadruplex-forming sequence using structural techniques including both low-resolution methods such as CD spectroscopy and high-resolution methods such as X-ray crystallography or NMR spectroscopy (Figure 81). From the structure, *in vitro* screenings are carried out on chemical libraries to identify possible hits, which are then validated by biophysical measurements such as spectroscopy and calorimetry. Preclinical testing (cell culture studies, animal studies, pharmacodynamics and pharmacokinetic determination) are carried out to characterize biological functions before clinical trials and the emergence of a new drug. As we demonstrated in Chapter II, the progression from a PQS to G-quadruplex structures can be severely hindered by the fact that the structural outcomes of G-quadruplex formation are highly polymorphic. In addition, this structural polymorphism can also affect the interpretation of biophysical studies of small molecules interaction with such sequences. As such, the current approach to drug discovery is not appropriate for G-quadruplex structures. We proposed a model-based approach to drug design (Figure 82) which used molecular modeling as a mean to validate pass experimental results and to guide future experiment designs. The experimental

approaches we presented are not limited in their applications either. The first and second approaches can be useful in investigating new QFOs. We have adapted the first approach to characterize a putative G-quadruplex-forming sequence from the *Zeb1* promoter. In our findings, we were successful in separating the monomer G-quadruplex species from the higher-order species. Furthermore, circular dichroism spectroscopy of the monomer G-quadruplex showed a signal characteristic of antiparallel G-quadruplex formation while the higher-order sample showed a signal characteristic of parallel G-quadruplex formation (Le *et al.*, unpublished data). In addition to its application in characterizing the structures of a QFO, the second approach can also be used to predict the structure of a new QFO. We have used this approach to predict structures for human telomere sequences forming two, three, four, and eight contiguous G-quadruplexes. Recently, we employed this approach to predict a structure for G-quadruplex formation in the *hTERT* promoter. In addition, we have adapted the second approach to be used with a software being developed to generate three dimensional G-quadruplex molecular models of guanine-rich DNA for hydrodynamic bead modeling. Lastly, we have adapted the biophysical experiments presented in the third approach into a validation platform to screen four large compound sets from NCI. Our screening was able to identify several known anti-cancer drugs as G-quadruplex-interacting agents. Overall, our findings showed that the approaches we presented can be used for the successful study of G-quadruplex structures.

Figure 81. Current strategy for the identifications of G-quadruplex-binding drugs
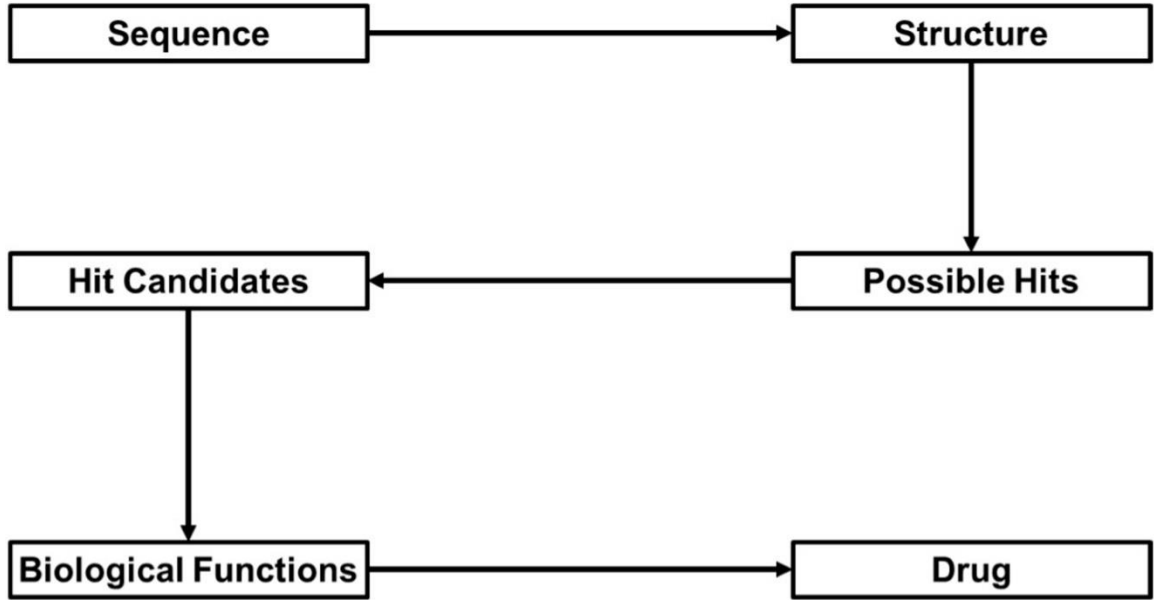
Figure 82. A model-based approach to G-quadruplex drug discovery.
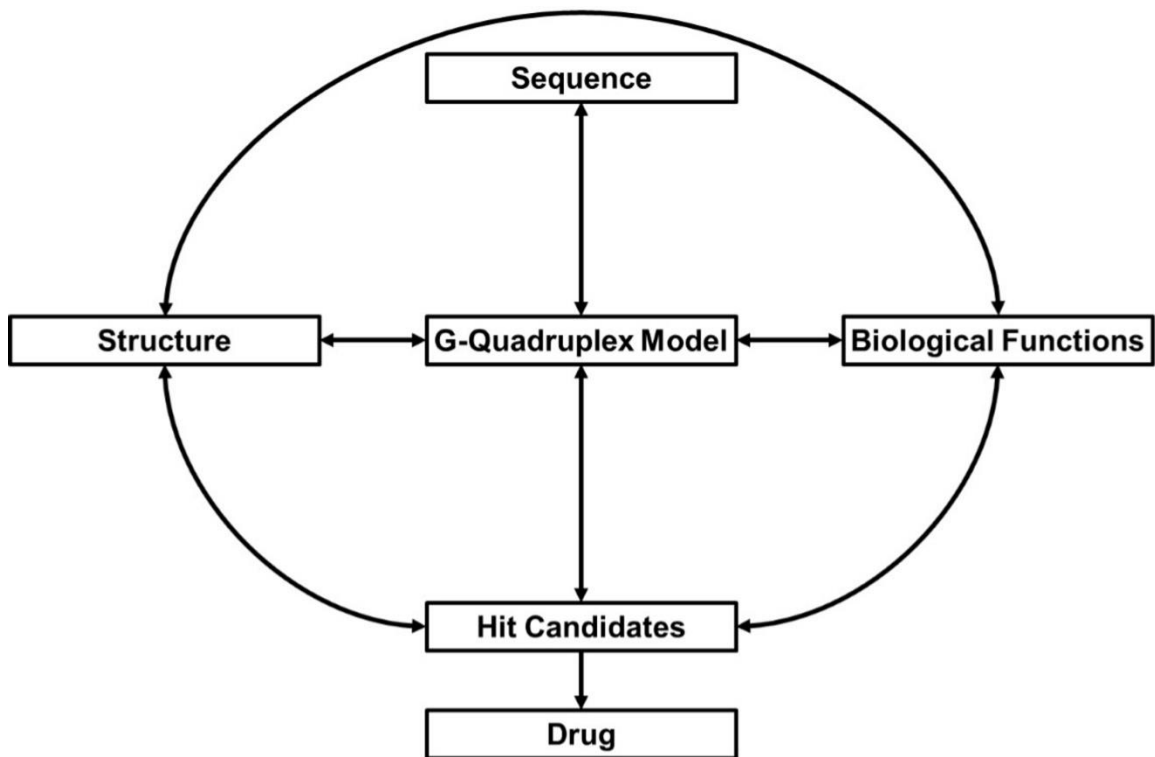
REFERENCES

Akhshi, P., G. Acton and G. Wu (2012). "Molecular Dynamics Simulations to Provide New Insights into the Asymmetrical Ammonium Ion Movement inside of the $[d(G_3T_4G_4)]_2$ G-Quadruplex DNA Structure." *The Journal of Physical Chemistry B* 116(31): 9363-9370.

Alcaro, S., C. Musetti, S. Distinto, M. Casatti, G. Zagotto, A. Artese, L. Parrotta, F. Moraca, G. Costa, F. Ortuso, E. Maccioni and C. Sissi (2013). "Identification and Characterization of New DNA G-Quadruplex Binders Selected by a Combination of Ligand and Structure-Based Virtual Screening Approaches." *Journal of Medicinal Chemistry* 56(3): 843-855.

Ambrus, A., D. Chen, J. X. Dai, R. Jones and D. Yang (2005). "Solution Structure of the Biologically Relevant G-Quadruplex Element in the Human *c-Myc* Promoter. Implications for G-Quadruplex Stabilization." *Biochemistry* 44(6): 2048-2058.

Anantha, N., M. Azam and R. Sheardy (1998). "Porphyrin Binding to Quadruplexed T_4G_4." *Biochemistry* 37(9): 2709-2714.

Arnott, S., R. Chandrasekaran and C. Marttila (1974). "Structures for Polyinosinic Acid and Polyguanylic Acid." *Biochemical Journal* 141(2): 537-543.

Arnott, S. and E. Selsing (1974). "Structures for the Polynucleotide Complexes Poly(dA) • Poly(dT) and Poly(dT) • Poly(dA) • Poly(dT)." *Journal of Molecular Biology* 88(2): 509-521.

Arora, A and B Suess (2011). "An RNA G-Quadruplex in the 3' UTR of the Proto-Oncogene *PIM1* Represses Translation." *RNA Biology* 8(5): 802-805.

Arora, A. and S. Maiti (2009). "Differential Biophysical Behavior of Human Telomeric RNA and DNA Quadruplex." *The Journal of Physical Chemistry B* 113(30): 10515-10520.

Balasubramanian, S. and S. Neidle (2009). "G-Quadruplex Nucleic Acids as Therapeutic Targets." *Current Opinion in Chemical Biology* 13(3): 345-353.

Balasubramanian, S., L. Hurley and S. Neidle (2011). "Targeting G-Quadruplexes in Gene Promoters: A Novel Anticancer Strategy?" *Nature Reviews Drug Discovery* 10(4): 261-275.

Bang, I. (1910). "Untersuchungen Über Die Guanylsäure." *Biochemische Zeitschrift.* 26: 293-311.

Baral, A., P. Kumar, R. Pathak and S. Chowdhury (2013). "Emerging Trends in G-Quadruplex Biology - Role in Epigenetic and Evolutionary Events." *Molecular BioSystems* 9(7): 1568-1575.

Bashkirov, V., H. Scherthan, J. Solinger, J.M. Buerstedde and W.Di. Heyer (1997). "A Mouse Cytoplasmic Exoribonuclease (Mxrn1p) with Preference for G4 Tetraplex Substrates." *The Journal of Cell Biology* 136(4): 761-773.

Biffi, G., D. Tannahill, J. McCafferty and S. Balasubramanian (2013). "Quantitative Visualization of DNA G-Quadruplex Structures in Human Cells." *Nature Chemistry* 5(3): 182-186.

Blackburn, E. (1991). "Structure and Function of Telomeres." *Nature* 350(6319): 569-573.

Bloomfield, V., W. Dalton and K. Van Holde (1967). "Frictional Coefficients of Multisubunit Structures. I. Theory." *Biopolymers* 5(2): 135-148.

Blume, S., V. Guarcello, W. Zacharias and D. Miller (1997). "Divalent Transition Metal Cations Counteract Potassium-Induced Quadruplex Assembly of Oligo(dG) Sequences." *Nucleic Acids Research* 25: 617-625.

Bochman, M., K. Paeschke and V. Zakian (2012). "DNA Secondary Structures: Stability and Function of G-Quadruplex Structures." *Nature Reviews Genetics* 13(11): 770-780.

Bonifacio, L., F. Church and M. Jarstfer (2008). "Effect of Locked-Nucleic Acid on a Biologically Active G-Quadruplex. A Structure-Activity Relationship of the Thrombin Aptamer." *International Journal of Molecular Sciences* 9(3): 422-433.

325

Bonnal, S., C. Schaeffer, L. Créancier, S. Clamens, H. Moine, A.C. Prats and S. Vagner (2003). "A Single Internal Ribosome Entry Site Containing a G Quartet RNA Structure Drives Fibroblast Growth Factor 2 Gene Expression at Four Alternative Translation Initiation Codons." *Journal of Biological Chemistry* 278(41): 39330-39336.

Brahms, J. and C. Sadron (1966). "Ionic and Hydrogen Bond Interactions Contributing to the Conformational Stability of Polynucleotides." *Nature* 212(5068): 1309-1312.

Broitman, S., D. Im and J. Fresco (1987). "Formation of the Triple-Stranded Polynucleotide Helix, Poly(A•A•U)*." *Proceedings of the National Academy of Sciences* 84(15): 5120-5124.

Brooks, T. and L. Hurley (2010). "Targeting *Myc* Expression through G-Quadruplexes." *Genes and Cancer* 1(6): 641-649.

Brooks, T., S. Kendrick and L. Hurley (2010). "Making Sense of G-Quadruplex and I-Motif Functions in Oncogene Promoters." *The FEBS Journal* 277(17): 3459-3469.

Brown, R., F. Danford, V. Gokhale, L. Hurley and T. Brooks (2011). "Demonstration That Drug-Targeted Downregulation of *Myc* in Non-Hodgkins Lymphoma Is Directly Mediated through the Promoter G-Quadruplex." *Journal of Biological Chemistry* 286(47):41018-41027.

Bryan, T. and P. Baumann (2011). "G-Quadruplexes: From Guanine Gels to Chemotherapeutics." *Molecular Biotechnology* 49(2): 198-208.

Bugaut, A. and S. Balasubramanian (2012). "5′-UTR RNA G-Quadruplexes: Translation Regulation and Targeting." *Nucleic Acids Research*.

Bugaut, A., R. Rodriguez, S. Kumari, S.T. Hsu and S. Balasubramanian (2010). "Small Molecule-Mediated Inhibition of Translation by Targeting a Native RNA G-Quadruplex." *Organic and Biomolecular Chemistry* 8(12): 2771-2776.

Burge, S., G. Parkinson, P. Hazel, A. Todd and S. Neidle (2006). "Quadruplex DNA: Sequence, Topology and Structure." *Nucleic Acids Research* 34: 5402-5415.

Buscaglia, R., M.C. Miller, W. Dean, R. Gray, A. Lane, J. Trent and J. Chaires (2013). "Polyethylene Glycol Binding Alters Human Telomere G-Quadruplex Structure by Conformational Selection." *Nucleic Acids Research* 41: 7934-7946.

Byron, O. (2008). Hydrodynamic Modeling: The Solution Conformation of Macromolecules and Their Complexes. *Methods in Cell Biology.* J. Correia and H.W. Detrich, III, Academic Press. 84: 327-373.

Cang, X., J. Šponer and T. Cheatham (2011). "Insight into G-DNA Structural Polymorphism and Folding from Sequence and Loop Connectivity through Free Energy Analysis." *Journal of the American Chemical Society* 133(36): 14270-14279.

Castillo-Gonzalez, D., G. Perez-Machado, A. Guedin, J.L. Mergny and M. Cabrera-Perez (2013). "FDA-Approved Drugs Selected Using Virtual Screening Bind Specifically to G-Quadruplex DNA." *Current Pharmaceutical Design* 19(12): 2164-2173.

Chaires, J. (2006). "A Thermodynamic Signature for Drug-DNA Binding Mode." *Archives of Biochemistry and Biophysics* 453(1): 26-31.

Chaires, J. (2010). "Human Telomeric G-Quadruplex: Thermodynamic and Kinetic Studies of Telomeric Quadruplex Stability." *The FEBS Journal.* 277(5): 1098-1106.

Chantot, J.F. and W. Guschlbauer (1972). Mechanism of Gel Formation by Guanine Nucleosides. *The Purines: Theory and Experiment; Proceedings of an International Symposium Held in Jerusalem, 4-8 April 1971.* E. Bergman. Jerusalem, Israel, Israel Academy of Sciences and Humanities: 205-214.

Cheatham, T. and P. Kollman (1997). "Molecular Dynamics Simulations Highlight the Structural Differences among DNA:DNA, RNA:RNA, and DNA:RNA Hybrid Duplexes." *Journal of the American Chemical Society* 119(21): 4805-4825.

Chen, S., J. Tan, T. Ou, S. Huang, L. An, H. Luo, D. Li, L. Gu and Z. Huang (2011). "Pharmacophore-Based Discovery of Triaryl-Substituted Imidazole as New Telomeric G-Quadruplex Ligand." *Bioorganic and Medicinal Chemistry Letters* 21(3): 1004-1009.

327

Cheung, I., M. Schertzer, A. Rose and P. Lansdorp (2002). "Disruption of *Dog-1* in *Caenorhabditis Elegans* Triggers Deletions Upstream of Guanine-Rich DNA." *Nature Genetics* 31(4): 405-409.

Christiansen, J., M. Kofod and F. Nielsen (1994). "A Guanosine Quadruplex and Two Stable Hairpins Flank a Major Cleavage Site in Insulin-Like Growth Factor II mRNA." *Nucleic Acids Research* 22(25): 5709-5716.

Chung, W., B. Heddi, M. Tera, K. Iida, K. Nagasawa and A.T. Phan (2013). "Solution Structure of an Intramolecular (3 + 1) Human Telomeric G-Quadruplex Bound to a Telomestatin Derivative." *Journal of the American Chemical Society* 135(36): 13495-13501.

Collie, G. and G. Parkinson (2011). "The Application of DNA and RNA G-Quadruplexes to Therapeutic Medicines." *Chemical Society Reviews* 40(12): 5867-5892.

Cosconati, S., L. Marinelli, R. Trotta, A. Virno, L. Mayol, E. Novellino, A. J. Olson and A. Randazzo (2009). "Tandem Application of Virtual Screening and Nmr Experiments in the Discovery of Brand New DNA Quadruplex Groove Binders." *Journal of the American Chemical Society* 131(45): 16336-16337.

Cosconati, S., A. Rizzo, R. Trotta, B. Pagano, S. Iachettini, S. De Tito, I. Lauri, I. Fotticchia, M. Giustiniano, L. Marinelli, C. Giancola, E. Novellino, A. Biroccio and A. Randazzo (2012). "Shooting for Selective Druglike G-Quadruplex Binders: Evidence for Telomeric DNA Damage and Tumor Cell Death." *Journal of Medicinal Chemistry* 55(22): 9785-9792.

Cuesta, J., M. Read and S. Neidle (2003). "The Design of G-Quadruplex Ligands as Telomerase Inhibitors." *Mini-Reviews in Medicinal Chemistry* 3(1): 11-21.

Dai, J., M. Carver, C. Punchihewa, R. Jones and D. Yang (2007a). "Structure of the Hybrid-2 Type Intramolecular Human Telomeric G-Quadruplex in K^+ Solution: Insights into Structure Polymorphism of the Human Telomeric Sequence." *Nucleic Acids Research* 35(15): 4927-4940.

Dai, J., M. Carver and D. Yang (2008). "Polymorphism of Human Telomeric Quadruplex Structures." *Biochimie* 90(8): 1172-1183.

Dai, J., D. Chen, R. Jones, L. Hurley and D. Yang (2006). "NMR Solution Structure of the Major G-Quadruplex Structure Formed in the Human *BCL2* Promoter Region." *Nucleic Acids Research* 34(18): 5133-5144.

Dai, J., C. Punchihewa, A. Ambrus, D. Chen, R. Jones and D. Yang (2007b). "Structure of the Intramolecular Human Telomeric G-Quadruplex in Potassium Solution: A Novel Adenine Triple Formation." *Nucleic Acids Research* 35(7): 2440-2450.

Dailey, M., C. Hait, P. Holt, J. Maguire, J. Meier, M.C. Miller, L. Petraccone and J. Trent (2009). "Structure-Based Drug Design: From Nucleic Acid to Membrane Protein Targets." *Experimental and Molecular Pathology* 86(3): 141-150.

Dailey, M., M.C. Miller, P. Bates, A. Lane and J. Trent (2010). "Resolution and Characterization of the Structural Polymorphism of a Single Quadruplex-Forming Sequence." *Nucleic Acids Research* 38: 4877-4888.

Dam, J. and P. Schuck (2004). Calculating Sedimentation Coefficient Distributions by Direct Modeling of Sedimentation Velocity Concentration Profiles. *Methods in Enzymology* L.J. Michael and B. Ludwig, Academic Press. 384: 185-212.

Đapić, V., V. Abdomerović, R. Marrington, J. Peberdy, A. Rodger, J. Trent and P. Bates (2003). "Biophysical and Biological Properties of Quadruplex Oligodeoxyribonucleotides." *Nucleic Acids Research* 31(8): 2097-2107.

Davies, D. (2005). "A Quiet Life with Proteins." *Annual Review of Biophysics and Biomolecular Structure* 34(1): 1-20.

Davis, J. (2004). "G-Quartets 40 Years Later: From 5′-GMP to Molecular Biology and Supramolecular Chemistry." *Angewandte Chemie International Edition* 43(6): 668-698.

De Armond, R., S. Wood, D. Sun, L. Hurley and S. Ebbinghaus (2005). "Evidence for the Presence of a Guanine Quadruplex Forming Region within a Polypurine Tract of the Hypoxia Inducible Factor 1 Alpha Promoter." *Biochemistry* 44(49): 16341-16350.

De Cian, A., L. Lacroix, C. Douarre, N. Temime-Smaali, C. Trentesaux, J. Riou and J.L. Mergny (2008). "Targeting Telomeres and Telomerase." *Biochimie* 90: 131-155.

329

De, S. and F. Michor (2011). "DNA Secondary Structures and Epigenetic Determinants of Cancer Genome Evolution." *Nature Structural and Molecular Biology* 18(8): 950-955.

Deng, J., Y. Xiong and M. Sundaralingam (2001). "X-Ray Analysis of an RNA Tetraplex (UGGGGU)$_4$ with Divalent Sr^{2+} Ions at Subatomic Resolution (0.61 Å)." *Proceedings of the National Academy of Sciences* 98(24): 13665-13670.

Downward, J. (2003). "Targeting *RAS* Signalling Pathways in Cancer Therapy." *Nature Reviews Cancer* 3(1): 11-22.

Drygin, D., A. Siddiqui-Jain, S. O'Brien, M. Schwaebe, A. Lin, J. Bliesath, C. Ho, C. Proffitt, K. Trent, J. Whitten, J. Lim, D. Von Hoff, K. Anderes, and W. Rice (2009). "Anticancer Activity of CX3543: A Direct Inhibitor of rRNA Biogenesis." *Cancer Research* 69(19): 7653-7661.

Du, Z., P. Kong, Y. Gao and N. Li (2007). "Enrichment of G4 DNA Motif in Transcriptional Regulatory Region of Chicken Genome." *Biochemical and Biophysical Research Communications* 354(4): 1067-1070.

Du, Z., Y. Zhao and N. Li (2008). "Genome-Wide Analysis Reveals Regulatory Role of G4 DNA in Gene Transcription." *Genome Research* 18(2): 233-241.

Du, Z., Y. Zhao and N. Li (2009). "Genome-Wide Colonization of Gene Regulatory Elements by G4 DNA Motifs." *Nucleic Acids Research* 37(20): 6784-6798.

Duca, M., P. Vekhoff, K. Oussedik, L. Halby and P. Arimondo (2008). "The Triple Helix: 50 Years Later, the Outcome." *Nucleic Acids Research* 36(16): 5123-5138.

Düchler, M. (2012). "G-Quadruplexes: Targets and Tools in Anticancer Drug Design." *Journal of Drug Targeting* 20(5): 389-400.

Eddy, J. and N. Maizels (2006). "Gene Function Correlates with Potential for G4 DNA Formation in the Human Genome." *Nucleic Acids Research* 34(14): 3887-3896.

Eddy, J. and N. Maizels (2008). "Conserved Elements with Potential to Form Polymorphic G-Quadruplex Structures in the First Intron of Human Genes." *Nucleic Acids Research* 36(4): 1321-1333.

Edmunds, C., L. Simpson and J. Sale (2008). "PCNA Ubiquitination and REV1 Define Temporally Distinct Mechanisms for Controlling Translesion Synthesis in the Avian Cell Line DT40." *Molecular Cell* 30(4): 519-529.

Egli, M., N. Usman, S. Zhang and A. Rich (1992). "Crystal Structure of an Okazaki Fragment at 2-Å Resolution." *Proceedings of the National Academy of Sciences* 89(2): 534-538.

Esposito, V., A. Randazzo, G. Piccialli, L. Petraccone, C. Giancola and L. Mayol (2004). "Effects of an 8-Bromodeoxyguanosine Incorporation on the Parallel Quadruplex Structure [d(TGGGT)](4)." *Organic and Biomolecular Chemistry* 2(3): 313-318.

Esposito, V., A. Virgilio, A. Randazzo, A. Galeone and L. Mayol (2005). "A New Class of DNA Quadruplexes Formed by Oligodeoxyribonucleotides Containing a 3 '-3 ' or 5 '-5 ' Inversion of Polarity Site." *Chemical Communications*(31): 3953-3955.

Esposito, V., A. Virgilio, A. Pepe, G. Oliviero, L. Mayol and A. Galeone (2009). "Effects of the Introduction of Inversion of Polarity Sites in the Quadruplex Forming Oligonucleotide TGGGT." *Bioorganic and Medicinal Chemistry* 17(5): 1997-2001.

Fadrná, E., N. Špačková, R. Štefl, J. Koča, T. Cheatham and J. Šponer (2004). "Molecular Dynamics Simulations of Guanine Quadruplex Loops: Advances and Force Field Limitations." *Biophys. Journal* 87(1): 227-242.

Felsenfeld, G., D. Davies and A. Rich (1957). "Formation of a Three-Stranded Polynucleotide Molecule." *Journal of the American Chemical Society* 79(8): 2023-2024.

Fernandes, M., A. Ortega, M. López Martínez and J. García de la Torre (2002). "Calculation of Hydrodynamic Properties of Small Nucleic Acids from Their Atomic Structure." *Nucleic Acids Research* 30(8): 1782-1788.

Franklin, R. and R. Gosling (1953a). "The Structure of Sodium Thymonucleate Fibres. I. The Influence of Water Content." *Acta Crystallographica* 6(8-9): 673-677.

Franklin, R. and R. Gosling (1953b). "The Structure of Sodium Thymonucleate Fibres. Ii. The Cylindrically Symmetrical Patterson Function." *Acta Crystallographica* 6(8-9): 678-685.

Franklin, R. and R. Gosling (1953c). "Molecular Configuration in Sodium Thymonucleate." *Nature* 171(4356): 740-741.

Franklin, R. and R. Gosling (1953d). "Evidence for 2-Chain Helix in Crystalline Structure of Sodium Deoxyribonucleate." *Nature* 172(4369): 156-157.

Fresco, J. and J. Massoulie (1963). "Polynucleotides V. Helix-Coil Transition of Polyriboguanylic Acid." *Journal of the American Chemical Society* 85(9): 1352-1353.

Fresco, J. and D. Su (1962). "Polynucleotides: IV. Synthesis of Polyriboguanylic Acid Catalyzed by Polynucleotide Phosphorylase." *Journal of Biological Chemistry* 237(10): PC3305-PC3306.

Freyer, M., R. Buscaglia, K. Kaplan, D. Cashman, L. Hurley and E. Lewis (2007). "Biophysical Studies of the *c-Myc* NHE III1 Promoter: Model Quadruplex Interactions with a Cationic Porphyrin." *Biophysical Journal* 92(6): 2007-2015.

Galeone, A., L. Mayol, A. Virgilio, A. Virno and A. Randazzo (2008). "A Further Contribution to the Extreme Variability of Quadruplex Structures from Oligodeoxyribonucleotides Containing Inversion of Polarity Sites in the G-Tract." *Molecular BioSystems* 4(5): 426-430.

Garbett, N., P. Ragazzon and J. Chaires (2007). "Circular Dichroism to Determine Binding Mode and Affinity of Ligand-DNA Interactions." *Nature Protocols* 2(12): 3166-3172.

García-Sosa, A. (2013). "Hydration Properties of Ligands and Drugs in Protein Binding Sites: Tightly-Bound, Bridging Water Molecules and Their Effects and Consequences on Molecular Design Strategies." *Journal of Chemical Information and Modeling* 53(6): 1388-1405.

García de la Torre, J. (2001). "Hydration from Hydrodynamics. General Considerations and Applications of Bead Modelling to Globular Proteins." *Biophysical Chemistry* 93(2-3): 159-170.

Gasparini, G. (2000). "Prognostic Value of Vascular Endothelial Growth Factor in Breast Cancer." *The Oncologist* 5(suppl 1): 37-44.

Gellert, M., M. Lipsett and D. Davies (1962). "Helix Formation by Guanylic Acid." *Proceedings of the National Academy of Sciences* 48(12): 2013-2018.

Ghosh, A. and M. Bansal (2003). "A Glossary of DNA Structures from A to Z." *Acta Crystallographica Section D* 59(4): 620-626.

González, V., K. Guo, L. Hurley and D. Sun (2009). "Identification and Characterization of Nucleolin as a *c-Myc* G-Quadruplex-Binding Protein." *Journal of Biological Chemistry* 284(35): 23622-23635.

Grand, C., H. Han, R. Muñoz, S. Weitman, D. Von Hoff, L. Hurley and D. Bearss (2002). "The Cationic Porphyrin TMPyP4 Down-Regulates *c-MYC* and Human Telomerase Reverse Transcriptase Expression and Inhibits Tumor Growth *in vivo*." *Molecular Cancer Therapeutics* 1(8): 565-573.

Gray, R. and J. Chaires (2011). "Linkage of Cation Binding and Folding in Human Telomeric Quadruplex DNA." *Biophysical Chemistry* 159(1): 205-209.

Gray, R., J. Li and J. Chaires (2009a). "Energetics and Kinetics of a Conformational Switch in G-Quadruplex DNA." *The Journal of Physical Chemistry B* 113(9): 2676-2683.

Gray, R., L. Petraccone, J. Trent and J. Chaires (2009b). "Characterization of a K^+-Induced Conformational Switch in a Human Telomeric DNA Oligonucleotide Using 2-Aminopurine Fluorescence." *Biochemistry* 49: 179-194.

Gros, J., A. Avino, J. de la Osa, C. Gonzalez, L. Lacroix, A. Perez, M. Orozco, R. Eritja and J.L. Mergny (2008). "8-Amino Guanine Accelerates Tetramolecular G-Quadruplex Formation." *Chemical Communications* (25): 2926-2928.

Guo, S., L. Colbert, M. Fuller, Y. Zhang and R. Gonzalez-Perez (2010). "Vascular Endothelial Growth Factor Receptor-2 in Breast Cancer." *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* 1806(1): 108-121.

Guschlbauer, W., J.F. Chantot and D. Thiele (1990). "Four-Stranded Nucleic Acid Structures 25 Years Later: From Guanosine Gels to Telomer DNA." *Journal of Biomolecular Structure and Dynamics* 8(3): 491-511.

333

Haider, S. and S. Neidle (2010). "Molecular Modeling and Simulation of G-Quadruplexes and Quadruplex-Ligand Complexes." *Methods in Molecular. Biology* P. Bauman, Humana Press 608: 17-37.

Haider, S., G. Parkinson and S. Neidle (2008). "Molecular Dynamics and Principal Components Analysis of Human Telomeric Quadruplex Multimers." *Biophysical Journal* 95(1): 296-311.

Halder, K., M. Wieland and J. Hartig (2009). "Predictable Suppression of Gene Expression by 5′-UTR-Based RNA Quadruplexes." *Nucleic Acids Research* 37(20): 6811-6817.

Hall, A., C. Marshall, N. Spurr and R. Weiss (1983). "Identification of Transforming Gene in Two Human Sarcoma Cell Lines as a New Member of the *Ras* Gene Family Located on Chromosome 1." *Nature* 303(5916): 396-400.

Han, H. and L. Hurley (2000). "G-Quadruplex DNA: A Potential Target for Anti-Cancer Drug Design." *Trends in Pharmacolical Sciences* 21(4): 136-142.

Hanahan, D. and R. Weinberg (2000). "The Hallmarks of Cancer." *Cell* 100(1): 57-70.

Hanahan, D. and R. Weinberg (2011). "Hallmarks of Cancer: The Next Generation." *Cell* 144(5): 646-674.

Hansel, R., S. Foldynova-Trantirkova, V. Dotsch and L. Trantirek (2013a). "Investigation of Quadruplex Structure under Physiological Conditions Using in-cell NMR." *Topics in Current Chemistry* J. Chaires and D. Graves, Springer Berlin Heidelberg 330: 47-65.

Hansel, R., F. Lohr, L. Trantirek and V. Dotsch (2013b). "High-Resolution Insight into G-Overhang Architecture." *Journal of the American Chemical Society* 135(7): 2816-2824.

Hänsel, R., F, Löhr, S. Foldynová-Trantírková, E. Bamberg, L. Trantírek and V. Dötsch (2011). "The Parallel G-Quadruplex Structure of Vertebrate Telomeric Repeat Sequences Is Not the Preferred Folding Topology under Physiological Conditions." *Nucleic Acids Research* 39(13): 5768-5775.

334

Haq, I., J. Trent, B. Chowdhry and T. Jenkins (1999). "Intercalative G-Tetraplex Stabilization of Telomeric DNA by a Cationic Porphyrin." *Journal of the American Chemical Society* 121(9): 1768-1779.

Harding, S. E. (2002). Frictional Coefficient, Ratio. *Encyclopedia of Molecular Biology*, John Wiley *and* Sons, Inc.

Hayward, S. and B. Groot (2008). Normal Modes and Essential Dynamics. *Molecular Modeling of Proteins*. Editor: A. Kukol, Humana Press. 443: 89-106.

Heddi, B. and A.T. Phan (2011). "Structure of Human Telomeric DNA in Crowded Solution." *Journal of the American Chemical Society* 133(25): 9824-9833.

Heer, K., H. Kumar, J. Read, J. Fox, J. Monson and M. Kerin (2001). "Serum Vascular Endothelial Growth Factor in Breast Cancer: Its Relation with Cancer Type and Estrogen Receptor Status." *Clinical Cancer Research* 7(11): 3491-3494.

Hélène, C. (1991). "The Anti-Gene Strategy: Control of Gene Expression by Triplex-Forming-Oligonucleotides." *Anti-Cancer Drug Design* 6(6): 569-584.

Henderson, E., C. Hardin, S. Walk, I. Tinoco Jr. and E. Blackburn (1987). "Telomeric DNA Oligonucleotides Form Novel Intramolecular Structures Containing Guanine•Guanine Base Pairs." *Cell* 51(6): 899-908.

Holmes, K., O. Roberts, A. Thomas and M. Cross (2007). "Vascular Endothelial Growth Factor Receptor-2: Structure, Function, Intracellular Signalling and Therapeutic Inhibition." *Cellular Signalling* 19(10): 2003-2012.

Holt, P., J. Chaires and J. Trent (2008). "Molecular Docking of Intercalators and Groove-Binders to Nucleic Acids Using Autodock and Surflex." *Journal of Chemical Information and Modeling* 48(8): 1602-1615.

Holt, P., P. Ragazzon, L. Strekowski, J. Chaires and J. Trent (2009). "Discovery of Novel Triple Helical DNA Intercalators by an Integrated Virtual and Actual Screening Platform." *Nucleic Acids Research* 37(4): 1280-1287.

Holt, P., R. Buscaglia, J. Trent and J. Chaires (2011). "A Discovery Funnel for Nucleic Acid Binding Drug Candidates." *Drug Development Research* 72(2): 178-186.

Hoogsteen, K. (1959). "The Structure of Crystals Containing a Hydrogen-Bonded Complex of 1-Methylthymine and 9-Methyladenine." *Acta Crystallographica* 12(10): 822-823.

Hoogsteen, K. (1963). "The Crystal and Molecular Structure of a Hydrogen-Bonded Complex between 1-Methylthymine and 9-Methyladenine." *Acta Crystallographica* 16(9): 907-916.

House, S., C. Bolte, M. Zhou, T. Doetschman, R. Klevitsky, G. Newman and J. Schultz (2003). "Cardiac-Specific Overexpression of Fibroblast Growth Factor-2 Protects against Myocardial Dysfunction and Infarction in a Murine Model of Low-Flow Ischemia." *Circulation* 108(25): 3140-3148.

Howard, F. and H.T. Miles (1982). "Poly(Inosinic Acid) Helixes: Essential Chelation of Alkali Metal Ions in the Axial Channel." *Biochemistry* 21(26): 6736-6745.

Howard, F., H.T. Miles and P. Ross (1995). "The Poly(dT)•2poly(dA) Triple Helix." *Biochemistry* 34(21): 7135-7144.

Hsu, S., P. Varnai, A. Bugaut, A. P. Reszka, S. Neidle and S. Balasubramanian (2009). "A G-Rich Sequence within the C-Kit Oncogene Promoter Forms a Parallel G-Quadruplex Having Asymmetric G-Tetrad Dynamics." *Journal of the American Chemical Society* 131(37): 13399-13409.

Huppert, J. and S. Balasubramanian (2005a). "Prevalence of Quadruplexes in the Human Genome." *Nucleic Acids Research* 33(9): 2908-2916.

Huppert, J. and S. Balasubramanian (2007a). "G-Quadruplexes in Promoters Throughout the Human Genome." *Nucleic Acids Research* 35(2): 406-413.

Huppert, J. (2008). "Four-Stranded Nucleic Acids: Structure, Function and Targeting of G-Quadruplexes." *Chemical Society Reviews* 37(7): 1375-1384.

Huppert, J., A. Bugaut, S. Kumari and S. Balasubramanian (2008). "G-Quadruplexes: The Beginning and End of UTRs." *Nucleic Acids Research* 36(19): 6260-6268.

Iball, J., C. Morgan and H. Wilson (1963). "Fibres of Guanine Nucleosides and Nucleotides." *Nature* 199(4894): 688-689.

Irwin, J. and B. Shoichet (2005). "Zinc – A Free Database of Commercially Available Compounds for Virtual Screening." *Journal of Chemical Information and Modeling* 45(1): 177-182.

Islam, B., M. Sgobba, C, Laughton, M. Orozco, J. Sponer, S. Neidle and S. Haider (2013). "Conformational Dynamics of the Human Propeller Telomeric DNA Quadruplex on a Microsecond Time Scale." *Nucleic Acids Research* 41(4): 2723-2735.

Jain, A. (2003). "Surflex: Fully Automatic Flexible Molecular Docking Using a Molecular Similarity-Based Search Engine." *Journal of Medicinal Chemistry* 46(4): 499-511.

Joachimi, A., A. Benz and J. Hartig (2009). "A Comparison of DNA and RNA Quadruplex Structures and Stabilities." *Bioorganic and Medicinal Chemistry* 17(19): 6811-6815.

Joung, I. and T. Cheatham (2008). "Determination of Alkali and Halide Monovalent Ion Parameters for Use in Explicitly Solvated Biomolecular Simulations." *The Journal of Physical Chemistry B* 112(30): 9020-9041.

Kang, C., X. Zhang, R. Ratliff, R. Moyzis and A. Rich (1992). "Crystal Structure of Four-Stranded Oxytricha Telomeric DNA." *Nature* 356(6365): 126-131.

Kar, R., P. Suryadevara, J. Jana, A. Bhunia and S. Chatterjee (2012). "Novel G-Quadruplex Stabilizing Agents: *in-silico* Approach and Dynamics." *Journal of Biomolecular Structure and Dynamics* 31(12): 1497-1518.

Karkkainen, M. and T. Petrova (2000). "Vascular Endothelial Growth Factor Receptors in the Regulation of Angiogenesis and Lymphangiogenesis." *Oncogene* 19(49): 5598-5605.

Karsisiotis, A., N. Hessari, E. Novellino, G. Spada, A. Randazzo and M. Webba da Silva (2011). "Topological Characterization of Nucleic Acid G-Quadruplexes by UV Absorption and Circular Dichroism." *Angewandte Chemie International Edition* 50(45): 10645-10648.

Knauert, M. and P. Glazer (2001). "Triplex Forming Oligonucleotides: Sequence-Specific Tools for Gene Targeting." *Human Molecular Genetics* 10(20): 2243-2251.

337

Kollman, P., I. Massova, C. Reyes, B. Kuhn, S. Huo, L. Chong, M. Lee, T. Lee, Y. Duan, W. Wang, O. Donini, P. Cieplak, J. Srinivasan, D. Case and T. Cheatham. (2000). "Calculating Structures and Free Energies of Complex Molecules: Combining Molecular Mechanics and Continuum Models." *Accounts of Chemical Resesearch* 33(12): 889-897.

Kruisselbrink, E., V. Guryev, K. Brouwer, D. Pontier, E. Cuppen and M. Tijsterman (2008). "Mutagenic Capacity of Endogenous G4 DNA Underlies Genome Instability in *FANCJ*-Defective *C. Elegans*." *Current Biology* 18(12): 900-905.

Kumar, N. and S. Maiti (2007). "Role of Locked Nucleic Acid Modified Complementary Strand in Quadruplex/Watson-Crick Duplex Equilibrium." *The Journal of Physical Chemistry B* 111(42): 12328-12337.

Lane, A., J. Chaires, R. Gray and J. Trent (2008). "Stability and Kinetics of G-Quadruplex Structures." *Nucleic Acids Research* 36(17): 5482-5515.

Lane, A. (2012). "The Stability of Intramolecular DNA G-Quadruplexes Compared with Other Macromolecules." *Biochimie* 94: 277-286.

Le, H., M.C. Miller, R. Buscaglia, W. Dean, P. Holt, J. Chaires and J. Trent (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." *Organic and Biomolecular Chemistry* 10(47): 9393-9404.

Lee, H., D. Chan, F. Yang, H. Lam, S. Yan, C. Che, D. Ma and C. Leung (2010). "Identification of Natural Product Fonsecin B as a Stabilizing Ligand of *c-Myc* G-Quadruplex DNA by High-Throughput Virtual Screening." *Chemical Communications* 46(26): 4680-4682.

Levene, P. and W. Jacobs (1909). "Über Guanylsäure." *Berichte der deutschen chemischen Gesellschaft* 42(2): 2469-2473.

Li, J, J. Correia, L. Wang, J. Trent and J. Chaires (2005). "Not So Crystal Clear: The Structure of the Human Telomere G-Quadruplex in Solution Differs from That Present in a Crystal." *Nucleic Acids Research* 33(14): 4649-4659.

Li, Q., J. Xiang, X. Li, L. Chen, X. Xu, Y. Tang, Q. Zhou, L. Li, H. Zhang, H. Sun, A. Guan, Q. Yang, S. Yang and G. Xu (2009). "Stabilizing Parallel G-Quadruplex DNA by a New Class of Ligands: Two Non-Planar Alkaloids through Interaction in Lateral Grooves." *Biochimie* 91(7): 811-819.

Lim, K., S. Amrane, S. Bouaziz, W. Xu, Y. Mu, D. Patel, K.N. Luu and A.T. Phan (2009). "Structure of the Human Telomere in K^+ Solution: A Stable Basket-Type G-Quadruplex with Only Two G-Tetrad Layers." *Journal of the American Chemical Society* 131(12): 4301-4309.

Lim, K., L. Lacroix, D. Jia En Yue, J.K.C. Lim, J.M.W. Lim and A.T. Phan (2010). "Coexistence of Two Distinct G-Quadruplex Conformations in the *hTERT* Promoter." *Journal of the American Chemical Society* 132(35): 12331-12342.

Lipay, J. and M.R. Mihailescu (2009). "NMR Spectroscopy and Kinetic Studies of the Quadruplex Forming RNA r(UGGAGGU)." *Molecular BioSystems* 5(11): 1347-1355.

Lipinski, C., F. Lombardo, B.Dominy and P. Feeney (2001). "Experimental and Computational Approaches to Estimate Solubility and Permeability in Drug Discovery and Development Settings." *Advanced Drug Delivery Reviews* 46(1-3): 3-26.

Lipsett, M. (1964). "Complex Formation between Polycytidylic Acid and Guanine Oligonucleotides." *Journal of Biological Chemistry* 239(4): 1256-1260.

Liu, H., M. Kanagawa, A. Matsugami, Y. Tanaka, M. Katahira and S. Uesugi (2000). "NMR Study of a Novel RNA Quadruplex Structure." *Nucleic Acids Symposium Series* 44(1): 65-66.

Liu, H., A. Kugimiya, A. Matsugami, M. Katahira and S. Uesugi (2002a). "Quadruplex Structures of RNA 14-Mer, r(GGAGGUUUUGGAGG) and DNA 14-Mer, d(GGAGGTTTTGGAGG)." *Nucleic Acids Symposium Series* 2(1): 177-178.

Liu, H., A. Kugimiya, T. Sakurai, M. Katahira and S. Uesugi (2002b). "A Comparison of the Properties and the Solution Structure for RNA and DNA Quadruplexes Which Contain Two GGAGG Sequences Joined with a Tetranucleotide Linker." *Nucleosides, Nucleotides and Nucleic Acids* 21(11-12): 785-801.

339

Liu, H., A. Matsugami, M. Katahira and S. Uesugi (2002c). "A Dimeric RNA Quadruplex Architecture Comprised of Two G:G(:A):G:G(:A) Hexads, G:G:G:G Tetrads and Uuuu Loops." *Journal of Molecular Biology* 322(5): 955-970.

London, T., L. Barber, G. Mosedale, G. Kelly, S. Balasubramanian, I. Hickson, S. Boulton and K. Hiom (2008). "FANCJ Is a Structure-Specific DNA Helicase Associated with the Maintenance of Genomic G/C Tracts." *Journal of Biological Chemistry* 283(52): 36132-36139.

Lopes, J., A. Piazza, R. Bermejo, B. Kriegsman, A. Colosio, M.P. Teulade-Fichou, M. Foiani and A. Nicolas (2011). "G-Quadruplex-Induced Instability During Leading-Strand Replication." *EMBO J.* 30: 4033-4046.

Luedtke, N. (2009). "Targeting G-Quadruplex DNA with Small Molecules." *Chimia* 63: 134-139.

Luu, K.N., A.T. Phan, V. Kuryavyi, L. Lacroix and D. Patel (2006). "Structure of the Human Telomere in K^+ Solution: An Intramolecular (3 + 1) G-Quadruplex Scaffold." *Journal of the American Chemical Society* 128(30): 9963-9970.

Lv, L., Z. Guo, J. Wang and E. Wang (2012). "G-Quadruplex as Signal Transducer for Biorecognition Events." *Current Pharmaceutical Design* 18(14): 2076-2095.

Ma, D., D. Chan, W. Fu, H. He, H. Yang, S. Yan and C. Leung (2012a). "Discovery of a Natural Product-Like *c-Myc* G-Quadruplex DNA Groove-Binder by Molecular Docking." *PLoS ONE* 7(8): e43278.

Ma, D., T. Lai, F. Chan, W. Chung, R. Abagyan, Y. Leung and K. Wong (2008). "Discovery of a Drug-Like G-Quadruplex Binding Ligand by High-Throughput Docking." *ChemMedChem* 3(6): 881-884.

Ma, D., V. Ma, D. Chan, K. Leung, H. Zhong and C. Leung (2012b). "*In Silico* Screening of Quadruplex-Binding Ligands." *Methods* 57(1): 106-114.

Manning, G. (1978). "The Molecular Theory of Polyelectrolyte Solutions with Applications to the Electrostatic Properties of Polynucleotides." *Quarterly Reviews of Biophysics* 11(02): 179-246.

Marathias, V., M. Sawicki and P. Bolton (1999). "6-Thioguanine Alters the Structure and Stability of Duplex DNA and Inhibits Quadruplex DNA Formation." *Nucleic Acids Research* 27(14): 2860-2867.

Marcel, V., P. Tran, C. Sagne, G. Martel-Planche, L. Vaslin, M.P. Teulade-Fichou, J. Hall, J.L. Mergny, P. Hainaut and E. Van Dyck (2011). "G-Quadruplex Structures in *TP53* Intron 3: Role in Alternative Splicing and in Production of p53 mRNA Isoforms." *Carcinogenesis* 32(3): 271-278.

Marshall, C., A. Hall and R. Weiss (1982). "A Transforming Gene Present in Human Sarcoma Cell Lines." *Nature* 299(5879): 171-173.

Martadinata, H. and A.T. Phan (2009). "Structure of Propeller-Type Parallel-Stranded RNA G-Quadruplexes, Formed by Human Telomeric RNA Sequences in K^+ Solution." *Journal of the American Chemical Society* 131(7): 2570-2578.

Martadinata, H. and A.T. Phan (2013). "Structure of Human Telomeric RNA (TERRA): Stacking of Two G-Quadruplex Blocks in K^+ Solution." *Biochemistry* 52(13): 2176-2183.

Mathad, R., E. Hatzakis, J. Dai and D. Yang (2011). "*c-MYC* Promoter G-Quadruplex Formed at the 5′-End of NHE III1 Element: Insights into Biological Relevance and Parallel-Stranded G-Quadruplex Stability." *Nucleic Acids Research* 39(20): 9023-9033.

Matsugami, A., K. Ouhashi, M. Kanagawa, H. Liu, S. Kanagawa, S. Uesugi and M. Katahira (2001). "New Quadruplex Structure of GGA Triplet Repeat DNA – An Intramolecular Quadruplex Composed of a G:G:G:G Tetrad and G(:A):G(:A):G(:A):G Heptad, and Its Dimerization." *Nucleic Acids Symposium Series* 1(1): 271-272.

Mekmaysy, C., L. Petraccone, N. Garbett, P. Ragazzon, R. Gray, J. Trent and J. Chaires (2008). "Effect of O6-Methylguanine on the Stability of G-Quadruplex DNA." *Journal of the American Chemical Society* 130(21): 6710-6711.

Mergny, J.L., J. Li, L. Lacroix, S. Amrane and J. Chaires (2005). "Thermal Difference Spectra: A Specific Signature for Nucleic Acid Structures." *Nucleic Acids Research* 33(16): e138.

Miles, H.T. and J. Frazier (1978). "Poly(I) Helix Formation. Dependence on Size-Specific Complexing to Alkali Metal Ions." *Journal of the American Chemical Society* 100(25): 8037-8038.

Miller, M.C., R. Buscaglia, J. Chaires, A. Lane and J. Trent (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* 132(48): 17105-17107.

Miller, M.C., H. Le, W. Dean, P. Holt, J. Chaires and J. Trent (2011). "Polymorphism and Resolution of Oncogene Promoter Quadruplex-Forming Sequences." *Organic and Biomolecular Chemistry* 9(22): 7633-7637.

Miller, M.C. and J. Trent (2001). Resolution of Quadruplex Polymorphism by Size-Exclusion Chromatography. *Current Protocols in Nucleic Acid Chemistry*, John Wiley *and* Sons, Inc.

Mitsui, Y., R. Langridge, B. Shortle, C. Cantor, R. Grant, M. Kodama and R. Wells (1970). "Physical and Enzymatic Studies on Poly D(I-C)•Poly D(I-C), an Unusual Double-Helical DNA." *Nature* 228(5277): 1166-1169.

Miyoshi, D., H. Karimata and N. Sugimoto (2007). "Hydration Regulates the Thermodynamic Stability of DNA Structures under Molecular Crowding Conditions." *Nucleosides Nucleotides Nucleic Acids* 26(6-7): 589-595.

Miyoshi, D., A. Nakao and N. Sugimoto (2001). "Structural Transition of d(G$_4$T$_4$G$_4$) from Antiparallel to Parallel G-Quartet Induced by Divalent Cations." *Nucleic Acids Research Supplement* (1): 259-260.

Monchaud, D., C. Allain, H. Bertrand, N. Smargiasso, F. Rosu, V. Gabelica, A. De Cian, J. L. Mergny and M. P. Teulade-Fichou (2008). "Ligands Playing Musical Chairs with G-Quadruplex DNA: A Rapid and Simple Displacement Assay for Identifying Selective G-Quadruplex Binders." *Biochimie* 90(8): 1207-1223.

Monchaud, D., C. Allain and M. P. Teulade-Fichou (2006). "Development of a Fluorescent Intercalator Displacement Assay (G4-FID) for Establishing Quadruplex-DNA Affinity and Selectivity of Putative Ligands." *Bioorganic and Medicinal Chemistry Letters* 16(18): 4842-4845.

Morgan, A. and R. Wells (1968). "Specificity of the Three-Stranded Complex Formation between Double-Stranded DNA and Single-Stranded RNA Containing Repeating Nucleotide Sequences." *Journal of Molecular Biology* 37(1): 63-80.

Morris, M., Y. Negishi, C. Pazsint, J. Schonhoft and S. Basu (2010). "An RNA G-Quadruplex Is Essential for Cap-Independent Translation Initiation in Human *VEGF* IRES." *Journal of the American Chemical Society* 132(50): 17831-17839.

Müller, W, H. Bünemann and N. Dattagupta (1975). "Interactions of Heteroaromatic Compounds with Nucleic Acids." *European Journal of Biochemistry* 54(1): 279-291.

Müller, W. and D. Crothers (1975). "Interactions of Heteroaromatic Compounds with Nucleic Acids." *European Journal of Biochemistry* 54(1): 267-277.

Müller, W. and F. Gautier (1975). "Interactions of Heteroaromatic Compounds with Nucleic Acids." *European Journal of Biochemistry* 54(2): 385-394.

Murat, Pierre and S. Balasubramanian (2014). "Existence and Consequences of G-Quadruplex Structures in DNA." *Current Opinion in Genetics and Development* 25(0): 22-29.

Neidle, S. (2010). "Human Telomeric G-Quadruplex: The Current Status of Telomeric G-Quadruplexes as Therapeutic Targets in Human Cancer." *The FEBS Journal* 277(5): 1118-1125.

Neidle, S. (2009). "The Structures of Quadruplex Nucleic Acids and Their Drug Complexes." *Current Opinion in Structural Biology* 19(3): 239-250.

Neidle, S. and M. Read (2000). "G-Quadruplexes as Therapeutic Targets." *Biopolymers* 56(3): 195-208.

Nesbit, C., J. Tersak and E. Prochownik (1999). "*MYC* Oncogenes and Human Neoplastic Disease." *Oncogene* 18(19): 3004-3016.

Niermann, M., M. Bolten and W. Eimer (1999). "Optimization of the Hydrodynamic Bead Model for the Analysis of DNA Conformations in Solution." *The Journal of Physical Chemistry B* 103(45): 10065-10074.

343

Nilsson, J. and J. Cleveland (2003). "*Myc* Pathways Provoking Cell Suicide and Cancer." *Oncogene* 22(56): 9007-9021.

Olsen, C., W. Gmeiner and L. Marky (2006). "Unfolding of G-Quadruplexes: Energetic, and Ion and Water Contributions of G-Quartet Stacking." *The Journal of Physical Chemistry B* 110(13): 6962-6969.

Ortega, A., D. Amorós and J. García de la Torre (2011). "Prediction of Hydrodynamic and Other Solution Properties of Rigid Proteins from Atomic- and Residue-Level Models." *Biophysical Journal* 101(4): 892-898.

Ou, T.M., Y. Lu, J. Tan, Z. Huang, K. Wong and L. Gu (2008). "G-Quadruplexes: Targets in Anticancer Drug Design." *ChemMedChem* 3(5): 690-713.

Ou, T.M., J. Lin, Y.J. Lu, J.Q. Hou, J.H. Tan, S.H. Chen, Z. Li, Y.P. Li, D. Li, L.Q. Gu and Z.S. Huang (2011). "Inhibition of Cell Proliferation by Quindoline Derivative (SYUIQ-05) through Its Preferential Interaction with *c-Myc* Promoter G-Quadruplex." *Journal of Medicinal Chemistry* 54(16): 5671-5679.

Paeschke, K., M. Bochman, P. Daniela Garcia, P. Cejka, K. Friedman, S. Kowalczykowski and V. Zakian (2013). "Pif1 Family Helicases Suppress Genome Instability at G-Quadruplex Motifs." *Nature* 497(7450): 458-462.

Paeschke, K., J. Capra and V. Zakian (2011). "DNA Replication through G-Quadruplex Motifs Is Promoted by the *Saccharomyces Cerevisiae* Pif1 DNA Helicase." *Cell* 145(5): 678-691.

Pagano, B., C. Mattia, L. Cavallo, S. Uesugi, C. Giancola and F. Fraternali (2008). "Stability and Cations Coordination of DNA and RNA 14-mer G-Quadruplexes: A Multiscale Computational Approach." *The Journal of Physical Chemistry B* 112(38): 12115-12123.

Pan, J. and S. Zhang (2009). "Interaction between Cationic Zinc Porphyrin and Lead Ion Induced Telomeric Guanine Quadruplexes: Evidence for End-Stacking." *Journal of Biological Inorganic Chemistry* 14(3): 401-407.

Parkinson, G., M. Lee and S. Neidle (2002). "Crystal Structure of Parallel Quadruplexes from Human Telomeric DNA." *Nature* 417(6891): 876-880.

344

Patel, D., A.T. Phan and V. Kuryavyi (2007). "Human Telomere, Oncogenic Promoter and 5'-UTR G-Quadruplexes: Diverse Higher Order DNA and RNA Targets for Cancer Therapeutics." *Nucleic Acids Research* 35(22): 7429-7455.

Petraccone, L., N. Garbett, J. Chaires and J. Trent (2010). "An Integrated Molecular Dynamics (MD) and Experimental Study of Higher Order Human Telomeric Quadruplexes." *Biopolymers* 93(6): 533-548.

Petraccone, L., J. Trent and J. Chaires (2008). "The Tail of the Telomere." *Journal of the American Chemical Society* 130(49): 16530-16532.

Petraccone, L., C. Spink, J. Trent, N. Garbett, C. Mekmaysy, C. Giancola and J. Chaires (2011). "Structure and Stability of Higher-Order Human Telomeric Quadruplexes." *Journal of the American Chemical Society* 133(51): 20951-20961.

Petrovic, A. and P. Polavarapu (2008). "Quadruplex Structure of Polyriboinosinic Acid: Dependence on Alkali Metal Ion Concentration, pH and Temperature." *Journal of Physical Chemistry B* 112(7): 2255-2260.

Phan, A.T. (2010). "Human Telomeric G-Quadruplex: Structures of DNA and RNA Sequences." *The FEBS Journal* 277(5): 1107-1117.

Phan, A.T., V. Kuryavyi, S. Burge, S. Neidle and D. Patel (2007a). "Structure of an Unprecedented G-Quadruplex Scaffold in the Human *c-kit* Promoter." *Journal of the American Chemical Society* 129(14): 4386-4392.

Phan, A.T., V. Kuryavyi, H. Yan Gaw and D. Patel (2005). "Small-Molecule Interaction with a Five-Guanine-Tract G-Quadruplex Structure from the Human *MYC* Promoter." *Nature Chemical Biology* 1(3): 167-173.

Phan, A.T., V. Kuryavyi, K.N. Luu and D. Patel (2007b). "Structure of Two Intramolecular G-Quadruplexes Formed by Natural Human Telomere Sequences in K$^+$ Solution." *Nucleic Acids Research* 35(19): 6517-6525.

Phan, A.T., Yasha S. Modi and Dinshaw J. Patel (2004). "Propeller-Type Parallel-Stranded G-Quadruplexes in the Human *c-Myc* Promoter." *Journal of the American Chemical Society* 126(28): 8710-8716.

Pohl, F. and T. Jovin (1972). "Salt-Induced Co-Operative Conformational Change of a Synthetic DNA: Equilibrium and Kinetic Studies with Poly(dG-dC)." *Journal of Molecular Biology* 67(3): 375-396.

Poornima, C. and P. Dean (1995a). "Hydration in Drug Design. 1. Multiple Hydrogen-Bonding Features of Water Molecules in Mediating Protein-Ligand Interactions." *Journal of Computer-Aided Molecular Design* 9(6): 500-512.

Poornima, C. and P. Dean (1995b). "Hydration in Drug Design. 2. Influence of Local Site Surface Shape on Water Binding." *Journal of Computer-Aided Molecular Design* 9(6): 513-520.

Poornima, C. and P. Dean (1995c). "Hydration in Drug Design. 3. Conserved Water Molecules at the Ligand-Binding Sites of Homologous Proteins." *Journal of Computer-Aided Molecular Design* 9(6): 521-531.

Praseuth, D., A. Guieysse and C. Hélène (1999). "Triple Helix Formation and the Antigene Strategy for Sequence-Specific Control of Gene Expression." *Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression* 1489(1): 181-206.

Qi, J. and R. Shafer (2007). "Human Telomere Quadruplex: Refolding and Selection of Individual Conformers Via RNA/DNA Chimeric Editing." *Biochemistry* 46(25): 7599-7606.

Rawal, P., V. Kummarasetti, J. Ravindran, N. Kumar, K. Halder, R. Sharma, M. Mukerji, S. Das and S. Chowdhury (2006). "Genome-Wide Prediction of G4 DNA as Regulatory Motifs: Role in *Escherichia Coli* Global Regulation." *Genome Research* 16(5): 644-655.

Read, M., R. Harrison, B. Romagnoli, F. Tanious, S. Gowan, A. Reszka, W. Wilson, L. Kelland and S. Neidle (2001). "Structure-Based Design of Selective and Potent G Quadruplex-Mediated Telomerase Inhibitors." *Proceedings of the National Academy of Sciences* 98(9): 4844-4849.

Record, M.T., C. Anderson and T. Lohman (1978). "Thermodynamic Analysis of Ion Effects on the Binding and Conformational Equilibria of Proteins and Nucleic Acids: The Roles of Ion Association or Release, Ccreening, and Ion Effects on Water Activity." *Quarterly Reviews of Biophysics* 11(2): 103-178.

Reshetnikov, R., J. Sponer, O. Rassokhina, A. Kopylov, P. Tsvetkov, A. Makarov and A. Golovin (2011). "Cation Binding to 15-TBA Quadruplex DNA Is a Multiple-Pathway Cation-Dependent Process." *Nucleic Acids Research* 39(22): 9789-9802.

Rich, A. (1958). "The Molecular Structure of Polyinosinic Acid." *Biochimica et Biophysica Acta* 29(3): 502-509.

Rich, A. (1993). "DNA Comes in Many Forms." *Gene* 135(1–2): 99-109.

Rich, A. (1960). "A Hybrid Helix Containing Both Deoxyribose and Ribose Polynucleotides and Its Relation to the Transfer of Information between the Nucleic Acids." *Proceedings of the National Academy of Sciences* 46(8): 1044-1053.

Rich, A. and D. Davies (1956). "A New Two Stranded Helical Structure: Polyadenylic Acid and Polyuridylic Acid." *Journal of the American Chemical Society* 78(14): 3548-3549.

Riou, J.F., L. Guittat, P. Mailliet, A. Laoui, E. Renou, O. Petitgenet, F. Mégnin-Chanet, C. Hélène and J.L. Mergny (2002). "Cell Senescence and Telomere Shortening Induced by a New Series of Specific G-Quadruplex DNA Ligands." *Proceedings of the National Academy of Sciences* 99(5): 2672-2677.

Roberts, R. and D. Crothers (1992). "Stability and Properties of Double and Triple Helices: Dramatic Effects of RNA or DNA Backbone Composition." *Science* 258(5087): 1463-1466.

Rodriguez, R., K. Miller, J. Forment, C. Bradshaw, M. Nikan, S. Britton, T. Oelschlaegel, B. Xhemalce, S. Balasubramanian and S. Jackson (2012). "Small-Molecule–Induced DNA Damage Identifies Alternative DNA Structures in Human Genes." *Nature Chemical Biology* 8(3): 301-310.

Rosenberg, P., M. Greene and B. Alter (2003). "Cancer Incidence in Persons with Fanconi Anemia." *Blood* 101(3): 822-826.

Ruttkay-Nedecky, B., J. Kudr, L. Nejdl, D. Maskova, R. Kizek and V. Adam (2013). "G-Quadruplexes as Sensing Probes." *Molecules* 18(12): 14760-14779.

347

Sagi, J. (2013). "G-Quadruplexes Incorporating Modified Constituents: A Review." *Journal of Biomolecular Structure and Dynamics* 32(3): 477-511.

Sanders, C. (2010). "Human Pif1 Helicase Is a G-Quadruplex DNA-Binding Protein with G-Quadruplex DNA-Unwinding Activity." *Biochemical Journal* 430(1): 119-128.

Sannohe, Y. and H. Sugiyama (2001). Overview of Formation of G-Quadruplex Structures. *Current Protocols in Nucleic Acid Chemistry*, John Wiley *and* Sons, Inc.: Unit 17.12.11-17.

Saretzki, G. (2003). "Telomerase Inhibition as Cancer Therapy." *Cancer Letters* 194(2): 209-219.

Sarkies, P., C. Reams, L. Simpson and J. Sale (2010). "Epigenetic Instability Due to Defective Replication of Structured DNA." *Molecular Cell* 40(5): 703-713.

Sarkies, P., P. Murat, L. Phillips, K.J. Patel, S. Balasubramanian and J. Sale (2012). "FANCJ Coordinates Two Pathways That Maintain Epigenetic Stability at G-Quadruplex DNA." *Nucleic Acids Research* 40(4): 1485-1498.

Schoeftner, S. and M. Blasco (2008). "Developmentally Regulated Transcription of Mammalian Telomeres by DNA-Dependent RNA Polymerase II." *Nature Cell Biology* 10(2): 228-236.

Schuck, P., M. Perugini, N. Gonzales, G. Howlett and D. Schubert (2002). "Size-Distribution Analysis of Proteins by Analytical Ultracentrifugation: Strategies and Application to Model Systems." *Biophysical Journal* 82(2): 1096-1111.

Seenisamy, J., S. Bashyam, V. Gokhale, H. Vankayalapati, D. Sun, A. Siddiqui-Jain, N. Streiner, K. Shin-ya, E. White, W.D. Wilson and L. Hurley (2005). "Design and Synthesis of an Expanded Porphyrin That Has Selectivity for the *c-MYC* G-Quadruplex Structure." *Journal of the American Chemical Society* 127(9): 2944-2959.

Seenisamy, J., E. Rezler, T. Powell, D. Tye, V. Gokhale, C.S. Joshi, A. Siddiqui-Jain and L. Hurley (2004). "The Dynamic Character of the G-Quadruplex Element in the *c-MYC* Promoter and Modification by TMPyP4." *Journal of the American Chemical Society* 126(28): 8702-8709.

Shay, J. and S. Bacchetti (1997). "A Survey of Telomerase Activity in Human Cancer." *European Journal of Cancer* 33(5): 787-791.

Shay, J. and W. Wright (2006). "Telomerase Therapeutics for Cancer: Challenges and New Directions." *Nature Reviews Drug Discovery* 5(7): 577-584.

Shimizu, K., M. Goldfarb, M. Perucho and M. Wigler (1983). "Isolation and Preliminary Characterization of the Transforming Gene of a Human Neuroblastoma Cell Line." *Proceedings of the National Academy of Sciences* 80(2): 383-387.

Siddiqui-Jain, A., C. Grand, D. Bearss and L. Hurley (2002a). "Direct Evidence for a G-Quadruplex in a Promoter Region and Its Targeting with a Small Molecule to Repress *c-MYC* Transcription." *Proceedings of the National Academy of Sciences* 99(18): 11593-11598.

Simonsson, T., M. Kubista and P. Pecinka (1998). "DNA Tetraplex Formation in the Control Region of *c-myc*." *Nucleic Acids Research* 26(5): 1167-1172.

Smith, C. (1988). "Estimation of Sedimentation Coefficients and Frictional Ratios of Globular Proteins." *Biochemical Education* 16(2): 104-106.

Smith, F. and J. Feigon (1992). "Quadruplex Structure of Oxytricha Telomeric DNA Oligonucleotides." *Nature* 356(6365): 164-168.

Smith, S. (2010). "Evolutionary Expansion of Structurally Complex DNA Sequences." *Cancer Genomics - Proteomics* 7(4): 207-215.

Spackova, N., E. Cubero, J. Sponer and M. Orozco (2004). "Theoretical Study of the Guanine $\rightarrow$ 6-Thioguanine Substitution in Duplexes, Triplexes, and Tetraplexes." *Journal of the American Chemical Society* 126(44): 14642-14650.

Štefl, R., T. Cheatham, N. Špačková, E, Fadrná, I. Berger, J. Koča and J. Šponer (2003). "Formation Pathways of a Guanine-Quadruplex DNA Revealed by Molecular Dynamics and Thermodynamic Analysis of the Substates." *Biophysical Journal* 85(3): 1787-1804.

349

Sun, D., K. Guo, J. Rusche and L. Hurley (2005). "Facilitation of a Structural Transition in the Polypurine/Polypyrimidine Tract within the Proximal Promoter Region of the Human *VEGF* Gene by the Presence of Potassium and G-Quadruplex-Interactive Agents." *Nucleic Acids Research* 33(18): 6070-6080.

Sundquist, W. and A. Klug (1989). "Telomeric DNA Dimerizes by Formation of Guanine Tetrads between Hairpin Loops." *Nature* 342(6251): 825-829.

Tai, K., T. Shen, U. Börjesson, M. Philippopoulos and J.A. McCammon (2001). "Analysis of a 10-ns Molecular Dynamics Simulation of Mouse Acetylcholinesterase." *Biophysical Journal* 81(2): 715-724.

Tai, K., T. Shen, R. Henchman, Y. Bourne, P. Marchot and J.A. McCammon (2002). "Mechanism of Acetylcholinesterase Inhibition by Fasciculin: A 5-ns Molecular Dynamics Simulation." *Journal of the American Chemical Society* 124(21): 6153-6161.

Tang, C. and R. Shafer (2006). "Engineering the Quadruplex Fold: Nucleoside Conformation Determines Both Folding Topology and Molecularity in Guanine Quadruplexes." *Journal of the American Chemical Society* 128(17): 5966-5973.

Tarsounas, M. and M. Tijsterman (2013). "Genomes and G-Quadruplexes: For Better or for Worse." *Journal of Molecular Biology* 425(23): 4782-4789.

Teeter, M. and D. Case (1990). "Harmonic and Quasiharmonic Descriptions of Crambin." *The Journal of Physical Chemistry* 94(21): 8091-8097.

Tong, X., W. Lan, X. Zhang, H. Wu, M. Liu and C. Cao (2011). "Solution Structure of All Parallel G-Quadruplex Formed by the Oncogene *RET* Promoter Sequence." *Nucleic Acids Research* 39(15): 6753-6763.

Tougard, P., J.F. Chantot and W. Guschlbauer (1973). "Nucleoside Conformations: X. An X-Ray Fiber Diffraction Study of the Gels of Guanine Nucleosides." *Biochimica et Biophysica Acta (BBA) - Nucleic Acids and Protein Synthesis* 308(1): 9-16.

Tran, P., A. Virgilio, V. Esposito, G. Citarella, J.L. Mergny and A. Galeone (2011). "Effects of 8-Methylguanine on Structure, Stability and Kinetics of Formation of Tetramolecular Quadruplexes." *Biochimie* 93(3): 399-408.

350

Trotta, R., S. De Tito, I. Lauri, V. La Pietra, L. Marinelli, S. Cosconati, L. Martino, M. Conte, L. Mayol, E. Novellino and A. Randazzo (2011). "A More Detailed Picture of the Interactions between Virtual Screening-Derived Hits and the DNA G-Quadruplex: NMR, Molecular Modelling and ITC Studies." *Biochimie* 93(8): 1280-1287.

Uesugi, S., H. Liu, A. Kugimiya, A. Matsugami and M. Katahira (2003). "RNA and DNA, Which Contain Two GGAGG Segments Connected with UUUU or TTTT Sequences, Form Entirely Different Quadruplex Structures." *Nucleic Acids Symposium Series* 3(1): 51-52.

Verma, A., K. Halder, R. Halder, V. Yadav, P. Rawal, R. Thakur, F. Mohd, A. Sharma and S. Chowdhury (2008). "Genome-Wide Computational and Expression Analyses Reveal G-Quadruplex DNA Motifs as Conserved *cis*-Regulatory Elements in Human and Related Species." *Journal of Medicinal Chemistry* 51(18): 5641-5649.

Virgilio, A., V. Esposito, A. Randazzo, L. Mayol and A. Galeone (2005a). "8-Methyl-2'-Deoxyguanosine Incorporation into Parallel DNA Quadruplex Structures." *Nucleic Acids Research* 33(19): 6188-6195.

Virgilio, A., V. Esposito, A. Randazzo, L. Mayol and A. Galeone (2005b). "Effects of 8-Methyl-2'-Deoxyadenosine Incorporation into Quadruplex Forming Oligodeoxyribonucleotides." *Bioorganic and Medicinal Chemistry* 13(4): 1037-1044.

Virgilio, A., V. Esposito, G. Citarella, A. Pepe, L. Mayol and A. Galeone (2012). "The Insertion of Two 8-Methyl-2′-Deoxyguanosine Residues in Tetramolecular Quadruplex Structures: Trying to Orientate the Strands." *Nucleic Acids Research* 40(1): 461-475.

Wang, A., S. Fujii, J. van Boom, G. van der Marel, S. van Boeckel and A. Rich (1982). "Molecular Structure of r(GCG)d(TATACGC): A DNA-RNA Hybrid Helix Joined to Double Helical DNA." *Nature* 299(5884): 601-604.

Wang, A., G. Quigley, F. Kolpak, J. Crawford, J. van Boom, G. van der Marel and A. Rich (1979). "Molecular Structure of a Left-Handed Double Helical DNA Fragment at Atomic Resolution." *Nature* 282(5740): 680-686.

Wang, J.M., F.C. Huang, M. Kuo, Z.F. Wang, T.Y. Tseng, L.C. Chang, S.J. Yen, T.C. Chang and J.J. Lin (2014). "Inhibition of Cancer Cell Migration and Invasion through Suppressing the Wnt1-Mediating Signal Pathway by G-Quadruplex Structure Stabilizers." *Journal of Biological Chemistry* 289(21):14612-14623.

Wang, Y. and D. Patel (1993). "Solution Structure of the Human Telomeric Repeat d[AG$_3$(T$_2$AG$_3$)$_3$] G-Tetraplex." *Structure* 1(4): 263-282.

Watson, J. and F. Crick (1953a). "Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid." *Nature* 171(4356): 737-738.

Watson, J. and F. Crick (1953b). "Genetical Implications of the Structure of Deoxyribonucleic Acid." *Nature* 171(4361): 964-967.

Wei, C., G. Jia, J. Yuan, Z. Feng and C. Li (2006). "A Spectroscopic Study on the Interactions of Porphyrin with G-Quadruplex DNAs." *Biochemistry* 45(21): 6681-6691.

Wei, C., G. Jia, J. Zhou, G. Han and Can Li (2009). "Evidence for the Binding Mode of Porphyrins to G-Quadruplex DNA." *Physical Chemistry Chemical Physics* 11(20): 4025-4032.

Wei, D., G. Parkinson, A. Reszka and S. Neidle (2012). "Crystal Structure of a *c-kit* Promoter Quadruplex Reveals the Structural Role of Metal Ions and Water Molecules in Maintaining Loop Conformation." *Nucleic Acids Research* 40: 4691-4700.

White, L., W. Wright and J. Shay (2001). "Telomerase Inhibitors." *Trends in Biotechnology* 19(3): 114-120.

Whitehouse, I. and D. Smith (2013). "Chromatin Dynamics at the Replication Fork: There's More to Life Than Histones." *Current Opinion in Genetics and Development* 23(2): 140-146.

Wilkins, M., A. Stokes and H. Wilson (1953). "Molecular Structure of Nucleic Acids: Molecular Structure of Deoxypentose Nucleic Acids." *Nature* 171(4356): 738-740.

Williamson, J. (1994). "G-Quartet Structures in Telomeric DNA." *Annual Review of Biophysics and Biomolecular Structure* 23(1): 703-730.

Williamson, J., M. Raghuraman and T. Cech (1989). "Monovalent Cation-Induced Structure of Telomeric DNA: The G-Quartet Model." *Cell* 59(5): 871-880.

Wright, W., V. Tesmer, K. Huffman, S. Levene and J. Shay (1997). "Normal Human Chromosomes Have Long G-Rich Telomeric Overhangs at One End." *Genes and Development* 11(21): 2801-2809.

Wu, P., D. Ma, C. Leung, S. Yan, N. Zhu, R. Abagyan and C. Che (2009). "Stabilization of G-Quadruplex DNA with Platinum(II) Schiff Base Complexes: Luminescent Probe and Down-Regulation of *c-myc* Oncogene Expression." *Chemistry* 15(47): 13008-13021.

Wu, Y. and R. Brosh (2010). "G-Quadruplex Nucleic Acids and Human Disease." *The FEBS Journal* 277(17): 3470-3488.

Wu, Y., K. Shin-ya and R. Brosh (2008). "FANCJ Helicase Defective in Fanconia Anemia and Breast Cancer Unwinds G-Quadruplex DNA to Defend Genomic Stability." *Molecular and Cellular Biology* 28(12): 4116-4128.

Xu, Y., K. Kaminaga and M. Komiyama (2008a). "G-Quadruplex Formation by Human Telomeric Repeats-Containing RNA in Na^+ Solution." *Journal of the American Chemical Society* 130(33): 11179-11184.

Xu, Y., K. Kaminaga and M. Komiyama (2008b). "Human Telomeric RNA in G-Quadruplex Structure." *Nucleic Acids Symposium Series* 52(1): 175-176.

Xu, Y. and H. Sugiyama (2006). "Formation of the G-Quadruplex and i-Motif Structures in Retinoblastoma Susceptibility Genes (*Rb*)." *Nucleic Acids Research* 34(3): 949-954.

Xue, Y., Z.Y. Kan, Q. Wang, Y. Yao, J. Liu, Y.H Hao and Z. Tan (2007). "Human Telomeric DNA Forms Parallel-Stranded Intramolecular G-Quadruplex in K^+ Solution under Molecular Crowding Condition." *Journal of the American Chemical Society* 129: 11185-11191.

353

Yang, D. and K. Okamoto (2010). "Structural Insights into G-Quadruplexes: Towards New Anticancer Drugs." *Future Medicinal Chemistry* 2(4): 619-646.

Youds, J., L. Barber, J. Ward, S. Collis, N. O'Neil, S. Boulton and A. Rose (2008). "DOG-1 Is the *Caenorhabditis Elegans* BRIP1/FANCJ Homologue and Functions in Interstrand Cross-Link Repair." *Molecular and Cellular Biology* 28(5): 1470-1479.

Yphantis, D. and D. Roark (1971). "Equilibrium Centrifugation of Nonideal Systems. Donnan Effect in Self-Associating Systems." *Biochemistry* 10(17): 3241-3249.

Zhang, D.H., T. Fujimoto, S. Saxena, H.Q. Yu, D. Miyoshi and N. Sugimoto (2010a). "Monomorphic RNA G-Quadruplex and Polymorphic DNA G-Quadruplex Structures Responding to Cellular Environmental Factors." *Biochemistry* 49(21): 4554-4563.

Zhang, Z., J. Dai, E. Veliath, R. Jones and D. Yang (2010b). "Structure of a Two-G-Tetrad Intramolecular G-Quadruplex Formed by a Variant Human Telomeric Sequence in K^+ Solution: Insights into the Interconversion of Human Telomeric G-Quadruplex Structures." *Nucleic Acids Research* 38(3): 1009-1021.

Zhao, Y., Z. Du and N. Li (2007). "Extensive Selection for the Enrichment of G4 DNA Motifs in Transcriptional Regulatory Regions of Warm Blooded Animals." *FEBS Letters* 581(10): 1951-1956.

Zhou, W., N. Brand and L. Ying (2011). "G-Quadruplexes – Novel Mediators of Gene Function." *Journal of Cardiovascular Translational Research* 4(3): 256-270.

Zimmerman, S., G. Cohen and D. Davies (1975). "X-Ray Fiber Diffraction and Model-Building Study of Polyguanylic Acid and Polyinosinic Acid." *Journal of Molecular Biology* 92(2): 181-192.

APPENDIX

This is a list of publications whose findings are presented in this dissertation:

1. **Huy T. Le,** Patrick A. Holt, Lynn W. DeLeeuw, Jonathan B. Chaires and John O. Trent. "From Cyberspace to Cancer Drug: Biophysical Characterization of a G-Quadruplex-Interacting Small Molecule Identified by Virtual Screening." (Manuscript to be Submitted)

2. **Huy T. Le,** William L. Dean, Robert Buscaglia, Jonathan B. Chaires and John O. Trent. "An Investigation of G-Quadruplex Structural Polymorphism in the Human Telomere Using a Combined Approach of Hydrodynamic Bead Modeling and Molecular Dynamics Simulation." *The Journal of Physical Chemistry B* 2014,118(20): 5390-5405.

   Reproduced with permission from the American Chemical Society.

3. **Huy T. Le,** M. Clarke Miller, Robert Buscaglia, William L. Dean, Patrick A. Holt, Jonathan B. Chaires and John O. Trent (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." *Organic and Biomolecular Chemistry* 2012, 10(47): 9393-9404.

   Reproduced by permission of The Royal Society of Chemistry.

4. **Huy T. Le,** Robert Buscaglia, William L. Dean, Jonathan B. Chaires and John O. Trent. Calculation of Hydrodynamic Properties for G-Quadruplex Nucleic Acid Structures from *in silico* Bead Models. *Topics in Current Chemistry: Quadruplex Nucleic Acids*. Jonathan B. Chaires and David Graves. Berlin, Springer Berlin Heidelberg. 330:179-210.

5. M. Clarke Miller, **Huy T. Le,** William L. Dean, Patrick A. Holt, Jonathan B. Chaires and John O. Trent. "Polymorphism and Resolution of Oncogene Promoter Quadruplex-Forming Sequences." *Organic and Biomolecular Chemistry* 2011, 9(22): 7633-7637.

CURRICULUM VITA

NAME: Huy Tuan Le

ADDDRESS: 3100 Beargrass Run Court
                  Unit 206
                  Louisville, KY 40220

DATE OF BIRTH: Ho Chi Minh City, Vietnam – October 2, 1986

EDUCATION
& TRAINING: Bachelor of Science (B.S.), 2008
                  Department: Chemistry (Concentration: Biochemistry)
                  College of Arts and Sciences
                  University of Louisville

                  Master of Science (M.S.), 2013
                  Department: Biochemistry and Molecular Biology
                  School of Medicine
                  University of Louisville

                  Doctor of Philosophy (Ph.D.), 2014
                  Department of Biochemistry and Molecular Biology
                  School of Medicine
                  University of Louisville

                  Doctor of Medicine (M.D.), Anticipated Date: 2016
                  School of Medicine
                  University of Louisville

ACADEMIC
HONORS: University of Louisville M.D./Ph.D. Program (2010 – present)

                  University of Louisville Honors Scholar (2008)

                  Mortar Board Senior Honors Society (2007 – 2008)

357

Omicron Delta Kappa Honors Society (2006 – 2008)

Phi Eta Sigma Honors Society (2006 – 2008)

Order of Omega Honors Society (2006 – 2008)

Golden Key Honors Society (2005 – 2008)

Alpha Epsilon Delta Honors Society (2004 – 2008)

Honors Program in College of Arts & Sciences
University of Louisville (2004 – 2008)

SCHOLARSHIPS
& FELLOWSHIPS: Summer Research Scholar Program Fellowship
School of Medicine, University of Louisville (2009)

Nathaniel G. Symonds Scholarship Leadership Award
Kappa Sigma Fraternity (2007)

Cancer Education Program Fellowship
National Cancer Institute (2006)

Undergraduate Summer Research Opportunity Program
University of Louisville (2006)

William S. McClintic Scholarship Leadership Award
Kappa Sigma Fraternity (2006)

Hallmark Scholar
University of Louisville Board of Trustees (2004 – 2008)

James E. Casey Scholar
UPS Foundation (2004 – 2008)

COMMUNITY
INVOLVEMENTS: Volunteer, Cathedral of the Assumption Soup Kitchen
Louisville, KY (2013 – present)

Organizer, Kentucky Future Business Leaders of America State
Conference, Louisville, KY (2008 – present)

Organizer, American Cancer Society "An Evening of Hope" Gala
Louisville, KY (2007 – 2012)

RESEARCH
EXPERIENCE: Graduate Research Assistant, M.D./Ph.D. Program (2010 – 2014)
Department of Biochemistry and Molecular Biology
School of Medicine
University of Louisville, Louisville, KY 40202
Mentor: John O. Trent

Summer Research Scholar Program (2009)
Department of Medicine
School of Medicine,
University of Louisville, Louisville, KY 40202
Mentor: John O. Trent

Undergraduate Research Assistant (2006 – 2009)
Hormone Receptor Laboratory
Institute of Molecular Diversity and Drug Design
Department of Biochemistry and Molecular Biology
School of Medicine
University of Louisville, Louisville, KY 40202
Co-Mentors: James L. Wittliff and John O. Trent

ABSTRACTS & POSTERS:

Huy T. Le, William L. Dean, Robert Buscaglia, Jonathan B. Chaires and John O. Trent. "Predicting G-Quadruplex Folding Topologies Using Hydrodynamic Bead Modeling." Midwest Meeting of the American Physician Scientists Association. Columbus, OH, 2013.

Huy T. Le, William L. Dean, Robert Buscaglia, Jonathan B. Chaires and John O. Trent. "An Investigation of G-Quadruplex Structural Polymorphism Using a Combined Approach of Hydrodynamic Bead Modeling and Molecular Dynamics Simulation." Research! Louisville. Louisville, KY, 2013.

Nicholas Siow, Yidi Huang, **Huy T. Le,** Jonathan M. Maguire and John O. Trent "Automated G-Quadruplex Generation: From Sequence To Structure." Research! Louisville. Louisville, KY, 2013.

Yidi Huang, Nicholas Siow, **Huy T. Le,** Jonathan M. Maguire and John O. Trent. "LoopGen: Modeling Connecting Loops in Unimolecular G-Quadruplexes." Research! Louisville. Louisville, KY, 2013.

Huy T. Le, William L. Dean, Robert Buscaglia, Jonathan B. Chaires and John O. Trent. "Calculations of Hydrodynamic Properties for G-Quadruplex Nucleic Acids Using Bead Modeling." Department of Biochemistry and Molecular Biology Retreat. Louisville, KY, 2013.

John D. Gettelfinger, **Huy T. Le,** Lynn W. Deleeuw, Jonathan B. Chaires and John O. Trent. "Off-Target Drug Effects: The Identification of Natural Products and FDA-Approved Cancer Drugs as G-Quadruplex DNA-Interacting Agents." Research! Louisville. Louisville, KY, 2012.

Huy T. Le, M. Clarke Miller, Patrick A. Holt, William L. Dean, Jonathan B. Chaires and John O. Trent. "Not All G-Quadruplexes are Created Equal: Investigating the Structural Polymorphism of the *c-Myc* G-Quadruplex-Forming Sequence and its Interaction with the Porphyrin meso-Tetra(N-methyl-4-pyridyl)porphine". Research! Louisville. Louisville, KY, 2011.

Michael T. Miao, **Huy T. Le** and John O. Trent. "Identifying Novel G-Quadruplex Selective Binding Agents by Screening the NCI Diversity Set II using a FRET Melting Assay." Research! Louisville. Louisville, KY, 2011.

Huy T. Le, M. Clarke Miller, Robert Buscaglia, William L. Dean, Patrick A. Holt, Jonathan B. Chaires and John O. Trent. "Separation of G-Quadruplex Structures Using Size-Exclusion Chromatography." Department of Biochemistry and Molecular Biology Retreat. Louisville, KY, 2011.

Huy T. Le and John O. Trent. "Drug Discovery with Unique Virtual Chemical Libraries." Research! Louisville. Louisville, KY, 2009.

James L. Wittliff, Sarah A. Andres, D. Alan Kerr II, Irina A. Smolenkova, Stephanie B. Bumpus and **Huy T. Le.** "Predicting Structural Requirements of Ligands with a Multi-Tiered Receptor-Based Approach to Assess Estrogen Mimicry." American Society for Biochemistry and Molecular Biology Annual Meeting Abstract. San Diego, CA, 2008.

James L. Wittliff, Sarah A. Andres, D. Alan Kerr II, Irina A. Smolenkova, Stephanie B. Bumpus and **Huy T. Le.** "Assessment of EDCs Exhibiting Estrogen Mimicry using a Multi-tiered Approach for Predicting Ligand Structural Requirements." The 6th International Conference on Pharmaceuticals and Endocrine Disrupting Chemicals. Costa Mesa, CA, 2007.

Huy T. Le, John O. Trent and James L. Wittliff. "Development of Models for Ligand Recognition of the Binding Domain of Human Estrogen Receptor Alpha." Clinical Ligand Assay Society Poster Abstract. Fajardo, Puerto Rico, 2007.

Huy T. Le, Stephanie B. Bumpus, Sarah A. Andres, John O. Trent and James L. Wittliff. "Determination of the Pharcacophore for Ligand Recognition by Human Estrogen Receptor-Alpha: An Assessment of GALAHAD Alignment Software." Summer Opportunity Research Program Abstract. Louisville, KY, 2006

PUBLICATIONS:

Huy T. Le, William L. Dean, John O. Trent, and Jonathan B. Chaires. "Density Contrasted Determination of Partial Specific Volumes for G-Quadruplex DNA." (Manuscript in Preparation)

Huy T. Le, Indira Chaudury, Jonathan M. Maguire, Nicholas Siow, Yidi Huang, and John O. Trent, "QuadGen: A Program for Generating 3D Models of G-quadruplex Structures from Input Sequence." (Manuscript in Preparation)

Huy T. Le, John D. Gettelfinger, Michael T. Miao, Rafael G. Del Villar, Lynn W. DeLeeuw, Jonathan B. Chaires, and John O. Trent. "Identification of FDA-Approved Oncology Drugs as G-Quadruplex-Interacting Agents." (Manuscript in Preparation)

Huy T. Le, Patrick A. Holt, Robert Buscaglia, Jonathan B. Chaires and John O. Trent. "Rational Drug Design using *in Silico* Virtual Screening and Biophysical Validation: Identification of a Novel G-Quadruplex Binding Agent with a Unique Chemical Scaffold." (Manuscript to be Submitted)

Huy T. Le, Patrick A. Holt, Lynn W. DeLeeuw, Jonathan B. Chaires and John O. Trent. "From Cyberspace to Cancer Drug: Biophysical Characterization of a G-Quadruplex-Interacting Small Molecule Identified by Virtual Screening." (Manuscript to be Submitted)

Huy T. Le, William L. Dean, Robert Buscaglia, Jonathan B. Chaires and John O. Trent. "An Investigation of G-Quadruplex Structural Polymorphism in the Human Telomere Using a Combined Approach of Hydrodynamic Bead Modeling and Molecular Dynamics Simulation." *The Journal of Physical Chemistry B* 2014,118(20): 5390-5405.

Huy T. Le, M. Clarke Miller, Robert Buscaglia, William L. Dean, Patrick A. Holt, Jonathan B. Chaires and John O. Trent. "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." *Organic and Biomolecular Chemistry* 2012, 10(47): 9393-9404.

M. Clarke Miller, **Huy T. Le,** William L. Dean, Patrick A. Holt, Jonathan B. Chaires and John O. Trent. "Polymorphism and Resolution of Oncogene Promoter Quadruplex-Forming Sequences." *Organic and Biomolecular Chemistry* 2011, 9(22): 7633-7637

BOOK CHAPTERS:

Huy T. Le, Robert Buscaglia, William L. Dean, Jonathan B. Chaires and John O. Trent. Calculation of Hydrodynamic Properties for G-Quadruplex Nucleic Acid Structures from *in silico* Bead Models. *Topics in Current Chemistry: Quadruplex Nucleic Acids*. Jonathan B. Chaires and David Graves. Berlin, Springer Berlin Heidelberg. 330:179-210