1-2020

Structural characterization and selective drug targeting of higher-order DNA G-quadruplex systems.

Robert Chandos Monsen
University of Louisville

STRUCTURAL CHARACTERIZATION AND SELECTIVE DRUG TARGETING OF HIGHER-ORDER DNA G-QUADRUPLEX SYSTEMS

By

Robert Chandos Monsen
B.S., University of Southern Indiana, 2014
M.S., University of Louisville, 2018

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

Doctor of Philosophy
in Biochemistry and Molecular Genetics

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

December 2020

STRUCTURAL CHARACTERIZATION AND SELECTIVE DRUG TARGETING OF HIGHER-ORDER DNA G-QUADRUPLEX SYSTEMS

By

Robert Chandos Monsen
B.S., University of Southern Indiana, 2014
M.S., University of Louisville, 2018

A Dissertation Approved on

November 20, 2020

by the following Dissertation Committee

_____
Dissertation Director
John O. Trent

_____
Jonathan B. Chaires

_____
Barbara J. Clark

_____
Steven R. Ellis

_____
Thomas M. Sabo

DEDICATIONS

This dissertation and all that it represents is dedicated to:

Jesse Alexander Moore and Eric Dewayne Glass (friends)
for their unwavering support and encouragement throughout the past 12 years,
without which I would likely still be making sandwiches.

Catherine Ellen Monsen (wife), Eli Joseph Wilson (stepson), and Larry (dog)
who during the most difficult times provided mental and emotional support and kept me from losing perspective of what is most important in life.

Katherine Kelly Biondini (mother)
as she is a living example of what I strive to be.
She has shown me through both actions and words what hard work, patience,
and persistence can achieve. Her encouragement and sacrifices throughout
the years have been instrumental in getting me to where I am today.

ACKNOWLEDGMENTS

The completion of this dissertation would not be possible with the following individuals. First, I thank my supervisor, John O. Trent, for his continuous support and excellent mentorship over these years. I will forever be grateful for the opportunity to be a member of his lab. I will always remember our dialogues which have directly shaped my understanding of what it is to do “good” science. I also thank my co-mentor, J. Brad Chaires, for his sage advice throughout the years. He has been an irreplaceable friend, teacher, and resource—especially during happy hours. I thank both for their financial support of my dissertation. My time spent working with these two individuals has undoubtably molded me into the best possible version of myself. I will be forever grateful.

I am also thankful to have worked with the other members of the Trent/Chaires lab. I thank Lynn DeLeuuw for all her guidance and time spent teaching me so many of the essential techniques used in this work—she was *instrumental* to my success. I also thank William “Bill” Dean and Robert “Bob” Gray for all the assistance and wisdom which they have imparted throughout our many (many!) discussions on feasibility and execution of experiments. I especially thank Bill for passing me the metaphorical analytical ultracentrifugation (AUC) baton. I look forward to someday mentoring others in AUC techniques (‘cause AUC don’t lie!). I also thank Jon Maguire for his extensive help and thoughtfulness throughout all the computational work.

In addition to the lab, I would like to express my thanks to my committee. I thank Steve Ellis for always asking the right questions, pushing me to find a deeper understanding in my work, and keeping me on my toes. I also appreciate Barbara Clark for her insightfulness and thoughtful questioning of my research, as well as her guidance throughout the PhD program in times uncertain. I also thank Mike Sabo for his continued willingness to collaborate on this work, as well as the discussions and instruction he has provided me during our NMR experiments. I appreciate them all for their commitment to my success, their agreement to be on my committee, and their thoughtful criticism and support of my research.

My time at the University of Louisville has allowed me to meet so many individuals whom I'm thankful to have known that have supported, collaborated, and celebrated with me. I thank William Gibson for being a wonderful friend throughout this entire process. I appreciate our talks and his persistent encouragement and support in dealing with problems students are so often faced with. I look forward to our future interactions as both scientists and friends. I am thankful that I had the opportunity to spend time Alfred B. Jenson, and I appreciate his wisdom, positivity, and thoughtful discussions. He is sorely missed. I thank Srinivas Chakravarthy for dedicating time to my research projects and collaborating on all the SAXS experiments. He has been essential to much of this work. I thank Joe Burlison and Nagaraju Miriyala for all their support and assistance in our drug discovery efforts. I thank David Samuelson for his expert guidance over the years. I thank Dillon Hofsommer for all his assistance in working with the EPR. I thank Saurabh Kumar, Müge Sak, Alexis Vega, Mark Dela Cerna, Timothy Audam, Emily Duderstadt, Justin Kos, Benjamin "Roody" Rood, Kaitlyn Shields, Stephanie Metcalf, Lindsey Reynolds, Rumeysa Biyik, and Josiah Hardesty for their comradery throughout the program.

ABSTRACT

STRUCTURAL CHARACTERIZATION AND SELECTIVE DRUG TARGETING OF HIGHER-ORDER DNA G-QUADRUPLEX SYSTEMS

Robert Chandos Monsen

November 20, 2020

There is now substantial evidence that guanine-rich regions of DNA form non-B DNA structures known as G-quadruplexes in cells. G-quadruplexes (G4s) are tetraplex DNA structures that form amid four runs of guanines which are stabilized via Hoogsteen hydrogen bonding to form stacked tetrads. DNA G4s have roles in key genomic functions such as regulating gene expression, replication, and telomere homeostasis. Because of their apparent role in disease, G4s are now viewed as important molecular targets for anticancer therapeutics. To date, the structures of many important G4 systems have been solved by NMR or X-ray crystallographic techniques. Small molecules developed to target these structures have shown promising results in treating cancer *in vitro* and *in vivo,* however, these compounds commonly lack the selectivity required for clinical success. There is now evidence that long single-stranded G-rich regions can stack or otherwise interact intramolecularly to form G4-multimers, opening a new avenue for rational drug design. For a variety of reasons, G4 multimers are not amenable to NMR or X-ray crystallography. In the current dissertation, I apply a variety of biophysical techniques in an integrative structural biology (ISB) approach to determine the primary conformation of two disputed higher-order G4 systems: (1) the extended human telomere G-quadruplex and (2) the G4-multimer formed within the *human telomerase reverse transcriptase* (*hTERT*) gene core promoter. Using the higher-order human

telomere structure in virtual drug discovery approaches I demonstrate that novel small molecule scaffolds can be identified which bind to this sequence *in vitro*. I subsequently summarize the current state of G-quadruplex focused virtual drug discovery in a review that highlights successes and pitfalls of *in silico* drug screens. I then present the results of a massive virtual drug discovery campaign targeting the *hTERT* core promoter G4 multimer and show that discovering selective small molecules that target its loops and grooves is feasible. Lastly, I demonstrate that one of these small molecules is effective in down-regulating hTERT transcription in breast cancer cells. Taken together, I present here a rigorous ISB platform that allows for the characterization of higher-order DNA G-quadruplex structures as unique targets for anticancer therapeutic discovery.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

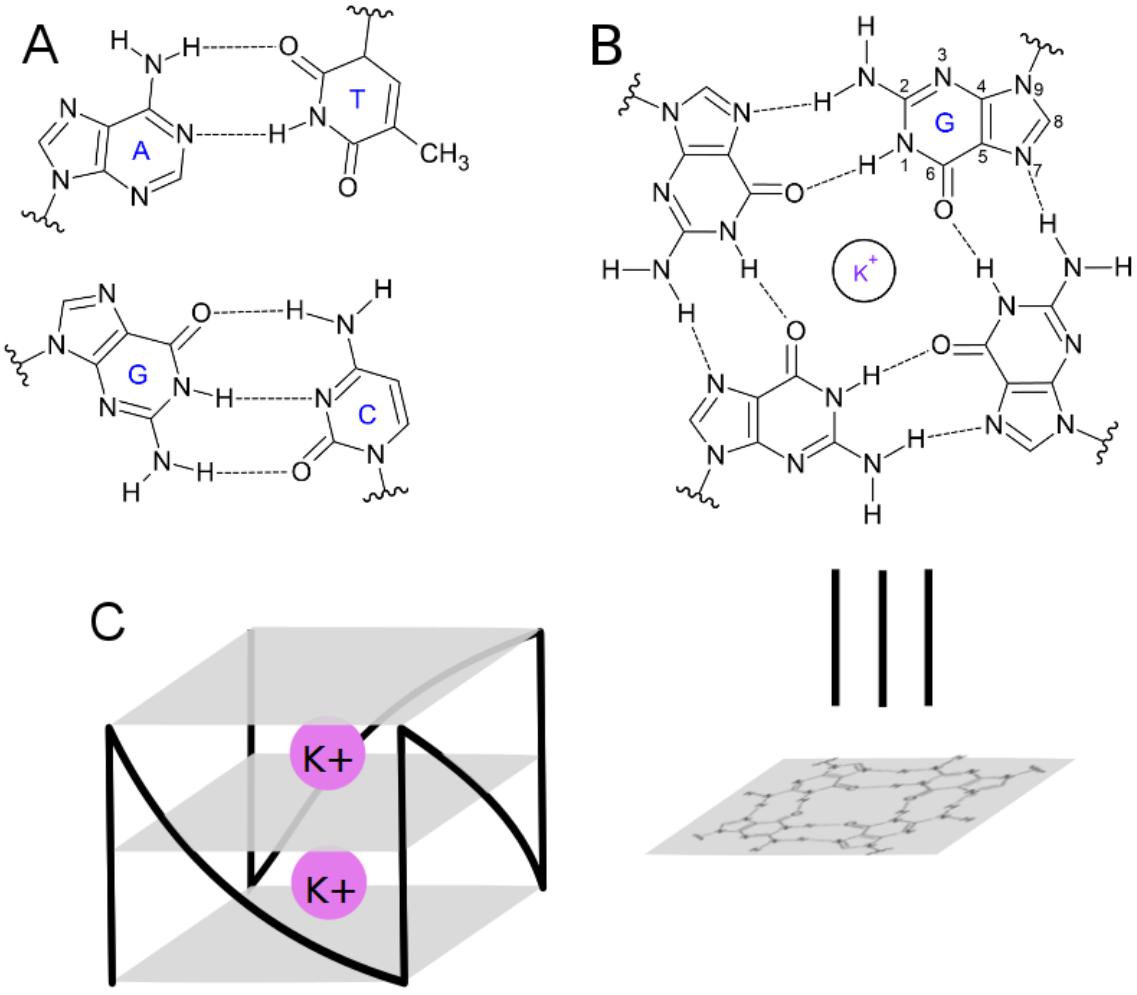
INTRODUCTION

In 1953, James Watson and Francis Crick provided a model of the B-form of DNA (1). This seminal work suggested, for the first time, the structural basis of genetic inheritance inferred from purine-pyrimidine base pairing in the double helix. The structure itself is arguably just as important. It is now known that the structure of DNA is critical for coherent interactions with proteins, without which would result in chaos within the cell. Since its discovery, a variety of non-B DNA structures have been revealed, each of which with their own sequence requirements for formation (2). In particular, DNA G-quadruplexes (G4s) have garnered serious attention over the past few decades as their role in disease has become increasingly evident.

Just as the DNA duplex is made up of constitutive base pairs (or “dyads”), the G-quadruplex is made up of guanine tetrads (G-tetrads) (**Figure 1A-B**) (3,4). G-tetrads are stabilized by hydrogen bonding across the O6 oxygen and N7 nitrogen of the first guanine with the N1 and N2 nitrogen of the adjacent guanine. This Hoogsteen bonding pattern repeats to form a square planar arrangement, allowing tetrads to stack atop one another as shown in **Figure 1C**. The “canonical” unimolecular G-quadruplex has between 2 and 4 contiguous stacks of G-tetrads. A variety of topologies exist for a single-stranded intra-molecular G-quadruplex, and these various forms are dictated primarily by phosphodiester backbone directionality and nucleoside conformations (*syn* or *anti)* with respect to the glycosidic bond. This implies that a unimolecular two-tetrad G-quadruplex has 26 possible loop configurations and 24 possible combinations of G-tetrad glycosidic bond orientations (5). The same number of loop combinations exist for a three-tetrad system, however, the number of possible G-tetrad glycosidic bond combinations increases, making the maximum total configurations 32 (5). Further, loop sizes, orientations, and interactions can vary considerably.

Thus, the conformational diversity of these structures has led to significant interest recently for their potential to be selectively targeted with small molecules.

Figure 1. Chemical structures of the duplex “dyads”, G-quadruplex tetrad, and a simplified schematic of a G-quadruplex. (A) Traditional base pairs, A:T and G:C, of the DNA duplex. (B) G-quadruplex tetrad, showing the coordination of a potassium ion in the central cavity. Broken bond symbols represent the disconnect from the deoxyribose (i.e. the glycosidic bond). (C) schematic representation of a basic three tetrad G-quadruplex. Gray squares represent the G-tetrads, black lines represent the phosphate backbone, and pink circles represent potassium ions.


A
A
T
CH3
G
C
B
G
K+
C
K+
K+

Since the discovery that G-quadruplex structures form spontaneously in biologically relevant ionic conditions, the field has exploded with reports of their wide array of biological roles. Early investigations of eukaryotic telomere sequences determined that G-quadruplex formation critically regulates the activity of human telomerase reverse transcriptase (hTERT), the enzyme that extends the telomere (6-8). This has obvious implications in cancer, as more than 85% of all cancers have reactivated hTERT, and its activity at the telomeres is essential in cellular immortality (9,10). Moreover, bioinformatic studies reveal that ~300,000 putative G-quadruplex forming sequences (PQSs) are scattered throughout the human genome non-randomly (11,12). PQSs are highly concentrated in regions such as immunoglobulin switch regions, mRNA 5' and 3' untranslated regions (UTRs), and regulatory regions in gene promoters (13,14), all of which are unique points for therapeutic intervention. Some promoter G-quadruplexes with relevance in disease have since been successfully targeted by small molecules (14,15). A notable example is the nuclease hypersensitivity element III_1 (NHE III_1) G-quadruplex of the *c-MYC* promoter. Stabilizing this G-quadruplex with the G4 ligand TMPyP4 resulted in marked repression of c-MYC expression—cementing the case for G-quadruplex-mediated transcriptional regulation (16). Currently there are about a thousand or more G-quadruplex interacting ligands which have been reported that bind to the telomeres, various promoters, and mRNA G-quadruplexes (17). However, just as selectivity is an issue with many protein targeting small molecules, the same issue now pertains to G-quadruplexes (18). G-quadruplex "monomers" (i.e. a single intra-molecular G-quadruplex unit consisting 2-4 G-tetrad stacks) are often amenable to high-resolution structural biology techniques, such as NMR and X-ray crystallography (although typically requiring stabilizing mutations), allowing for their unambiguous structural characterization. This has positioned them as prime targets in the pursuit of G-quadruplex selective drugs. Unfortunately, efforts are hampered by the common feature among all monomeric G-quadruplexes: their 3' and 5' G-tetrad faces. Binding, or "end-pasting", to these sites is the primary reason for ligand promiscuity (18).

Long single-stranded regions of G-rich DNA containing multiple G-quadruplex motifs have the potential to form G-quadruplex "multimers", that is, monomeric G-quadruplexes which can stack atop or otherwise interact with one another to form a tertiary arrangement (19,20). It is apparent

that such structures offer a more specific target for rational drug discovery whereby loop, groove, and G-quadruplex stacking junctions form specific pockets useful for rational drug discovery techniques (18,20). Unfortunately, these multimeric forms are difficult to characterize, which stems from their size, inherent heterogeneity in solution, guanine imino spectral overlap in NMR, and propensity to favor (often not relevant) conformations in X-ray diffraction. Therefore, a critical limitation in understanding the role these structures play in the cell, as well as in developing new selective anticancer therapeutics, resides in their structural characterization.

This dissertation provides an integrative structural biology (ISB) platform that allows for detailed characterization of multimeric G-quadruplexes in their wild type (WT) state and under physiologically relevant conditions. Using a suite of robust biophysical tools, such as size-exclusion chromatography (SEC), small-angle X-ray scattering (SAXS), analytical ultracentrifugation (AUC), circular dichroism (CD), nuclear magnetic resonance (NMR), and molecular dynamics (MD), I show that it is possible to determine the tertiary structure of higher-order G-quadruplexes at 20-40 Å resolution. I begin with an example of this approach, whereby a rigorous investigation of the higher-order human telomere reveals that it is composed of alternating hybrid-2 and hybrid-1 G4 topologies, and that it maximizes its usage of G-tracts in solution. The derived telomere G4 multimer structures are subsequently targeted at their inter-quadruplex junctional regions; whereby novel ligand scaffolds are identified. I next apply this characterization approach to the *hTERT* core promoter G-quadruplex and show that it preferentially adopts an all-parallel stacked G-quadruplex arrangement with multiple unique loop sites useful in drug discovery efforts. Following this, I describe and summarize the current state of G-quadruplex virtual screening in a review, pointing out where improvements could be made to increase selectivity and novelty of virtually screened small molecules. Last, using the derived model of the *hTERT* core promoter G-quadruplex, I show that unique, drug-like small molecules can be discovered which preferably bind to it over duplex, triplex, and monomeric G-quadruplex topologies. Additionally, I provide biological evidence that one of these small molecules can repress *hTERT* transcription in breast cancer cells, and so is a suitable lead for development as an anticancer therapeutic.

A Brief History of G-quadruplex DNA

When one thinks about deoxyribonucleic acid (DNA) in the cell, we traditionally think of the classical double helix (1), discovered more than 60 years ago by Watson, Crick, Franklin, and Wilkins. The double helix structure provided a platform by which we could relate structure to function, where base-pairing provided a mechanistic basis for the transfer of genetic information, and codons provided the blueprints for making proteins. Of course, we know now that DNA is more than just a medium for information transfer, it is dynamic and structurally diverse. For instance, the majority of DNA in the cell is negatively supercoiled, which favors the formation of non-B DNA structures (21). Less than 10 years after the duplex was elucidated, four-stranded guanylic acid structures were observed by fiber diffraction which indicated that guanines could form stable, planar tetrads via hydrogen bonding of their Hoogsteen faces (**Figure 1B**) (3,4). At the time, guanine tetrads were viewed as mere test tube oddities [and even "nuisances" (22)]. It wasn't until the late 80's when it was discovered that under physiological cation conditions, *in vitro,* guanine-rich telomere sequences spontaneously formed discrete four-stranded structures, now known as G-quadruplexes (G4s) (6,23,24). Although speculation of their *in vivo* relevance was widespread at the time, the discovery of these spontaneously formed G4s prompted Nobel Laurette Aaron Klug to remark "If G-quadruplexes form so readily *in vitro*, Nature will have found a way of using them *in vivo*" (25).

The 2000's hailed a new era for G-quadruplex biology. It was the advent of G4-targeting antibodies that visualized, for the first time, the formation of G4 DNA structures in cells (26-30). Their visualization across the genome was simultaneously verified by advances in G-quadruplex sequencing techniques (31) and bioinformatic inquiries (11,12). These technological advances paved the way for studies demonstrating that G-quadruplexes can be targeted and stabilized by small molecules in cells (27,30). A variety of quadruplex-interacting proteins have since been verified to interact with biological G-quadruplexes, such as telomerase (32), helicases (27,30,33), transcription factors (34), and chromatin remodeling complexes (35). For an excellent review of the newfound diversity of G-quadruplex functions in biology see that of Spiegel, Santosh, and

Balasabrumanian (36). Clearly, G-quadruplexes are important structures *in vivo,* and their roles in biology and disease are of great interest to the biomedical community.

Monomeric G-quadruplex Structure

The discovery that guanylic acids could form higher order structure dates back to 1910 when Ivar Bang reported that, upon heating and cooling, guanylic acid formed an extremely viscous, clear gel (37). It wasn't until 1962 when this phenomenon was further investigated. Ralph and colleagues showed that guanylic acid, but not other nucleotide derivatives, formed aggregates or other higher-order species in the presence of physiological buffers using analytical ultracentrifugation experiments (38). This same year, Gellert et al. reported the unique optical properties and arrangement of the G-quartet using X-ray diffraction, showing unambiguously the formation of a guanine helix (G-quadruplex) (3). A decade later the geometry of the G-quadruplex tetrad stack was elucidated using X-ray diffraction (4,39).

The basic structural motif of the G-quadruplex is the G-tetrad (or G-quartet), which consists of four cyclically Hoogsteen hydrogen-bonded guanines situated in a planar square shape (**Figure 1**). Bonding is through the protons of the nitrogen N7 and oxygen O6 atoms with that of the adjacent base nitrogen N1 and N2. The O6 atoms project towards the interior of the central G-tetrad helical axis and, when stacked onto an adjacent tetrad, the cavity formed among the eight central O6 atoms is large enough to house a variety of monovalent cations. Stabilization of the stacked G-tetrads is primarily through overlap of π-orbitals of stacked guanines and cation coordination (40). Monovalent cation coordination within the central cavity imparts a major stabilizing effect by neutralization of the partial negative charges from the inward projecting O6 oxygens (40). Commonly, the coordination ions are either K^+ or Na^+ but other ions of appropriate size can also fulfil this stabilizing role (41).

Two or more contiguous G-tetrad stacks constitute a G-quadruplex. The features of G-quadruplexes, which are used in describing their overall topology, are glycosidic conformations of guanines, strand orientations, groove widths, and loop arrangements (42,43). It is the arrangement

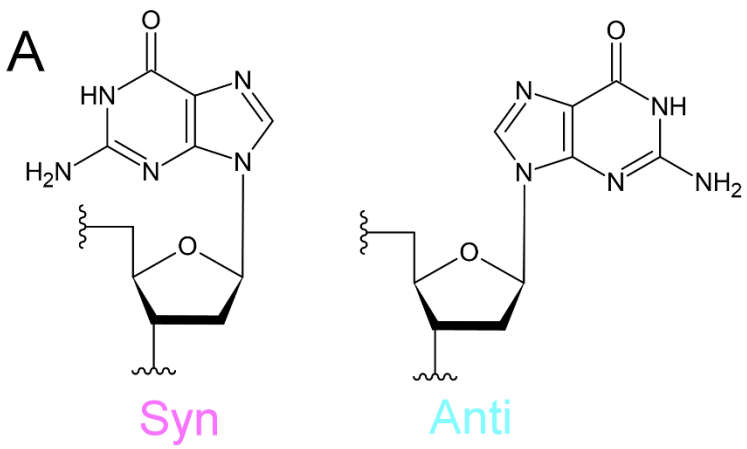
of nucleobases within a G-quadruplex that give rise to their diagnostic circular dichroism signatures (44). The glycosidic bond angle (GBA) is the orientation of the guanosine about the *N*-glycosidic bond of the deoxyribose, and this can be *syn* or *anti* (with *anti* being most prevalent in B-DNA). The specific GBAs of guanines within G-tetrads define the G-quadruplex's groove (43). The three types of grooves are narrow (n), medium (m), and wide (w) (**Figure 2**), and only G-tetrads with the same groove combinations may stack to form a G-quadruplex (43). The three loop types (diagonal, propeller, and lateral) are entwined with GBA. The lateral loop (a.k.a. edgewise loop) links guanines within the same tetrad that share hydrogen bonds. Conversely, diagonal loops link guanines within the same tetrad but do not share hydrogen bonds. In both cases the loops link guanines of different GBA. G-quadruplexes with these types of loops are said to be "hybrid" or "antiparallel" (**Figure 3**). The third loop type, now commonly known as propeller (but earlier called "double chain reversal"), links guanines which are not in the same tetrad, but within the same groove, and this groove is always medium since the bases of propeller loops always have the same glycosidic bond angle (43). G-quadruplexes which have all propeller type loops (and thus all bases with the same *anti* GBA) are said to be "parallel", as all the runs of guanines in the G-tetrads point in the same 5' to 3' direction (**Figure 3A**). There are primarily three types of loop "directionalities" observed in G-quadruplexes. These are known as parallel (↑↑↑↑), hybrid 3+1 (↑↑↓↑ or ↑↓↑↑), and anti-parallel or basket (↑↓↑↓) (depending on looping), where the arrows are from 5' (left) to 3' (right) and indicate the phosphodiester backbone direction (**Figure 3**) (42).

The G-quadruplex topological descriptors above (parallel, hybrid, and antiparallel) are the informal classes of G4s, and these descriptions are not sufficient to fully describe G4 conformation. For instance, the human telomere G4-forming sequence(s), depending on flanking nucleotides and solution conditions (45,46) can adopt six variations of these topologies (e.g. parallel, basket, two different hybrid forms, and two different antiparallel forms) (45-48). To this end, the Webba da Silva lab has created a formalism for describing quadruplex folding based on the loop orientation and GBA relative to the starting nucleotide closest to the 5' end of the G4 strand (49). The authors propose that, starting with the 5'-most guanine (which is typically *anti* in conformation), the looping type and direction can be described as '-' or '+' for anti-clockwise and clockwise progression,

respectively. The anti-clockwise '-' looping is most commonly observed in G4s and is akin to the right-handed B-DNA double helical turn [although clockwise '+' progressions have recently been observed, which are akin to Z-DNA and are deemed "Z-G4" (50,51)]. Further, the type of loop, parallel 'p', lateral 'l', and diagonal 'd', can be assigned as such. Thus, these formal assignments are more adequate at describing looping type and directionality (although diagonal loops do not have '-' or '+' because they do not connect in an anti-clockwise or clockwise fashion). For instance, a parallel G-quadruplex with loops that run anti-clockwise is -p-p-p for three anti-clockwise propeller loops, and +p+p+p for three clockwise propeller loops (**Figure 3**). An antiparallel G4 with a diagonal loop (such as the human telomere in sodium, PDB ID: 143D) would be +ld-l (**Figure 3D**). For lateral loops, an additional descriptor can be added for even more specificity. The designation of a subscript 'm', 'n', or 'w' for medium, narrow, or wide grooves can be appended after 'l' (52). For the antiparallel (143D) example above, its descriptor would become $+l_wd\text{-}l_n$. Clearly, this naming convention is powerful compared with the traditional parlance.

Altogether, the structural diversity of two- and three-tetrad G-quadruplexes is immense, whereby each can theoretically adopt 26 different looping configurations and 32 different combinations of *anti* and *syn* guanines (43). This diversity is further compounded when considering that many biologically relevant sequences contain more than four runs of G-tracts which are often longer than two consecutive guanines, and can be variable in number (53). While this greatly complicates G-quadruplex structure determination, it is also the reason why these structures are so enticing from a drug discovery perspective.

Figure 2. Guanine glycosidic bonds and G-tetrad groove definitions. The two possible conformations, *syn* and *anti,* of guanines about their glycosidic bonds (A), and the G-tetrad with groove designations based on guanine glycosidic bonds and spatial arrangements (B). Pink ovals indicate guanines with a *syn*-conformation and cyan ovals indicate guanines with *anti*-conformations. The grooves are marked as medium, wide, or narrow, as determined by the arrangement of guanines and their respective glycosidic bond orientations.


A
Syn
Anti

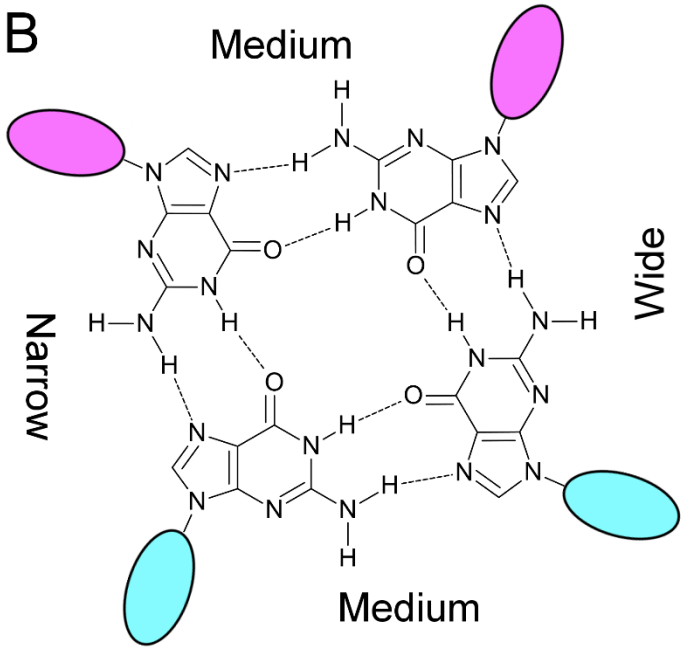
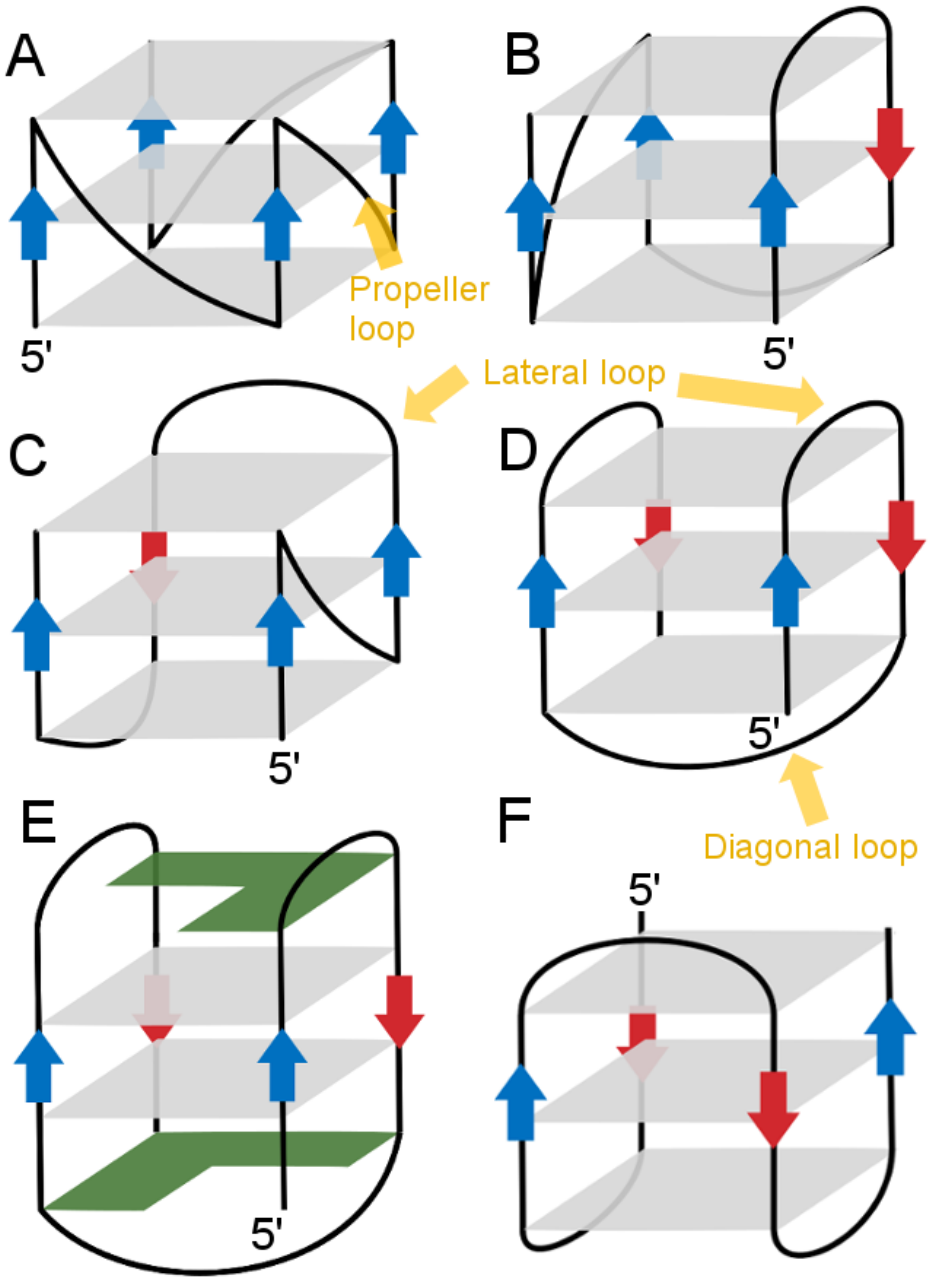

B
Medium
Wide
Narrow
Medium

Figure 3. G-quadruplex conformations. In each diagram grey squares represent G-tetrads, black lines indicate the phosphate backbone, blue and red arrows highlight backbone directionality, yellow text and arrows indicate the three main loop types, and green represents any non-tetrad participating nucleotide. (A) Parallel fold with three propeller loops -p-p-p. (B & C) two different hybrid 3 + 1 folds, each with two lateral (or edgewise) loops, and a single propeller loop (-l-l-p and -p-l-l, respectively). (D) Antiparallel fold with two lateral loops and one diagonal loop -ld+l. (E) A two-tetrad antiparallel fold with the same loop types as in D but showing how additional nucleotides can potentially stack on the 5' and 3' G-tetrad faces -ld+l. (F) A "chair" formation, with three consecutive lateral loops +l+l+l.

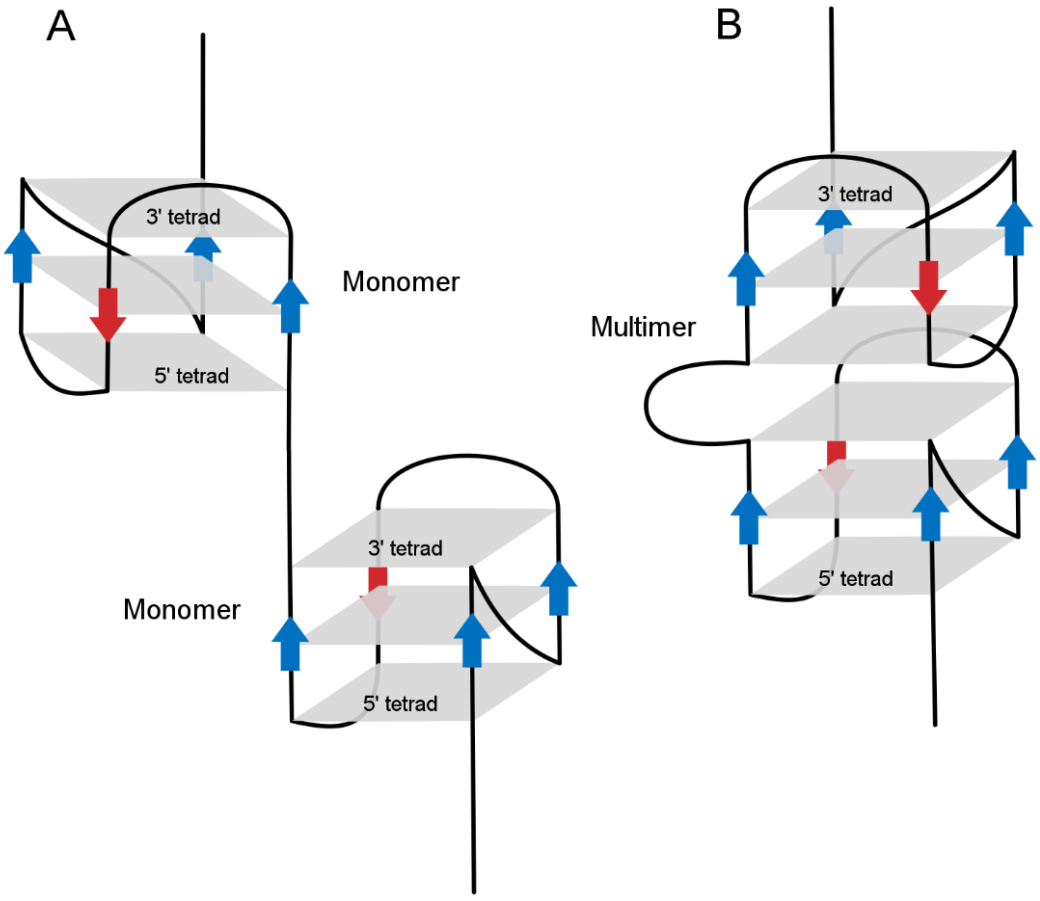

A
B
Propeller loop
5'
5'
Lateral loop
C
D
5'
5'
Diagonal loop
E
F
5'
5'

Multimeric G-quadruplex Structure

The diversity of G-quadruplex structures extends well past that of single "monomeric" intra-stranded structures. Multimeric G-quadruplexes also exist, and these can be inter-stranded G-quadruplexes such as "dimers", "trimers", and "tetramers", as well as oligomer and polymer type "G-wires". In this dissertation the emphasis will remain primarily on intra-strand (unimolecular) G-quadruplexes and, for clarity, the term "monomer" will hence forth refer only to a single intra-strand G-quadruplex. In the context of long single-stranded DNA (ssDNA), such as would be found in a cell during replication or transcription, it could be envisioned that multiple monomer G-quadruplexes could form on the same strand. These G4s could either act independently (known as "beads-on-a-string") (**Figure 4A**) or interact with each other by direct contact. In the latter case, the G-quadruplex (i.e. the entire G-rich strand of DNA) will be referred to as a G-quadruplex "multimer" (**Figure 4B**) (20).

Just as duplex dyads and G-quadruplex tetrads are stabilized through nucleotide stacking, monomeric G-quadruplexes can stack atop on another under physiologically relevant DNA concentrations (as low as a few micromolar strand concentration) (20). G-quadruplex multimers form in primarily two ways: direct G-tetrad stacking of terminal (5' or 3') tetrad faces, or via "sandwiching" of non-tetrad-participating loop residues between tandem monomeric G-quadruplex units (54), although other inter-domain interactions have been proposed (55). Direct G-tetrad face interactions of monomer quadruplexes of 3' to 5' and 5' to 5' occur most readily, as the 3' to 3' stacking interface is thought to be energetically unfavorable without facilitation by small molecule ligands or sandwiched adenine residues (54). A variety of monomeric G-quadruplex stacking interfaces have been reported using both NMR and X-ray crystallographic techniques (hundreds of structures of unimolecular and stacked dimer DNA G-quadruplexes in the Protein Data Bank: https://www.rcsb.org/). There are only a few multimers with high-resolution structures available: a dimer of the *c-MYB* promoter (56), a synthetic aptamer two-stacked two-tetrad G4 (57), and two, two-stacked two-tetrad multimers containing left-handed "Z-G4s" (50,51).

Figure 4. Schematic showing the “beads-on-a-string” and “multimer” higher-order G-quadruplex structures. Schematics are represented as in **Figure 3**, with labeling of the 5’ and 3’ tetrad faces. (A) G-quadruplex monomers can spontaneously fold adjacent to one another on the same DNA strand but not necessarily interact, forming the beads-on-a-string structure. These quadruplexes could also contain 2-, 4-, or more tetrads. (B) Two monomers on the same strand of DNA interacting to form a G-quadruplex multimer.


A
3' tetrad
5' tetrad
Monomer
3' tetrad
5' tetrad
Monomer
B
3' tetrad
5' tetrad
Multimer

G-quadruplex Functions

Telomere G-quadruplexes

At the end of eukaryotic chromosomes is a specialized, non-coding sequence of nucleotides known as the telomere which consists of tandem repeats of the sequence 5'-TTAGGG-3' (58). The telomeres have long been associated with human disease, such as cancer (59), telomeropathies (60), the aging process (61), and genome (in)stability (62). The common etiology among these conditions is dysregulated telomere length control. In normal human somatic cells telomeres are approximately 5-25 kb in length and have an extended single-stranded 3' overhang of ~35-600 bases (63). The entire region is coated in telomere associated proteins known as the shelterin complex, which protect the exposed ends from nucleolytic cleavage, end-to-end fusion events, and unintentional activation of the DNA damage response pathway (64,65).

In most normal somatic (non-germ) cells, progressive shortening of the telomeres occurs with each round of cell division due to the end replication problem (66). This replicative "clock" has long been thought to be a protective mechanism against tumorigenesis (65). Once a critical number of replications have occurred the telomeres become sufficiently small and trigger a DNA damage response by "uncapping" of the telomere shelterin proteins, leading to cellular senescence (65,66). Cancer cells have found mechanisms which circumvent natural telomere attrition. The two main mechanisms used in cancer to restore telomere length are DNA recombination (known as the alternative lengthening of telomeres [ALT] pathway), and re-activation of human telomerase reverse transcriptase (hTERT). hTERT is the catalytic protein component of the telomerase ribonucleoprotein which extends the telomere using its reverse transcriptase activity (67). The ALT pathway is only observed in a small fraction of cancers (~10-15%), with the majority (85-90%) of cancer cells exhibiting aberrant telomerase activity (10,68,69). Thus, by extending the telomere, cancer cells can divide indefinitely, and so telomeres and their associated proteins are promising targets for anticancer therapeutics (59).

In 1987, while studying the telomere sequences in Tetrahymena, Henderson and Blackburn discovered that the G-rich single-stranded telomeres spontaneously form non-Watson-Crick type intramolecular structures (6). Two years later the telomere G-quadruplex was confirmed by DNA foot-printing experiments, which showed protection of the G-tetrad guanine N7 groups exclusively when folded in the presence of Na^+ and K^+ salts (24,70). These discoveries suggested an *in vivo* role for G-quadruplexes, and along with advances in oligonucleotide synthesis schemes (71), renewed interest in their study. In 1992 the first atomic resolution crystal structure of a telomere DNA G-quadruplex was reported using sequences from *Oxytricha* (72), followed soon after by the NMR solution structure of the same sequence (73). One year later, the first atomic structure of the human telomere G-quadruplex was reported by Wang and Patel, which showed that it adopted an antiparallel "basket" type fold (like **Figure 3D**) in a sodium buffer (74). A variety of monomer telomeric G-quadruplexes have since been solved by X-ray crystallography or NMR techniques: parallel (75), hybrid 3+1 (76,77), antiparallel (74), and a two-tetrad antiparallel (47). Not all telomere topologies are biologically relevant, as some appear to be the result of either non-physiological cation conditions or unnatural physical forces, such as those in crystal packing (78). Extensive investigations have led to the conclusion that the two prominent forms of the human telomere G-quadruplex, under physiologically relevant ion concentrations *in vitro,* are the hybrid 3+1 topologies, called "hybrid-1" (↓↓↑↓) (76) and "hybrid-2" (↓↑↓↓) (77). In 2019, Bao et al. found that the hybrid-1 and -2 type conformations exist within HeLa cells by monitoring the folding of an injected 22-nt long telomere sequence, $AGGG(TTAGGG)_3$, using a state-of-the-art in-cell ^{19}F-NMR technique (79).

Initial investigations of telomeric G-quadruplexes demonstrated that they act as a steric inhibitor of telomerase (7,32), and therefore are a transient protection mechanism in cells with exposed 3' single-stranded ends. Indeed, when shelterin proteins are absent (or "uncapped"), the exposed ssDNA folds into a G-quadruplex and elicits a specific DNA damage response (80). There may also be a role for G-quadruplexes in mediating the so-called "T-loop" protective structure. The T-loop is formed when the telomere end forms a protective circle which requires a strand-invasion by the free 3' overhang (65). Thus, the T-loop requires a free 3' overhang for interaction with the

shelterin proteins TRF1 and TRF2 (TTAGGG Repeat binding Factors 1 and 2) (81,82). Indeed, when telomere G-quadruplexes are stabilized by small molecules a DNA damage response is observed in cells (83), which is a result of sequestering the free 3' $(TTAGGG)_n$ overhang from both telomerase and TRF1/2 (82). Another shelterin protein, human protection of telomeres 1 (hPOT1), binds directly to the free 3' terminus and acts as a critical regulator of its length by both unfolding telomere G4s for hTERT to gain access and preventing aberrant extension by blocking access to the free 3' end (84,85). Direct inhibition of hPOT1 results in reduced cellular proliferation of cells requiring telomerase activity (86). Thus, targeting the telomeric G-quadruplex is a promising route for anti-cancer therapeutic development (87).

Non-telomere G-quadruplexes

The first glimpse of non-telomeric G-quadruplex functionality came from a report in 1988 in which evidence of an inter-stranded tetramer G-quadruplex was described in immunoglobulin switch regions based on mobility shift assays (23). This discovery was essential in explaining the peculiar rearrangements observed within the human insulin gene promoter reported six years prior (88). It was subsequently discovered that the Bloom's syndrome helicase is able to unwind DNA G-quadruplexes *in vitro* (89), prompting sequencing studies of G4 helicase knock out cells which ultimately confirmed that G-quadruplexes play a direct role in recombination events (90). In 1994, Woodford and colleagues observed the first biologically relevant intra-strand G-quadruplex in the promoter of the chicken *β*-globin gene (91). Using a standard linear polymerase reaction, and clever mutational analyses, the authors showed that this tetraplex formation stalled DNA polymerase in a potassium-dependent manner and, further, demonstrated that the effect was not observed when Hoogsteen hydrogen bonding was disrupted through mutation with 7-deaza-dGTP. These discoveries were paramount for understanding, in part, the "obscure" mechanism by which *c-MYC* transcription is regulated.

The *c-MYC* gene is a part of the MYC family of oncogenes which are best known for their central role in cell growth and proliferation (92). MYC is overexpressed in as much as half of all human cancers, and its ability to initiate and maintain tumorigenesis has long made it an attractive

anticancer target (93). However, it has thus far been “undruggable” using traditional protein-centric techniques due to the nature of its protein structure, and so alternative strategies for its inhibition have been in the works for decades (94). One of the major regulatory control elements of *c-MYC* transcription, located from -115 to -142 bp upstream of the promoter, known as the nuclease-hypersensitive element III_1 (NHE-III_1), accounts for as much as 85% of total transcription (95,96). The NHE-III_1 is interesting because of its strand asymmetry, meaning that one strand is nearly exclusively homopurine. Further, *in vitro* this sequence is able to adopt a non-helical, atypical DNA structure (97). In their 1998 seminal work, Simonsson et al. showed that the *c-MYC* NHE-III_1 sequence forms a unique intra-strand antiparallel G-quadruplex, indicating for the first time the possibility of a G-quadruplex-mediated transcriptional control mechanism (98). Four years later, transcriptional repression of c-MYC via G-quadruplex stabilizing small molecules was confirmed, sparking a much wider interest in the G-quadruplex as a potential new class of receptor (99).

To date, *In vitro* studies conducted with G4-specific antibodies and fluorescent probes (26-30), G-quadruplex sequencing (31), and targeted bioinformatic inquiries (11,12) have unearthed hundreds of thousands of places in the human genome where G-quadruplexes putatively form (putative quadruplex forming sequences, PQSs). PQSs are non-randomly distributed, conserved between species, and primarily reside in functionally important regions (100). Further, these motifs are significantly associated with oncogene promoters and somatic copy number alterations related to cancer development (31). To date, a variety of *monomeric* oncogenic promoter G-quadruplexes have been validated both *in vitro* and in functional cell-based assays: *c-MYC (16), KRAS (101)*, *HRAS (102), HIF-1α (103), VEGF (104)*, and *hTERT (105,106)*.

Genome-wide mapping studies have now identified what appears to be a universal epigenetic mechanism of genomic G-quadruplexes. PQSs are highly susceptible to the formation of 8-oxoguanine (8-oxoG) as the result of oxidative insult (107). These lesions are recognized by 8-oxoG DNA glycosylase (OGG1) an initiator of the base excision-repair (BER) pathway (108). Excision by OGG1 of the oxidized base results in an apurinic (AP) site, which in turn, leads to the recruitment of AP endonuclease 1 (APE1) (108,109). Using an innovative “AP-seq” technique, along with traditional ChIP- and G4-sequencing, Roychoudhury and colleagues have now shown

that APE1, in response to 8-oxoG induced AP site generation, is essential in facilitating the formation of G-quadruplexes throughout the genome (109). Further, they show that the recruitment of APE1 to 8-oxoG G4 sites directly increases transcription factor loading onto the promoter, providing essential insight into the role of G-quadruplexes in transcriptional control.

Aside from telomere protection, recombination events, and transcriptional regulation, G-quadruplexes serve a variety of other biological roles. For instance, in mammals, the telomere is transcribed into a long non-coding G-rich stretch of RNA ($UUAGGG_n$) known as "TERRA" (Telomeric Repeat-containing RNA) (110). TERRA forms parallel stacked RNA multimeric G-quadruplexes (111), and is currently under investigation as an essential component in telomere maintenance, genome stability, and heterochromatin formation (110). Other "structural" RNA G-quadruplexes have been identified in ribosomal RNA (112), recruitment sites for histone modifiers (113), and transiently in other RNA transcripts across the cell (114). RNA G-quadruplexes play important roles in mRNA translation, processing, and targeting (25). One early example is the identification of a G-quadruplex motif within the insulin-like growth factor II mRNA that flanks the major cleavage site within its 3' untranslated region (3' UTR) and appears essential in its processing (115). In fact, many such examples exist of mRNA with G-quadruplex formation within their 5' and 3' UTRs, and this mode of regulation appears to be a common mechanism [see ref (116) for a recent review]. For instance, the NRAS proto-oncogene mRNA transcript contains a G-quadruplex motif in its 5' UTR 222 nucleotides upstream of the translation start site that acts as a repressor of its translation (117). Altogether, G-quadruplexes are diverse and important nucleic acid secondary structures involved in a wide array of essential biological processes, making them exciting new targets in drug discovery (**Figure 5**).

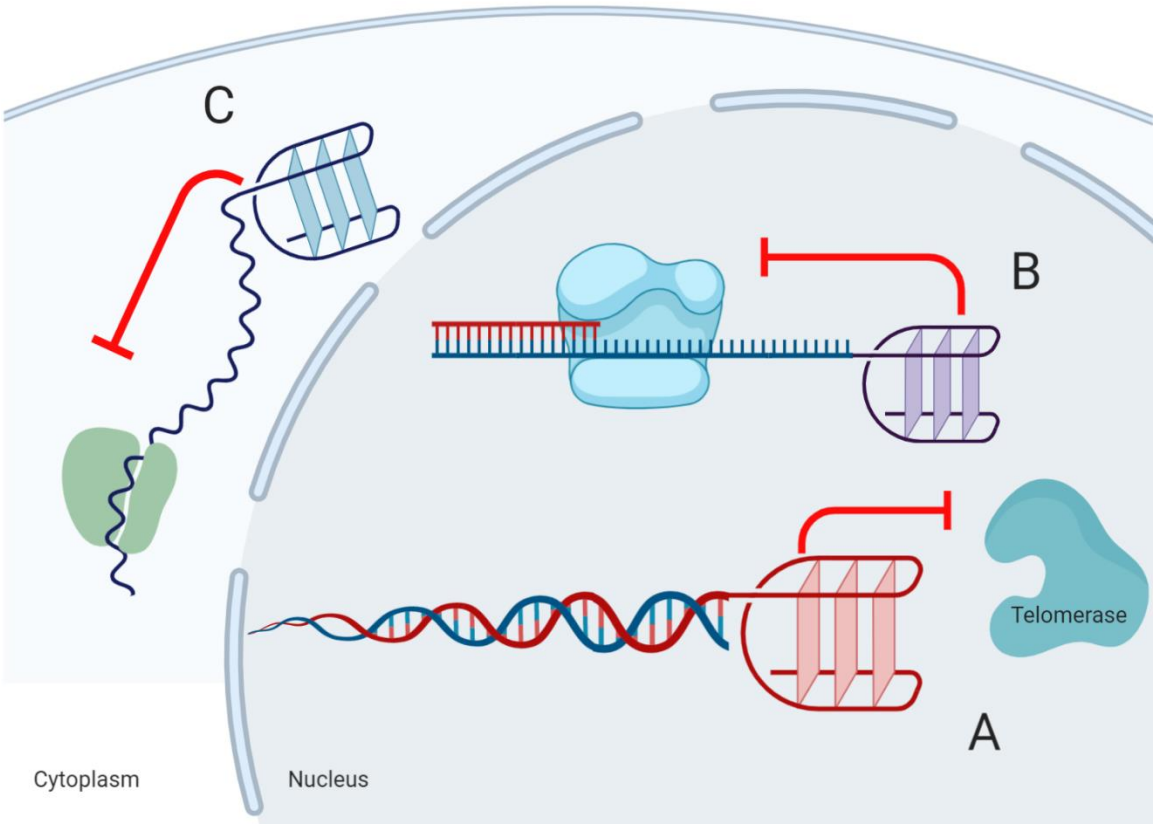
G-quadruplexes in nanotechnology and therapeutics

G-quadruplex based systems have applications outside of their biological functionality, such as in nanotechnology (118) and as therapeutics (119). When Ivar Bang made the discovery that concentrated guanosine monophosphate solutions would spontaneously form "gels", he suggested that the gel was the result of polymerization of the nucleotides (37). This polymerization

results in long stacking interactions of hundreds of G-tetrad units, now known as G4-wires (120). Due to their extensive π overlap, structural rigidity, and guanine having the lowest ionization potential of all nucleotides, G4-wires are now being assessed for their use in nanoelectronics applications (121,122). Further, the formation of G-quadruplexes can be rapid and reversible, making them well suited as components or “building blocks” of nanodevices (118). For instance, G4s are now being used as biosensor logic gates, where the formation of a G4 in response to a given analyte results in the complexation of hemin with the G4, which in turn, leads to an amplified redox reaction that can be monitored spectroscopically (123).

G-quadruplexes have also made their way into the clinic as macromolecular drugs called aptamers (119). DNA aptamers are short oligonucleotides that bind specifically to biological targets to exert their functions. Currently, DNA G-quadruplex aptamers are being developed for the purposes of anticoagulation, anti-cancer, antiviral, antibacterial, antifungal, and as treatments for a variety of human maladies such as inflammatory diseases, prion diseases, and thyroid disorders (119,124). A notable example is the G-quadruplex aptamer AS1411 discovered by Bates et al. (125,126). Originally, AS1411 was shown to exert its anti-cancer activity by selectively binding to nucleolin expressed on the cell surfaces of cancer cells (as nucleolin is absent from normal cell surfaces) (126). At the time it was believed that nucleolin acted as a cancer-specific receptor for AS1411. Once internalized, the aptamer could interfere with nucleolin’s pro-survival functions, resulting in cancer cell death (126). However, recent work has called this mechanism into question (127). AS1411 entered clinical trials but did not progress past phase 2 (127). Overall, it was well tolerated by patients and exhibited an excellent safety profile. The response rate in patients was low, although 3 patients with renal cell carcinoma and 4 with acute myeloid leukemia had favorable and lasting responses (127). Since its clinical assessment AS1411 has spurred great interest in anti-cancer G-quadruplex aptamer development.

Figure 5. Locations and functions of G-quadruplexes in cells. G4s are non-randomly distributed in the genome, and specifically occupy telomeres (A), gene promoters or replication forks (B), and 5' UTRs of mRNAs (C). Red T-bars indicate their putative functions (e.g. prevention of telomere extension by telomerase [A] or blockage of transcription [B], replication [B], or translation [C]).

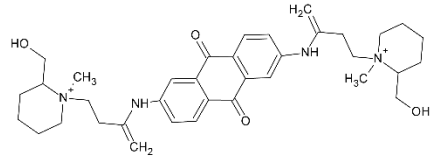

C
B
Telomerase
A
Cytoplasm
Nucleus

Targeting G-quadruplexes

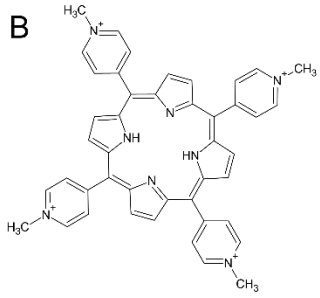
Due to the wealth of conformational diversity, the various monomeric G-quadruplex conformations have been targeted extensively for over two decades with hopes of finding selective small molecules with *in vivo* efficacy. The first report of a telomere-interacting small molecule was in 1997 when Sun et al. demonstrated that a di-substituted anthraquinone derivative was able to stabilize the telomere G-quadruplex and effectively inhibit telomerase's ability to extend *in vitro* (101). Soon after, the cationic porphyrin molecule TMPyP4 was identified which had a 2-fold preference for binding G-quadruplex over duplex DNA (128). Since then, a variety of other G4 selective molecules have been developed with drug-like affinities in the low nanomolar range, notably: BRACO-19 (129), Telomestatin (130), Pyridostatin (131), Phen-DC3 (132), CX-5461 (133), and CX-3543 (a.k.a. "Quarfloxin") (134) (**Figure 6**). TMPyP4, BRACO-19, Telomestatin, and Pyridostatin all interact with the telomeres, and induce a variety of biological responses, such as: telomerase inhibition, induction of DNA damage responses, telomere shortening, selective induction of apoptosis in cancer cells, uncapping of shelterin proteins, induction of double-stranded breaks, and inhibition of oncogene transcription (via promoter G-quadruplex stabilization) (135). CX-5461 was initially shown to block the initiation of ribosomal RNA (rRNA) synthesis, and this was believed to be its anticancer mechanism (136). However, it was recently revealed that CX-5461 interacts strongly with G-quadruplexes, and in doing so induces double-stranded DNA breaks which effectively kill DNA repair-deficient cancer cells (133). Currently, CX-5461 is in phase I clinical trials as a general cancer treatment (Trial NCT02719977, opened May 2016), although the trial is not currently recruiting patients. Conversely, Quarfloxin has completed Phase II clinical trials and was shown effective against neuroendocrine tumors, carcinoid tumors, and lymphoma (134). Interestingly, the mechanism of Quarfloxin appears to be the disruption of nucleolin binding to ribosomal DNA, causing nucleolin to reroute to the *c-MYC* promoter G-quadruplex to repress it. However, Phase III clinical trials have stopped due to its high serum albumin binding (18). A variety of other G4 ligands can be found in the G-quadruplex ligands database, or "G4LDB" (17), which was created in 2012 and contains nearly 1,000 confirmed G4 ligands.

Figure 6. Structures of some common high affinity G4 ligands. (A) 2,6-diamidoanthraquinone, (B) TMPyP4, (C) Phen-DC3, (D) BRACO-19, (E) Telomestatin, (F) CX-3543, and (G) CX-5461 (Quarfloxin).

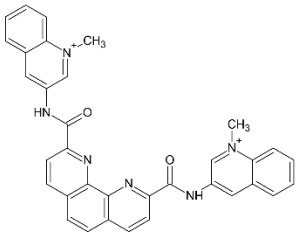
A



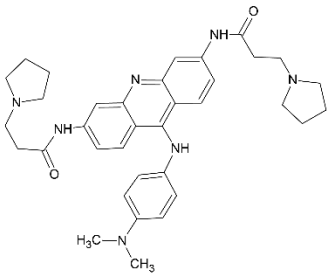
B



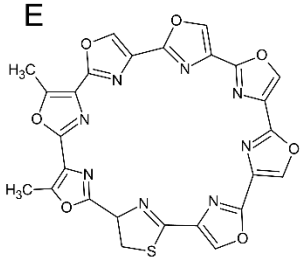
C



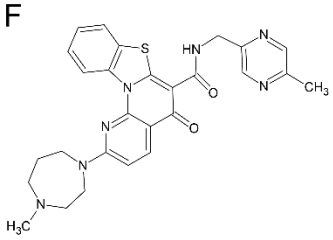
D



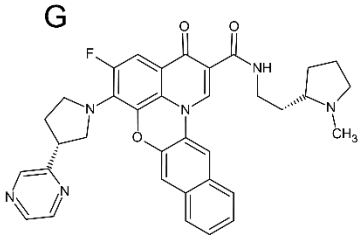
E



F

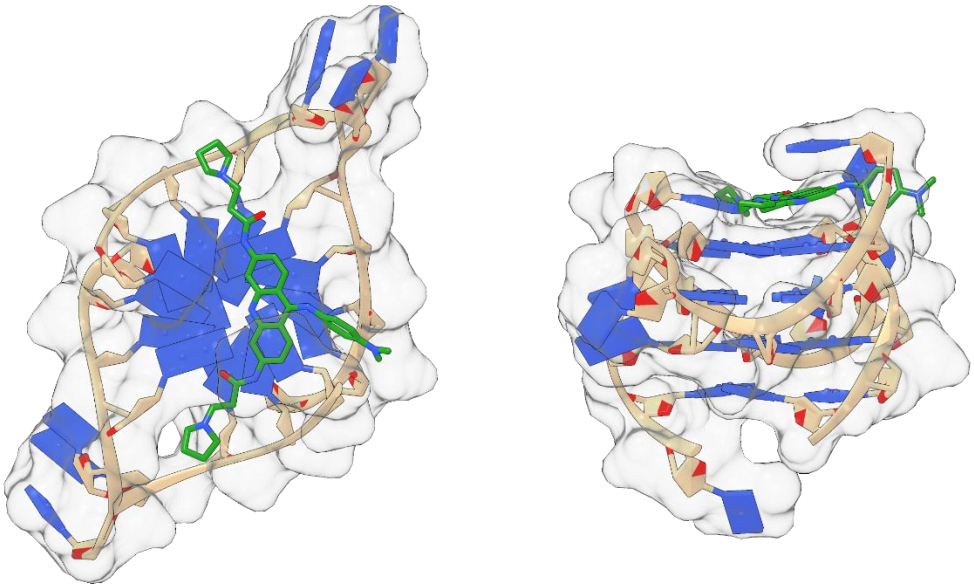


G



G-quadruplex stabilizing ligands are clearly viable anti-tumor agents. Why then have the majority of identified G4 ligands not succeeded clinically? The answer, in part, is two-fold: (1) monomeric G4s share too similar a structural landscape [selective targeting of the G-tetrad face is akin to the protein kinase promiscuity problem (15)] and, (2) attempts at improving the selectivity of small molecules towards these monomeric G4s, which is typically achieved by adding loop interacting chemical moieties, tend to render them less “drug-like” or impact their bioavailability (18,137). For instance, BRACO-19, which inhibits telomerase with an IC_{50} (half maximal inhibitory concentration) of 115 nM, was developed from a core disubstituted acridine scaffold (129). *In vitro*, BRACO-19 binds tightly to the telomere and leads to uncapping of the shelterin protein complex, ultimately eliciting a severe telomere DNA damage response (83). A crystal structure of BRACO-19 in complex with the human telomere sequence $d(TAGGGTTAGGGT)_2$ was reported, showing that BRACO-19 preferentially end-pastes onto the terminal G-tetrad face (**Figure 7**) (138). BRACO-19 has demonstrable anti-tumor activity in cells derived from epidermoid carcinoma, colorectal cancer, uterus carcinoma, and prostate cancer (although also cardiotoxic) (135). The acridine scaffold is well known to end-paste on the terminal tetrads of the telomere G-quadruplexes (139). The addition of the pyrrolidine rings to the core acridine scaffold in BRACO-19 increases its selectivity via interaction with the G4 loops. However, the increase in both molecular weight (MW) and charge from the protonated tertiary amines reduces its ability to cross membranes (140), rendering it less bioavailable and drug-like. While this is not the rule (See Quarfloxin above), it emphasizes the problem of selectivity faced by the G4 drug discovery community (18).

Figure 7. Top and side views of BRACO-19 bound to the human telomere sequence d(TAGGGTTAGGGT)$_2$ via an end-pasting mechanism. BRACO-19 is shown in green and the telomere G-quadruplex is shown as a ribbon model with semi-transparent space-fill.



Most rational drug discovery studies targeting G4s to date have focused on targeting the monomeric forms of G-quadruplexes—owing to the wealth of atomic structures available (>250 in the Protein Data Bank). It has been evident since the 1980's [but observed well before then (37)] that guanine-rich DNAs can multimerize under physiologically relevant buffer conditions and oligonucleotide concentrations (19,20). There is now evidence for the occurrence of G4 multimers in the telomeres (141,142) and in oncogene promoters [*hTERT* (143-145), *c-MYB* (146), *kRAS* (147), *c-MYC* (148), and *c-KIT* (149)]. The few multimers with high-resolution structures available are the putative T:H:H:T dimer of the *c-MYB* promoter (56), a synthetic aptamer two-stacked two-tetrad G4 (57), and two, two-stacked two-tetrad multimers containing left-handed "Z-G4s" (50,51). Importantly, G4 multimers offer unique binding sites among their loop and groove interface that may allow circumvention of the non-specificity problem encountered with monomers (141). Indeed, this potential for selective-targeting of a multimer over monomer G4 has already been demonstrated using a non-drug-like chiral metallo-supramolecular complex targeting the higher-order telomere sequence (150). Altogether, these findings highlight the dire need for a better understanding of multimeric G-quadruplexes as specific targets for rational drug discovery campaigns.

Summary of Dissertation Works

The ISB approach to solve the structure and spatial organization of higher order protein, RNA, protein-RNA, and protein-DNA systems has been a paradigm for over 20 years (151), yet no such approach has been applied to higher-order G-quadruplex DNA structures. Here I have outlined a clear need for tools which will enable a detailed description of these multimeric systems which putatively reside in both telomeres and oncogene promoters. To this end, the major goal of this work is to apply the ISB approach to G-quadruplex multimers for their use in rational drug discovery.

In Chapter II I have characterized the major structure of the higher-order telomere G-quadruplex. I first applied size-exclusion chromatography coupled with small-angle X-ray scattering

(SEC-SAXS) to the telomere sequences Tel48 (d[TTAGGG]$_8$), Tel72 (d[TTAGGG]$_{12}$), and Tel96 (d[TTAGGG]$_{16}$) which allowed for their qualitative and quantitative structural descriptions. I subsequently showed that the higher-order telomere sequences preferentially maximize their G4 formation. I then rigorously dissected the major structural components of the intramolecular "dimer", Tel48, and a variety of mutant sequences by circular dichroism (CD) spectroscopy. This mutational study revealed that the higher-order human telomere favors the hybrid-1 and hybrid-2 topologies, adopting a ~25:75 ratio of hybrid-1 and hybrid-2. Next, by combining monomeric atomic coordinate files from previous NMR studies with MD simulations, I constructed and refined the first ever all-atom molecular models of the extend human telomere at the highest resolution to date. From these models, multiple unique sites among junctions and grooves were revealed that could be used as receptors in future rational drug discovery campaigns. This work has been submitted to *Nucleic Acids Research* and is in review as of the drafting of this dissertation.

Chapter III uses the telomere G-quadruplex multimer models described above as targets in a massive virtual drug discovery campaign. In this chapter I applied the program Surflex-Dock v2.11 to screen millions of compounds virtually against G4-junctional sites formed between the hybrid-1 and hybrid-2 moieties. From this screen I purchased 37 compounds and, using CD melting and analytical ultracentrifugation (AUC) binding studies, found that one compound, C37, binds in a 1:1 stoichiometry with the G4-juntions (e.g. 1:1 with the telomere dimer, tel48, 2:1 with the telomere trimer, tel72. This binding is specific for the higher-order telomere G4s over monomer telomere G-quadruplexes. C37 is currently being investigated for its binding mode and potential anticancer effects in cells.

In Chapter IV I applied a similar ISB approach to the *hTERT* core promoter G-quadruplex. In this study I used circular dichroism to characterize the full-length and truncated *hTERT* core promoter sequences. ^{1}H-NMR spectroscopy was employed to investigate the extent of G-tetrad formation as well as investigate the possibility of hairpin formation. I then utilized a novel circular dichroism-monitored DNase I reaction to confirm that no hairpin forms within the hTERT WT sequence, but does in the artificial hairpin control. Hydrodynamic investigations were then conducted in combination with hydrodynamic bead model calculations from MD-derived atomistic

models to investigate the size and organization of G4 domains. I then used SEC-SAXS to probe the overall higher-order assemblies which confirmed that the WT promoter is an all-parallel stacked G4 system. This work is published in *Nucleic Acids Research* and is reproduced in Chapter IV.

G4 DNA virtual drug discovery is still in its infancy, and so in light of this Chapter V is a review of the contemporary G-quadruplex virtual screening (VS) techniques. In this chapter I have presented the various methods used in pharmacophore and docking-based screens, library size and preparation, scoring functions, and the future of computational VS approaches. I also discuss past successes and failures in the field and point out pitfalls that can be avoided in future campaigns. Lastly, I provide recommendations on how to properly report on virtual screening campaigns by describing guidelines for future investigators. This work is published in the journal *Biochimie* and is reproduced in Chapter V.

In my final experimental section, I applied our recommendations outlined in Chapter V to targeting the *hTERT* core promoter G-quadruplex multimer characterized in Chapter IV. I began with a massive virtual screening campaign using the docking program Surflex-Dock v2.11, whereby I targeted 12 different loop and groove pockets of the hTERT G4 *in silico*. I then clustered the resulting top hits using a hierarchical clustering algorithm and identified 69 unique molecular scaffolds which were available for purchase. Multiple rounds of a high-throughput thermal denaturation screens and orthogonal biophysical assays were then used to assess drug selectivity for the hTERT multimer over duplex, triplex, and monomer G-quadruplex DNA topologies. I subsequently utilized an in-house automated docking to MD simulation pipeline to investigate the potential binding modes of the top hits, revealing multiple preferential loop and groove binding modes. Lastly, I employed quantitative real-time polymerase chain reaction (qRT-PCR) and proliferation assays to assess compound efficacy in human breast cancer cells, confirming that one compound, 3B1, reduces *hTERT* transcription. Compound 3B1 is currently being investigated extensively in breast cancer cells as a lead molecule.

CHAPTER II

THE SOLUTION STRUCTURES OF HIGHER-ORDER HUMAN TELOMERE G-QUADRUPLEX MULTIMERS

Human telomeres contain the repeat DNA sequence 5'(TTAGGG), with duplex regions that are several kilobases long terminating in a 3' single-stranded overhang. The structure of the single-stranded overhang is not known with certainty, with disparate modes proposed in the literature. We report here the results of an integrated structural biology approach that combines small-angle X-ray scattering, circular dichroism (CD), analytical ultracentrifugation, size-exclusion column chromatography and molecular dynamics simulations that provide the most detailed characterization to date of the structure of the telomeric overhang. We find that the single-stranded sequences 5'$(TTAGGG)_n$, with n=8, 12, and 16, fold into multimeric structures containing the maximal number (2, 3, and 4, respectively) of contiguous G4 units with no long gaps between units. The G4 units are a mixture of hybrid-1 and hybrid-2 conformers. In the multimeric structures, G4 units interact, at least transiently, at the interfaces between units to produce distinctive CD signatures. Global fitting of our hydrodynamic and scattering data to a worm-like chain (WLC) model indicates that these multimeric G4 structures are semi-flexible, with a persistence length of about 34 Å. Investigations of its flexibility using MD simulations reveal stacking, unstacking, and coiling movements, which yield unique sites for drug targeting.

Introduction

Telomeres are structures found at the end of eukaryotic chromosomes which protect genomic DNA from degradation, end-to-end fusion, and homologous recombination (64,65). The human telomere consists of the repeat $d(TTAGGG)_n$, and ranges from 5-25 kb in length with an extended single-stranded 3' overhang of a few hundred bases in non-germ cells (63). This locus has long been associated with human diseases, such as cancer (59) and telomeropathies (60), as well as aging (61) and general genome homeostasis (62). In normal somatic cells, each round of cellular division results in a shortening of the telomere due to the so-called end replication problem—a mechanism believed to be protective against uncontrolled replication (66). Once the telomere has become critically short in normal (non-stem) cells, a DNA damage response is triggered, resulting in uncapping of the telomere-bound shelterin proteins and, eventually, apoptosis (65,66). Cancer cells avoid this fate by utilizing mechanisms that restore telomere length. In more than 85% of cancers, this is accomplished by reactivating human telomerase reverse transcriptase (hTERT), a ribonucleoprotein that extends the telomere 3' overhang (10,68,69). G-quadruplex (G4) formation in the telomere overhang can inhibit hTERT binding and extension function (7). Treating cells with telomere G4-specific small molecules leads to uncapping of the shelterin proteins and a sequestering of the free single-stranded telomere overhang, ultimately resulting in a telomere-specific DNA damage response (82,83,87). These findings have made telomere G4 an attractive cancer target (87).

G-quadruplexes form in guanine-rich sequences, in which guanine tracts interact to form square planar tetrads (G-tetrads) that stack atop one another and are stabilized by coordinating cations, pi-stacking interactions, and a Hoogsteen hydrogen bonding network (40). Many telomere G4 topologies have been characterized at the atomic level by X-ray crystallography and NMR studies. These studies have demonstrated that the monomeric form of the human telomere can exist as parallel (75), hybrid 3+1 (76,77), antiparallel (74), and two-tetrad antiparallel (47) structures under various ionic and crowding conditions. The Yang lab (76,77,152), Patel lab (153,154), and

we (78) have since shown that the wild-type telomere adopts primarily the hybrid-1 and hybrid-2 topologies in physiologically relevant solution conditions. The Yang lab has shown by NMR that *in vitro* the wild-type monomeric telomere sequence of the form $(TTAGGG)_4T$ exists in a dynamic equilibrium of hybrid-2 (~75%) and hybrid-1 (~25%) (46).

Telomere G-quadruplexes have also been observed directly in cells. *In vivo,* G4-specific antibodies and fluorescent ligands have confirmed the formation of telomere G4s (26-28). Using the sequence $AGGG(TTAGGG)_3$, Hong-Liang and colleagues used ^{19}F-NMR cell studies to show that the hybrid-1, -2, and a two-tetrad anti-parallel type (hybrid-3), but not the parallel or antiparallel "basket" topologies, spontaneously form when injected into live HeLa cells (79). Altogether, these studies demonstrate that the most physiologically and thermodynamically relevant monomeric telomere conformations are of the hybrid type.

Although the monomeric telomere G-quadruplex has been extensively studied, there is little structural information on longer telomere sequences forming higher-order telomere structures. Conservative estimates of the length of the single-stranded overhang of the human telomere in fibroblasts indicate that the sequence exceeds the ~30 nucleotides necessary for formation of a single telomere G-quadruplex. Estimates of "normal" single-stranded overhangs range from ~50 to >600 nucleotides (58,63), supporting the possibility of multiple G4s forming in tandem. There have been few attempts to characterize these systems at the atomic level because of the difficulties involving guanine imino overlap and structural polymorphism which hamper NMR studies (46), and the difficulty of obtaining quality crystals for X-ray diffraction (78). Elucidating this higher-order structure is important, as its role in mediating interactions with shelterin proteins, single-stranded binding proteins, and telomerase is critical in maintaining genomic integrity (64,155,156).

To date, only a few low-resolution molecular models and characterizations were reported for the long telomere sequences. In 2006, using a combination of gel electrophoresis, CD, and UV-melting, Yu *et al*. proposed that the telomere multimer of the form $(TTAGGG)_n$, where n is 4, 8, or 12, maximizes its usage of G-tracts by forming a "beads-on-a-string" assembly of a variety of telomere topologies (parallel, antiparallel, and hybrid) (55). In 2009 Renčiuk and colleagues, using CD and PAGE experiments, came to the same general conclusion that the higher-order telomere

is capable of folding into multiple conformations in K^+ buffers (parallel, antiparallel, or hybrid) but also that they have the potential to stack, depending on the amount of macromolecular crowding (157). The same year Xu *et al*. demonstrated the formation of higher-order G-quadruplex formation in 96 nucleotide (nt) long telomere sequences by atomic force microscopy (AFM) (158). While the authors arrived at a similar conclusion about the overall higher order assembly (e.g. maximized G4 formation and potential for G4-G4 interactions), they did not report on the topologies of the G4 subunits. A later AFM investigation of a 96 nt long telomere sequence, $(TTAGGG)_{16}$, by Wang and colleagues reported the presence of gaps between G4 units, and suggested that the extended sequences "rarely" maximize G-tract usage (159). A similar conclusion was drawn from low-resolution studies using electron microscopy (EM), single-molecule magnetic tweezers, and single-molecule force ramp assays (160,161). Although, these studies may suffer from insufficient sample annealing protocols or equilibration times. Our prior biophysical studies investigating the secondary and tertiary structure of the higher-order telomere sequences $(TTAGGG)_n$ and $(TTAGGG)_nTT$, where n = 4, 8, 12, 16 and 32, gave evidence that these sequences preferentially maximize G-tract usage, and preferentially form a mixture of the hybrid-1 and hybrid-2 conformations (141,162,163). Subsequent investigations by molecular dynamics (MD) simulations, analytical ultracentrifugation (AUC) (163), and differential scanning calorimetry (DSC) (162) studies indicated that, overall, the extended telomere G4s adopt compact, somewhat rod-like structures via stacking interactions between G4 subunits and intervening TTA linkers (162). The best-fit models from these analyses were alternating (5') hybrid-1 (3') hybrid-2, referred to as hybrid-12 and hybrid-121, for n = 8 and n = 12 runs, respectively. Interestingly, thermodynamic studies of these two higher-order systems revealed that "each quadruplex in the higher-order structures is not independent and identical but is thermodynamically unique and is influenced by its neighbors" (162). Clearly, there is no consensus on the higher-order telomere's behavior in solution. Low-resolution imaging and single-molecule studies would suggest a very flexible beads-on-a-string arrangement with large gaps occurring between G-quadruplexes, whereas the latter investigations suggest a more rigid structure, with maximal G-quadruplex formation.

Using an integrative structural biology approach (164,165), which combines CD, hydrodynamics, molecular dynamics, and small-angle X-ray scattering (SAXS), we show that the telomeric sequences form the maximal number of G4 units without any long gaps. Modeling the hydrodynamic and scattering-derived properties of sequences from 24 nt to 96 nt to a worm-like chain (WLC) model reveals a persistence length of ~34 Å, which is in between that of single-stranded DNA (ssDNA) (~22 Å) (166) and double-stranded DNA (dsDNA) (~550 Å) (167), indicating that the extended telomere G4 is semi-flexible. This flexibility is consistent with MD simulations, which show transient stacking interfaces that create potentially unique binding grooves useful in drug targeting. We follow this with an extensive sequence analysis of the sequence $d(TTAGGG)_8$ to determine the major constituent G4 topologies. Using CD and mutational analyses we show that the higher-order human telomere is composed of a ratio of hybrid-1 (~25%) and hybrid-2 (~75%) topologies. Our results are in excellent agreement with prior hydrodynamic and NMR analyses of the human telomere sequences (77,154,163). The resulting structural ensembles provide the first "medium-resolution" look at the conformational heterogeneity and dynamics of the higher-order telomere G-quadruplex.

Materials and Methods

Oligonucleotides

Oligonucleotide sequences were purchased from IDT (Integrated DNA Technologies, Coralville, IA) with standard desalting. Upon receipt, stock oligos were dissolved in MilliQ ultrapure water (18.2 MΩ x cm at 25°C) at 1 mM and stored at -20.0°C until use. All experiments were carried out in a potassium phosphate buffer (6 mM Na_2HPO_4, 2 mM NaH_2PO_4, 185 mM KCl, 1 mM Na_2EDTA, pH 7.2). Folding was achieved by diluting stock oligos into buffer and boiling in a water bath for 20 minutes, followed by slow cooling overnight. Purification was achieved using size exclusion chromatography (SEC) as detailed previously (168). Briefly, oligos were annealed at concentrations of 40-60 μM, filtered through 0.2 μm filters, and injected onto an equilibrated Superdex 75 16/600 SEC column (GE Healthcare 28-9893-33) using a Waters 600 HPLC system.

The flow rate was maintained at 0.5 mL/min and sample fractions were collected every 2 minutes from 100 to 180 minutes run time. The molecular weights of fractionated species were estimated based on a regression analysis of elution time vs. log(MW) of protein standards (Sigma #69385), the major folded species were visually evident as symmetric peaks when monitored at 260 nm (or 280 nm for protein standards). Fractionated samples were pooled and stored at 4°C prior to concentration. Where applicable, pooled fractions were concentrated using Pierce protein concentration devices with 3k MWCO (Thermo #88512, #88515, and #88525) which were rinsed free of glycerol. For AUC and SEC-SAXS experiments, samples were dialyzed after concentration using Spectra/Por Float-A-Lyzers G2 3.5 kDa (Sigma #Z726060) in order to buffer match. Concentrations were determined using molar extinction coefficient given in **Table 1**.

Table 1. Names, properties, and sequences of oligonucleotides used in this study.

Name	Sequence	Length (nt)	MW (kDa)	E260 ($M^{-1}cm^{-1}$)
2JSL	TAGGGTTAGGGTTAGGGTTAGGGTT	25	7.9	253100
Tel48	$(TTAGGGTTAGGGTTAGGGTTAGGG)_2$	48	15.2	489100
Tel72	$(TTAGGGTTAGGGTTAGGGTTAGGG)_3$	72	22.8	733400
Tel96	$(TTAGGGTTAGGGTTAGGGTTAGGG)_4$	96	30.5	977800
Tel49	TTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGT	49	15.5	497500
Tel50	TTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTT	50	15.8	505600
M1	TAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGG	47	14.9	480900
M2	TAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTT	49	15.5	497500
hybrid-12*	TT_GGGTTAGGGTTAGGGTTAGGGATAGGGTTAGGGTTAGGGTTAGGGT	48	15.2	488000
hybrid-11*	TT_GGGTTAGGGTTAGGGTTAGGGAT_GGGTTAGGGTTAGGGTTAGGGA	47	14.9	479200
hybrid-21*	TTAGGGTTAGGGTTAGGGTTAGGGTT_GGGTTAGGGTTAGGGTTAGGGA	48	15.2	488700
M3	AGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTT	48	15.2	489500
hybrid-32*	AGGGTTAGGGTTAIGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTT	48	15.2	489850
M4	GGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTT	47	14.9	476000
M5	AGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGT	47	14.9	481100
hybrid-33*	AGGGTTAGGGTTAIGGTTAGGGTTAGGGTTAGGGTTAIGGTTAGGGT	47	14.9	482100

_ = removed base, I = Inosine sub, * = sequence has internal stabilizing modification

Size exclusion chromatography (SEC) determination of Stokes radii

Elution times from re-injections of SEC purified fractions were used in the method of Irvine (169) to determine Stokes radii, which were converted to translational diffusion (D_t) coefficients for use in Multi-HYDFIT hydrodynamic modeling (see below). Stokes radii were determined from a regression analysis of elution time vs. log(MW) of protein standards (Sigma #69385).

Analytical ultracentrifugation (AUC)

Sedimentation velocity (SV) experiments were performed in a Beckman Coulter ProteomeLab XL-A analytical ultracentrifuge (Beckman Coulter Inc., Brea, CA) at 20.0°C and 40,000 rpm in standard 2-sector cells using either an An60Ti or An50Ti rotor. Samples were equilibrated in the rotor at 20.0°C for at least 1 hour prior to the collection of 100 scans over an 8-hour period. Initial analyses were performed in SEDFIT (170) using the continuous C(s) model with resolution 100 and S range from 0 to 10. A partial specific volume of 0.55 mL/g for DNA G-quadruplexes was used as previously determined (163). The Tel72 and Tel96 sequence sedimentation coefficients were additionally corrected for any concentration-dependence using three separate concentrations.

Circular dichroism

CD spectra were collected on a Jasco-710 spectropolarimeter (Jasco Inc. Eason, MD) equipped with a Peltier thermostat regulated cell holder equilibrated to 20.0°C. Spectra were collected using the following instrument parameters: 1 cm path length quartz cuvettes, 1.0 nm step size, 200 nm/min scan rate, 1.0 nm bandwidth, 2 second integration time, and 4 scan accumulation. Spectra were corrected by subtracting a buffer blank and normalized to molar circular dichroism ($\Delta\varepsilon$, $M^{-1}cm^{-1}$) based on DNA strand concentration using the following equation:

$$\Delta\varepsilon = \theta/(32982cl)$$

where θ is ellipticity in millidegrees, c is molar DNA concentration in mol/L, and l is the path length of the cell in cm. Comparison or fitting of CD spectra with their monomer theoretical spectra was done manually in Microsoft Excel using spectra from a previously reported database (171).

Residual sum of squares (RSS) analysis of the CD ΔΔε “residuals” was carried out and plotted in Origin 2020.

Size exclusion chromatography resolved small angle X-ray scattering (SEC-SAXS)

SAXS was performed at BioCAT (beamline 18ID at the Advanced Photon Source, Chicago) with in-line size exclusion chromatography. Samples in BPEK buffer (6 mM Na_2HPO_4, 2 mM NaH_2PO_4, 185 mM KCl, 1 mM Na_2EDTA, pH 7.2) were loaded onto an equilibrated Superdex 75 10/300 GL column, which was maintained at a constant flow rate of 0.7 mL/min using an AKTA Pure FPLC (GE Healthcare Life Sciences) and the eluate after it passed through the UV monitor was directed through the SAXS flow cell, which consists of a 1 mm ID quartz capillary with 50 µm walls. A co-flowing buffer sheath was used to separate the sample from the capillary walls, helping to prevent radiation damage (172). Scattering intensity was recorded using a Pilatus3 1M (Dectris) detector which was placed 3.5 m from the sample giving access to a q-range of 0.004 $Å^{-1}$ to 0.4 $Å^{-1}$. A series of 0.5 second exposures was acquired every 2 seconds during elution and data was reduced using BioXTAS RAW 1.6.3 (173). Buffer blanks were created by averaging regions flanking the elution peak and subtracted from exposures selected from the elution peak to create the I(q) vs. q curves used for subsequent analyses. More information on SAXS data collection, reduction and interpretation can be found in **Table 2**. SAXS sample preparation, analysis, data reduction, and data presentation has been done in close accordance with recent guidelines (174).

Table 2. Tabulated collection parameters, data reduction methods, and data analyses for small-angle X-ray scattering data.

(a) Sample Details.

	2JSL	Tel48	Tel72	Tel96
Organism	synthetic	synthetic	synthetic	synthetic
Source	IDT	IDT	IDT	IDT
Extinction coefficient (nearest neighbor approximation) (M^{-1} cm^{-1})	253100	489000	733400	974870
vbar (cm^3/g) (estimate)	0.55	0.55	0.55	0.55
M from chemical composition (Da)	7879	15212	22849	30486
SEC-SAXS column, 10 x 300 Superdex 75				
Loading concentration (mg/mL)	7.0	13.0	10.0	6.0
Injection volume (µL)	300	300	250	440
Flow rate (mL/min)	0.7	0.7	0.7	0.7
Solvent (solvent blanks taken from SEC flow through prior to elution of protein)	6 mM Na2HPO4, 2 mM NaH2PO4, 185 mM KCl, 1 mM Na2EDTA, pH 7.2	6 mM Na2HPO4, 2 mM NaH2PO4, 185 mM KCl, 1 mM Na2EDTA, pH 7.2	6 mM Na2HPO4, 2 mM NaH2PO4, 185 mM KCl, 1 mM Na2EDTA, pH 7.2	6 mM Na2HPO4, 2 mM NaH2PO4, 185 mM KCl, 1 mM Na2EDTA, pH 7.2

(b) SAXS data-collection parameters.

Instrument/data processing	BioCAT facility at the Advanced Photon Source beamline 18ID with Pilatus3 1M (Dectris) detector
Wavelength (Å)	1.033
Beam size (µm)	150 (h) x 25 (v)
Camera length (m)	3.5
q measurement range ($Å^{-1}$)	0.004-0.4
Absolute scaling method	N/A
Normalization	To incident intensity, by ion chamber counter
Monitoring for radiation damage	Automated frame-by-frame comparison of relevant regions
Exposure time, number of exposures	0.5 s exposure time with a 2s total exposure period (0.5 s on, 1.5 s off) of entire SEC elution
Sample configuration	SEC-SAXS. Size separation by an AKTA Pure with a Superdex 75 Increase 10/300 GL column. SAXS data measured in a 1.5 mm ID quartz capillary
Sample temperature (°C)	20

(c) Software employed for SAXS data reduction, analysis, and interpretation.

SAXS data reduction	Radial averaging; frame comparison, averaging, and subtraction done using BioXTAS RAW 1.6.3 (Hopkins et al. 2017 (172))
Extinction coefficient estimate	Nearest neighbor approximation
Basic analyses: Guinier, P(r), V_p	Guinier fit, Kratky analysis, and molecular weight using BioXTAS RAW 1.6.3, P(r) function using PRIMUSqt (ATSAS v2.8.4 (173))

Shape/bead modelling	DAMMIF (Franke & Svergun, 2009) via ATSAS online (https://www.embl-hamburg.de/biosaxs/atsas-online/)
Atomic structure modelling	CRYSOL from PRIMUSqt in ATSAS v2.8.4(Svergun et al., 1995 (175))
Three-dimensional graphic model representations	UCSF Chimera v1.11

(d) Structural parameters.

Guinier analysis	2JSL	Tel48	Tel72	Tel96
	0.00957 ±	0.0318 ±	0.003742 ±	0.0202 ±
I(0) (cm^{-1})	0.00002	0.00002	0.000005	0.0000406
R_g (Å)	12.41 ± 0.05	19.23 ± 0.03	25.37 ± 0.06	31.68 ± 0.13
qmin ($Å^{-1}$)	0.014	0.009	0.007	0.007
qR_g max	1.27	1.33	1.32	1.17
Coefficient of correlation, R^2	0.972	0.998	0.996	0.996
M from volume of correlation, V_c (ratio to predicted)	6600 (0.84)	16100 (1.06)	23500 (1.03)	30900 (1.01)
P(r) analysis (GNOM)				
	0.00954 ±		0.00376 ±	0.0203 ±
I(0) (cm^{-1})	0.00002	0.032 ± 0.00002	0.000005	0.00004
Rg (Å)	12.33 ± 0.03	19.69 ± 0.03	26.01 ± 0.06	32.65 ± 0.10
Dmax (Å)	38	65	87	109
χ^2	0.95	1.76	0.84	1.1
Porod volume ($Å^3$) (ratio V_p/calculated M)	9040 (1.15)	16100 (1.06)	26300 (1.15)	32700 (1.07)

(e) Shape model-fitting results

	2JSL	Tel48	Tel72	Tel96
Ambimeter (default parameters)				
Number of compatible shape categories, ambiguity score	19, 1.279	634, 2.802	712, 2.852	644, 2.809
3D reconstruction	potentially unique	highly ambiguous	highly ambiguous	highly ambiguous
DAMMIF (default parameters, 20 calculations)				
q range for fitting ($Å^{-1}$)	-	-	0.0088-0.3146	0.0054-0.2539
Symmetry, anisotropy assumptions	-	-	P1, prolate	P1, prolate
NSD (standard deviation), No. of clusters	-	-	1.204 (0.082), 4	1.126 (0.095), 11
χ^2	-	-	1.166	1.167
Resolution (from SASRES) (Å)			31 ± 3	38 ± 3
DAMMIN (default, slow)				
q range for fitting ($Å^{-1}$)	0.014-0.3488	0.0115-0.3488	-	-

Symmetry, anisotropy assumptions	P1, none	P1, prolate	-	-
χ^2, CORMAP P-values	1.508, 0.1322	1.713, 0.249	-	-
Constant adjustment to intensities	8.46E-05	0.00E+00	-	-

(f) Atomistic modelling.

Crystal structures/atomic coordinate files	PDB ID: 2JSL	Modeled	Modeled	Modeled
q range for modelling	0.01-0.3	0.006-0.3	0.007-0.3	0.0054-0.2500
EOM GAJOE 2.1 (min ensembles = 1, max = 20, default parameters)				
χ^2	-	1.81	1.09	1.154
Rflex (random) / Rsigma	-	79.26 (88.83) / 0.62	79.96 (85.17) / 0.97	79.80 (86.06) / 1.39
Constant subtraction	-	0	0	0
No. of representative structures	-	6	4	4
Final ensemble Rg (Å), Dmax (Å)	-	19.58, 65.62	25.78, 82.65	32.11, 103.18
CRYSOL (single model, default parameters)				
χ^2	1.20	1.82	1.81	2.08
Predicted Rg (Å)	12.3	19.82	25.74	32.63
Dro (optimal hydration shell contrast), Ra (optimal atomic group radius (Å)	0.060, 1.760	0.065, 1.800	0.045, 1.400	0.075,1.400

(g) SASBDB IDs for data and models.

ID	SASDKF3	SASDKG3	SASDKH3	SASDKJ3

Molecular dynamics simulations and hydrodynamic calculations

Molecular dynamics simulations were carried out on Tel48, Tel72, and Tel96 constructs created previously (141), or modeled based on their solution NMR structures from the Protein Data Bank using the following IDs: 2GKU (hybrid-1), 2JSL (hybrid-2). Base modifications and optimization of starting configurations were performed in UCSF Chimera v1.12 (176) or Maestro v11.8 (177). The partial negative charges of carbonyls at the center of tetrads were neutralized with coordinated potassium counter-ions added manually in Maestro with subsequent minimization prior to simulation. The PDB structures created were then imported into the xleap module of AMBER 2018 (178), neutralized with K^+ ions, and solvated in a rectangular box of TIP3P water molecules with a 12 Å buffer distance. All simulations were equilibrated using sander at 300 K and 1 atm using the following steps: (1) minimization of water and ions with weak restraints of 10.0 kcal/mol/Å on all nucleic acid residues (2000 cycles of minimization, 500 steepest decent before switching to conjugate gradient) and 10.0 Å cutoff, (2) heating from 0 K to 100 K over 20 ps with 50 kcal/mol/Å restraints on all nucleic acid residues, (3) minimization of the entire system without restraints (2500 cycles, 1000 steepest decent before switching to conjugate gradient) with 10 Å cutoff, (4) heating from 100 K to 300 K over 20 ps with weak restraints of 10.0 kcal/mol/Å on all nucleic acid residues, and (5) equilibration at 1 atm for 100 ps with weak restraints of 10.0 kcal/mol/Å on nucleic acids. The resulting coordinate files from equilibration were then used as input for 100 ns of unrestrained, solvated MD simulations using pmemd with GPU acceleration in the isothermal isobaric ensemble (P = 1 atm, T = 300 K) with DNA OL15 and TIP3P water force fields. Periodic boundary conditions and PME were used. 2.0 fs time steps were used with bonds involving hydrogen frozen using SHAKE (ntc = 2). For the Tel48 constructs, an additional 100 ns of accelerated MD (aMD) simulation were carried out using the average torsional and potential energies from the end of the standard 100 ns simulations as input for calculating the “boosting” of both whole potential and torsional terms (iamd=3). Trajectories were analyzed using the CPPTRAJ module in the AmberTools18 package. Hydrodynamic properties were calculated as average and standard deviation of equally spaced trajectory snapshots (i.e. every 100 ps) using the program HYDROPRO10 (179) with the recommended parameters for G-quadruplexes (180). Clustering of

the trajectories was performed using the DBSCAN method in the CPPTRAJ module of Amber (minpoints = 5, epsilon = 1.7, sieve 10, rms residues 1-48 over atoms P, O3', and O5'). Electrostatic calculations for visualization were performed using PDB2PQR software on the APBS web server (http://server.poissonboltzmann.org/) (181,182) with AMBER force field and pH set to 7.2. All molecular visualizations were performed in UCSF Chimera v1.12 (176).

Ensemble optimization method (EOM)

Telomere ensembles were derived using the Ensemble Optimization Method 2.1 (183) program from the ATSAS suite of tools. For the Tel48 constructs, which included the four combinations of hybrid-1 and hybrid-2 topologies (i.e. hybrid-11, hybrid-12, hybrid-21, and hybrid-22), a total of 2,000 PDB snapshots were derived from the 100 ns of MD and aMD trajectories stripped of water and K^+ and pooled, totaling 8,000 coordinate files. GAJOE was used in pool "-p" mode, with maximum curves per ensemble set to 30, minimum curves per ensemble set to 1, constant subtraction allowed, curve repetition allowed, and the genetic algorithm (GA) repeated 200 times. Where noted, the minimum curves were increased to higher numbers, and the curve repetition was disallowed. The same process was repeated for the Tel72 (hybrid-122, -121, -212, and -221) and Tel96 (hybrid-1222, -2122, -2212, -2221) constructs with a total of 4,000 pooled structures. In brief, EOM takes a large pool of macromolecules covering as much conformational space as possible (and reasonable) and selects from this pool a sub-ensemble of conformers that best recapitulate the experimental scattering. The best fitting ensemble is the subset of weighted theoretical curves from conformations that minimizes the discrepancy χ^2:

$$\chi^2 = \frac{1}{K-1}\sum_{j=1}^{K}\left[\frac{\mu I(s_j) - I_{exp}(s_j)}{\sigma(s_j)}\right]^2$$

where $I_{exp}(s_j)$ is the experimental scattering, $I(s_j)$ is the calculated scattering, K is the number of experimental points, $\sigma(s_j)$ are standard deviations, and μ is a scaling factor (184).

***Ab initio* model generation and single model validation**

P(r) distributions obtained from GNOM (185) using scattering data from 2JSL, Tel48, Tel72, and Tel96 were submitted using the ATSAS online servers (https://www.embl-hamburg.de/biosaxs/atsas-online/) for either *DAMMIN* or *DAMMIF* bead model generation. Relevant parameters, anisotropy assumptions, normalized spatial discrepancy values (NSDs), χ^2 values, and resolutions are given in **Table 2**. Single best fit models for each telomere construct were determined using the initial pool of conformers derived from MD simulations (or NMR structures for 2JSL) and calculated using CRYSOL (186). The best fit structure was determined by minimization of a χ^2 function:

$$\chi^2(r_o, \delta_\rho) = \frac{1}{N_p}\sum_{i=1}^{N_p}\left(\frac{I_{exp}(q_i) - cI(q_i, r_o, \delta_\rho)}{\sigma(q_i)}\right)^2$$

where $I_{exp}(q_i)$ and $I(q_i)$ are the experimental and computed profiles, respectively, $\sigma(q_i)$ is the experimental error of the measured profile, N_p is the number of points in the profile, and c is the scaling factor. Two other parameters, r_o and δ_ρ, are fitted and represent the effective atomic radius and the hydration layer density, respectively.

Flexibility analyses by swollen Gaussian chain and WLC models

Fitting of the radii of gyration, as measured by SEC-SAXS for 2JSL, Tel48, Tel72, and Tel96 was performed as outlined recently by Capp at al. (187) using the following relationship describing the stiffness and conformational space of a swollen Gaussian coil:

$$R_g = l_p\sqrt{\frac{N^{2\nu}}{(2\nu+1)(2\nu+2)}}$$

Where l_p is the persistence length and ν is the Flory coefficient. The R_g values with their respective errors were plotted against their G4 number and fit using a non-linear least squares fitting procedure in Origin 2020 (OriginLab Corporation, Northampton, MA, USA).

For the worm-like chain (WLC) modeling, a global analysis of three properties was used in the program Multi-HYDFIT (188,189). Measurements of two other properties, diffusion coefficient, D_t (calculated from measured Stokes radii, R_s, via the Stokes-Einstein equation), and corrected

sedimentation coefficient, $S_{20,w}$, were obtained for each sequence using SEC (average ± S.D. of 4 measurements each) or AUC (concentration series extrapolated to infinite dilution ± standard error from regression analysis), respectively. Each value, with respective weighting, and molecular weight (MW) was used as the input for the Multi-HYDFIT program. Multi-HYDFIT uses comparisons of the so-called equivalent radii and ratios of radii to calculate theoretical values of R_g, D_t, and $S_{20,w}$, which are then compared to that of the measured values. The ratios of radii are directly related to the ratios of length to diameter (L/d) and length to persistence length (L/l_p). With starting estimates of l_p, d, and mass per unit length (M_L), the Multi-HYDFIT procedure seeks to minimize a target function (189):

$$\Delta^2(l_p, M_L, d) = \frac{1}{N_s}\sum_{i=1}^{N_s}\left[\left(\sum_Y w_Y\right)^{-1}\sum_Y w_Y\left(\frac{a_{Y(cal)} - a_{Y(exp)}}{a_{Y(exp)}}\right)^2\right]$$

where N_s is the number of samples of different MW, w_Y is the weighting, and a_Y is the ratio of radii for each property. In this equation, the outermost sum runs over the N_s samples and the inner most sum runs over the available properties of each sample. The Δ^2 is a mean-square relative deviation for the data, and 100Δ is the percent difference between experimental and theoretical values over the entire set. Additional information is required for the calculation, such as temperature (here 20.0°C was used), solvent viscosity (0.00995 poise), starting guesses for diameter, d (10 to 100 angstrom), mass per unit length, M_L (10 to 300 Da/angstrom), and persistence length, l_p (20 to 100 angstrom). Intrinsic viscosities calculated from best-fit models using HYDROPRO10 were also included with modest weighting, as they were not empirically determined but rather derived from SAXS best-fit models. The goal of the procedure is to determine the best-fit values of the latter properties, which are given in **Table 3**.

Table 3. Table of properties derived from Multi-HYDFIT fitting of the higher-order telomere experimental properties to a worm-like chain model.

HYDFIT Worm-like Chain results	
Diameter (Å)	40 (±5)
Persistence length (Å)	33 (±3)
Mass per unit length (Da/Å)	163 (±15)
Deviation from exp. equiv. radii (%)	3.7

Force of bending curves were calculated using the literature persistence length values for single- and double-stranded DNA at cationic conditions similar to used here, based on the relationship (190):

$$F_{bend} = \frac{1}{2} k_B T L_p R^{-2}$$

where F_{bend} is the bending force in piconewtons, $k_B$is the Boltzmann constant, T is temperature in Kelvin, L_p is the persistence length in meters, and R is the radius of the arc of a curve. The data was plotted such that the values on the X-axis correspond to the end-to-end length of the polymer curved 180° around the arc of a semi-circle.

Molecular visualizations

All molecular visualizations of MD trajectories and models and RMSD calculations were performed in UCSF Chimera v1.11 (176).

Results

Small-angle X-ray scattering reveals G4 maximization and indicates that the higher-order telomere G4s are semi-flexible

To verify that the extended telomere sequences are in fact maximizing their G-tract usage we employed size-resolved small-angle X-ray scattering (SEC-SAXS) to assess each sequence for size, shape, and compactness (191,192). The results of the SEC-SAXS analysis for sequences 2JSL (hybrid-2), Tel48, Tel72, and Tel96 (**Table 1**) are shown in **Figures 8**, **9**, and **Table 2**. **Figure 8A** shows the scattering intensity as a function of momentum transfer (q) on a log-log scale for each sequence. Each scattering profile proceeds horizontally to the Y-axis at low values of q, indicating the absence of inter-particle interactions or repulsions (191). Scattering from 2JSL shows a distinct smooth curvature at higher q values which is indicative of a globular particle, whereas the extended telomere sequences deviate from this curvature between about 0.05 and 0.2 q, suggesting a non-globular structure (192).

Figure 8. SEC-SAXS analysis of 2JSL (gray), Tel48 (red), Tel72 (blue), and Tel96 (green). (A) Log-log plot of the scattering intensity vs. scattering vector, q. (B) Dimensionless Kratky plots of data in A. (C) Pair distribution function plots of data in A normalized to I(0). (D) Scatter plot of the radii of gyration from each sequence as a function of G-quadruplex motif fit to a swollen Gaussian chain polymer model (see methods) with (inset) derived persistence length (L_p) and Flory coefficient (ν). (E) DAMMIN and DAMMIF *ab initio* space-filling models from the data in C.

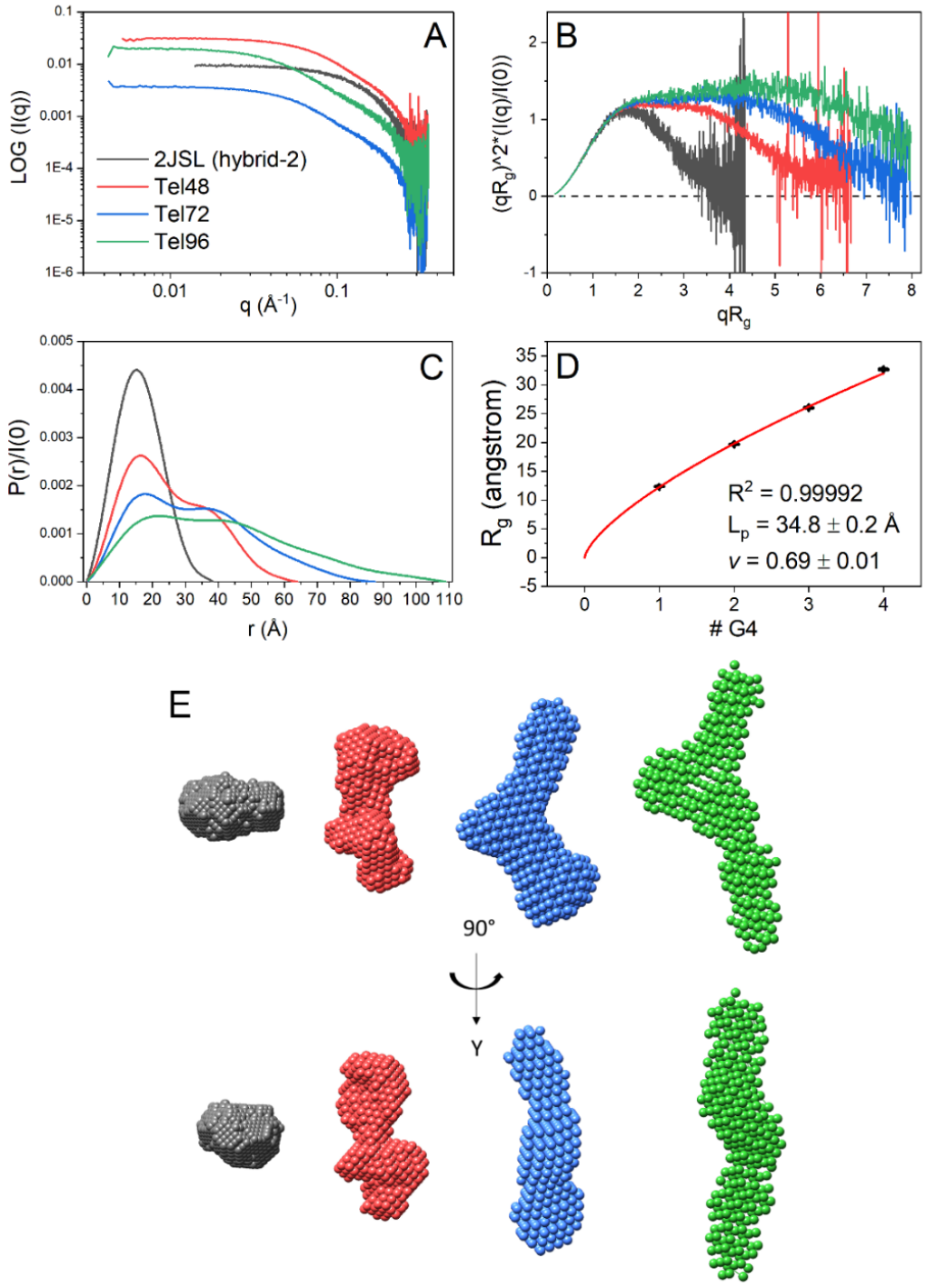
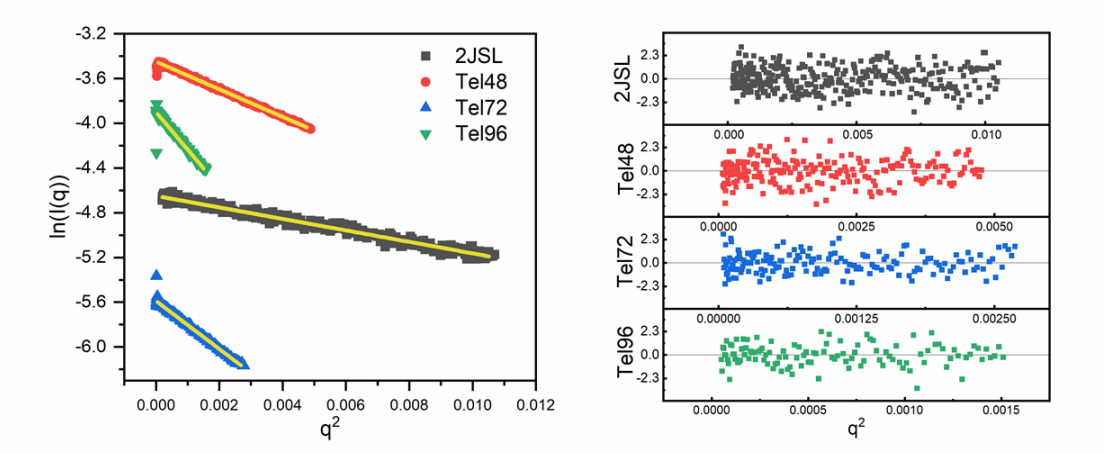

A
LOG (I(q))
q (Å^{-1})
2JSL (hybrid-2)
Tel48
Tel72
Tel96
B
(qR$_g$)^2*(I(q)/I(0))
qR$_g$
C
P(r)/I(0)
r (Å)
D
R$_g$ (angstrom)
G4
R^2 = 0.99992
L$_p$ = 34.8 ± 0.2 Å
ν = 0.69 ± 0.01
E
90°
Y

Figure 9. Guinier analyses of 2JSL, Tel48, Tel72, and Tel96 (left) with fit overlaid in yellow for each sequence and (right) residuals of fits. Guinier fit results are tabulated in **Table 2**.


2JSL
Tel48
Tel72
Tel96
ln(I(q))
q²
-3.2
-3.6
-4.0
-4.4
-4.8
-5.2
-5.6
-6.0
0.000
0.002
0.004
0.006
0.008
0.010
0.012
2JSL
Tel48
Tel72
Tel96
2.3
0.0
-2.3
0.000
0.005
0.010
0.0000
0.0025
0.0050
0.00000
0.00125
0.00250
0.0000
0.0005
0.0010
0.0015
q²

Two useful transformations of the scattering data are the Kratky plot and distance distribution, P(r), plots (**Figures 8B and 8C**), which allow for qualitative appraisal of compactness and overall structure, respectively (192). In **Figure 8B** the dimensionless Kratky plot shows that 2JSL (gray) exhibits a nearly perfect Gaussian distribution that returns to baseline at high qR_g, confirming that it is globular and folded (192). The Tel48 and Tel72 sequences also approach baseline at high qR_g, indicating that they are folded and do not contain significant amounts of flexibility (192). The higher-order sequences also exhibit distinct plateau regions above 2 qR_g, indicating that they have non-globular shapes and are likely multi-domain, consistent with tandem G4 domains. However, Tel96 (green) exhibits a slight rise in its plateau towards higher qR_g, indicating that it is flexible relative to 2JSL, Tel48, and Tel72. **Figure 8C** shows the corresponding P(r) distributions (normalized to scattering intensity, I[0]), which are probability distributions of the inner-atomic distances within each macromolecule (191). 2JSL (gray) again exhibits a symmetric distribution, indicative of a globular molecule (192). Conversely, the extended sequences are all multi-modal. Tel48 exhibits a biphasic distribution (red) indicating a characteristic dumbbell-like tertiary arrangement (192), consistent with two G4 domains separated by a small linker region. Tel72 and Tel96 have tri- and tetra-phasic curves, respectively, which we take as indicating three and four contiguous globular domains in tandem, respectively.

P(r) distributions also allow for quantitative characterization of macromolecules. The point on the X-axis at which each sequence converges to zero is the maximum diameter, D_{max}, which is the diameter of the particle's longest axis (192). The D_{max} of each sequence increases approximately linearly with a ~24 Å increase with each additional G4 motif. Any substantial amount of telomere species with gaps, or non-maximization of G4s, would likely result in a non-linearity (as well as large upticks in the Kratky curves at high qR_g). The radius of gyration, R_g, is the root mean square distance of the macromolecule's parts from its center of mass and reflects the particle's size (191). The R_g can be calculated by either the Guinier approximation (from plots shown in **Figure 9**) or directly from its P(r) distribution, the latter of which is thought to be more representative in cases where flexibility is assumed (although both values should be in general agreement)(192). D_{max} and R_g values for each sequence are reported in **Table 2**. Shown in **Figure 8D** is a plot of each

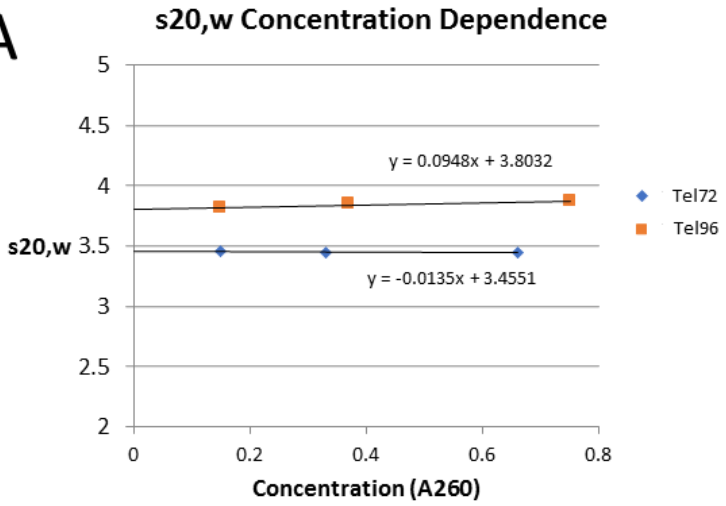
sequence's radii of gyration plotted against G4 number. Each additional G4 motif leads to an approximate R_g increase of 6.7 Å.

The extended tails of the P(r) distributions and upward trend in plateau regions of the Kratky plots (**Figure 8B**) signify flexibility. Although plotting the radii of gyration *versus* the putative number of G4 subunits appears entirely linear (R^2 = 0.9985), a better fit is obtained when fitting to a swollen Gaussian polymer model (R^2 = 0.9999, **Figure 8D**). The non-linear least-squares fit to this model allows for the estimation of two parameters: persistence length, L_p, and the Flory exponent, v. The persistence length represents the distance along the telomere G4 polymer which behaves as a rigid rod. At lengths much greater than this, the polymer behaves as a flexible Gaussian chain. The Flory coefficient (also known as the excluded volume parameter) varies between 0.5 and 1.0 and describes the degree of flexibility of the system. A Flory coefficient for a theoretical freely jointed flexible chain is 0.5 (maximum flexibility), whereas that of a rigid rod is 1.0. For reference, the empirical value of chemically denatured proteins is v = ~0.588 (193). Fitting to this model we find that the telomere G4 has a persistence length of 34.8 ± 0.2 Å and Flory coefficient of 0.69 ± 0.01. The persistence length is approximately the size of a single telomere G4 (~32 Å, calculated from PDB 2JSL less the flanking nucleotides), which indicates that the TTA linkers may provide a point of flexibility. This persistence length is about 50% greater than ssDNA (L_p = ~22 Å under similar ionic conditions (194)). As an independent method of estimating the persistence length, we used the hydrodynamic modeling program Multi-HYDFIT (188). This program integrates multiple independently measured properties, such as sedimentation coefficients ($S_{20,w}$) from AUC (**Figure 10**) (163), translational diffusion coefficients (D_t) from SEC, and radii of gyration (R_g) from SEC-SAXS, for a series of macromolecules of given molecular weight (*MW*), and uses these values to find the optimum values of the model parameters for a worm-like chain (WLC) model (189). In total, we fit 12 independent properties from three independent techniques with their respective weights (estimated from standard deviations of multiple measurements), yielding a persistence length of 33 ± 3 Å (**Table 3**), in excellent agreement with the L_p estimated from the swollen Gaussian chain model. Altogether, these results, along with the qualitative information from Kratky and P(r) distributions, suggest that the extended telomere maximizes G4 formation, is closely packed, and

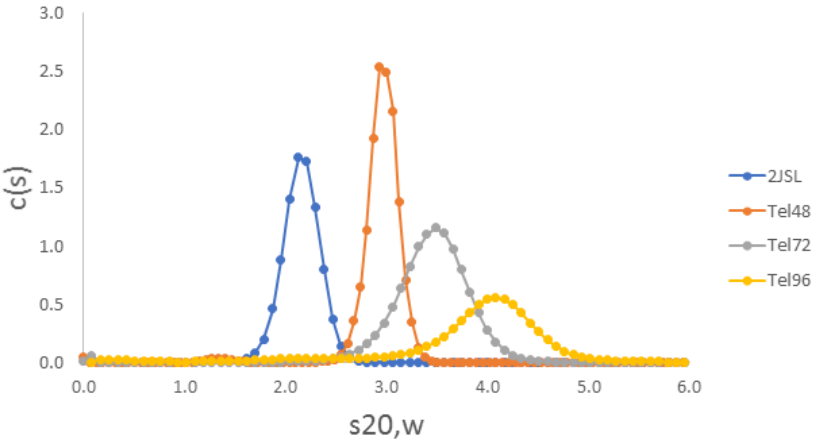
is moderately flexible. The flexibility is consistent with rigid G4 units linked by flexible, hinged, interfaces.

Figure 10. Sedimentation velocity analysis of higher-order telomere sequences. (Top) Analysis of Tel72 and Tel96 concentration dependence on sedimentation. Extrapolation to the Y-axis gives the infinite dilution $S_{20,w}$ values. (Bottom) representative C(s) vs. $S_{20,w}$ distributions for 2JSL, Tel48, Tel72, and Tel96.

A


s20,w Concentration Dependence
5
4.5
4
3.5
3
2.5
2
s20,w
y = 0.0948x + 3.8032
y = -0.0135x + 3.4551
Tel72
Tel96
0
0.2
0.4
0.6
0.8
Concentration (A260)

B


3.0
2.5
2.0
1.5
1.0
0.5
0.0
c(s)
0.0
1.0
2.0
3.0
4.0
5.0
6.0
s20,w
2JSL
Tel48
Tel72
Tel96

***Ab initio* and atomistic modeling reveals an ensemble of conformations ranging from entirely stacked and condensed to a coiled "beads-on-a-string" configuration**

The above analyses suggest a flexible system which would render the higher-order SAXS data unsuitable for use in *ab initio* bead reconstruction methods. However, upon seeing the resulting space-filling models we were compelled to include them. **Figure 8E** shows the resulting DAMMIN and DAMMIF space-filling models of 2JSL, Tel48, Tel72, and Tel96 created based on the P(r) data in **Figure 8C** (with corresponding fit results tabulated in **Table 2**). Consistent with predictions from the Kratky and P(r) distribution plots Tel48 looks like a dumbbell with two domains roughly the size of the 2JSL reconstruction with a small linker region in the middle. Similarly, Tel72 and Tel96 have what appear to be three and four G-quadruplex domains (indicated by their distinct "bends"), although their resolution is not quite as high as the Tel48 reconstruction (**Table 2**). The similar overall shape and curvature coupled with the flexibility assessment above indicates a non-rod-like structure for telomere sequences with more than two G4 motif repeats. These shapes are generally in accord with previous hydrodynamic investigations based on rigid structures (163), but offer a more detailed and nuanced characterization because the flexibility of the structures can be taken into account.

We next employed an ensemble modeling approach that combined explicit solvent MD-derived models with the ensemble optimization tool GAJOE (of the EOM 2.0 suite) (184). A CD analysis that will follow indicated that the telomere sequences are best represented by a combination of hybrid-1 and hybrid-2 topologies. However, the order in which they occur is not evident, and it may be that the extended sequences are dynamic and interconvert on timescales much longer than is accessible by standard MD simulations (>1 ms). Therefore, we modeled every combination of the simplest multimer system, Tel48. Using the PDB atomic structures for hybrid-1 (PDB ID: 2GKU) and hybrid-2 (PDB ID: 2JSL) we generated each of the four possible combinations: hybrid-11, hybrid-12, hybrid-21, and hybrid-22. Each structure was subjected to 100 ns of both standard MD and accelerated MD (aMD) simulations to produce a pool of 8,000 conformations for use in minimal ensemble and single structure modeling efforts. In the GAJOE ensemble

optimization method, a pool of PDB atomic coordinate files are generated that cover as much conformational space as possible and utilized in calculating theoretical scattering profiles. Next, a genetic algorithm acts on these scattering profiles to minimize a fitness function by weighting each scattering profile and comparing combined profiles to the experimental (see methods). The output is an ensemble of conformers which best recapitulate the experimental scattering profile based the minimized χ^2 value. An ensemble is considered a better fit than a single conformer when its χ^2 value is reduced relative to the single best-fit conformation.

Figures 11 and 12 shows the results of modeling efforts with the Tel48 constructs. **Figures 11A and 11B** are scatter plots which show the calculated radii of gyration (Y-axis) and corrected sedimentation coefficients (X-axis) (**Figure 10**), for each of 2,000 frames across both MD (light gray) and aMD (dark gray) trajectories for the hybrid-12, -21, -11, and -22 constructs. These plots indicate that both hybrid-12 and -21 sample conformations which agree with either the experimental R_g, $S_{20,w}$, or intersect both values. The hybrid-11 and hybrid-22 constructs rarely sampled conformations that corresponded with the experimental values (see **Figure 12**). Interestingly, although hybrid-12 extensively samples conformations which agree with both hydrodynamic and scattering-derived measurements, the best fit model by CRYSOL analysis was found to be a highly extended hybrid-21 conformation (cyan dot and curve in **Figures 11B & 11C**). Because this conformation appeared unnatural (e.g. maximally extended) and did not agree very well with the P(r)-derived R_g and D_{max} values, we speculated that this configuration may be biased simply by our initial start configurations. The hybrid-21 clearly tended towards an overall more compact structure as indicated by the histograms. Therefore, we next asked what the maximum number of curves could be which could reconstruct the experimental scattering without worsening the χ^2 value. We found that an ensemble of six conformations gave approximately the same χ^2 value (magenta dots in **Figures 11A** and **11B**, magenta curve **Figure 11C**) and agreed much better with the experimental R_g and D_{max} values from the P(r) analysis ($R_{g,cal}$ = 19.58 Å *vs*. $R_{g,exp}$ = 19.69 Å and $D_{max,calc}$ = 66 Å *vs*. $D_{max,exp}$ = 65 Å, **Table 2**). The resulting topologies were a 50/50 mix of hybrid-12 and -21 (**Figure 11D**), which sampled conformations ranging from extended to fully stacked. The

flexibility of the ensemble was only marginally lower than the pool, as judged by EOM's Rflex flexibility analysis, supporting semi-flexibility. Interestingly, we found that one of the hybrid-12 conformers (bottom right of **Figure 11D**) was nearly identical in conformation to our previously reported hybrid-12 model (141), with an RMSD of just 1.6 Å over all residue pairs (**Figure 13**).

Figure 11. Results of Tel48 SAXS atomistic modeling efforts. (A-B) scatter plots of calculated radii of gyration and sedimentation coefficients for hybrid-12 (A) and hybrid-21 (B) with MD-derived values shown in light gray and aMD-derived values in dark gray. The inset dashed red and blue lines represent the experimentally measured values for sedimentation coefficient and radius of gyration, respectively. The outer histograms represent the distributions of values from both MD and aMD snapshots combined. The cyan dot represents the single best-fit model (hybrid-21) as determined by CRYSOL (top left model in D). Magenta dots represent the six conformers in the best fit ensemble (all six models in D). (C) Experimental SAXS scattering data with fits from single (cyan) or ensemble (magenta) calculated scattering overlaid with χ^2 values inset. (D) Single best fit model (hybrid-21, top left model) and best fit ensemble of six conformers (top row hybrid-21, bottom row hybrid-12). Models are oriented with their 5' ends at the top.

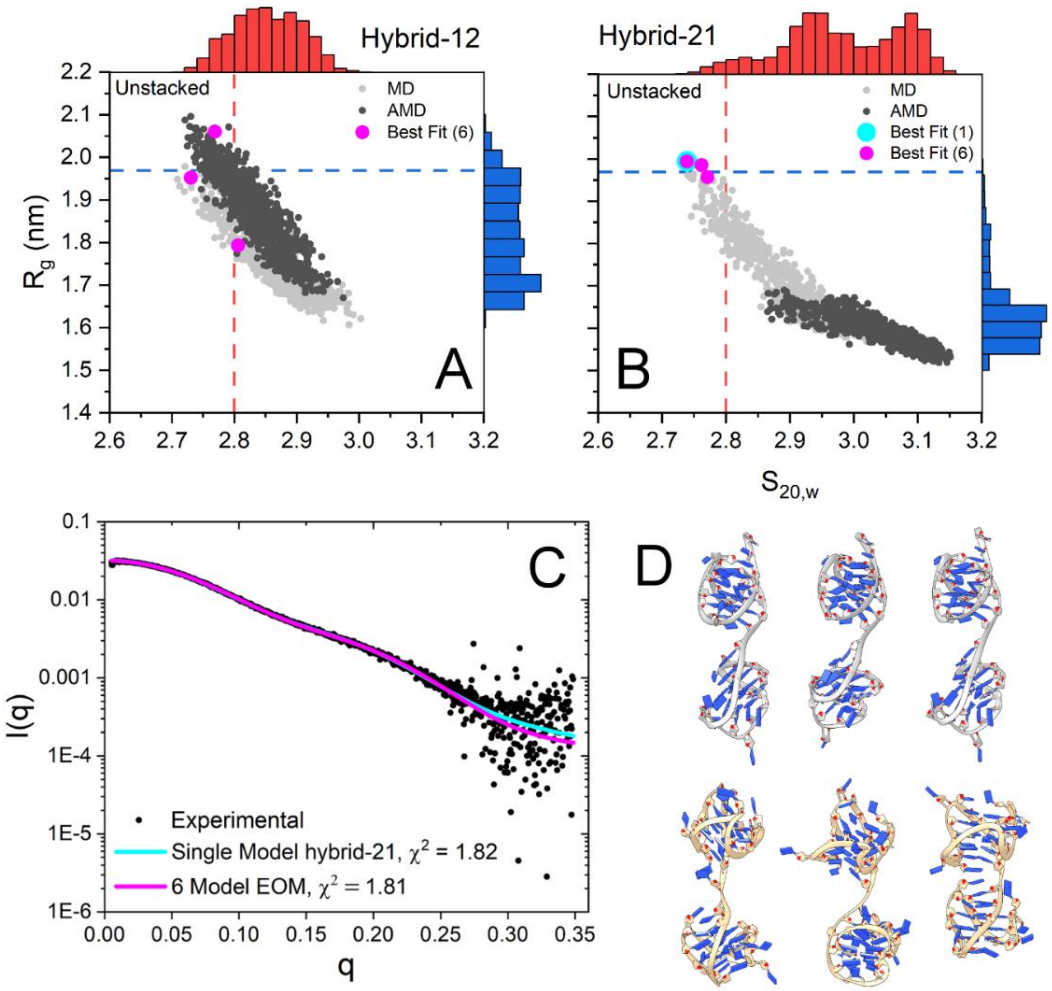

Hybrid-12
Hybrid-21
Unstacked
MD
AMD
Best Fit (6)
Best Fit (1)
A
B
R_g (nm)
$S_{20,w}$
C
D
I(q)
q
Experimental
Single Model hybrid-21, χ^2 = 1.82
6 Model EOM, χ^2 = 1.81

Figure 12. Additional results of Tel48 SAXS atomistic modeling efforts shown in **Figure 11**. (A-B) scatter plots of calculated radii of gyration and sedimentation coefficients for hybrid-11 (A) and hybrid-22 (B) with MD-derived values shown in light gray and aMD-derived values in dark gray. The inset dashed red and blue lines represent the experimentally measured values for sedimentation coefficient and radius of gyration, respectively. The outer histograms represent the distributions of values from both MD and aMD snapshots combined. The histograms indicate that the major sampled conformations in both cases are much more compact than would be expected from either SAXS or AUC analyses.

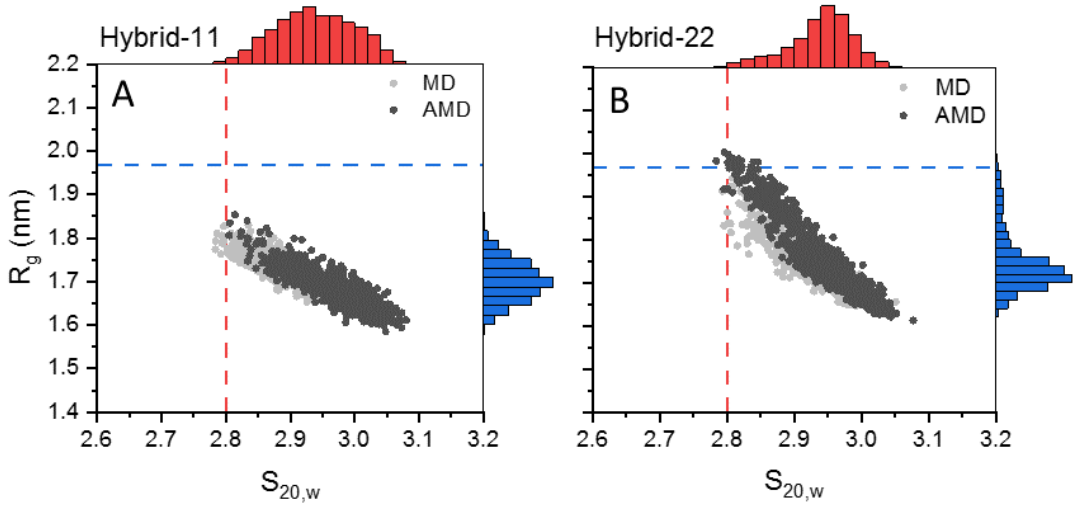
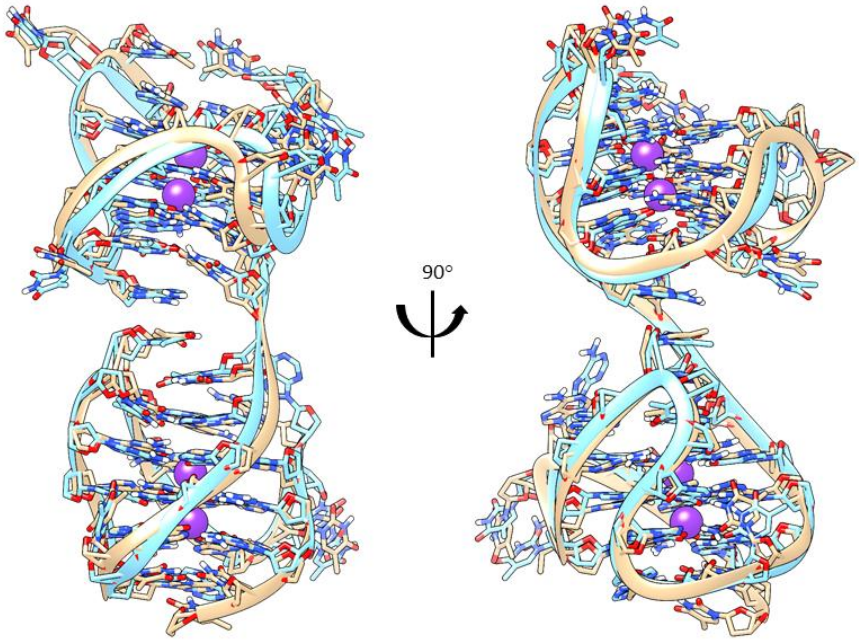

Hybrid-11
A
MD
AMD
R_g (nm)
2.2
2.1
2.0
1.9
1.8
1.7
1.6
1.5
1.4
2.6
2.7
2.8
2.9
3.0
3.1
3.2
$S_{20,w}$
Hybrid-22
B
MD
AMD
2.6
2.7
2.8
2.9
3.0
3.1
3.2
$S_{20,w}$

Figure 13. Comparison of the Tel48 (cyan) and Tel50 (tan) hybrid-12 conformers. The tan hybrid-12 conformer was taken from an earlier report by Petraccone et al.(141) and the cyan hybrid-12 is the model derived here from EOM (bottom right-most conformer in **Figure 11**). The pair-wise residue RSMD is 1.6 Å as determined by the matchmaker module of UCSF Chimera v1.12. Potassium is shown as purple spheres and is derived from the Petraccone model (EOM hybrid-12 potassium is hidden).


90°

Next, we investigated the Tel72 and Tel96 constructs in the same manner as above but only using standard MD. Models were created to reflect the ratio of hybrid-1/-2 (25/72) as determined by our later CD analyses of telomere mutants (e.g. hybrid-121, -122, -212, and -221 for Tel72 and hybrid-1222, -2122, -2212, and -2221 for Tel96). The hybrid-121 was included because it was proposed previously (141). Each model was then subjected to explicit solvent MD and simulated for a total of 100 ns. From these trajectories, 1,000 equally spaced frames were used as the pool for EOM's GAJOE program. For the Tel72 the best fit was obtained with an ensemble of four conformers ($\chi^2_{ensemble}$ = 1.09 *vs.* $\chi^2_{single\text{-}model}$ = 1.81) which was composed of the hybrid-212 (89.7%) and hybrid-221 (10.3%) (**Figure 14**). Surprisingly, this configuration agrees well with Tel48 having a 5' preference for hybrid-2 followed by hybrid-1. The R_g and D_{max} values of the Tel72 ensemble (**Figure 14C**) agree with the experimental values ($R_{g,calc}$ = 25.8 Å *vs.* $R_{g,exp}$ = 26.0 Å and $D_{max,calc}$ = 83 Å *vs.* $D_{max,exp}$ = 87 Å), indicating that the ensemble is an excellent solution. Similarly, Tel96 scattering was best recapitulated by an ensemble of four conformers ($\chi^2_{ensemble}$ = 1.15 *vs.* $\chi^2_{single\text{-}model}$ = 2.08) but was composed entirely of different conformations of the hybrid-2122 (**Figure 15**). Again, there is an agreement with a 5' hybrid-2 followed by hybrid-1. The R_g and D_{max} values of this conformer ensemble are also in agreement with the experimental values ($R_{g,calc}$ = 32.1 Å *vs.* $R_{g,exp}$ = 32.7 Å and $D_{max,calc}$ = 103 Å *vs.* $D_{max,exp}$ = 109 Å), indicating that this ensemble is reasonable. In both cases, the EOM Rflex quantification of flexibility indicates that the ensembles are only marginally less flexible than the pool (**Table 2**), consistent with the system semi-flexibility. This flexibility is also illustrated by the conformer ensembles themselves (**Figures 14C and 15**). Further, docking of each ensemble into their respective *ab initio* space-filling models from **Figure 8E** reveals excellent fits for the models of topological sequence 5'-hybrid-2,-1,-2,-2 (**Figure 16**). Collectively, these analyses indicate that in physiological buffer conditions the extended telomeres maximize their formation of G4 subunits, prioritize hybrid-2 at the 5' end, and are semi-flexible.

Figure 14. Results of Tel72 SAXS atomistic modeling efforts. (A) scatter plot of calculated radii of gyration and sedimentation coefficients for the hybrid-212 model from 100 ns of standard MD simulation. The inset dashed red and blue lines represent the experimentally measured values for sedimentation coefficient and radius of gyration, respectively. The outer histograms represent the distributions of values. The cyan dot represents the single best-fit model as determined by CRYSOL. Magenta dots represent the four conformers in the best fit ensemble. (B) Experimental SAXS scattering data with fits from single (cyan) or ensemble (magenta) calculated scattering overlaid with χ^2 values inset. (C) Conformations of the three hybrid-212 configurations (not showing the hybrid-221) from the best fit ensemble as determined by EOM. Models are oriented with their 5' ends at the top.

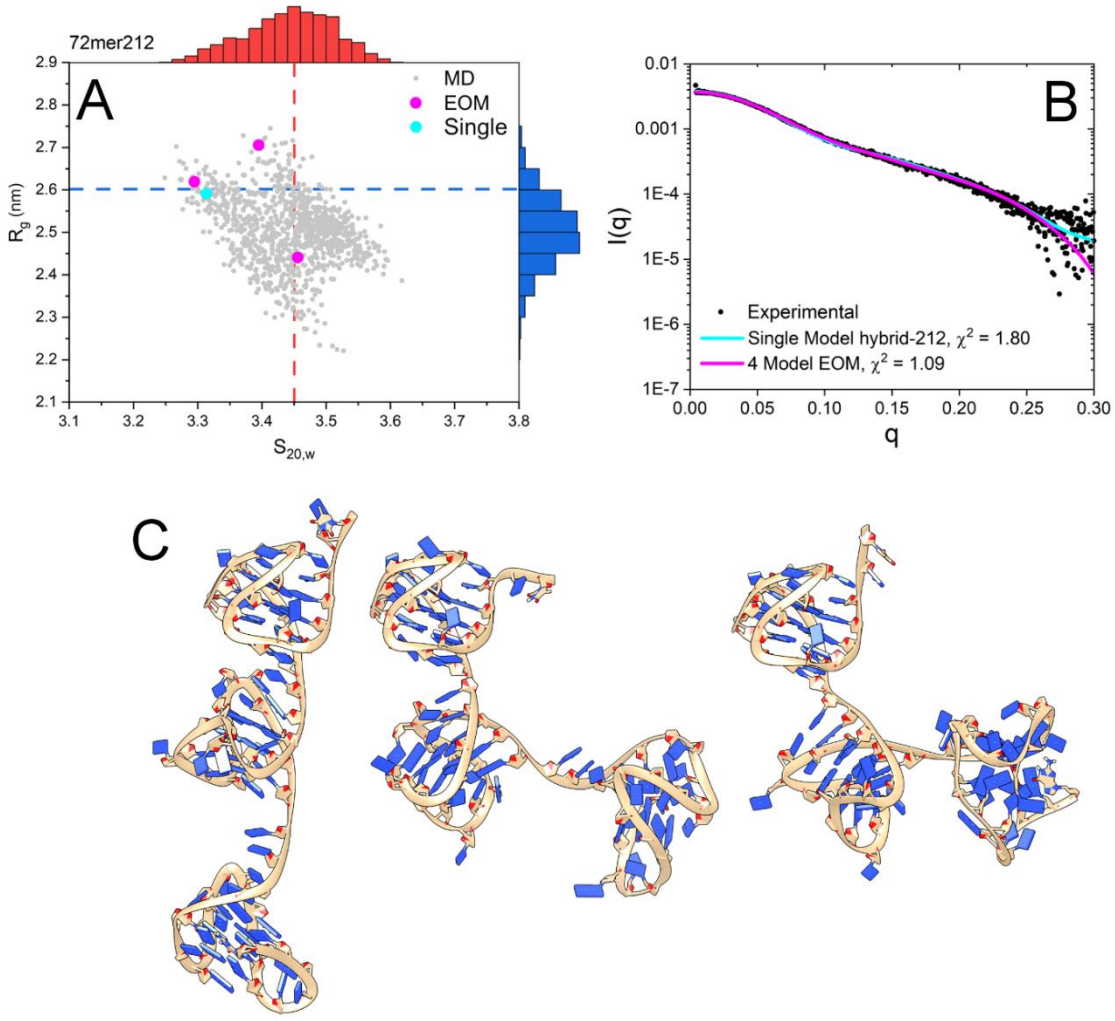

72mer212
A
MD
EOM
Single
R_g (nm)
$S_{20,w}$
B
I(q)
q
Experimental
Single Model hybrid-212, χ^2 = 1.80
4 Model EOM, χ^2 = 1.09
C

Figure 15. Results of Tel96 atomistic modeling efforts. Depicted are the four hybrid-2122 conformers derived from EOM analysis of the 100 ns MD simulations with their respective weights (as % of the reconstructed scattering curve). All conformers are arranged with their 5' ends at the top of the figure.

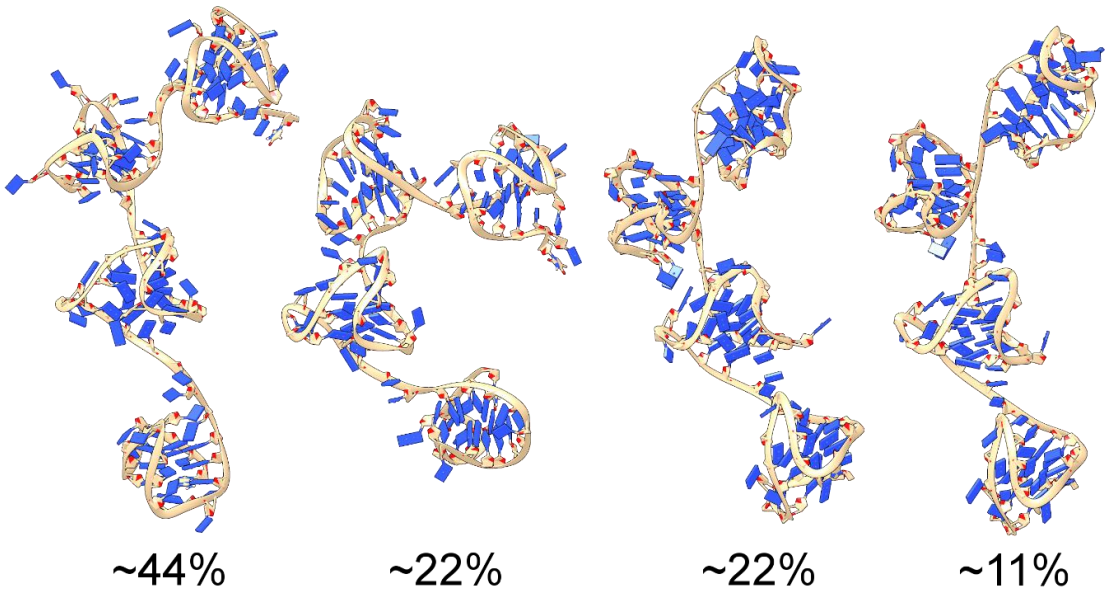
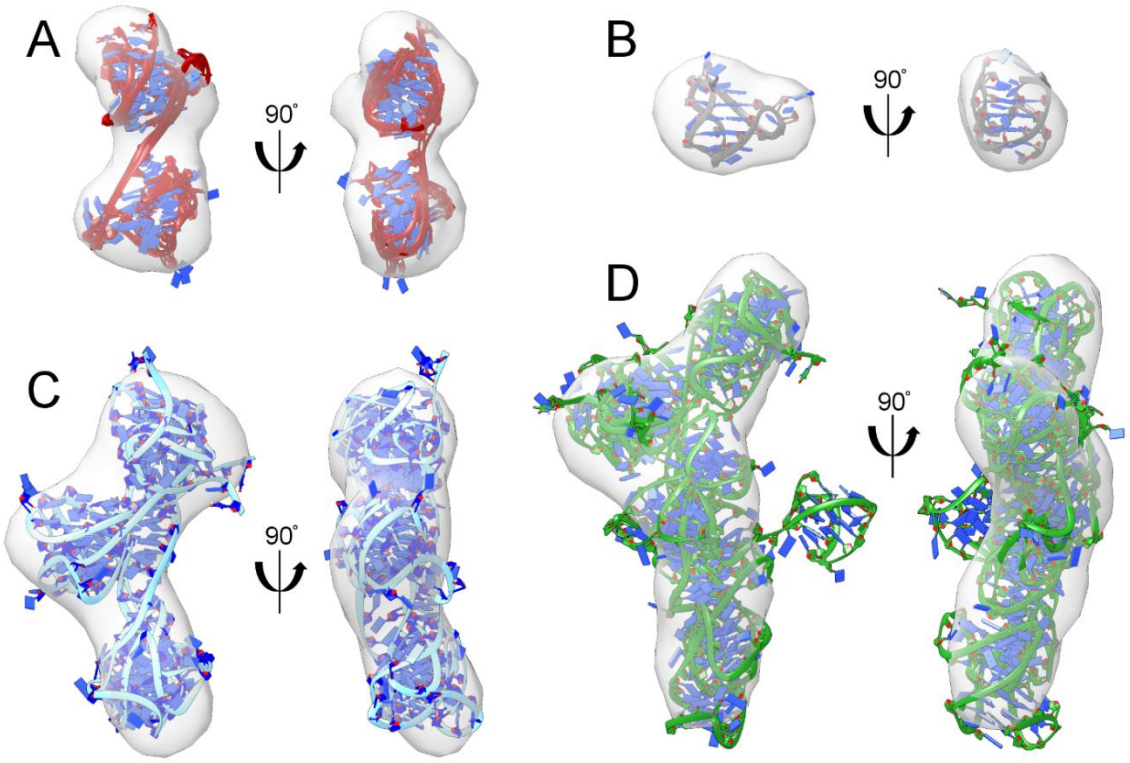

~44%
~22%
~22%
~11%

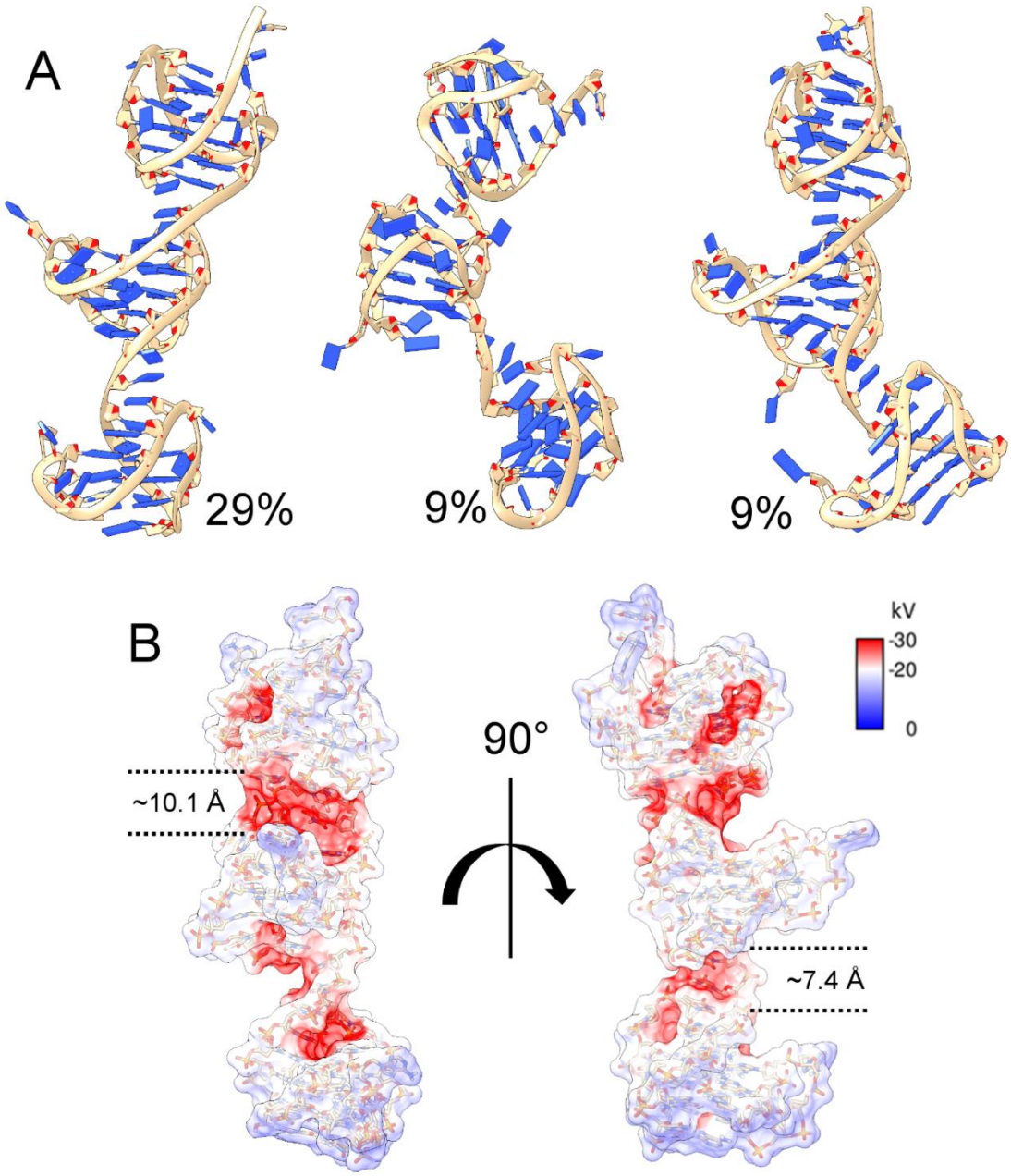
Figure 16. Telomere G4 ensembles from EOM GAJOE analysis docked into *ab initio* space-filling reconstructions from DAMMIN/DAMMIF. (A) Tel48 hybrid-21 conformers, (B) 2JSL with single best-fit NMR-derived model, (C) Tel72 hybrid-212 conformers (the same as in **Figure 14C**), (D) Tel96 hybrid-2122 models (the same as in **Figure 15**).


A
90˚
B
90˚
C
90˚
D
90˚

The Tel72 hybrid-212 preferentially samples a stacked conformation, forming unique electronegative binding pockets useful for drug targeting

The Tel72 ensembles reflect well the SAXS-derived properties, R_g and D_{max}. However, they do not agree as well with their measured sedimentation coefficients. SAXS scattering is exquisitely sensitive to changes in particle volume (conformation in this case). Systems which exist in an equilibrium of stacked and unstacked, or coiled and beads-on-a-string, will have a scattering profile which is composed of a continuous distribution of conformations (as the scattering intensity, I[0], is directly related to the volume of the scattering particle) (195). Therefore, we next endeavored to find the most frequently sampled conformation from the MD trajectory of the Tel72 hybrid-212 construct. **Figure 17A** shows the top three most frequently sampled conformations across the 100 ns simulations with their respective weighting (% of MD frames). This analysis suggests that approximately 47% of the frames from simulation sampled a configuration which was partially (middle) or entirely stacked (left and right models). The major stacked conformation sampled by hybrid-212 has a calculated sedimentation coefficient which is in excellent agreement with the experimental ($S_{20,w(calc)}$ = 3.45 *versus* $S_{20,w(exp)}$ = 3.46, **Figure 10**) although the calculated radius of gyration is slightly lower ($R_{g(calc)}$ = 2.45 nm *versus* $R_{g(exp)}$ = 2.60). Electrostatic calculations of the most prominent form reveal highly electronegative grooves, which are appropriately sized for small molecules (**Figure 17B**). Overall, these analyses show that the hybrid-212 model of the Tel72 is consistent with all available spectroscopic, hydrodynamic, X-ray scattering, and MD-based analyses, and forms unique junctional grooves for selective small molecule targeting.

Figure 17. Results of MD clustering analysis of the Tel72 hybrid-212. (A) Top three representative centroids of DBSCAN clusters accounting for ~47% of frames across the entire 100 ns trajectory. (B) space-fill electrostatic APBS map of the first model from A with dashed lines indicating the approximate sizes of each groove created at the two stacking interfaces.


A
29%
9%
9%
B
kV
-30
-20
0
90°
~10.1 Å
~7.4 Å

The major topologies of the extended telomere G4 are hybrid-1 and hybrid-2

Simultaneous with our structural investigations above, we investigated the conformations of G4 units within higher-order structures by using systematic sequence variations of the wild type (WT) Tel48 sequence and observing changes in circular dichroism spectra. These sequence ”mutants” were created with variation in terminal nucleotides or by changes in internal sequences that favor the various hybrid topologies (e.g. hybrid-1 and hybrid-3) (**Table 1**). The Tel48 spectrum (black line in **Figure 18A**) has a main peak at ~290 nm, pronounced shoulder from 265-275 nm, and a trough at 235 nm, indicating that it is primarily composed of hybrid type folds (46). Comparing sequence variants of the form $T_nAGGG(TTAGGG)_7T_m$, where *n* = 1 or 2, and *m* = 0, 1, or 2, we found no major spectral differences (**Figure 19A**), indicating that changes in these flanking nucleotides have no effect on the overall topology. Removal of the 5’-end thymine residues led to a modest reduction in the shoulder at ~270 nm and peak at 290 nm when compared to the WT sequences of similar length (**Figure 19B**, red and blue lines). We speculated that this could be due to the formation of hybrid-3 in the 5’-most G4 unit, which has been reported in shorter sequences lacking the 5’ thymine residues (47,79). Indeed, when an inosine is introduced to favor the hybrid-3 topology in the 5’-most putative G4 the CD changes observed at 270 and 290 nm become more pronounced (**Figure 19C**, blue solid line). Importantly, this suggests that the hybrid-3 topology is not a major topology, as the extended telomere (in the cell) will always include 5’ thymine residues. The spectral change due to hybrid-3 incorporation is made more evident when stabilized in both 5’ and 3’ G4 units (**Figure 19C**, purple), which is of the same shape but approximately 2x the magnitude of hybrid-3. Thus, the hybrid-3 is not likely to exist in the context of the extended telomere, aside from as a potential folding intermediate (47).

Figure 18. Normalized circular dichroism spectra of Tel48 mutants and theoretical monomer G4 spectra. (A) CD spectra comparison of the WT Tel48 G-quadruplex (black) with constructs created to favor the hybrid-1 form in the second (red), first and second (blue), or first position (green). (B) Comparison of the Tel48 CD spectrum with theoretical monomer CD combinations of hybrid-22 (red dashed), hybrid-12 with a 30/70 weighting (blue dashed), and a hybrid-11 (purple dashed).

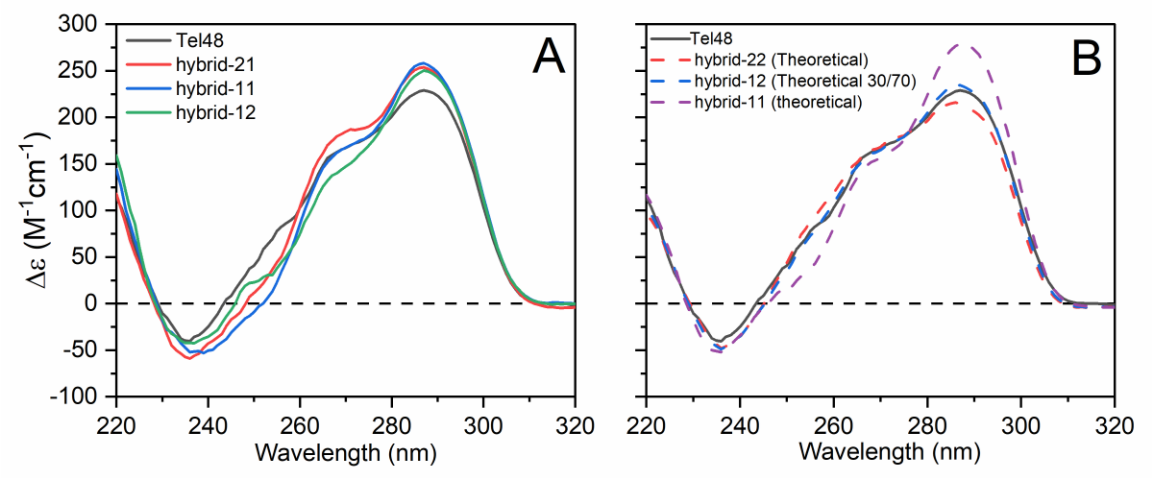
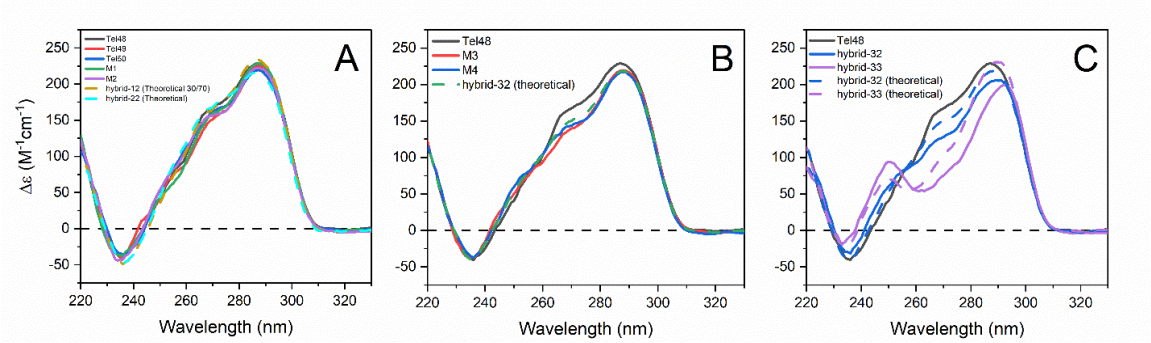

A
Tel48
hybrid-21
hybrid-11
hybrid-12
Δε (M-1cm-1)
Wavelength (nm)
B
Tel48
hybrid-22 (Theoretical)
hybrid-12 (Theoretical 30/70)
hybrid-11 (theoretical)
Wavelength (nm)

Figure 19. Normalized CD spectra of Tel48 compared to various flanking residue and internal mutant sequences.


A
Tel48
Tel49
Tel50
M1
M2
hybrid-12 (Theoretical 30/70)
hybrid-22 (Theoretical)
Δε (M-1cm-1)
Wavelength (nm)
B
Tel48
M3
M4
hybrid-32 (theoretical)
Wavelength (nm)
C
Tel48
hybrid-32
hybrid-33
hybrid-32 (theoretical)
hybrid-33 (theoretical)
Wavelength (nm)

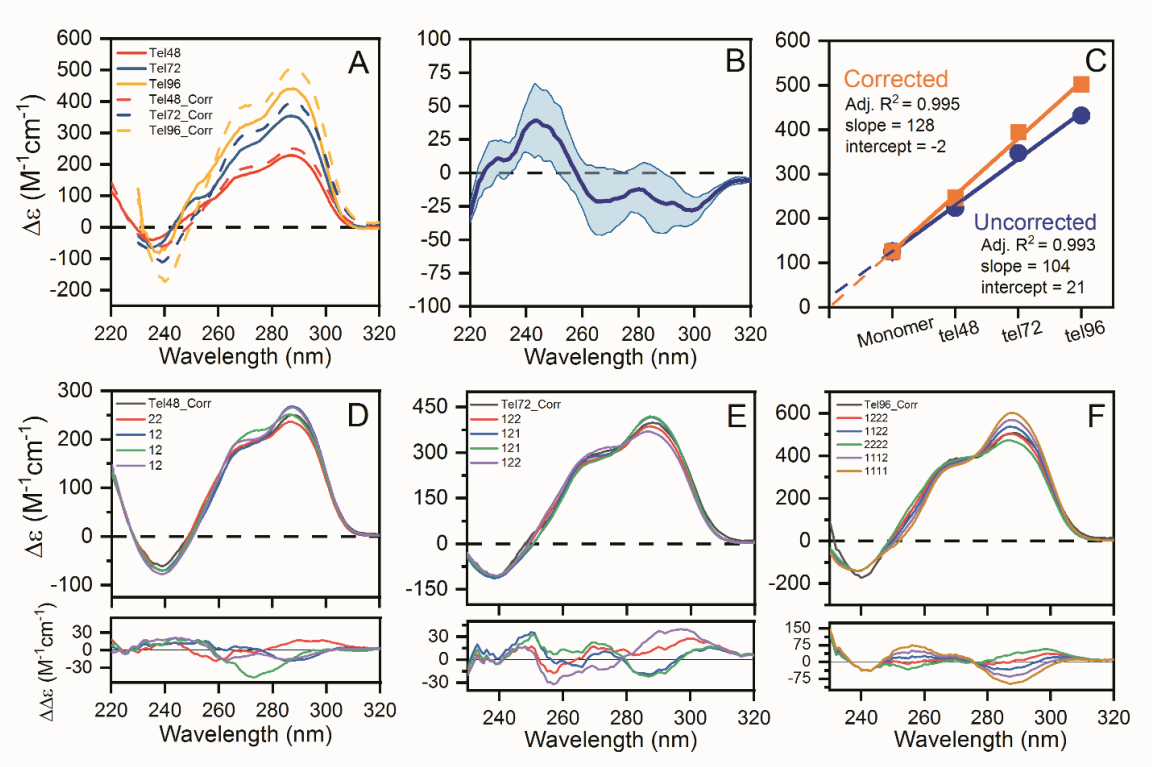
Comparison of the hybrid-21, -11, and -12 sequences to Tel48 revealed subtle differences in each case (where hybrid-2 is assumed as the major conformation in unadulterated telomere sequence flanked by thymine at both ends) (**Figure 18A**). Overall, each spectrum was consistent in shape, but varied in magnitude at various wavelengths. We suspected that this may indicate a preference for the hybrid-22 form. A theoretical hybrid-22 spectrum overlaid nicely with Tel48 (**Figure 18B,** red dashed line). In contrast, a theoretical hybrid-11 (**Figure 18B,** purple dashed line) had a greatly increased 290 nm peak and slight reduction at ~250 nm and looked similar to the mutant hybrid-11 spectrum from **Figure 18A**. Based on the reported 25/75 ratio of hybrid-1 and hybrid-2 for the monomer telomere sequence flanked at both ends by thymine (45), we next tested a variety of computed weighted combinations of hybrid-1 and -2 spectra and found that a 30/70 ratio best reflects the Tel48 spectrum (compare blue dashed line with black in **Figure 18B**). Collectively, these results indicate a preference for both hybrid-1 and -2 topologies in the extended telomere sequence, consistent with our EOM analysis of Tel48.

CD indicates that the higher-order telomere sequences converge on a 25/75 ratio of hybrid-1 and hybrid-2 with maximization of G4 formation

Strand-normalized circular dichroism spectra are the sum of constituent secondary and tertiary structure (196). Thus, just as above we expect that the spectra of higher-order telomere G4s could be recreated by addition of their measured “monomer” spectra. However, comparisons of the various monomer spectra (hybrid-1,-2,-3, and basket forms) to our higher-order telomere spectra resulted in non-negligible “difference” spectra of roughly the magnitude expected for di- or tri-nucleotides. As prior studies suggest, and we have shown here, the extended telomere sequences maximize their G4 potential by leaving no G-tract gaps. The resulting stacking junctions, or other inter-G4 interactions that constrain the loop regions, could potentially give rise to a “junctional” CD signal (196). Thus, to generate the theoretical “junctional” spectrum, we utilized the Tel48 mutant spectra from above and subtracted from them theoretical constituent monomer spectra as appropriate. The resulting spectrum is shown in **Figure 20B**. The junction spectrum has a peak at 240 nm and troughs at ~260 and ~285 nm, which is consistent with the known CD spectra

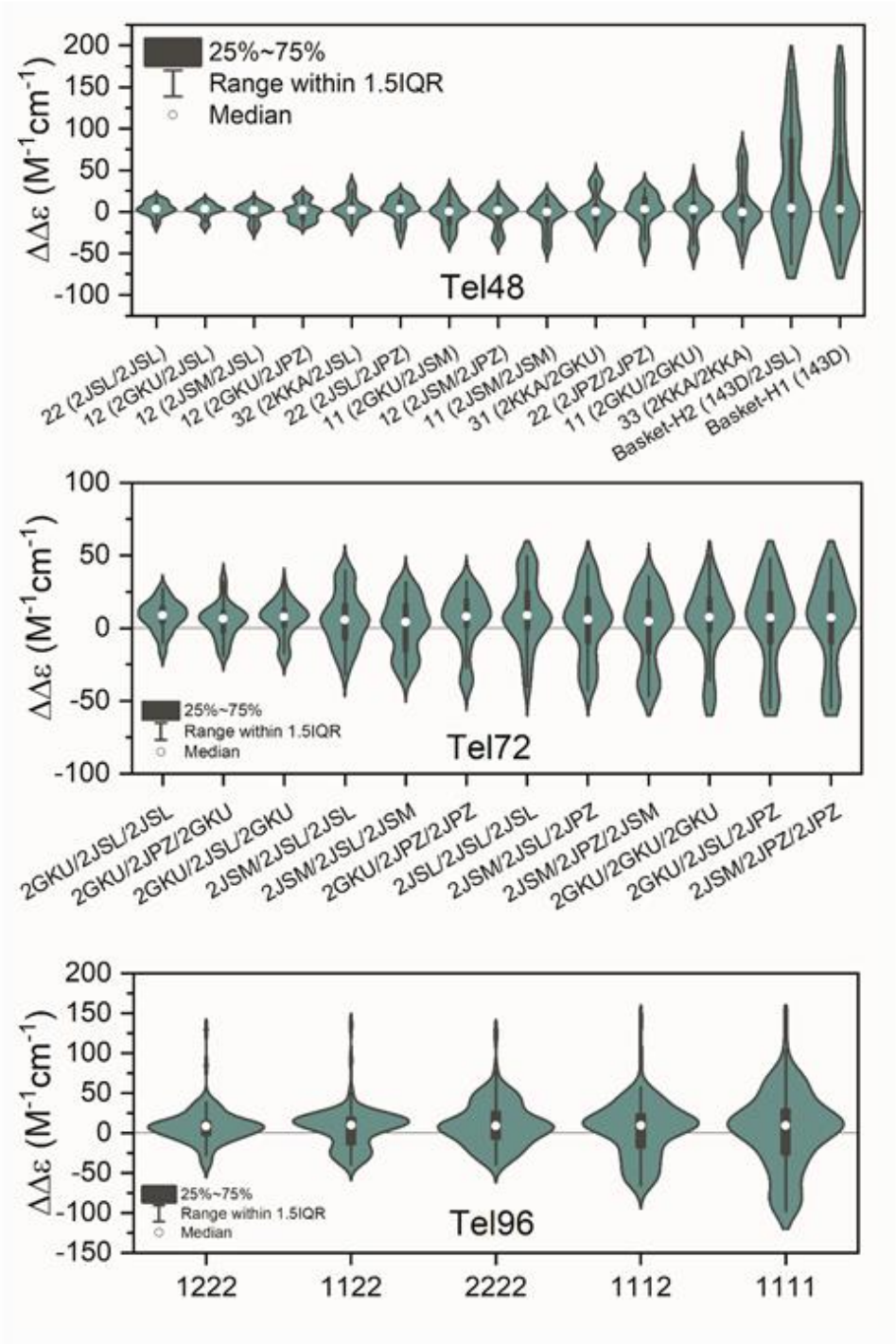
of adenine and thymine polynucleotides (197). This spectrum was then used as a correction factor, and was subtracted from the Tel48, Tel72, and Tel96 spectra (**Figure 20A**). A plot of $\Delta\varepsilon_{290}$ *versus* the putative number of G4s yields a linear regression with a near zero Y-intercept, which is more physically relevant than the regression without the correction (**Figure 20C**). The slope of the corrected regression data indicates that each additional putative G4 increases $\Delta\varepsilon_{290}$ by ~128 $M^{-1}cm^{-1}$, in excellent agreement with the average $\Delta\varepsilon_{290}$ obtained from hybrid (3+1) monomers. These corrected spectra should now be a composite spectrum of monomer G-quadruplex components. A novel finding here is that the CD spectra of higher-order G4 structures contain discernable contributions from G4-G4 interactions.

Figure 20. Circular dichroism analysis of the higher-order telomere G-quadruplexes. (A) Pre- and Post-corrected ("Corr") CD spectra of the Tel48, Tel72, and Tel96 sequences by subtraction of the "junctional" spectrum in B. (B) The average (dark blue line) and range (light blue space fill) of "junctional" CD spectra derived from deconstruction of the Tel48 sequences in **Figures 18 and 19** using constituent monomer spectra. (C) Regression analysis of the uncorrected and corrected $\Delta\varepsilon_{290nm}$ values as a function of the number of G4 motifs. (D-F) Corrected CD spectra of the Tel48, Tel72, and Tel96 sequences with overlaid theoretical spectra derived from the linear addition of monomer spectra. The red spectrum in each plot is the best fit as judged by RSS analysis. Residuals are shown below each figure.


A
Tel48
Tel72
Tel96
Tel48_Corr
Tel72_Corr
Tel96_Corr
Δε (M-1cm-1)
Wavelength (nm)
B
Wavelength (nm)
C
Corrected
Adj. R2 = 0.995
slope = 128
intercept = -2
Uncorrected
Adj. R2 = 0.993
slope = 104
intercept = 21
Monomer
tel48
tel72
tel96
D
Tel48_Corr
22
12
12
12
ΔΔε (M-1cm-1)
Wavelength (nm)
E
Tel72_Corr
122
121
121
122
Wavelength (nm)
F
Tel96_Corr
1222
1122
2222
1112
1111
Wavelength (nm)

The hybrid-1, -2, and -3, as well as the antiparallel basket monomer spectra were systematically compared with the Tel48, Tel72, and Tel96 corrected spectra. **Figures 20D-F** show the corrected spectra in black with "_Corr" indicating the corrected spectrum. Shown below are residuals from the best fit combinations of monomer spectra. In each plot the red spectrum is the best fit, followed by blue, etc. based on residual sum of squares (RSS) analysis (**Figure 21**). The optimal fit to Tel48_Corr is hybrid-22 (in agreement with **Figure 18A**), followed by various hybrid-1/-2 combinations; Tel72_Corr is best fit by a combination of hybrid-1, -2, and -2; Tel96 is best fit by a hybrid-1, -2, -2, -2 (not necessarily in that order as shown above). See **Figure 21** for the exhaustive residual sum of squares (RSS) ordering, PDB IDs, and distributions of CD residuals for each fit. These results suggest an overwhelming preference for hybrid-2 in the longer sequences, although some other hybrid-1/-2 combinations also yield comparable fits. Altogether, the above analyses confirm that the primary two topologies making up the WT telomere higher-order G4s are hybrid-1 and hybrid-2, with proportions approaching a 25% hybrid-1, 75% hybrid-2. Further, the linearity and slope of the regression analysis and excellent agreement with theoretical fits indicates that no long gaps exist in the higher-order human telomere.

Figure 21. Violin plots of residuals obtained as the difference between experimental and theoretical CD reconstruction curves for Tel48, Tel72, and Tel96. The residuals are plotted from best (left) to worst (right) based on residual sum of squares analysis in Origin 2020.


25%~75%
Range within 1.5IQR
Median
ΔΔε (M-1cm-1)
Tel48
22 (2JSL/2JSL)
12 (2GKU/2JSL)
12 (2JSM/2JSL)
12 (2GKU/2JPZ)
32 (2KKA/2JSL)
22 (2JSL/2JPZ)
11 (2GKU/2JSM)
12 (2JSM/2JPZ)
11 (2JSM/2JSM)
31 (2KKA/2GKU)
22 (2JPZ/2JPZ)
11 (2GKU/2GKU)
33 (2KKA/2KKA)
Basket-H2 (143D/2JSL)
Basket-H1 (143D)
Tel72
2GKU/2JSL/2JSL
2GKU/2JPZ/2GKU
2GKU/2JSL/2GKU
2JSM/2JSL/2JSL
2JSM/2JSL/2JSM
2GKU/2JPZ/2JPZ
2JSL/2JSL/2JSL
2JSM/2JSL/2JPZ
2JSM/2JPZ/2JSM
2GKU/2GKU/2GKU
2GKU/2JSL/2JPZ
2JSM/2JPZ/2JPZ
Tel96
1222
1122
2222
1112
1111

Discussion

These results provide the most detailed characterization of extended human telomere G-quadruplex structures in solution to date. We combine circular dichroism, hydrodynamics, and small-angle X-ray scattering experiments integrated with available high-resolution NMR studies on monomeric G4 structures (46) to build medium resolution higher-order structures. For dimeric structures containing two contiguous G4 units (Tel48), the best model is one featuring a 5' hybrid-2, followed by hybrid-1. For longer sequences with three and four G4 units, a mixture of hybrid-2 and hybrid-1 conformations seems to be present in an approximate 3:1 ratio. Our results show unequivocally that for all sequences up to 96 nt in length, the human telomere sequence maximizes its G4 formation, leaving no gaps—in direct contrast to prior EM, AFM, and single-molecule force ramping investigations (159-161). Our results provide the first quantitative estimates of the rigidity of folded telomeric DNA through determination of its persistence length (L_p = ~34 Å). We find that the semi-flexibility of the telomere G-quadruplex is best modeled by an ensemble of configurations which fluctuate between an entirely stacked multimer and unstacked monomeric G4s, as observed by MD simulations, providing potentially unique sites for small molecule targeting in the junctional regions. This model suggests that rigid G4 units are connected by a short, dynamic, interfacial hinge. That interface constitutes a unique structural element to target in drug design efforts.

The WT human telomere monomer sequences exhibit a high degree of polymorphism *in vitro* (46). Under physiologically relevant K^+ solution conditions the WT telomere G4 adopts a hybrid type conformation, favoring hybrid-2 over hybrid-1 (77,152). This conformational bias is seemingly dictated by the presence of 5' or 3' flanking nucleotides. Addition of 5'-TTA to the core sequence, $GGG(TTAGGG)_3$, leads to the favoring of hybrid-1 by a 5'-end capping adenine triplet, whereas addition of one or two thymine to the 3'-end results in a favoring of the hybrid-2 form via a T:A:T triplet stack on the 3'-end (46). The extended, end-flanked sequence, $(TTAGGG)_4T$, forms a major configuration of hybrid-2 (~75%), with minor amounts of hybrid-1 (~25%) (45). This implies that the

energy barrier between the two forms may be small. Consistent with this, our mutational analysis by CD and modeling studies agree with a 75/25 ratio of hybrid-2/-1 for the higher-order WT species. A significant result from our higher-order CD analyses is the unique junctional spectrum (**Figure 20B**). This spectrum was useful in providing a rationale for why the higher-order species exhibited CD signatures that were lower than expected for maximized G4 formation. Moreover, both SAXS and MD modeling studies revealed favorable, but dynamic, stacking interfaces between G4 moieties, in agreement with thermodynamic analyses (162).

Prior NMR investigations of the WT telomere sequence, $(TTAGGG)_4T$, indicate a dynamic equilibrium of conformations (46). If a similar equilibrium exists in the higher-order telomere, then an ensemble of both tertiary conformation and secondary structure would be required to explain both CD and SAXS results. Consistent with this, the Tel48 scattering is modeled well by an ensemble with a 50/50 ratio of hybrid-12 and hybrid-21 conformers. The single model hybrid-21 fit is comparable to the ensemble, and so this solution is, overall, somewhat ambiguous; although, the lack of thymine residues at the 3'-end would, in theory, favor hybrid-1 in the second position, giving us confidence in a preferential hybrid-21 model. Our previous investigation of the WT Tel50 sequence (141) (which differs in sequence by two additional thymine residues at the 3' end) proposed that the major form is hybrid-12. We used steady-state fluorescence measurements of 2-aminopurine-substitutions to assess the solvent accessibility for each adenine site. From this it was found that residue A15 is the least solvent exposed, which agreed with SASA calculations of the hybrid-12 model (in this conformation A15 is buried in the stacking junction between the two G-quadruplex units). Coincidentally, by having the EOM algorithm increase the number of conformers in the Tel48 ensemble, we find that part of the new solution is a hybrid-12 conformer that is almost identical to the previously proposed Tel50 hybrid-12 model (**Figure 13**). Thus, the collective experimental observables support an equilibrium of hybrid-1 and hybrid-2 in either position.

We have also investigated the possibility of a hybrid-3 form, which is a two-tetrad antiparallel G-quadruplex that has been characterized *in vitro* in potassium containing solution (47), and confirmed as existing in a cellular environment by Bao et al. (79). In both instances the telomere

sequences used were lacking 5' thymine residues. The 5' thymine residue destabilizes the hybrid-3 structure and, ultimately, favors the hybrid-1 (47). In the cell the 5'-flanking thymine is always present. We have confirmed that the extended WT telomere sequences do not favor the hybrid-3 in solution by mutational analysis, showing that it may only occur in the 5'-most position when thymine is removed (**Figure 19**). Thus, the hybrid-3 topology may only function as a folding intermediate (47), rather than as a major constituent topology of the higher-order telomere G4.

The single-stranded telomere overhang is a critical regulator of genomic integrity. Spanning the junction of the duplex and single-stranded telomere region is a protective protein complex known as shelterin (64,198). Shelterin is composed of the proteins TRF1, TRF2, RAP1, TIN2, TPP1, and POT1. POT1 (protection of telomeres 1) is essential in sequestering the free 3' overhang, shielding it from eliciting aberrant single-stranded DNA damage responses, preventing homologous recombination, and regulating the activity of telomerase (156). POT1 binds directly to the 3' single-stranded overhang with high affinity and in a highly sequence specific manner (64,155,199). EM micrographs have revealed that POT1-TPP1 complexes coat the entirety of the extended telomere 3' overhang, forming compact, ordered complexes of ssDNA-POT1-TPP1 without gaps (200). Importantly, disruption of POT1's shielding of the single-stranded overhang elicits an ATR-dependent DNA damage response through the promiscuous ssDNA binding protein RPA (156,201). A recent AFM investigation of the Tel96 sequence with POT1 by the Opresko lab (159) found that maximization of G4 formation "rarely" occurs, and that POT1 associates by simply recognizing the resulting gaps. We, and others (55,85,157,163), find this conclusion at odds with solution-based results. Accessible ssDNA gaps between G4s would allow RPA to compete unimpeded with POT1 binding. Indeed, RPA outcompetes POT1-TPP1 binding to single-stranded telomere DNA *in vitro* (202). Further, G-quadruplex secondary structure enhances POT1-TPP1's protection against RPA in physiologically relevant levels of K^+ (150 mM) (203). We recently showed that POT1 unfolds and binds to telomere G4s using a conformational selection mechanism (85) and demonstrated that the kinetics of unfolding the Tel48 sequence is essentially the same as the Tel24 monomeric G4. Importantly, this suggests that a maximization of G4 formation does not impede POT1 binding. Taken together, the physiological significance of telomeric G4 maximization

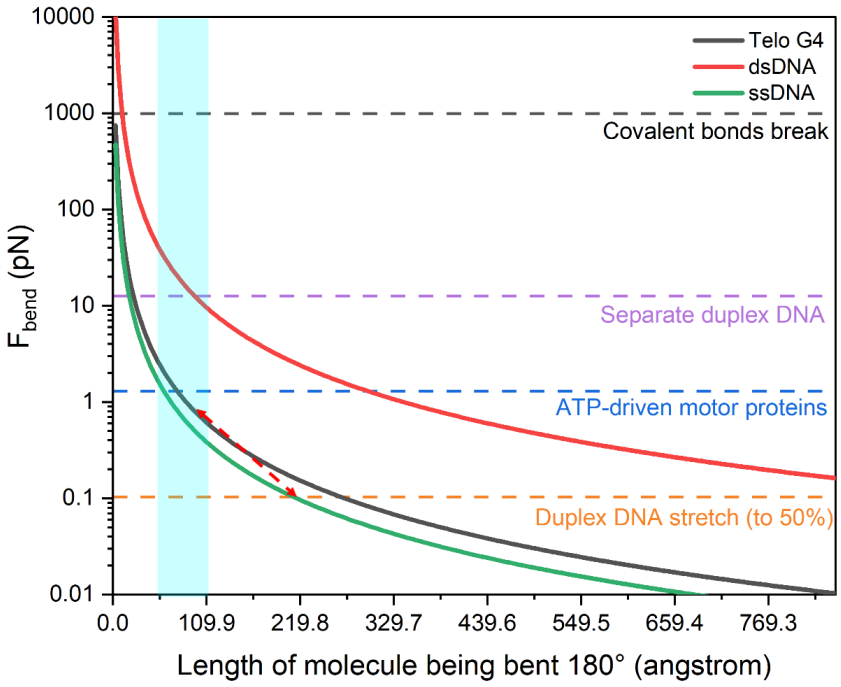
is two-fold: (1) G-quadruplex secondary structure prevents promiscuous RPA binding and, (2) the G4 secondary structure promotes exclusive interaction with shelterin through specific POT1 unfolding and binding, tilting the scale in favor of POT1 over RPA.

The mechanistic details of how the shelterin complex orchestrates the sequestration of the single-stranded 3' end are still not entirely understood, but are of great importance in drug discovery (198,204). Recently, the Cech laboratory conducted a thorough investigation of co-expressed and isolated complexes of the shelterin proteins *in vitro* (198). Based on their results a shelterin "load & search" model was proposed, whereby TRF2 and POT1 localize the shelterin complex to the telomere by specifically recognizing and binding to the single-stranded/double-stranded (ss/ds) telomere junction. The authors propose that POT1 searches for its optimal binding sequence, TTAG, at the 3' end by a scanning search mechanism, eventually looping the 3' end back forming a loop bridged by shelterin proteins that is "unlike a T-loop". An earlier report from the Cech laboratory found that POT1 and POT1-TPP1 completely coat the long ssDNA telomere repeats *in vitro* (200). Thus, the "normal" sequestration mechanism of the human telomere 3' overhang is seemingly a POT1-coated loop structure anchored to the ss/ds telomere junction.

DNA looping is a common theme in the cell (190). From a physical stand-point, DNA looping has been extensively studied for its relationship with genetic packaging into nucleosomes and effects on transcription (190,205). A commonly reported measure of the structural rigidity of a biological polymer is the persistence length, L_p, that defines the length over which a polymer remains unbent in solution. In this work, we have applied both SAXS and hydrodynamic modeling methods to derive an estimate of the telomere G4 persistence length. Using our telomere G4 L_p estimate, along with values reported for single- and double-stranded DNA, we can compare the relative forces required to bend each 180° around the arc of a semi-circle of a given radius (**Figure 22**) (190). From this plot we find that, in the case of single-stranded telomere DNA of length >200 angstroms (~63 nt) the force required to bend the polymer 180° (in the shape of a semi-circle) is negligible—energy requirements on the order of thermal fluctuations. However, if that same stretch of 63 nucleotides were to form maximal G-quadruplexes (decreasing polymer length to <100

angstrom), the increase in energy to bend increases by an order of magnitude—now requiring energy input equivalent to ATP hydrolysis. Although somewhat intuitive, this implies that in the absence of significant energy input, short telomere G4s must be made single-stranded in order to bridge the terminal 3' TTAG repeat (capped by POT1) with the shelterin complex. More importantly, this figure indicates that small molecules which bind the inter-G4 grooves, thereby increasing its effective persistence length, could shift the force curve to the right (towards the dsDNA curve, solid red in **Figure 22**) and subsequently drive up the energy cost for associating the POT1-bound 3' end with the shelterin loop. Indeed, during the drafting of this manuscript Gao *et al.* have demonstrated that a small molecule targeting the wild-type Tel48 can shift the distribution of conformations to favor a more compact, likely stacked, conformation (206)—a transition that would directly affect persistence length. It is well established that telomere G-quadruplex interacting small molecules are able to displace shelterin components, uncap the telomeres, and ultimately, inhibit telomerase *in vitro* and *in vivo* (83,207). Thus, there is abundant rationale for future work targeting these unique G4 junctional sites with stabilizing small molecules.

Figure 22. DNA force of bending plot for single-stranded (green), double-stranded (red), and G4 telomere DNA (black). Force curve calculations were performed similar to reference (190) using literature values of persistence length for ssDNA (L_p = 22 Å), dsDNA (L_p = 550 Å), and Telomere G4 (L_p = 34.8 Å) as measured here. The Y-axis is the estimated force (in pN) to bend a length of DNA (X-axis) 180° about the arc of a semi-circle (i.e. if you have a 330 angstrom long single-stranded DNA it will require a force of ~0.05 pN to bend it into a semi-circle). Dashed horizontal lines are visual references to common biological forces found in the cell (orange indicates the approximate range of force from thermal fluctuations). The light blue region highlights the range in which short telomere G4s would be found, indicating that a large force would be required to bend short telomeres (≤96 nt). The dashed red arrow illustrates that if a ~200 Å long ssDNA telomere (approximately 63 nt) were to spontaneously fold into a contiguous G4 structure, the resulting bending force required for a 180° turn increases by an order of magnitude. The increase in bending force is comparable to the same length of DNA in duplex form (330 Å long duplex requires external forces equivalent to ATP hydrolysis to bend 180°). In the case of duplex DNA, the energy requirement of “tight” bending is usually compensated for by the highly positive charge on histones.


10000
1000
100
10
1
0.1
0.01
F_{bend} (pN)
0.0
109.9
219.8
329.7
439.6
549.5
659.4
769.3
Length of molecule being bent 180° (angstrom)
Telo G4
dsDNA
ssDNA
Covalent bonds break
Separate duplex DNA
ATP-driven motor proteins
Duplex DNA stretch (to 50%)

Utilizing a robust integrative approach, we have presented here the highest-resolution view of the higher-order telomere G4 to date. SAXS refinement of MD-derived models constructed from high-resolution techniques is now a mainstay in structural biology. However, SAXS refinement of MD generated atomistic models, while excellent for discarding unrealistic topologies and conformations, is not necessarily definitive when conformational and topological polymorphism presents itself. Thus, we await higher-resolution techniques that can inform on the distributions of topologies in the higher-order telomere G-quadruplex.

CHAPTER III

STRUCTURE-BASED TARGETING OF THE HIGHER-ORDER HUMAN TELOMERE G-QUADRUPLEX

Disrupting telomere maintenance and homeostasis has emerged as a new avenue of anti-cancer therapy. Telomere shrinkage appears essential as a natural mechanism of cellular aging, as once a critical limit is reached the cells undergo senescence. In cancer, the primary mechanism used to avoid this fate is the re-activation of telomerase. In concert with the shelterin complex, telomerase uses its reverse transcriptase functionality to restore telomere length leading to unrestricted cell proliferation. Recently, the human telomeres have been shown to fold into non-B DNA structures known as G-quadruplexes (G4s). Small molecules that bind to monomeric telomere G4s with high affinity have been identified and show clear inhibition of telomerase in cells by sequestering of the free 3' telomere overhang. We have recently revealed that the telomere G4 multimer contains potentially novel inter-G4 junctional regions which could be targeted with structure-based drug discovery approaches. Herein we present the results of a massive virtual screening campaign targeting the telomere G4 multimer inter-quadruplex junctions with small molecules. Using circular dichroism melting studies as a screen we have identified a small molecule scaffold that interacts with the higher-order telomere G4. Further, using orthogonal biophysical methods, we determine a binding stoichiometry of 1:1 with the number of G4 junctions in higher order telomere G4s, and no binding to monomeric G4s, making it a promising lead molecule for selectively targeting the telomere.

Materials and Methods

Virtual drug screening

Virtual screening was performed using Surflex-Dock 2.11 on the Kentucky Dataseam Grid utilizing over 53 million virtual ligands from the ZINC 2014 (24,877,119 molecules), 2016 (17,244,856 molecules), and 2018 (11,154,739 molecules) drug-like libraries. Docking was performed on 27 unique sites across three different telomeric G-quadruplex models created previously (163) with Surflex-Dock's '-pgeom' parameter set. Surflex-Dock protomols were generated at residues within the G4-G4 junctions, loops, and grooves using standard Surflex-Dock procedures with 'bloat' and 'thresh' set to default values. In total, >53 million virtual small molecules were docked at each site. The top 1% scoring molecules across all sites, models, and small molecule libraries were pooled and analyzed in Schrödinger's (Schrödinger, Inc., New York, NY) Canvas application using a hierarchical clustering algorithm to cluster molecules based on binary fingerprints and Tanimoto similarity criteria. From this analysis, the top 100 centroid molecules (most representative scaffolds of each clade) were then chosen for purchasing. This process was also repeated using the top 5% of small molecules from the 2014 ZINC library alone. Duplicates were removed from the final list of molecules considered for purchase. In total, 32 visually assessed molecules were purchased from Molport.com for initial testing and given the designation "C#". After initial screening, 5 addition small molecules were also purchased using a structural similarity search on compound C21.

Buffers and Compounds

All experiments were conducted in a potassium phosphate buffer with varying levels of KCl (6 mM Na_2HPO_4, 2 mM NaH_2PO_4, 0-185 mM KCl, 1 mM Na_2EDTA, pH 7.2). Compounds were purchased from Molport.com and were dissolved to 10 mM in DMSO upon receiving. Compound stocks were stored at -20°C until use.

Preparative Size Exclusion Chromatography (SEC)

Oligonucleotides were purified using SEC as detailed previously (168). Briefly, oligonucleotides were annealed at concentrations of 40-100 μM in their respective buffers, filtered through 0.2 μm filters, and injected onto an equilibrated Superdex 75 16/600 SEC column (GE Healthcare 28-9893-33) using a Waters 600 HPLC system. The flow rate was maintained at 0.5 mL/min and sample fractions were collected every 2 minutes from 100 to 180 minutes run time. The molecular weights of fractionated species were estimated based on a regression analysis of elution time vs. log(MW) of protein standards (Sigma #69385), with elution profiles monitored at 260 nm and 280 nm. Purifications were carried out at room temperature and fractionated samples were stored at -20°C or 4°C prior to downstream analysis.

Circular Dichroism Spectroscopy (CD)

CD melting studies and spectra were collected on a Jasco-710 spectropolarimeter (Jasco Inc. Eason, MD) equipped with a Peltier thermostat regulated cell holder and magnetic stirrer. CD and melting spectra were collected using the following instrument parameters: 1 cm path length quartz cuvette, 240 to 340 nm wavelength range, 1.0 nm step size, 200 nm/min scan rate, 1.0 nm bandwidth, 2 second integration time, and 3 scan accumulation. Spectra were recorded at 20.0°C and melting spectra were collected over a range of 20°C to 98°C with 2°C step intervals, 4°C/min ramp speed, and a 2-minute equilibration time at each temperature before acquisition. Spectra were corrected by subtracting a buffer blank. Spectra were normalized to molar circular dichroism (Δε) based on DNA strand concentration using the following equation:

$$\Delta\varepsilon = \theta/(32982 \times c \times l)$$

where θ is ellipticity in millidegrees, c is molar DNA strand concentration in mol/L, and l is the path length of the cell in cm. Fitting of melting curves was performed using least-squares fitting of a Boltzmann function in Origin 2020.

Fluorescence thermal shift assays (FTSA)

Small molecule screening by FTSA was performed on an Applied Biosystems StepOnePlus Real-Time polymerase chain reaction (PCR) instrument in 96-well plates as an

adaption of previous work (208). Briefly, 10 mM compound stock solutions in DMSO were used to create 96-well stock solution plates by diluting each compound to 2x final concentration in potassium phosphate buffer. The same volume of DMSO was used as a control. The 5' 6-FAM (Fluoroscein) and 3' TAMRA (Carboxytetramethylrhodamine) FRET-labeled DNA, post-annealing, were quantified by UV-Vis and diluted to 2x final concentration. FTSA reaction mixes were made up in 96-well Applied Biosystems MicroAmp PCR plates by mixing 10 µL of 2x compound solution (or buffer/DMSO control) with 2x FRET-labeled DNA to yield 20 µL of 1x reaction mix. Plates were then spun down at 1250 rpm for 2-3 minutes in a benchtop centrifuge to remove bubbles. Samples were denatured by ramping the temperature from 20.0°C to 99.8°C in 0.2°C increments at a rate of approximately 0.7°C/min. Fluorescence quenching of 6-FAM was monitored at each 0.2°C step using the instrument's onboard FAM filter over the entire reaction yielding a melt curve. Melting temperatures (T_m) were determined from the 1st derivative of the normalized melting curves (209), and differences in melting temperatures (ΔT_m) were determined by taking the difference of control and sample wells:

$$\Delta T_m = T_{m,sample} - T_{m,control}$$

Where $T_{m,sample}$ and $T_{m,control}$ are the melting temperatures of the sample and control, respectively. Measurements are averages of triplicate experiments repeated on 3 separate days unless otherwise specified.

Analytical ultracentrifugation (AUC)

Sedimentation velocity (SV) experiments were performed in a Beckman Coulter ProteomeLab XL-A analytical ultracentrifuge (Beckman Coulter Inc., Brea, CA) at 20.0°C and 40,000 rpm in standard 2-sector cells using An50Ti or An60Ti rotors. 100 scans were collected over an 8-hour period and analyzed in Sedfit (170) using the continuous C(s) model with a partial specific volume of 0.55 mL/g for DNA. AUC drug binding experiments were carried out as detailed previously (210), with a final compound concentration of 100 µM and 10 µM DNA (10:1

[compound]:[DNA]). All compounds with reported stoichiometry from AUC experiments were monitored at a wavelength well above 300 nm to ensure DNA had no contribution to estimated drug concentrations.

Results and Discussion

We began *in vitro* screening of the initial 32 compounds against the Tel72 sequence (**Table 4**) using circular dichroism melting experiments with ratios of 1:50, 1:100, or 1:200 [Tel72]:[compound]. This screen resulted in only one compound, C20, which increased the T_m (**Figure 23A**). C20 contains a known G4 interacting scaffold (17) and so was not pursued further. Instead, we investigated compound C21, which caused a substantial decrease in the T_m and had a unique molecular scaffold (not shown). We confirmed that C21 could bind and stabilize the monomeric telomere-derived G-quadruplexes 143D, 2GKU, and 2JSL, as well as the parallel c-MYC promoter-derived 1XAV by FTSA assays (**Figure 24**). Surprisingly, treatment with C21 led to T_m increases of 5-12°C with monomeric G4s. Thus, C21 stabilizes monomer G4s but destabilizes the telomere G4 multimers. Binding of C21 was subsequently verified by AUC studies which showed an approximate 2:1 binding ratio with Tel72 (**Figure 23C**).

Table 4. Oligonucleotide sequences used in this study.

Name	Sequence	Length	MW (kDa)	E260 ($M^{-1}cm^{-1}$)
2JSL	TAGGGTTAGGGTTAGGGTTAGGGTT	25	7.9	253100
2GKU	TTGGGTTAGGGTTAGGGTTAGGGA	24	7.6	244300
143D	AGGGTTAGGGTTAGGGTTAGGG	22	7.0	228500
Tel48	$(TTAGGGTTAGGGTTAGGGTTAGGG)_2$	48	15.2	489100
Tel72	$(TTAGGGTTAGGGTTAGGGTTAGGG)_3$	72	22.8	733400
1XAV	TGAGGGTGGGTAGGGTGGGTAA	22	7.0	190394

Figure 23. Screening results for compounds C20 and C21. (A) CD melting plots showing the fractional change in ellipticity at 290 nm of Tel72 versus temperature in the presence or absence of compounds. Tel72 was at a concentration of 1 μM and the given compounds were at 100 μM. BRACO-19 served as a positive control. (B) Representative AUC C(s) *versus* S distributions of Tel72 alone (black, $S_{20,w}$ = 3.45), C21 alone (red), and Tel72 in the presence of C21 (green).

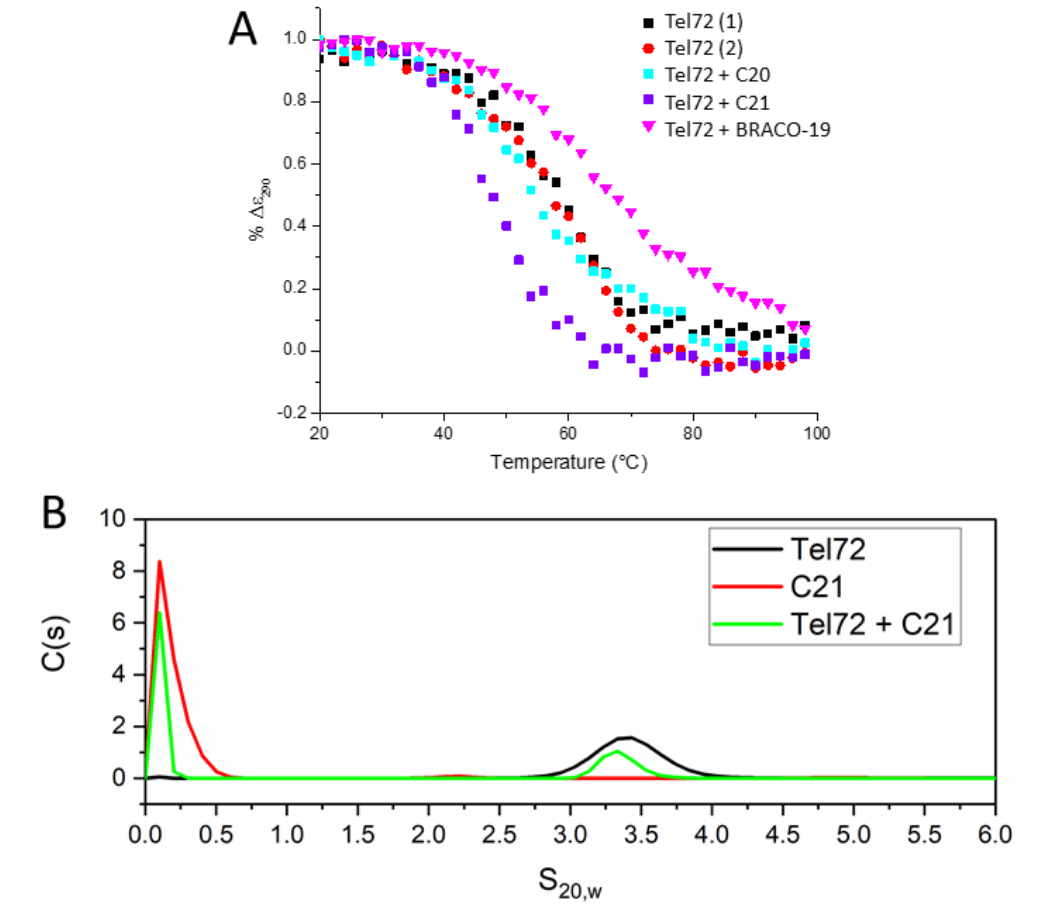
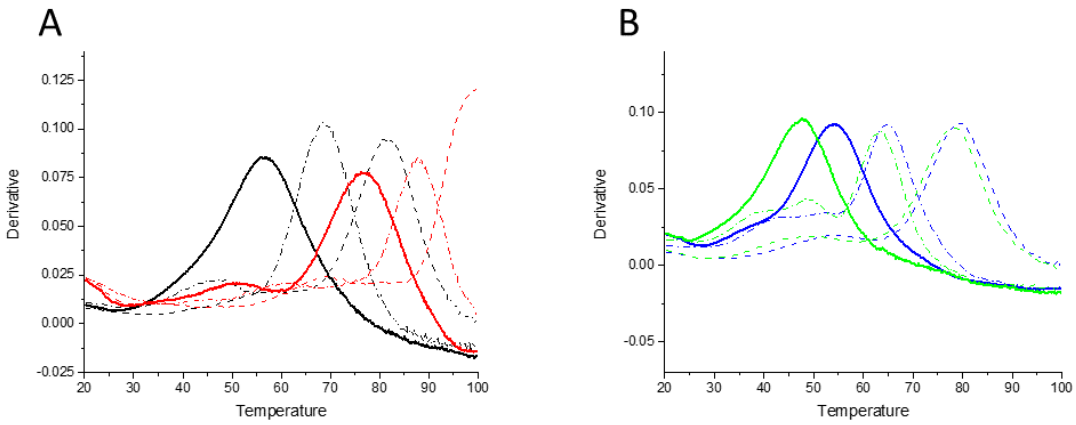

A
Tel72 (1)
Tel72 (2)
Tel72 + C20
Tel72 + C21
Tel72 + BRACO-19
% Δε290
Temperature (°C)
B
Tel72
C21
Tel72 + C21
C(s)
S20,w

Figure 24. FTSA assay results for compounds C20 and C21 with various G4s. (A) derivative curves of 143D (black) or 1XAV (red) in the absence (solid lines) or presence (dashed lines) of C20 and C21. (B) derivative curves of 2GKU (blue) or 2JSL (green) in the absence (solid lines) or presence (dashed lines) of C20 and C21.


A
Derivative
0.125
0.100
0.075
0.050
0.025
0.000
-0.025
20
30
40
50
60
70
80
90
100
Temperature
B
Derivative
0.10
0.05
0.00
-0.05
20
30
40
50
60
70
80
90
100
Temperature

Based on C21's confirmed interaction with Tel72 and monomer G4s, we next used Molport.com's similarity search tool to identify 5 compounds with similar scaffolds to test for improved selectivity. These compounds were screened by CD melting analysis as above, with no clear differences in CD spectra or T_m. However, by FTSA assay, we observed interaction of C37 with the Tel48 sequence, but not with monomeric telomere sequences (**Figure 25**). This finding was subsequently confirmed by AUC binding studies that showed no binding to the monomeric telomere sequences 143D, 2JSL, or 2GKU, and ratios of binding to the higher-order telomere G-quadruplexes of 1.3:1 [compound]:[Tel48] and 2.3:1 [compound]:[Tel72] (**Figure 26B**).

Due to the lack of interaction with the hybrid-1 (2GKU) or hybrid-2 (2JSL) monomeric telomere sequences, the above binding stoichiometries indicate that C37 prefers binding between the constituent G4s at the junctions, which are absent from the monomer forms. Thus, C37 exhibits a unique interaction mechanism that could be beneficial in inhibiting telomere homeostasis. Further, C37 is a unique molecular scaffold, making it an excellent lead as a selective telomere binding small molecule. Cell-based and selectivity investigations with C37 are ongoing.

Figure 25. FTSA melting analysis of various monomer G4s and Tel48 (black) in the presence of C36 (red) and C37 (blue). (A) Duplicate results of 143D. (B) Duplicate results of 2GKU. (C) Duplicate results of 2JSL. (D) Duplicate results of 1XAV. (E and F) Results of Tel48 melting with two different fluorescence emission filters.

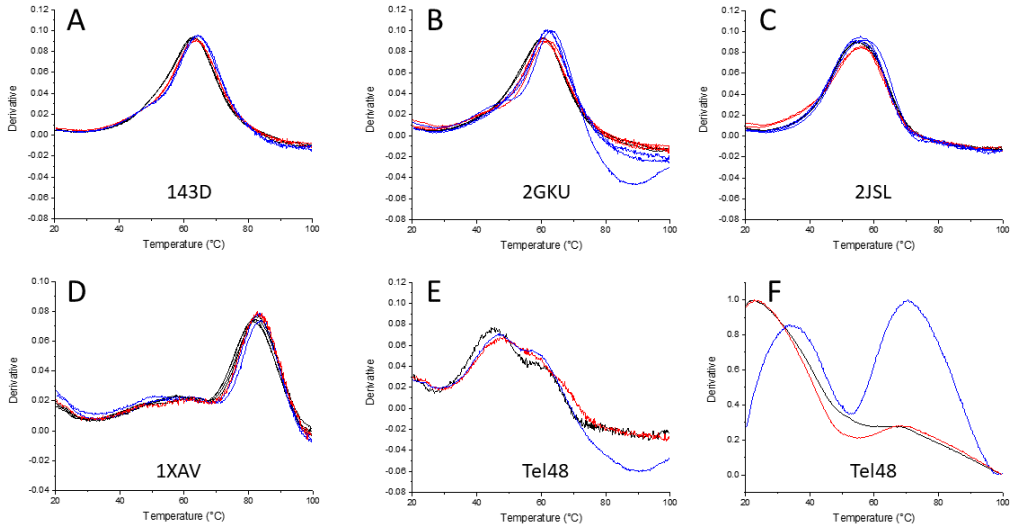
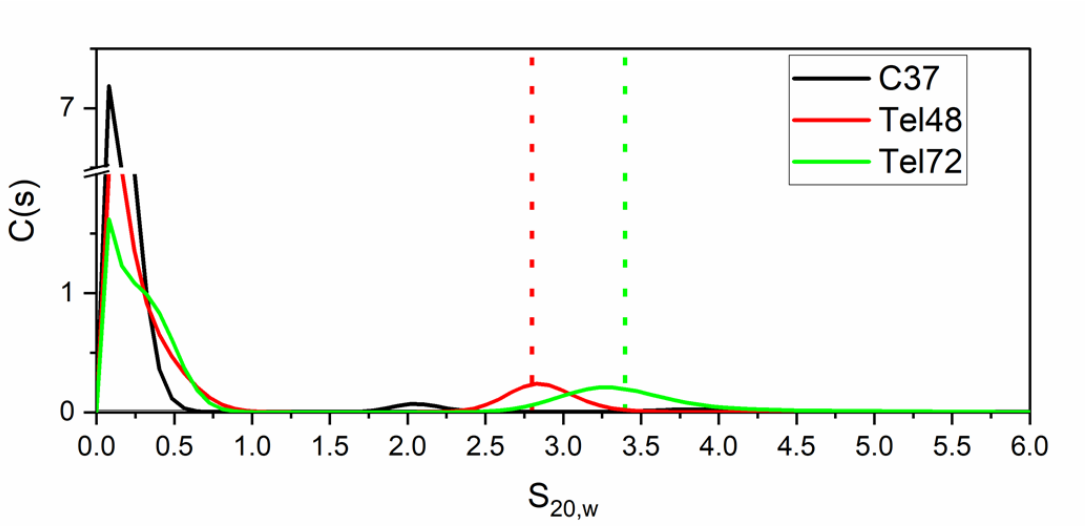

A
143D
B
2GKU
C
2JSL
D
1XAV
E
Tel48
F
Tel48
Derivative
Temperature (°C)

Figure 26. AUC C(s) *versus* S distributions of Tel48 and Tel72 with compound C37. Compound C37 alone is shown in black and exhibits a small peak around 2.1 S, potentially due to molecule aggregation. C37 in the presence of Tel48 (red) or Tel72 (green) are shown with dashed lines indicating the measured sedimentation coefficient for each species without compounds. C37 was measured at a wavelength of 350 nm to avoid any overlap with DNA's intrinsic absorption.


C37
Tel48
Tel72
C(s)
7
1
0
0.0
0.5
1.0
1.5
2.0
2.5
3.0
3.5
4.0
4.5
5.0
5.5
6.0
S20,w

Altogether, in this study we have demonstrated the validity of using atomistic models derived from integrated structural biology techniques in structure-based drug discovery campaigns. We have identified a novel molecular scaffold, C37, which exhibits a stoichiometric ratio of binding to the higher-order telomere G4s suggestive of inter-G4 junctional binding. Further, we demonstrate that it does not bind monomeric G4s, making it a unique and compelling lead molecule for future investigations.

CHAPTER IV

THE HTERT CORE PROMOTER FORMS THREE PARALLEL G-QUADRUPLEXES

The structure of the 68 nt sequence with G-quadruplex forming potential within the hTERT promoter has been disputed. One model featured a structure with three stacked parallel G-quadruplex units, while another featured an unusual duplex hairpin structure adjoined to two stacked parallel and antiparallel quadruplexes. We report here the results of an integrated structural biology study designed to distinguish between these possibilities. As part of our study, we designed a sequence with an optimized hairpin structure and show that its biophysical and biochemical properties are inconsistent with the structure formed by the hTERT wild-type sequence. By using circular dichroism, thermal denaturation, nuclear magnetic resonance spectroscopy, analytical ultracentrifugation, small-angle X-ray scattering, molecular dynamics simulations and a DNase I cleavage assay we found that the wild type hTERT core promoter folds into a stacked, three-parallel G-quadruplex structure. The hairpin structure is inconsistent with all of our experimental data obtained with the wild-type sequence. All-atom models for both structures were constructed using molecular dynamics simulations. These models accurately predicted the experimental hydrodynamic properties measured for each structure. We found with certainty that the wild-type hTERT promoter sequence does not form a hairpin structure in solution, but rather folds into a compact stacked three-G-quadruplex conformation.

Introduction

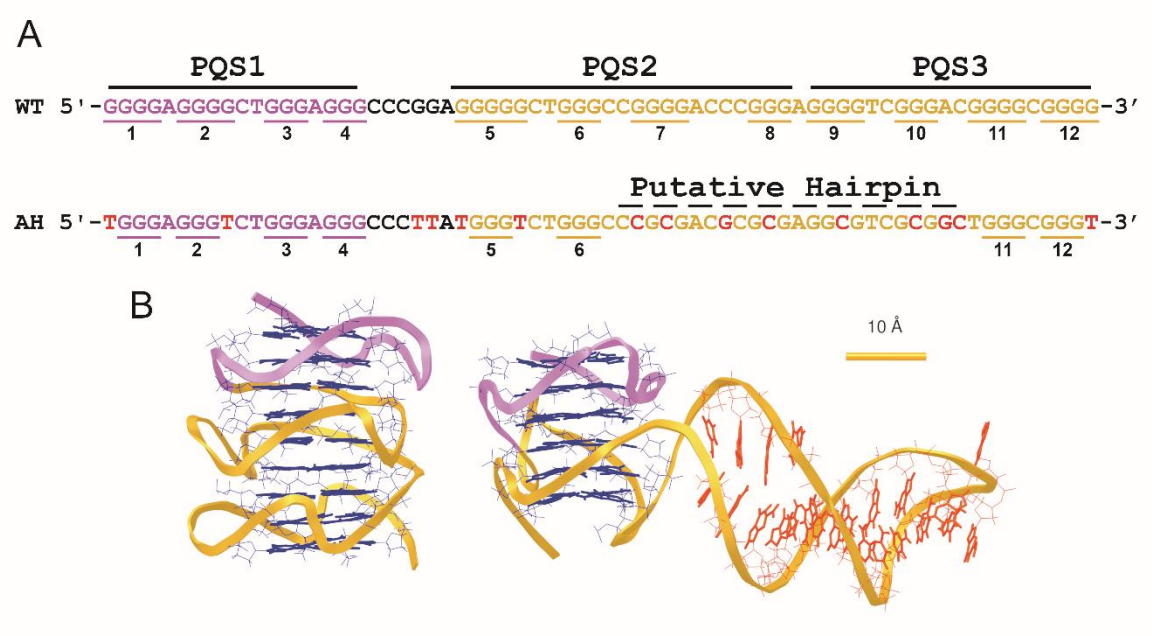
G-quadruplexes (G4s) are four-stranded non-B DNA structures formed from Hoogsteen hydrogen bonding of guanines to form stacked quartets. G-quadruplexes are known to form in the telomeres of a variety of eukaryotic organisms where their role is primarily in telomere homeostasis (25,211). Bioinformatic analyses have shown that G-quadruplex sequence motifs are concentrated in oncogene promoters (11,12,15), and these promoter G-quadruplexes have been under investigation for their ability to modulate gene expression (15). Many promoter G-quadruplexes are currently being investigated for their potential in modulating their respective gene products: c-MYC (16), KRAS (101), HRAS (102), HIF (103), and VEGF (104).

Human telomerase reverse transcriptase (hTERT) is the catalytic subunit of telomerase, the enzyme primarily responsible for the immortality of cancer cells. hTERT is an important oncogene with G4 motifs within its promoter (105,106,144). The *hTERT* gene encodes the reverse-transcriptase component of the human telomerase ribonucleoprotein complex (67). Telomerase (TERT) is responsible for maintenance of telomeres, and this activity is thought to be vital in cellular immortalization (212,213). TERT is normally undetectable in somatic cells (except for stem cells), and its aberrant expression is associated with 85-90% of cancers investigated (10,68,69). The nearly exclusive expression of TERT in cancer cells has been acknowledged for more than two decades as a target for anti-cancer therapies. Many contemporary techniques which target telomerase, such as small molecule inhibitors, gene therapy, anti-sense oligonucleotides, and immunotherapies, have demonstrated TERT inhibition as a viable mechanism in cancer treatment (214). Unfortunately, no direct inhibitors of telomerase have been clinically successful (215). Some of the more promising direct inhibitors exhibit severe toxicity in hematopoietic stem cells (216). This provides a strong rationale for investigating alternative mechanisms to prevent telomerase activity in cancer.

The wild type (WT) hTERT core promoter region (approximately -180 to +1 of transcription start site) (217) contains twelve tracts of three or more guanines which enable formation of G-

quadruplexes (105,106,144,145). Functional genetic studies have identified point mutations within these G-tracts that are directly linked to increased expression of TERT (218). Two mutations, G124A or G146A, are found in 60-80% of urothelial carcinomas (219), 71% of melanomas (220), 83% of glioblastomas (221), as well as a variety of other cancers. These mutations result in *de novo* formation of E-twenty-six (ETS) transcription factor binding sites and confer a selective advantage to cancer cells by allelic recruitment of the transcription factor GABP (219,222). These mutations occur within G-tracts 5 and 8, the terminal G-tracts of the second putative quadruplex sequence (PQS2) (**Figure 27**) and have been suggested to impact G-quadruplex transcriptional silencing (106,145). This has been supported by a G-quadruplex-stabilizing small molecule targeting the hTERT promoter in MCF-7 breast cancer cells (223). Thus, further investigation of the secondary structure formed by the promoter DNA sequence is warranted.

Figure 27. Comparison of WT and AH sequences and contemporary models. (A) (Top) The wild-type hTERT core promoter sequence and (bottom) the modified antiparallel hairpin (AH) sequence with PQS-1, -2, and -3 indicating the “putative quadruplex forming sequences”, and the artificially strengthened hairpin region shown with a dashed line. The purple and gold colors correspond to the purple and gold regions in B. Red nucleotides indicate residues that were modified from WT to force the formation of the parallel-antiparallel stacked hairpin model as in B. (B) The two current models proposed for the secondary structure formed in the hTERT core promoter, three parallel stacked (left) and a parallel stacked onto an antiparallel with 8 bp hairpin. The sugar phosphate backbone is shown in ribbon, guanines in G-tetrads in blue, and nucleotides involved in the hairpin shown in orange. Extraneous loop bases were removed for clarity. The purple “PQS1” region reportedly adopts the same parallel conformation in both models, and this sequence has been solved (105). The gold region highlights the major difference in the two models, herein referred to as the “PQS23” region.


A
PQS1
PQS2
PQS3
WT 5'-GGGGAGGGGCTGGGAGGGCCCGGAGGGGGCTGGGCCGGGGACCCGGGAGGGGTCGGGACGGGGCGGGG-3′
1 2 3 4 5 6 7 8 9 10 11 12
Putative Hairpin
AH 5'-TGGGAGGGTCTGGGAGGGCCCTTATGGGTCTGGGCCCGCGACGCGCGAGGCGTCGCGGCTGGGCGGGT-3′
1 2 3 4 5 6 11 12
B
10 Å

The original three studies (105,106,144) of G-quadruplex forming capability of the hTERT core promoter utilized a sliding frame approach to identify G-quadruplex formation and stability. From left to right (5' to 3') in **Figure 27A** we have designated these putative quadruplex forming segments (PQS) as PQS1, PQS2, and PQS3. Isolated PQS1 has been shown to exist as a mixture of both parallel and anti-parallel (3+1) structures by nuclear magnetic resonance (NMR) spectroscopy (105). Isolated PQS3 was shown to adopt a parallel conformation (105,144), albeit with slightly lower stability than PQS1. Further support for the formation of both PQS1 and PQS3 quadruplexes in the context of the full-length sequence was observed by Taq polymerase stop assays (106,144). The PQS2 segment alone does not appear to readily form a G-quadruplex. However, using an inverted Taq polymerase stop assay, Micheli et al. (144) observed that the PQS2-PQS3 ("PQS23") region could potentially form stacked parallel G-quadruplexes, implying that the inter-quadruplex stacking interface provided a stabilizing effect. This observation is substantiated by the large circular dichroism (CD) signal at 260 nm for the WT PQS2-3 sequence (223). Micheli et al. (144) proposed a model of "self-organization" that featured three contiguous stacked parallel quadruplexes, which stems from stabilization of the PQS2 through terminal G-quadruplex stacking interactions. That model was supported by subsequent biophysical studies combined with molecular dynamics simulations (145) (**Figure 27B**, left model). Alternatively, Palumbo et al. (106) proposed a model based on dimethylsulphate (DMS) foot-printing techniques that featured a parallel PQS1 stacked onto an antiparallel/hybrid G-quadruplex with 8-bp hairpin loop (**Figure 27B**, right model). A later study on the same sequence proposed a different structure (again based on DMS foot-printing) with a longer hairpin joining two parallel G-quadruplexes (224). In both cases (106,224), the CD spectra shown for the folded hTERT sequence lack the signature features expected for structures containing a significant amount of hairpin duplex component. In addition, both proposed hairpin structures contain several thermodynamically destabilizing features including mismatches and bulges.

The structure of the wild-type hTERT promoter sequence thus remains ambiguous. It is important to characterize its structure since it is now considered a target for potential cancer drugs.

For any rational structure-based design of drug candidates, it is essential to know the structure being targeted with certainty. The goal of our study is to clarify the hTERT promoter structure.

A challenge in the determination of the structures of long multimeric quadruplex-forming sequences is that conventional high-resolution NMR or crystallographic methods have yet to be successful, necessitating the use of lower-resolution methods. Herein we report the results of an integrated structural biology (225) investigation of the full-length hTERT promoter sequence using a battery of biophysical and biochemical approaches. In addition, we implemented what can be called a *falsum figura* ("false shape") strategy in which we designed and optimized a non-biological sequence that is forced into the unusual hairpin structure proposed by Palumbo et al. (106). We show that the biophysical and biochemical properties of that structure are unambiguously distinct from the structure formed by the wild-type hTERT sequence, indicating that such a structure is not the predominate folded form of the native sequence. We used classical spectroscopic techniques, hydrodynamic studies, small-angle X-ray scattering, and DNase I digestion as a biochemical probe for duplex DNA to characterize the structures, and we built all-atom models using molecular dynamics simulations to predict testable experimental properties to distinguish structural models. We conclude that the wild-type hTERT promoter sequence forms a compact structure containing three stacked parallel G-quadruplex units and that such a structure is the most appropriate target for any rational drug design effort.

Materials and Methods

Oligonucleotides

Oligonucleotides are given in Table 1. Oligos were purchased from either IDT (Integrated DNA Technologies, Coralville, IA) or Eurofins Genomics (Louisville, KY) with standard desalting unless otherwise specified. Upon receipt, stock oligos were dissolved in MilliQ ultrapure water (18.2 MΩ x cm at 25°C) at concentrations between 0.1 and 1 mM and stored at -20.0°C until use. Folding was achieved by diluting stock oligos into their respective buffer and heating to 99.9°C in a water bath for 20 minutes, followed by slow cooling overnight and subsequent storage at 4°C.

Table 5. Oligonucleotide sequences used in this study.

NAME	OLIGONUCLEOTIDE SEQUENCE 5' TO 3'	LENGTH	MW	E260 ($M^{-1}CM^{-1}$)
WT	GGGGAGGGGCTGGGAGGGCCCGGAGGGGGCTGGGCC GGGGACCCGGGAGGGGTCGGGACGGGGCGGGG	68	21633	672671
WT PQS1	AGGGGAGGGGCTGGGAGGGC	20	6369	202900
WT PQS12	AGGGGAGGGGCTGGGAGGGCCCGGAGGGGGCTGGGC CGGGGACCCGGGA	49	15523	478700
WT PQS23	AGGGGGCTGGGCCGGGGACCCGGGAGGGGTCGGGAC GGGGCGGGG	45	14278	436500
OP	ATGGGTGGGTGGGTGGGCCCTTAGGGTGGGTGGGTCG GGATGGGTGGGTGGGTGGGT	57	18145	553100
AH	TGGGAGGGTCTGGGAGGGCCCTTATGGGTCTGGGCCC GCGACGCGCGAGGCGTCGCGGCTGGGCGGGT	68	21289	628400
AH PQS23	TGGGTCTGGGCCCGCGACGCGCGAGGCGTCGCGGCTG GGCGGGT	44	13721	398100
AH Hairpin	CCCGCGACGCGCGAGGCGTCGCGGCT	26	7975	230296
1XAV	TGAGGGTGGGTAGGGTGGGTAA	22	6992	228700
Hairpin Duplex	GCATATATAGGACCCGCGAGCGGTCCTATATATGC	35	10756	339998

Buffers

All buffer reagents, unless otherwise specified, were purchased from Sigma-Aldrich. TBAP folding buffer (10 mM tetrabutylammonium dihydrogen phosphate, 200 mM KCl, 1 mM EDTA, pH 7.0) was prepared by dissolving 3.4 g of tetrabutylammonium phosphate monobasic, 14.9 g KCl, and 292 mg of acid EDTA in 10 mL of tetrabutylammonium hydroxide 40% solution in 900 mL of MilliQ ultrapure water and adjusted to pH 7.0 before bringing to 1 L (density = 1.0081 g/cm^3, viscosity = 0.01038 poise). Phosphate (PO_4) buffer (8 mM phosphate, 200 mM KCl, pH 7.2) was prepared by dissolving 1.0 g K_2HPO_4, 272 mg KH_2PO_4, and 13.9 g KCl in 900 mL of MilliQ ultrapure water and adjusting pH to 7.2 before bringing to 1 L (calculated density = 1.0081 g/cm^3, calculated viscosity = 0.00996 poise). DNase I reaction buffer (4x) (80 mM Tris, 8 mM $MgCl_2$, 40 mM KCl, pH 7.2) was prepared by dissolving 967 mg Tris base, 76 mg $MgCl_2$, and 298 mg KCl in 90 mL of MilliQ ultrapure water and adjusted to pH 7.2 before bringing to 100 mL. All buffers were filtering through 0.2 µm filters before use.

Preparative Size Exclusion Chromatography (SEC)

Oligonucleotide purification was achieved using SEC as detailed previously (168). Briefly, oligonucleotides were annealed at concentrations of 40-100 µM in their respective buffers, filtered through 0.2 µm filters, and injected onto an equilibrated Superdex 75 16/600 SEC column (GE Healthcare 28-9893-33) using a Waters 600 HPLC system. The flow rate was maintained at 0.5 mL/min and sample fractions were collected every 2 minutes from 100 to 180 minutes run time. The molecular weights of fractionated species were estimated based on a regression analysis of elution time vs. log(MW) of protein standards (Sigma #69385), with elution profiles monitored at 260 nm and 280 nm. Purifications were carried out at room temperature and fractionated samples were stored at 4°C prior to concentration and downstream analysis.

DNase I Degradation Assay

Amplification Grade DNase I was purchased from ThermoFisher and used without further modification (ThermoFisher, #18068015). PAGE purified oligonucleotides for the hTERT WT and

AH sequences were annealed in TBAP folding buffer (without EDTA) before being concentrated to ~50 μM in Pierce protein concentrators (ThermoFisher, #88515). The oligonucleotides were subsequently diluted to 160 ng/μL in TBAP buffer and mixed in a 2:1:1 with DNase I reaction buffer and nuclease free H2O (DNA:RXN-buffer:H2O) to give a final concentration of 80 ng/μL DNA in 10 μL of reaction mix. The reactions were initiated by the addition of 1 μL DNase I (at 1 unit/μL DNase I), incubated at room temperature, and stopped at the indicated time points by the addition of 1 μL of 50 mM EDTA solution. The DNase I cleavage products were then resolved on a 5% agarose gel (~2.5 hours at ~7 V/cm) with visualization by ethidium bromide or SYBR green stain. Gels were imaged using a PharosFX Plus imaging system (BioRad).

Circular Dichroism Spectroscopy (CD)

CD melting studies and spectra were collected on a Jasco-710 spectropolarimeter (Jasco Inc. Eason, MD) equipped with a Peltier thermostat regulated cell holder and magnetic stirrer. CD and melting spectra were collected using the following instrument parameters: 1 cm path length quartz cuvette, 210 or 240 to 340 nm wavelength range, 1.0 nm step size, 200 nm/min scan rate, 1.0 nm bandwidth, 2 second integration time, and 4 scan accumulation. Spectra were recorded at 20.0°C and melting spectra were collected over a range of 4°C to 98°C with 2°C step intervals, 4°C/min ramp speed, and a 2-minute equilibration time at each temperature before acquisition. Spectra were corrected by subtracting a buffer blank. In the case of DNase I degradation assays, the blank included DNase I. Spectra were normalized to molar circular dichroism (Δε) based on DNA strand concentration using the following equation:

$$\Delta\varepsilon = \theta/(32982 \times c \times l)$$

where θ is ellipticity in millidegrees, c is molar DNA concentration in mol/L, and l is the path length of the cell in cm.

For CD monitored DNase I reactions, oligonucleotides were prepared as in the standard DNase I reactions and diluted to a final strand concentration of 3 μM in 500 μL by mixing in the same v/v ratio of DNA, dH_2O, and 4x DNase I RXN buffer. Reactions were initiated by adding 50 μL of DNase I (at 1 unit/μL DNase I) and mixing by pipetting 15 times. Reactions were monitored

over a total of 4 hours in a 0.5 cm path length quartz cuvette. Four hours after DNase I addition, 20 µL of a 100 µM $CaCl_2$ solution was added (for a final concentration of ~3.5 µM Ca^{2+}), and the measurements were resumed to ensure that the DNase I was active.

^{1}H Nuclear Magnetic Resonance (^{1}H-NMR) Spectroscopy

1D ^{1}H-NMR spectroscopy was performed on a Bruker Avance Neo 600-MHz instrument equipped with a nitrogen-cooled Prodigy TCI cryoprobe. Experiments were performed at 25.0 or 40.0ºC using standard 3- or 5-mm NMR tubes. Minimization of water signal was achieved using a water flip-back pulse sequence. For each measurement, 1024 complex points were collected with an acquisition time of 86 ms. Total scans for each spectrum are as follows: WT (4,096), AH (128), OP (128), WT PQS2-3 (128), AH PQS2-3 (4). Samples were prepared by annealing in PO_4 folding buffer and purified by preparative SEC. Fractions were pooled and concentrated using pre-rinsed Pall 3K MWCO concentrators, followed by addition of 5% v/v D_2O. Final concentrations at time of analysis were: WT (150 µM), WT-XL (200 µM), AH (225 µM), optimized parallel (140 µM), WT PQS2-3 (285 µM), and AH PQS2-3 (285 µM). After concentration an aliquot was removed from each sample and analyzed by CD and AUC to ensure that there were no conformational changes due to concentration.

Analytical Ultracentrifugation (AUC)

Sedimentation velocity (SV) experiments were performed in a Beckman Coulter ProteomeLab XL-A analytical ultracentrifuge (Beckman Coulter Inc., Brea, CA) at 20.0°C and 40,000 rpm in standard 2-sector cells using An50Ti or An60Ti rotors. 100 to 150 scans over an 8-hour period were collected and analyzed in Sedfit using the continuous C(s) model. For the hTERT oligonucleotides (“WT” and “WT PQS23”) a concentration series of purified oligonucleotide was used to correct for any non-ideal concentration-dependency in sedimentation. This was done using the following concentrations and respective wavelengths: 2.5 mg/mL (306 nm), 1.25 mg/mL (302 nm), 0.5 mg/mL (298 nm), 0.125 mg/mL (290 nm), 0.05 mg/mL (272 nm), and 0.01 mg/mL (260 nm) at 20.0°C and 40k rpm. Buffer densities and viscosities used in the SV analyses are provided

in the buffers section, and the partial specific volume was held constant at 0.55 mL/g. All large oligo sequences (>~45 nt) regardless of purity at time of purchase had a propensity to aggregate under normal annealing conditions. Thus, all SV experiments were conducted directly after SEC purification, except when correcting for concentration effects. Estimation of weight averaged frictional ratios (f/f_o) of the monomeric WT sequences were carried using Sedfit's C(s,ff_o) model, with a frictional ratio resolution set to 10 and sedimentation coefficient resolution of 100.

SEC Resolved Small-angle X-ray Scattering (SEC-SAXS)

SAXS was performed at BioCAT (beamline 18ID at the Advanced Photon Source, Chicago) with in-line size exclusion chromatography. Samples in a modified PO_4 buffer (8 mM PO_4, 185 mM KCl, 15 mM NaCl, 1 mM EDTA, pH 7.2) were loaded onto a Superdex 75 10/300 GL column, which was run at 0.7 ml/min using an AKTA Pure FPLC (GE Healthcare Life Sciences) and the eluate after it passed through the UV monitor was directed through the SAXS flow cell, which consists of a 1 mm ID quartz capillary with 50 µm walls. A co-flowing buffer sheath was used to separate the sample from the capillary walls, helping prevent radiation damage (172). Scattering intensity was recorded using a Pilatus3 1M (Dectris) detector which was placed 3.5 m from the sample giving access to a q-range of 0.004 Å-1 to 0.4 Å-1. A series of 0.5 second exposures were acquired every 2 seconds during elution and data was reduced using BioXTAS RAW 1.6.3 (173). Buffer blanks were created by averaging regions flanking the elution peak and subtracted from exposures selected from the elution peak to create the I(q) vs. q curves used for subsequent analyses. More information on SAXS data collection, reduction and interpretation can be found in **Table 6**.

Table 6. Tabulated collection parameters, data reduction methods, and data analyses for small-angle X-ray scattering data.

(a) Sample Details.

Sequence name	WT	OP	AH	WT PQS23	AH PQS23
Organism	synthetic	synthetic	synthetic	synthetic	synthetic
Source	IDT	IDT	IDT	IDT	IDT
Extinction coefficient (nearest neighbor approximation) ($M^{-1}cm^{-1}$)	672671	553100	628400	436500	398100
vbar (cm^3/g)	0.55	0.55	0.55	0.55	0.55
M from chemical composition (Da)	21633	18145	21289	14278	13721
SEC-SAXS column, 10 x 300 Superdex 75					
Loading concentration (mg/ml)	4.0	3.7	13.3	5.2	6.5
Injection volume (µL)	300	300	240	250	260
Flow rate (ml/min)	0.75	0.75	0.75	0.75	0.75
Solvent (solvent blanks taken from SEC flow-through prior to elution of protein)	8 mM PO_4, 185 mM KCl, 1 mM EDTA, pH 7.2	8 mM PO_4, 185 mM KCl, 1 mM EDTA, pH 7.2	8 mM PO_4, 185 mM KCl, 1 mM EDTA, pH 7.2	8 mM PO_4, 185 mM KCl, 1 mM EDTA, pH 7.2	8 mM PO_4, 185 mM KCl, 1 mM EDTA, pH 7.2

(b) SAXS data-collection parameters.

Instrument/data processing	BioCAT facility at the Advanced Photon Source beamline 18ID with Pilatus3 1M (Dectris) detector
Wavelength (Å)	1.033
Beam size (µm)	150 (h) x 25 (v)
Camera length (m)	3.5
q measurement range ($Å^{-1}$)	0.004-0.4
Absolute scaling method	N/A
Normalization	To transmitted intensity by beam-stop counter
Monitoring for radiation damage	Automated frame-by-frame comparison of relevant regions
Exposure time, number of exposures	0.5 s exposure time with a 2 s total exposure period (0.5 s on, 1.5 s off) of entire SEC elution
Sample configuration	SEC-SAXS with sheath-flow cell(172), effective path length 0.542 mm. Size based separation by an AKTA Pure with a superdex 75 10/300 GL column
Sample temperature (°C)	23

(c) Software employed for SAXS data reduction, analysis, and interpretation.

SAXS data reduction	Radial averaging; frame comparison, averaging, and subtraction done using BioXTAS RAW 1.6.3(173)
Extinction coefficient estimate	Nearest neighbor approximation
Basic analyses: Guinier, P(r), V_p	Guinier fit, Kratky analysis, and molecular weight using BioXTAS RAW 1.6.3, P(r) function using PRIMUSqt(175)
Shape/bead modelling	DAMMIF(226) via ATSAS online (https://www.embl-hamburg.de/biosaxs/atsas-online/)
Atomic structure modelling	CRYSOL from PRIMUSqt in ATSAS v2.8.4(186)
Three-dimensional graphic model representations	UCSF Chimera v1.11(176)

(d) Structural parameters.

Guinier analysis					
$I(0)$ (cm^{-1})	0.000359 ± 0.0000011	0.000989 ± 0.0000035	0.00395 ± 0.0000045	0.00131 ± 0.0000032	0.00322 ± 0.0000034
R_g (Å)	21.93 ± 0.12	16.46 ± 0.12	24.78 ± 0.052	17.05 ± 0.08	20.38
qmin ($Å^{-1}$)	0.0092	0.0068	0.0043	0.0111	0.0074
qR_g max (qmin = 0.0066 $Å^{-1}$)	1.33	1.3	1.27	1.3	1.27
Coefficient of correlation, R^2	0.981	0.956	0.992	0.983	0.996
M from V_c (ratio to predicted)	23100 (1.07)	15200 (0.84)	23300 (1.09)	16400 (1.15)	16400 (1.20)
P(r) analysis					
$I(0)$ (cm^{-1})	0.00036 ± 0.000001	0.00099 ± 0.0000028	0.004 ± 0.0000041	0.0013 ± 0.0000024	0.0032 ± 0.0000032
R_g (Å)	21.97 ± 0.47	16.51 ± 0.24	24.77 ± 0.09	17.03 ± 0.21	20.36 ± 0.09
D_{max} (Å)	79	53	79	47	66
χ^2 (total estimate from GNOM)	0.79	0.75	0.98	0.85	0.94
Porod volume ($Å^{-3}$) (ratio V_p/calculated M)	27900 (1.29)	17100 (0.94)	28100 (1.32)	17700 (1.24)	16800 (1.22)
(e) Shape model-fitting results					
DAMMIF (default parameters, 15 calculations)					
q range for fitting ($Å^{-1}$)	0.01-0.35	0.006-0.35	0.007-0.32	0.006-0.35	0.006-0.35
Symmetry, anisotropy assumptions	P1, none	P1, none	P1, none	P1, none	P1, none
NSD (standard deviation), No. of clusters	0.88 (0.07), 4	0.51 (0.02), 5	1.18 (0.12), 3	0.82 (0.8), 2	1.00 (0.08), 3
χ^2 range	0.964-0.965	1.153-1.157	1.089-1.091	1.043-1.045	1.509-1.520
Constant adjustment to intensities	Unable to determine	Unable to determine	Unable to determine	Unable to determine	Unable to determine
Resolution (from SASRES) (Å)	26 ± 2	17 ± 2	30 ± 2	23 ± 2	27 ± 2
M estimate as 0.5 x volume of models (Da) (ratio to expected)	18338 (0.85)	11156 (0.61)	18770 (0.88)	11988 (0.84)	12256 (0.89)
(f) SASBDB IDs for data and models.					
ID	SASDHM3	SASDHP3	SASDHN3	SASDHQ3	SASDHR3

Molecular Dynamics simulations

Molecular dynamics simulations were carried out on the hTERT models created previously (145) with modifications of bases where necessary using the “swapna” command in UCSF Chimera 1.12 (176) and manual alterations in starting atomic configurations using Schrödinger’s Maestro 11.8 (https://www.schrodinger.com/). Coordinating counter ions (K^+) were manually added between G-quartet stacks and minimized prior to simulations. The PDB structures were imported into the xleap module of AMBER16 (https://ambermd.org/), neutralized with K^+ ions, and solvated in a rectangular box of TIP3P water molecules with a 15 Å buffer distance. All simulations were equilibrated using sander at 300 K and 1 atm using the following steps: (1) minimization of water and ions with weak restraints of 10.0 kcal/mol/Å on all nucleic acid residues (2000 cycles of minimization, 500 steepest decent before switching to conjugate gradient) and 10.0 Å cutoff, (2) heating from 0 K to 100 K over 20 ps with 50 kcal/mol/Å restraints on all nucleic acid residues, (3) minimization of entire system without restraints (2500 cycles, 1000 steepest decent before switching to conjugate gradient) with 10 Å cutoff, (4) heating from 100 K to 300 K over 20 ps with weak restraints of 10.0 kcal/mol/Å on all nucleic acid residues, and (5) equilibration at 1 atm for 100 ps with weak restraints of 10.0 kcal/mol/Å on nucleic acids. The output from equilibration was then used as the initial.rst input file for 100 ns of unrestrained MD simulations using pmemd with GPU acceleration in the isothermal isobaric ensemble (P = 1 atm, T = 300 K). Periodic boundary conditions and PME were used. 2.0 fs time steps were used with bonds involving hydrogen frozen using SHAKE (ntc = 2). The hairpin structure was manually placed in three different starting configurations and simulated three separate times until convergence was obtained. Trajectories were analyzed using the CPPTRAJ module in the AmberTools16 package (https://ambermd.org/). Hydrodynamic properties were calculated as average and standard deviation over 100 equally spaced trajectory snapshots, unless otherwise specified, using the program HYDROPRO10 (179) with the recommended parameters (180). G-quartet associated potassium ions were included (and added to the molecular weight) in the hydrodynamic calculations. Values are reported as average and standard deviations across 100 evenly spaced snapshots of the trajectories. Clustering of the

hTERT trajectory was performed as described (http://www.amber.utah.edu/AMBER-workshop/London-2015/Cluster/) using the CCPTRAJ module of Amber.

Molecular Visualizations

All molecular visualizations of MD trajectories and models were performed in UCSF Chimera v1.12 (176).

Results

Sequence design and logic

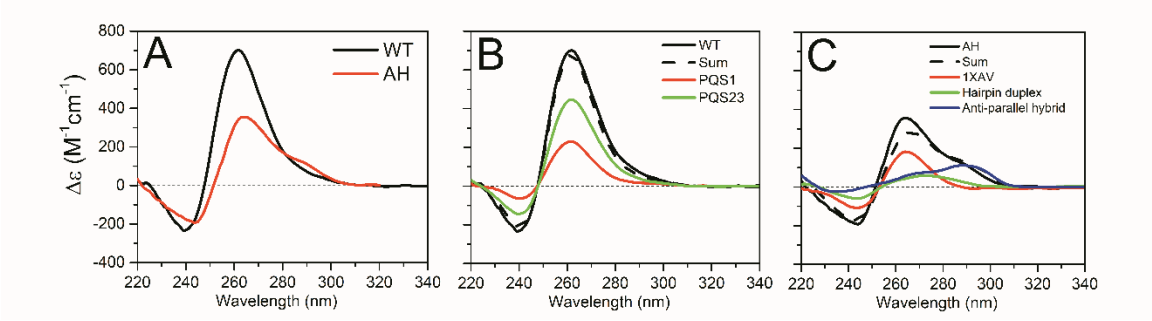
In the 2014 report from Chaires et al. (145) the authors noted that a parallel-antiparallel-hairpin structure, such as proposed by Palumbo et al. (106), should have a CD spectrum distinctly different from that observed for the WT sequence. To validate this assertion, we modified the WT hTERT sequence to generate an optimal sequence that would fold into the hairpin structure proposed by Palumbo et al. (106) to contrast its properties with that of the WT sequence (**Table 5**, AH). The AH sequence was designed to include base substitutions (G>T) to restrict G-quadruplex formation to runs of only 3 guanines and modified residues in the putative hairpin region to maximize hairpin formation and concomitantly disfavor G-quadruplex formation. It is worth noting that a significant number of modifications (15 out of 68 WT bases) were required to stabilize the hairpin structure, as fewer mutations resulted in mixtures of two or more species. We also created an optimized all-parallel structure (**Table 5**, OP) using three runs of a canonical parallel G-quadruplex motif which retains a 6-nucleotide modified non-guanine-containing loop sequence similar to the WT (**Figure 27A**, segment between PQS1 and PQS2). OP was designed to minimize possible G-register exchange (227) and thereby minimize heterogeneity while retaining the three-stack G4 structure. In addition, several truncated sequences were designed to contain isolated structural elements of the longer wild-type form. These represent the first (PQS1), first and second (PQS12), and second and third (PQS23) G4 forming regions of hTERT WT. A high-resolution structure of PQS1 has been reported that can be integrated into our structural models (105). For

the AH sequence analogs, a hairpin-G4 structure (AH-PQS2-3) and an isolated duplex hairpin (AH-hairpin) were designed. **Table 5** contains the complete list of sequences used in this study.

Circular dichroism and DNase I cleavage assays reveal only G-quadruplex moieties within the wild type hTERT core promoter sequence.

We began our structural investigations using CD spectroscopy in potassium buffer. The differences in the spectra of folded WT and AH are unambiguous (**Figure 28A**). Consistent with earlier reports (106,144,145), the strand-normalized hTERT WT spectrum exhibited strong positive molar ellipticity at 260 nm, a trough at 240 nm, and a small trailing shoulder at 290 nm. This large 260 nm amplitude is consistent with a high degree of anti-anti guanine base steps (44,196), typical of parallel quadruplexes, and could only arise from stacking of a large number of G-quartets. Indeed, the CD amplitude at 260 nm is linearly correlated with the number of stacked, parallel, G-quartets (228), so we can estimate from these data that the folded WT sequence contains 9 stacked quartets. In contrast, the AH spectrum has only a modest peak at 260 nm, a trough at 245 nm, and a relatively larger shoulder at 290 nm when compared with the WT. We found that the WT CD spectrum could be reconstructed by addition of spectra obtained with the truncated sequences, PQS1 and PQS23 that are known to form only parallel quadruplex structures (**Figure 28B**, **Table 5**). In contrast, the AH spectrum can only be reconstructed using a combination of parallel G4 [PDB ID: 1XAV (229)], antiparallel-hybrid G4 (196), and duplex hairpin CD spectra (**Figure 28C**). The low amplitude at 260 nm for AH is consistent with the contribution of three stacked G-quartets in the parallel conformation, while the pronounced shoulder near 290 nm arises from a three-quartet antiparallel contribution. The assumed B-form duplex contributes comparatively little to the CD spectra (230).

Figure 28. CD spectra for WT, AH, and their putative component spectra. (A) Strand-normalized CD spectra of the WT and AH sequences annealed in the presence of 200 mM K^+, showing distinct differences in the troughs (~240 nm vs. 245 nm), peak height at 260 nm, and shoulder at 290 nm. (B) The WT sequence can be faithfully reconstructed (dashed line “sum”) from the addition of the PQS1 (red) and PQS23 (green) fragment spectra which adopt parallel topologies in 200 mM K^+ buffer (as shown below). (C) The AH spectrum can be reconstructed (dashed line “sum”) from the addition of the parallel 1XAV (red), a hairpin (green), and an antiparallel G-quadruplex spectrum (blue).


A
WT
AH
Δε (M⁻¹cm⁻¹)
800
600
400
200
0
-200
-400
220 240 260 280 300 320 340
Wavelength (nm)
B
WT
Sum
PQS1
PQS23
220 240 260 280 300 320 340
Wavelength (nm)
C
AH
Sum
1XAV
Hairpin duplex
Anti-parallel hybrid
220 240 260 280 300 320 340
Wavelength (nm)

The Hurley group recently re-investigated using DMS foot-printing the WT hTERT sequence (224) and suggested formation of a different, larger hairpin which putatively forms within the internal PQS2 region, sandwiched between two outer parallel G4s (PQS1 and PQS3). Other DNA conformational forms might contribute to CD spectra. We reasoned that it may be possible for the hairpin to adopt an A-form duplex (44), and that perhaps this might contribute to the CD magnitude at 260 nm, compensating for the magnitude expected for a parallel quadruplex forming within PQS2, complicating interpretation of the CD spectra. As an independent, selective, test for the presence of duplex regions we used an enzymatic assay to probe the structures. Any proposed hairpin structure, whether in B- or A-form, should be susceptible to cleavage by deoxyribonuclease I (DNase I), while parallel G4 structures should remain undigested due to the occluded phosphate backbone (231,232). Upon treatment with DNase I (**Figure 29**) we found that the WT sequence is entirely protected from DNase I cleavage while AH is degraded into discrete components (**Figure 29A & B**). Further, the AH cleavage bands observed after 10 minutes at ~14 and ~8 base pairs are approximately the sizes expected for cleavage of the antiparallel hairpin (**Figure 30**) (233). A scaled-up DNase I cleavage reaction was also monitored using CD spectroscopy, revealing that there was no discernable change in the WT CD spectrum but a significant alteration of the AH CD (**Figure 29C & D**). The shapes of the difference spectra observed in **Figure 29D** for the digestion of AH over the course of the DNase I digestion are consistent with the degradation of a B-form duplex domain within the structure. Overall, these results clearly demonstrate that a DNase I-susceptible hairpin is not detectable in the WT hTERT core promoter sequence.

Figure 29. DNase I cleavage susceptibility assay. (A) Representative agarose gel showing DNase I treatment times (NT = no treatment) in minutes for WT or AH sequences after annealing in TBAP (without EDTA). The gel shows that the WT sequence is not sensitive to nuclease cleavage, whereas the AH sequence is cleaved into discrete bands by 10 minutes treatment time. (B) Densitometry of the AH lanes from A with dashed lines showing the appearance of discrete bands at ~14 bp (or ~28 nt) and ~8 bp (or ~16 nt). (C & D) CD difference plots showing the change in CD signal over the DNase I treatment time course. In both windows the left Y-axis corresponds to the black curves, which is the strand-normalized ellipticity of the WT or AH sequence pre-DNase I treatment. The right Y-axis corresponds to the colored spectra which (from blue to red) indicate the change in CD (difference spectra) from the original spectrum due to DNase I treatment. (C - Inset) Color scale representing time interval from addition of DNaseI (Blue, time = 0) to end of experiment (Red, time = 4 hours). The right Y-axis is scaled such as to emphasize where changes are occurring.

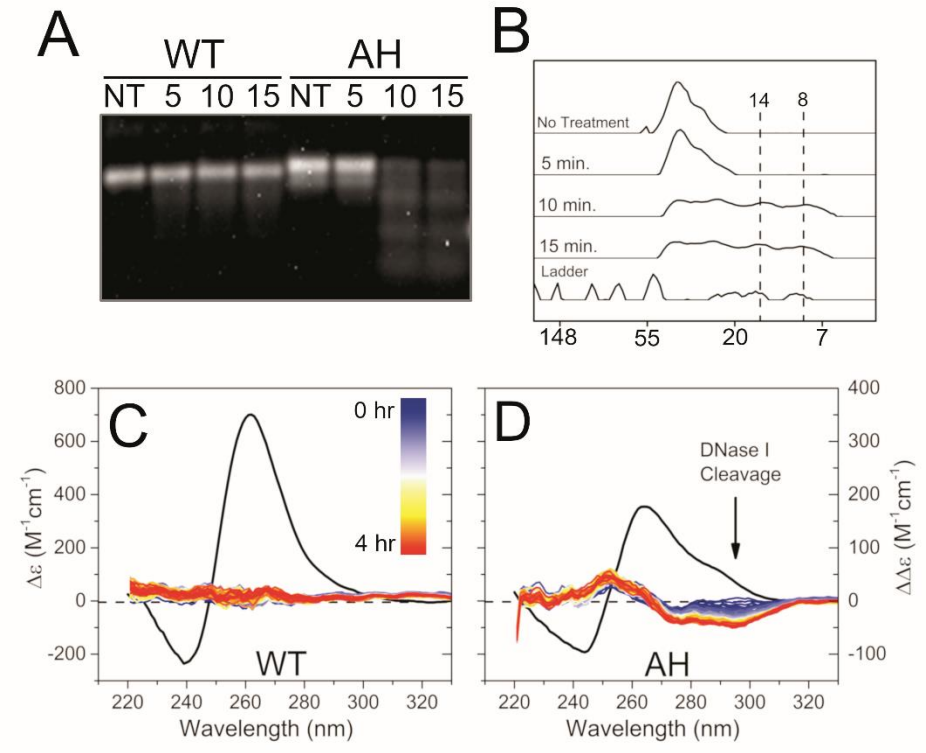
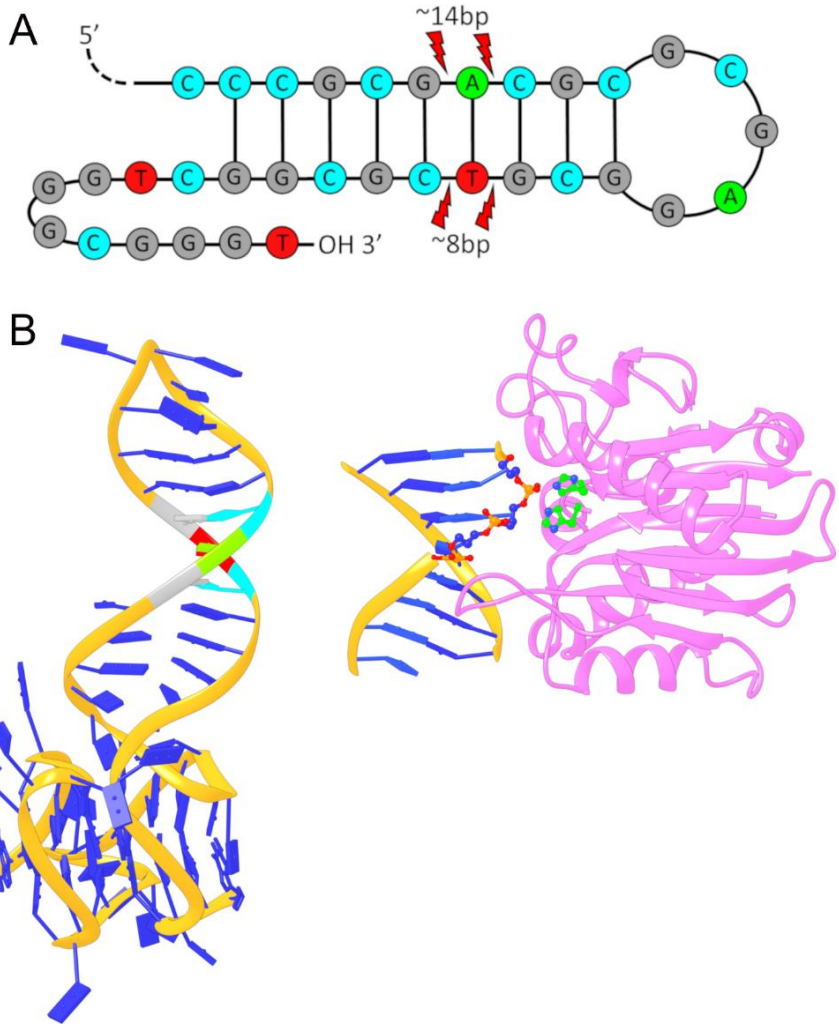
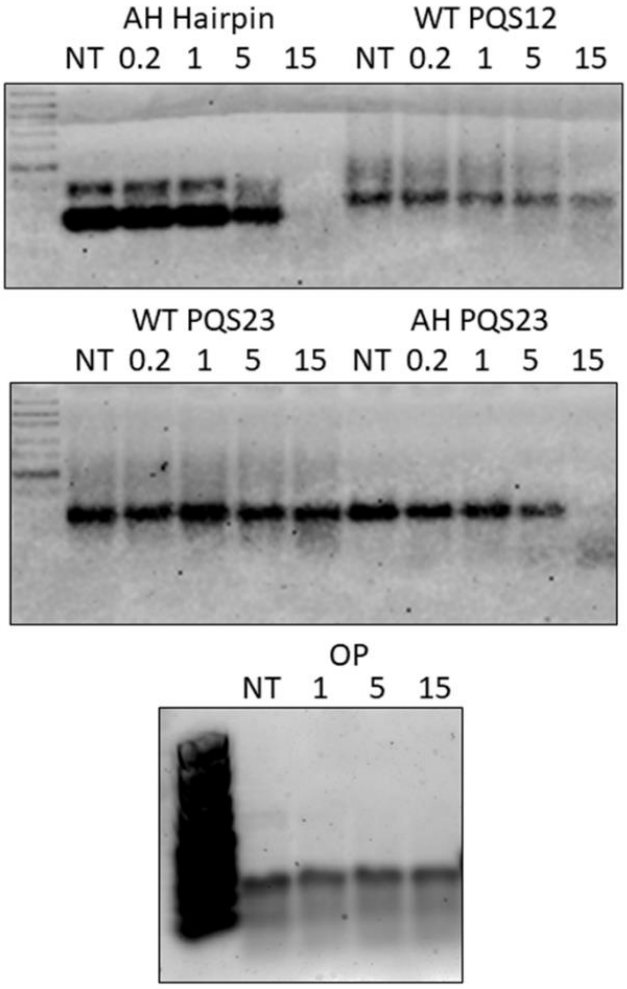

A
WT
AH
NT 5 10 15 NT 5 10 15
B
14 8
No Treatment
5 min.
10 min.
15 min.
Ladder
148 55 20 7
C
0 hr
4 hr
WT
Δε (M-1cm-1)
Wavelength (nm)
D
DNase I
Cleavage
AH
ΔΔε (M-1cm-1)
Wavelength (nm)

Figure 30. Proposed mechanism of hairpin cleavage by DNase I. (A) Schematic representation of the optimized hairpin region of the antiparallel hairpin (AH) sequence, with putative cleavage sites and fragment sizes indicated by red lightning bolts. (B) (Left) Structure of the antiparallel hairpin derived from MD simulations with GAC/CTG cleavage site colored as in A. (B, right) X-Ray crystallography-derived structure of DNase I (orchid ribbon) bound to the d(GGTATACC)2 duplex (PDB ID: 1DNK (234)) with phosphate backbone cleavage site represented as atoms and catalytic histidine residues displayed in green (some ribbon and residues have been omitted for clarity).


A
5′
~14bp
OH 3′
~8bp
B

Treatment of the optimized, G4-stacked, OP structure, or the truncated WT sequence ("WT PQS23") with DNase I had no cleavage effect, consistent with the absence of any duplex structure. A slight degradation over time was observed in the WT PQS12. In contrast, the AH truncated sequence ("AH PQS23") and the control AH hairpin sequence alone ("HP") were completely degraded by 15 minutes (**Figure 31**).

Figure 31. DNase I protection assays of truncated sequences. Agarose gel separation of DNase I treated nucleotide sequences with oligo name and time (in minutes) of DNase I treatment indicated above each lane. The ladder is an ultra-low range DNA ladder that spans from 300 to 10 bp.


AH Hairpin
NT 0.2 1 5 15
WT PQS12
NT 0.2 1 5 15
WT PQS23
NT 0.2 1 5 15
AH PQS23
NT 0.2 1 5 15
OP
NT 1 5 15

^{1}H-NMR confirms that the WT hTERT sequence forms only parallel quadruplexes.

We next analyzed the WT and modified sequences at the atomic level using ^{1}H-NMR to gain a better understanding of the secondary structures giving rise to the observed CD spectra (**Figure 28**). **Figure 32A** shows the Hoogsteen and Watson-Crick imino proton region of the 1H-NMR spectra (~10-14.6 ppm) for the WT, AH, OP, and truncated PQS23 sequences (**Figure 32A & C**), along with their respective CD spectra (**Figure 32B & D**). In the WT spectrum (**Figure 32A**, black) we observed a very broad envelope encompassing the guanine Hoogsteen imino protons between ~10 and 12 ppm, indicative of G-quadruplex formation, along with a slight, almost negligible, signal at ~12.9 ppm. In contrast, we find that the spectrum of AH exhibits the expected Watson Crick-like (W.C.) base-pairing interactions for a hairpin in the region from 12.6-13.6 ppm, along with a broad envelope in the Hoogsteen G-quadruplex region from 10.8-12 ppm, confirming the presence of both duplex and G-quadruplex structures. The contrasting behavior of WT and AH show that there is no appreciable duplex base pairing in the unmodified hTERT sequence. Additional sequences were also studied by NMR. The OP sequence displays only G-quadruplex imino proton shifts in the range of 10-12 ppm, with no signals within the W.C. range. The OP sequence was created as a reference “idealized” three-parallel G-quadruplex system with minimal loops connecting the stacked G4 units. **Figure 32B** shows that the WT and OP sequences exhibit nearly identical magnitude and shape from 220 to 270 nm by CD, the only difference being the shoulder at ~290 nm in the WT spectrum. Importantly, we find that the strand-normalized integrated intensity in the Hoogsteen imino proton signals for the OP and WT sequences is 1.1:1 (**Figure 33**).

The truncated WT PQS2-3 and AH PQS23 sequences also show a clear difference by NMR and CD (**Figure 32C & D**). With the removal of PQS1 (and ~6 nt connecting loop region), we find that the AH PQS23 imino proton spectra resolves ~12 G-quadruplex peaks, the number expected for a single three-tetrad G-quadruplex, with the same number W.C. imino peaks as the full-length AH construct. This is in clear contrast to the WT PQS23 segment, which displays almost twice as many G-quadruplex imino peaks by integration (1.8:1). Taken with their respective CD spectra, this clearly demonstrates that the full-length and truncated WT sequences preferentially form only parallel G-quadruplexes.

Figure 32. 1H-NMR spectra and corresponding CD spectra of WT, AH, OP, and truncated sequences. (A & C) Proton imino spectra from 10 to 14.6 ppm showing Watson-Crick and Hoogsteen type imino proton shifts for the OP, AH, WT, and truncated sequences WT PQS23 and AH PQS23. Intensities in A are only approximate, whereas concentrations and intensities in C are the same. (B & D) Strand-normalized CD spectra corresponding to spectra in A and C, respectively.

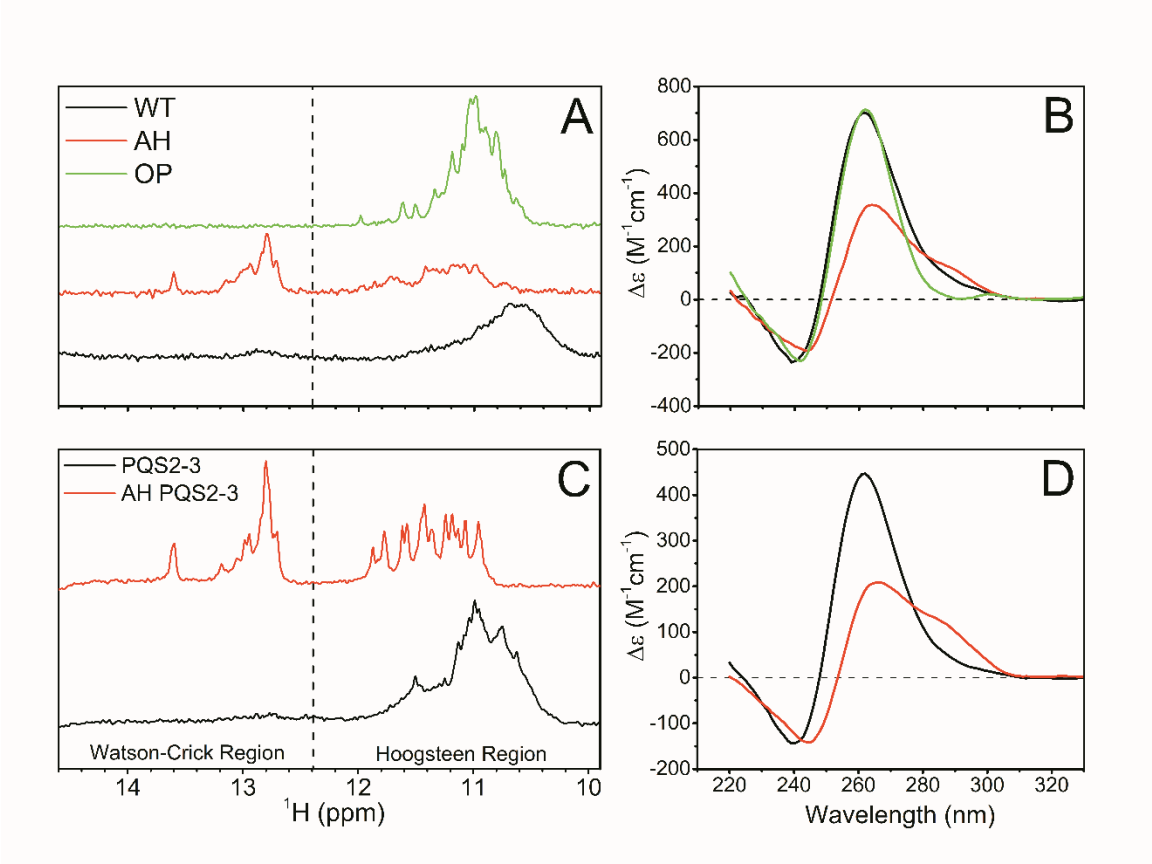
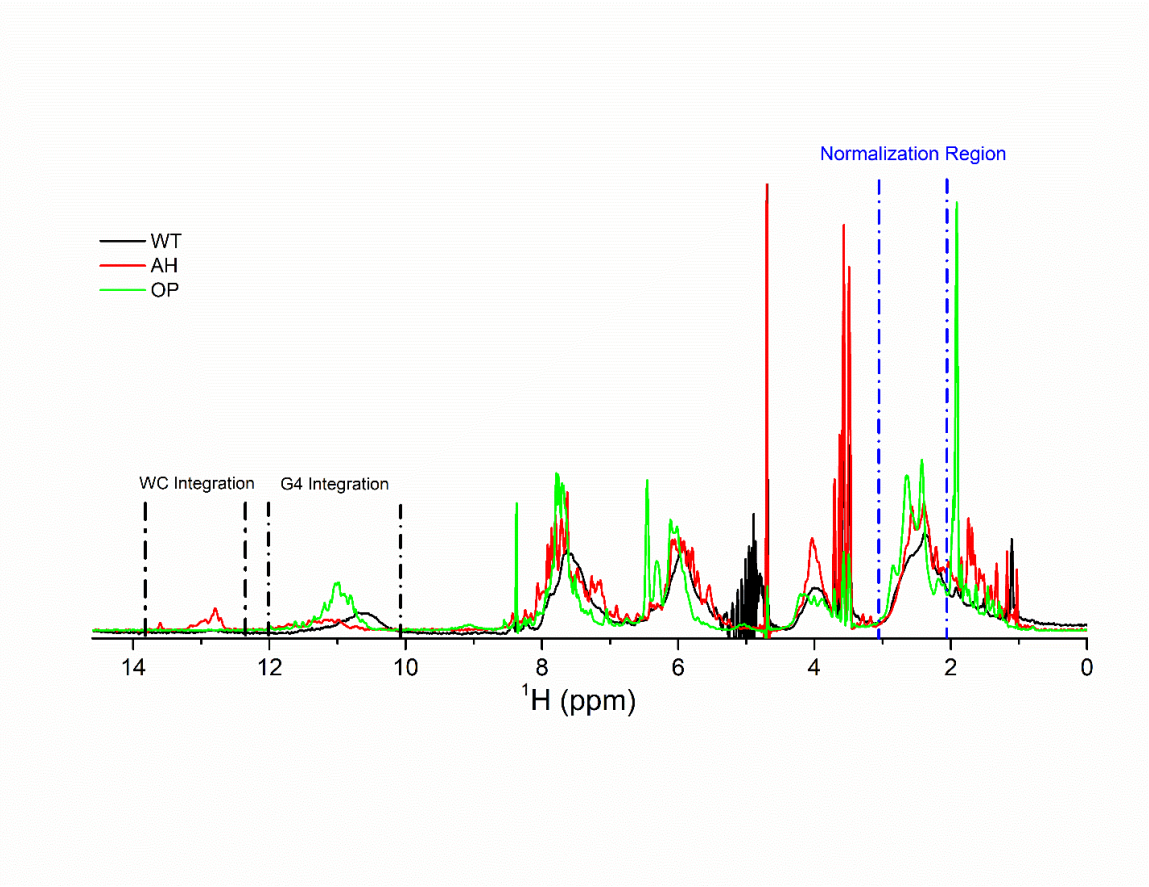

WT
AH
OP
A
B
800
600
400
200
0
-200
-400
Δε (M-1cm-1)
PQS2-3
AH PQS2-3
C
D
500
400
300
200
100
0
-100
-200
Watson-Crick Region
Hoogsteen Region
14
13
12
11
10
1H (ppm)
220
240
260
280
300
320
Wavelength (nm)

Figure 33. Full 1H-NMR spectra comparing the WT, AH, and OP and regions for scaling. Scaling of the Hoogsteen imino region peak intensities are approximate. The blue and black dashed regions represent the areas used for normalizing W.C. or Hoogsteen imino proton peaks to strand concentration.


Normalization Region
WT
AH
OP
WC Integration
G4 Integration
14
12
10
8
6
4
2
0
^{1}H (ppm)

Hydrodynamic size and shape of the WT sequence indicates a very compact and globular structure.

Hydrodynamic experimentation and modeling methods (235) have gone hand-in-hand in structural biology for over 75 years (236). It is now routine in our laboratory to use experimental hydrodynamic properties of nucleic acids to infer, and iteratively refine, their structural models using molecular dynamics modeling simulations (78,163,237). These techniques have allowed for the study of a wide array of DNA conformations in their native state and under biologically-relevant conditions (238). We used this powerful approach to examine the annealed and SEC-purified WT, AH and truncated hTERT promoter sequences by hydrodynamic, X-ray scattering, molecular dynamics and bead modeling methods.

To discern differences in the overall hydrodynamic shapes formed by the sequences listed in **Table 5** we employed analytical ultracentrifugation sedimentation velocity (AUC-SV) experiments to determine sedimentation coefficients from which molecular weights and frictional coefficients may be easily calculated. The results are tabulated in **Table 7**. **Figure 34** shows the significant difference in sedimentation coefficient distributions for the folded WT and AH sequences by C(s) species analysis (170). There is very little overlap between the WT and AH c(s) distributions, indicating different hydrodynamic shapes. The corrected $S_{20,w}$ values for WT and AH sequences were 3.86 ± 0.01 and 3.25 ± 0.09, respectively (**Table 7**)—the former WT value being consistent with our earlier report (145). (We note that the C(s) distribution of the WT shows a very slight heterogeneity with a few percent of the sample sedimenting between 2-3 S.)

Figure 34. Experimental and calculated sedimentation coefficient distributions. The red and black curves are representative SEDFIT C(s) distributions of the WT (black) and AH (red) sequences corrected to density and viscosity of water at 20.0°C ($S_{20,w}$). The overlaid histograms are of sedimentation coefficient values ($S_{20,w}$) obtained from hydrodynamic calculations of PDB frames extracted from 100 ns of explicit solvent MD trajectories for the stacked parallel model (black) or the hairpin model (red).

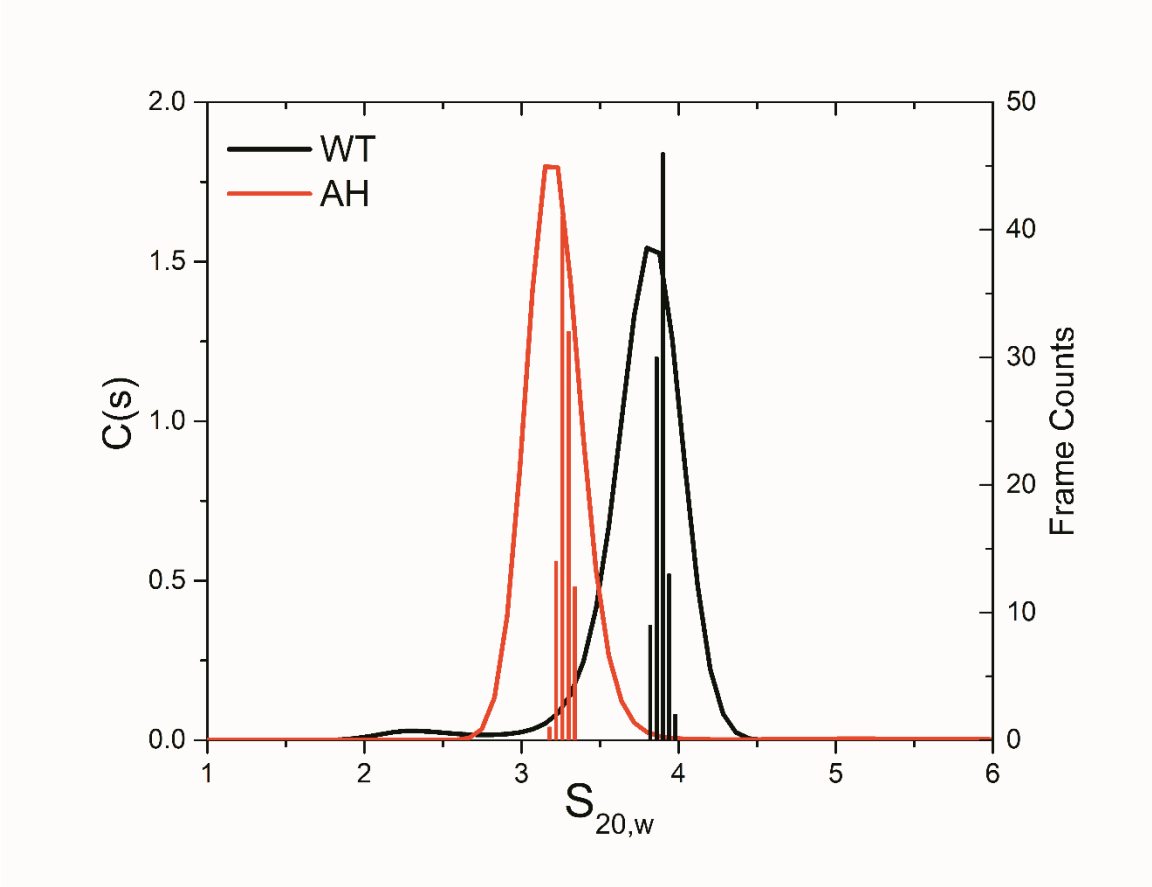

WT
AH
C(s)
Frame Counts
$S_{20,w}$
2.0
1.5
1.0
0.5
0.0
50
40
30
20
10
0
1
2
3
4
5
6

Table 7. Comparison of hydrodynamic properties measured by AUC-SV experiments with values calculated from molecular dynamics trajectories of given models. The table is organized such that the models with the best agreement from calculations are nearest their respective experimental values.

	WT		WT PQS2-3		AH		AH PQS2-3	
Property	Exp.	Parallel Stacked (calc.)	Exp.	Parallel Stacked (calc.)	Exp.	Hairpin (calc.)	Exp.	Hairpin (calc.)
Sedimentation Coefficient, $S_{20,w}$ ($x10^{-13}$ S)	3.86 (±0.01)	3.89 (±0.03)	2.95 (±0.03)	2.96 (±0.05)	3.25 (±0.09)	3.30 (±0.05)	2.70 (±0.01)	2.64 (±0.03)
Molecular Weight (kDa)	24.0 (±0.13)	22.0	16.5 (±0.15)	14.5	24.4 (±1.44)	21.5	16.8 (±0.68)	14.0
Stokes Radius, R_s (nm)	2.46 (±0.01)	2.22 (±0.02)	2.22 (±0.04)	1.96 (±0.03)	2.56 (±0.07)	2.58 (±0.03)	2.49 (±0.07)	2.14 (±0.02)
Frictional Ratio, f/f_o	1.42	1.26	1.45	1.26	1.7	1.54	1.6	1.45

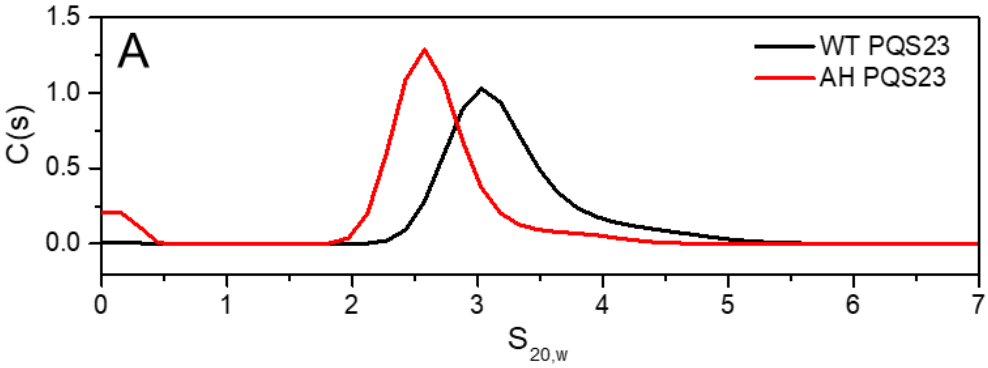
A similar trend was also observed for the truncated sequences WT PQS2-3 and AH PQS2-3 ($S_{20,w}$ values of 3.0 and 2.7, respectively) (**Figure 35**). Since the molecular weights of the pairs of sequences in question are approximately the same (given the limit in experimental accuracy), the measured sedimentation coefficients directly report the particle frictional coefficients that reflect differences in shape. These data indicate that all WT sequences are much more compact than their AH counterparts.

To rationalize these differences in size and shape, as well as the secondary structure derived from CD and NMR studies, we used molecular dynamics simulations and hydrodynamic calculations to refine the most plausible structural models (237). These models were constructed based on the proposed secondary structures (**Figure 27**) in conjunction with coordinates from the protein databank (PDB). The WT all-parallel stacked and hairpin models were used from previous work (145), with the latter modified to reflect the AH sequence with optimized duplex base pairing given in **Table 5**. Truncated models were created simply by removal of the PQS1 and 6 nt loop region. Each model was subjected to 100 ns of unrestrained, fully solvated molecular dynamics simulation. The resulting MD trajectories were then used to calculate the hydrodynamic properties of each system (**Table 7**) using HYDROPRO10 (179), which calculated the sedimentation and diffusion coefficients. These experimentally accessible measures were extracted from PDB coordinates of structures in evenly spaced frames throughout the trajectories to obtain statistically meaningful ensemble values (visualized as the histogram in **Figure 34** and averages in **Table 7**).

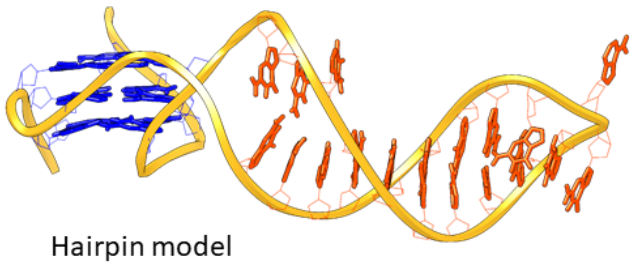
We found that the sedimentation coefficient for the all-parallel stacked molecular dynamics-derived model agreed extremely well with the WT experimental value (calculated $S_{20,w}$ = 3.89 ± 0.01 vs. experimental $S_{20,w}$ = 3.86 ± 0.03). The AH hairpin model also agreed well with the AH experimental sedimentation coefficient (calculated $S_{20,w}$ = 3.30 ± 0.05 vs. experimental $S_{20,w}$ = 3.25 ± 0.09) (**Figure 34**, **Table 7**). The agreement between WT PQS23 with a two-parallel stacked model was even closer (calculated: 2.96 ± 0.05 vs. experimental: 2.95 ± 0.03). The AH-PQS23 model was also consistent with experimental data (calculated $S_{20,w}$ = 2.64 ± 0.03 vs. experimental $S_{20,w}$ = 2.70 ± 0.01) (**Table 7**, **Figure 35**). The calculated and experimental $S_{20,w}$ values were effectively within experimental error in all cases. These hydrodynamic studies show that the folded hTERT WT

promoter is too compact to contain an extended 8 bp hairpin. In addition, it must contain fully stacked G4 units, since the more extended models with a displaced terminal G4 unit predict a sedimentation coefficient distinctly different from the observed experimental value.

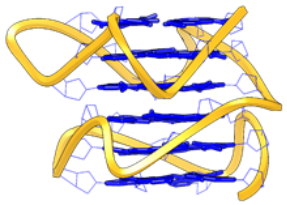
Figure 35. AUC-SV analysis and models of the WT and AH truncated oligonucleotides. (A) SEDFIT C(s) distributions for the WT (black) and AH (red) PQS23 oligonucleotides in potassium buffer showing a distinct difference in sedimentation coefficients. (B) Molecular dynamics-derived models of the hairpin (left) and parallel stacked (right) PQS23 oligonucleotide with extraneous loop bases removed for clarity. Yellow represents the sugar phosphate backbone, blue represents guanines involved in G-tetrad formation, and orange indicates nucleotides involved in hairpin formation.



B



Hairpin model



Parallel Stacked model

Small-angle X-ray scattering (SAXS)

Figure 36 shows the results of SEC-SAXS experiments obtained with folded hTERT constructs. **Table 6** provides complete details of SAXS experiments, with all information suggested by recent publication guidelines (174). Several general qualitative conclusions can be drawn by inspection of these plots (192,195). **Figure 36A-B** show the primary SAXS data. The double logarithmic plots of the scattering intensity for the WT, OP and AH structures (**Figure 36B**), revealing distinct differences between the structures. It is clear that scattering of WT is different from AH. The use of in-line size exclusion chromatography ensured the absence of contaminating species and sample monodispersity, which was also demonstrated by the Guinier plots shown in **Figure 36C**, which are linear for all samples (residuals plot for linear fits for all samples are shown to the right of panel C). Parameter estimates for the radii of gyration (R_g) obtained by analysis of Guinier plots are given in **Table 6**.

The differences between the WT, AH and OP are further illustrated by the pair-distance distribution (**Figure 36D**). The hairpin structure is inconsistent with the observed scattering of the folded wild-type hTERT sequence. For a homogeneous structure, the exact character of the P(r) plot depends on particle shape (e.g. globular, prolate, oblate) and the domain structure of the particle (239). The shape of the AH P(r) curve (**Figure 36D**, red), with a pronounced multimodal character, suggests the presence of multiple domains within the structure, consistent with the presence of an extend hairpin coupled to G4 units. In contrast, the WT P(r) curve (**Figure 36D**, black) is more symmetrical, indicative of a more compact structure. OP was designed to optimize the stacked G4 structure. Accordingly, the P(r) curve for OP (**Figure 36D**, green) is more symmetric (but still with a trailing edge at larger distances), consistent with a compact three-stacked G4 structure with an elongated shape. All P(r) plots yielded radii of gyration which were within 0.1 of those derived from Guinier approximation (**Table 6**).

Kratky plots (**Figure 36E**) for WT, AH and OP provide a qualitative appraisal of the degree of unfolding and the flexibility of samples (240). Compact, fully folded particles are expected to exhibit a Gaussian shaped curve, while unfolded or flexible particles would show nonzero plateau regions at high q. The major observation from the data in **Figure 36E** is that WT and OP (**Figure**

36E black and green, respectively) are clearly distinguishable from AH (**Figure 36E,** red line). A hairpin structure is inconsistent with their scattering data. WT and OP curves reach zero at high q, with nearly Gaussian shapes, indicating that they are fully folded and globular. In contrast, AH shows a more complex curve. While it seems to be nearly fully folded as judged by its intercept on the x-axis, the data for AH show a distinct plateau in the q range spanning 0.05-0.15 that indicates particle flexibility. We attribute this flexibility to the region linking the hairpin to the G-quadruplex domain, which was observed in our MD simulations (not shown). These data demonstrate that the WT sequence folds into a distinctly different compact globular structure. **Figure 37** shows the scattering behavior of the partial hTERT constructs WT-PQS23 and AH-PQS23. The general behavior and trends are similar to what was seen for the full-length constructs. Importantly, the more asymmetric hairpin-G4 structures can be clearly distinguished from compact, globular G4 structures.

Figure 36. SEC-SAXS results for the WT (black), AH (red), and OP (green) oligonucleotides. (A) I(q) versus q as log-linear and (B) log-log plots. (C) Guinier plots (with fits shown in blue) for $qR_g < 1.3$, along with corresponding residual plots (right). (D) P(r) versus r profiles from the data in (a and b) normalized to equal areas. (E) Normalized Kratky plots for the data in (a and b). Collection parameters, I(0), R_g, D_{max}, and other values can be found in **Table 6**.

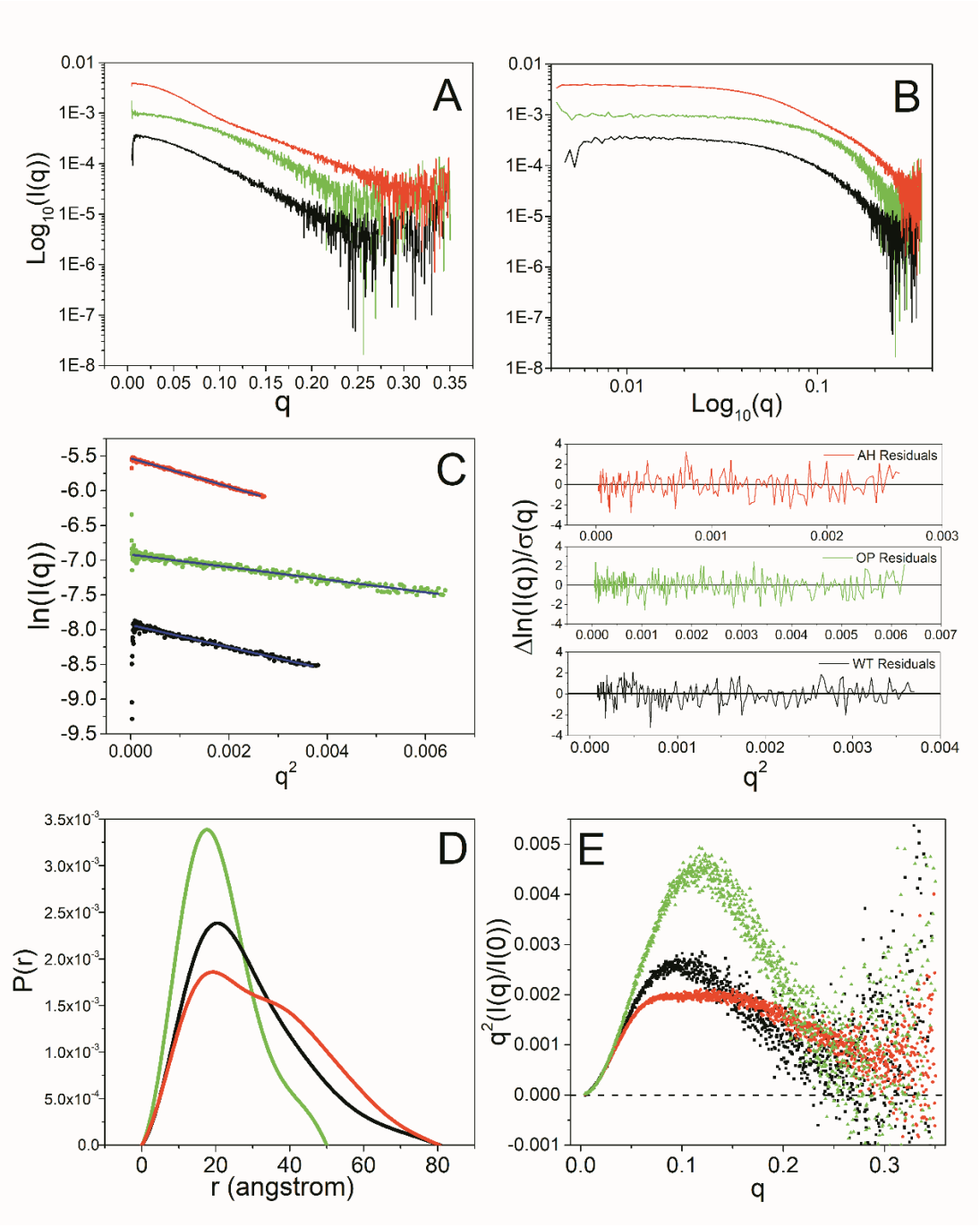
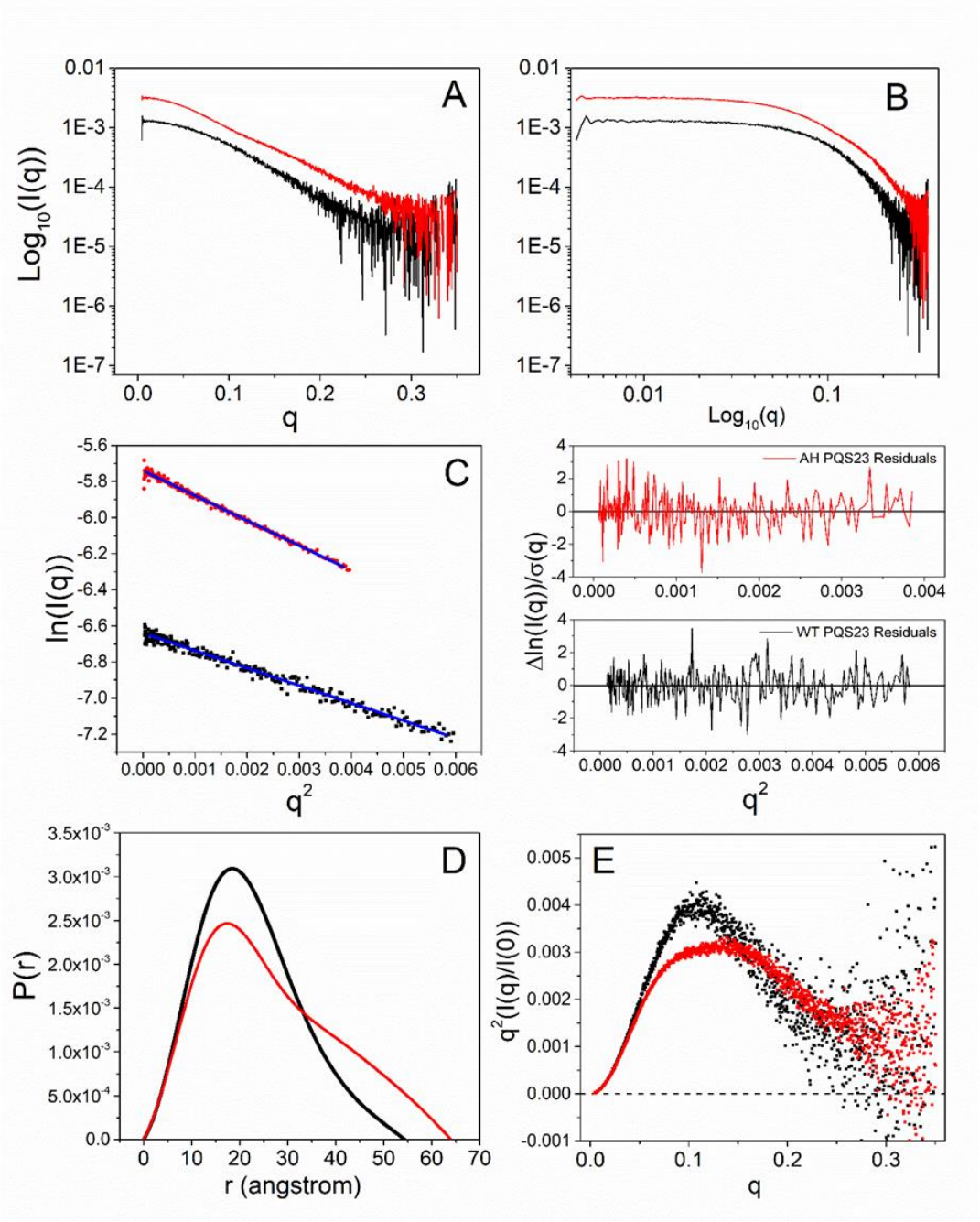
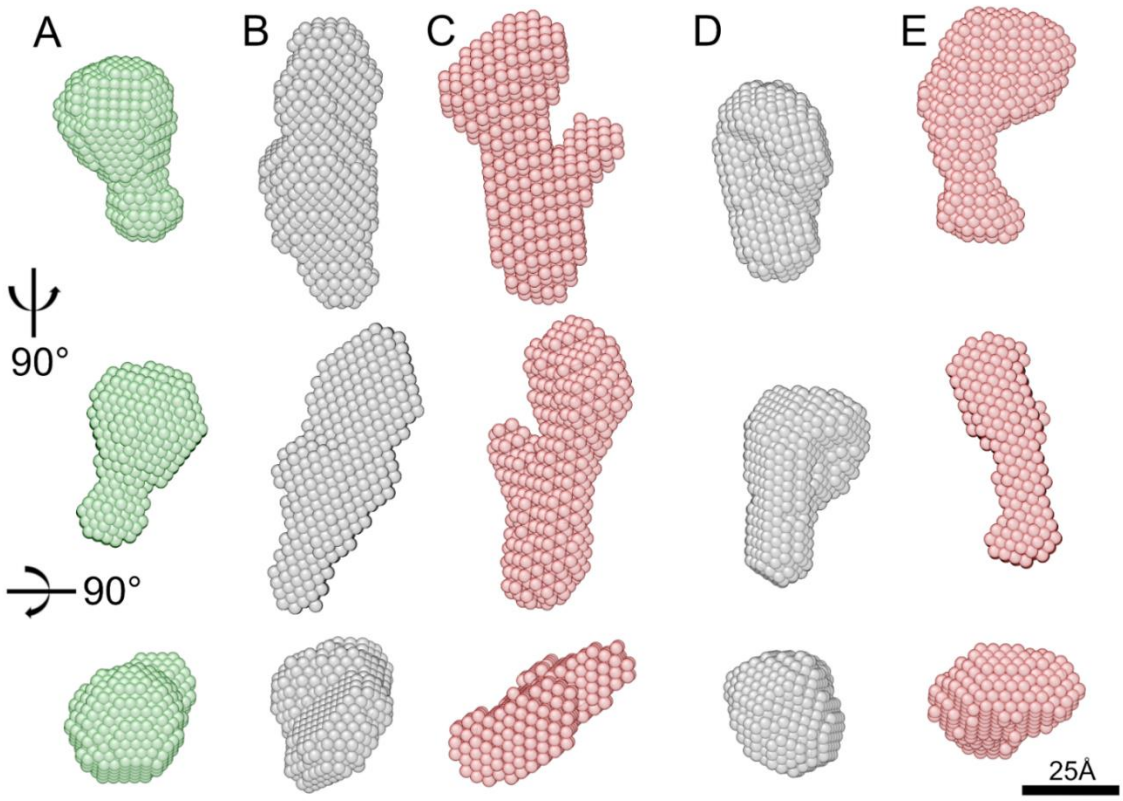

A
B
C
D
E
Log10(I(q))
q
Log10(q)
ln(I(q))
q2
ΔIn(I(q))/σ(q)
AH Residuals
OP Residuals
WT Residuals
P(r)
r (angstrom)
q2(I(q)/I(0))
q

Figure 37. SEC-SAXS results for the WT (black) and AH (red) PQS23 truncated oligonucleotides. (A) I(q) versus q as log-linear and (B) log-log plots. (C) Guinier plots (with fits shown in blue) for $qR_g < 1.3$ along with corresponding residual plots (right). (D) P(r) versus r profiles from the data in (a and b) normalized to equal areas. (E) Normalized Kratky plots for the data in A and B.


A
B
C
D
E
Log10(I(q))
q
Log10(q)
ln(I(q))
q2
AH PQS23 Residuals
WT PQS23 Residuals
ΔIn(I(q))/σ(q)
P(r)
r (angstrom)
q2(I(q)/I(0))

Ab initio bead models for WT, OP and AH were obtained using the DAMMIF program (226), with the results shown in **Figure 38**. The key point from inspection of these shapes is that both hairpin structures, AH and AH PQS23 (**Figures 38C & 38E**), feature clear protuberances that are absent from all other structures. These protuberances are most likely the hairpin duplex domain. We are aware of the utility of SAXS data in more detailed atomistic structural modeling of macromolecules with conformational heterogeneity (164,183,184). Efforts in that direction are currently underway in our laboratory.

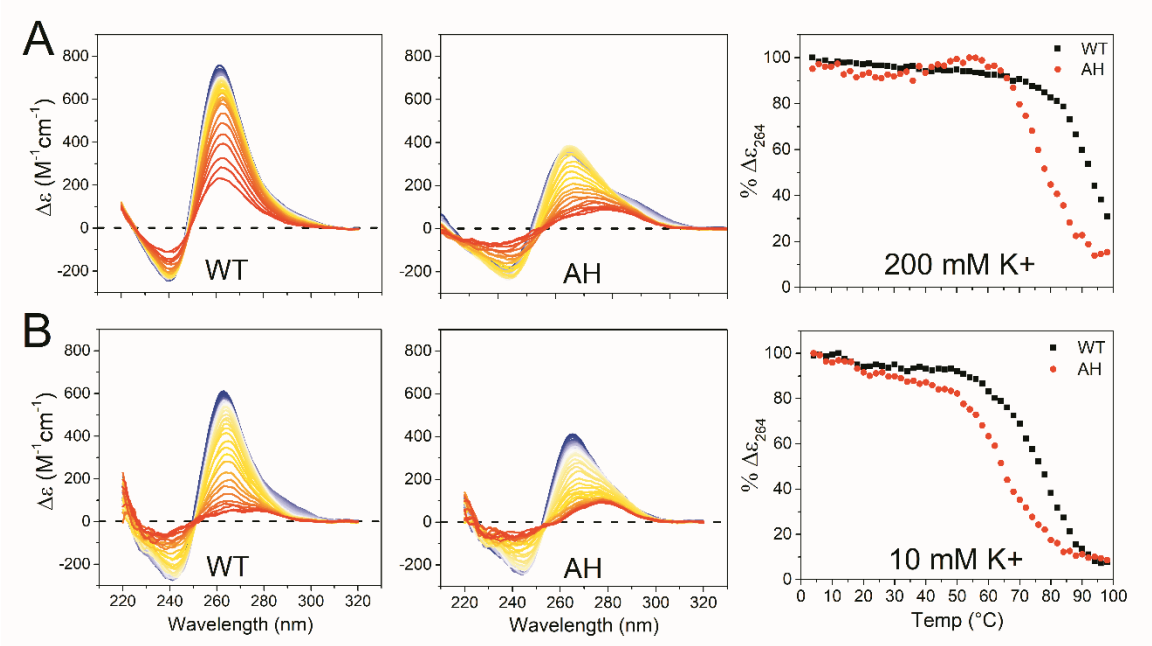
Figure 38. *Ab initio* bead model results for the WT, OP, AH, and truncated oligonucleotides. Averaged and filtered DAMMIF bead models for the OP (A), WT (B), AH (C), WT PQS23 (D), and AH PQS23 (E) oligonucleotides. All models are displayed at the same scale.


A
B
C
D
E
90°
90°
25Å

Thermal denaturation of hTERT structures

Figure 39 shows thermal denaturation studies, monitored by CD, for WT and AH structures. There are clear differences between the two structures, with hairpin-containing AH noticeably less stable. In 200 mM KCl (**Figure 39A**, right most panel), melting of WT is incomplete even at 98°C, while AH is clearly less thermally stable with a T_m near 79°C. By lowering the KCl concentration to 10 mM, complete thermal denaturation curves for both structures were obtained (**Figure 39B**). The right most panel shows that AH is thermodynamically less stable than WT. The apparent T_m values for WT and AH are 82.5 and 65.2°C, respectively, in 10 mM KCl. The 17.3 °C difference in T_m shows unambiguously that the hairpin-containing structure is thermodynamically less stable. A more detailed thermodynamic analysis of the thermal stabilities of WT and AH using CD and differential scanning calorimetry will be the subject of a manuscript that is in preparation.

Figure 39. CD thermal denaturation profiles of the hTERT WT and antiparallel hairpin sequences. (A) Strand-normalized CD spectra from 220 to 320 nm over 4 to 98 °C for the WT and AH sequences annealed in potassium phosphate buffer with 200 mM KCl (left) with normalized melting curves (right). (B) Strand-normalized CD spectra (left) and melting curves (right) for WT and AH sequences annealed in phosphate buffer with 10 mM KCl.

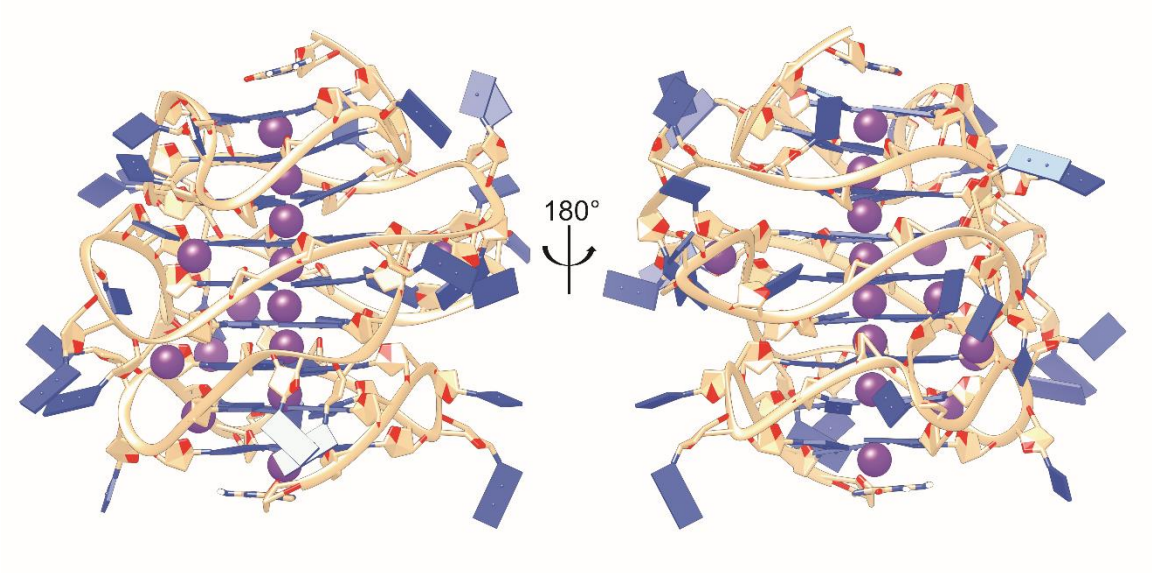

A
B
Δε (M-1cm-1)
WT
AH
% Δε264
200 mM K+
10 mM K+
Wavelength (nm)
Temp (°C)

Discussion

The results of our integrated structural biology approach show conclusively that the 68-nt wild-type hTERT promoter DNA sequence folds into a compact G-quadruplex structure that lacks any hairpin duplex domain. A structure with three stacked G4 units in the parallel conformation is both qualitatively and quantitatively consistent with our biophysical and biochemical data (circular dichroism, thermal denaturation, analytical ultracentrifugation, SEC-small angle X-ray scattering, nuclear magnetic resonance, DNase I cleavage assays, and molecular dynamics). An optimized hairpin-containing structure based on the model proposed by Palumbo et al. (106) shows unambiguously different biophysical and biochemical properties from the wild-type sequence.

Figure 40 shows a detailed model of the hTERT promoter structure obtained by our molecular dynamics simulations. This model integrates the high-resolution structure of the PQS1 region that was obtained by NMR by Phan and coworkers (105). The model features nine stacked G-quartets (arising from three stacked parallel G4 units), consistent with the large experimental CD amplitude observed in **Figure 28**. This model accurately predicts the experimentally observed WT hTERT sedimentation coefficient (**Figure 34**) and is qualitatively consistent with the SAXS data in **Figure 36**. Several features of the structure are of interest. Apart from the G4 stacking interactions, the structure shows several stabilizing loop-loop interactions, which are consistent with slight contributions to the WT ^{1}H-NMR of base-base interactions (small peak at ~12.9 ppm) and the lack of these in the designed OP ^{1}H-NMR (**Figure 32**). In addition to the K^{+} ion coordination sites within G-quartets, the loop topology presents additional specific K^{+} binding sites. This structure presents unique groove and interfacial geometries for small molecule binding interactions. Overall, the model in **Figure 40** represents an excellent target structure for rational drug discovery and development.

Figure 40. MD-derived model of the three-stacked parallel hTERT G-quadruplex. All-atom model of the stacked hTERT system (5’ top, 3’ bottom) showing phosphate backbone in tan, nucleotides in blue and potassium atoms in purple. This structure was derived from clustering over 100 ns of explicit solvent MD. The potassium ions observed in the central tetrad cavity, loops and grooves were observed in 66% of all frames used in clustering.


180°

One of the major difficulties encountered in studying G-rich DNA sequences in their biological context is in dealing with G-tracts with greater than three guanines or with numbers of G-tract greater than four. These sequences are known for forming multiple isoforms through G-register exchange and changes in loop directionality, and often require base modifications such as inosine or thymine substitutions to select for only one structure (42). This phenomenon was observed in the hTERT PQS1 sequence (105,241). The physical consequence of G-register exchange, as shown by the Mittermaier lab (227), is an entropic stabilization of the folded state (albeit an ensemble of folded states). In addition to the five runs of three guanines, the full-length hTERT core promoter sequence has six runs of four guanines (G_4) and one run of five guanines (G_5) (**Figure 27**). This equates to a theoretical 192 isomers ([6 x G_4] x [1 x G_5] which is 26 x 31) when in the all-parallel stacked conformation, whereas this number decreases substantially to only ~48 if in the hairpin conformation proposed by Palumbo et al. (106,227). This implies that the all-parallel stacked WT sequence would have an inherent entropic advantage over the hairpin structure, as well as increased thermodynamic stability, which is reflected in **Figure 39**.

Such conformational heterogeneity is also evident in structural characterizations. While the heterogeneity complicates interpretation of biophysical data, it in fact represents the reality of wild-type sequences whose complexities must be considered instead of being expeditiously simplified by arbitrary sequence modifications. In **Figures 32** and **33** we observed a clear broadening of all peaks in ^{1}H-NMR measurements of the WT sequence, suggesting the presence of parallel G-register isomers. This broadening was not exhibited by OP, which is by design a single, all-parallel stacked conformer. Moreover, we observed minor amounts of hTERT species with differing sedimentation coefficients (**Figure 34**, shoulder at ~2.4 S). A possible explanation for this discrepancy is that there is a dynamic equilibrium between parallel stacked and unstacked structures. Alternatively, this could be attributed to a mixture of slow and fast rearrangements (such as G-register sliding or folding), which depending on the timescales, could easily complicate analyses (240,242). Overall, however, the major form appears consistent with the stacked conformation as in **Figure 40**.

G4 stacking and multimerization is now a well-known phenomenon (20). Various G-quadruplex stacking interfaces have been characterized (243-245), and the physical forces involved investigated (54). All of these studies support a 5'-3' (head to tail) stacking mode, consistent with our model. However, our understanding of the biological relevance of G4 stacking is lacking. In promoters, stacked G4 structures are now speculated to be involved in a variety of roles, primarily as unique recognition sites for proteins, enhanced sensing of ligands through cooperativity, or as concentration-dependent G4 biological switches (20). In line with this is the unique opportunity of selective gene expression modulation via small molecules which stabilize or disrupt these stacking interfaces.

We have shown that these higher-order structures can be successfully examined using an integrated structural biology approach, coupled with judicious sequence design to create additional test structures with contrasting or confirmatory features. There are thousands of sequences in promoter regions in the human genome that have greater than four runs of multi-guanine tracks, yet almost all remain uncharacterized. An in-depth examination and understanding of these potential multimeric quadruplex structures will lead to the identification of unique binding sites and a potentially more selective way to target these G4 regions in the genome for transcription regulation.

CHAPTER V

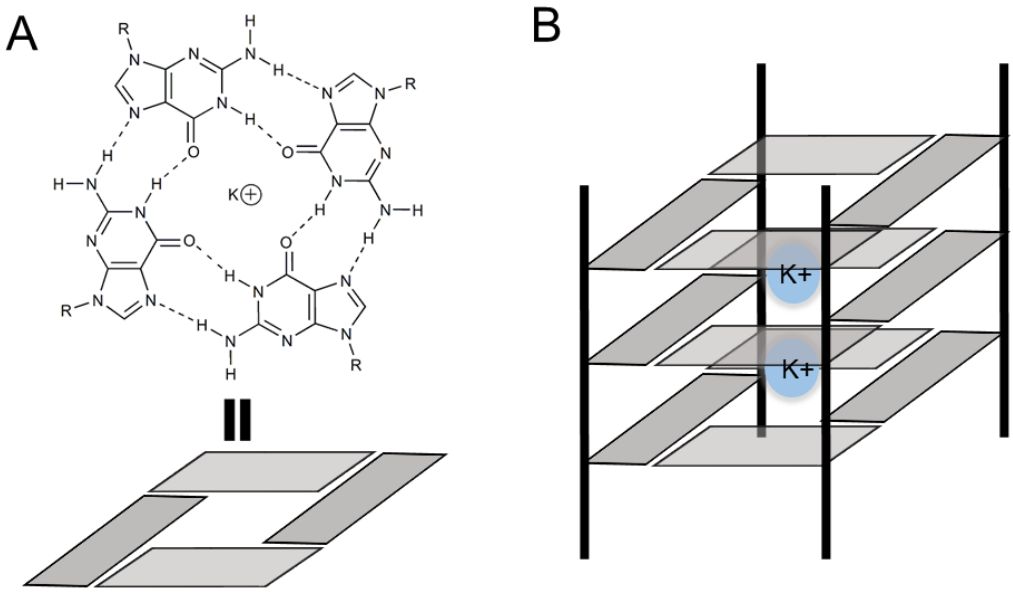
G-QUADRUPLEX VIRTUAL DRUG SCREENING: A REVIEW

Over the past two decades biologists and bioinformaticians have unearthed substantial evidence supporting a role for G-quadruplexes as important mediators of biological processes. This includes telomere damage signaling, transcriptional activity, and splicing. Both their structural heterogeneity and their abundance in oncogene promoters makes them ideal targets for drug discovery. Currently, there are hundreds of deposited DNA and RNA quadruplex atomic structures which have allowed researchers to begin using *in silico* drug screening approaches to develop novel stabilizing ligands. Here we provide a review of the past decade of G-quadruplex virtual drug discovery approaches and campaigns. With this, we introduce relevant virtual screening platforms followed by a discussion of best practices to assist future G4 VS campaigns.

Introduction

G-quadruplexes (G4s) are secondary structures which occur in both DNA and RNA under physiologically relevant conditions (25). G4s contain 2 or more stacks of 4 coplanar guanine residues stabilized via Hoogsteen hydrogen bonding. The stacking interaction is also facilitated by monovalent cations, such as sodium and potassium, as well as π-stacking of the purine bases (**Figure 41**) (40). Although it is unclear what promotes G4 formation *in vivo*, they are increasingly implicated in important biological events such as telomere maintenance, transcription regulation, mRNA translation, and replication (25,116,246-250). More recently, chromatin immunoprecipitation and high through-put sequencing analyses have provided *in vivo* evidence for the presence of ~9,000 non-telomeric G-quadruplexes that reside in nucleosome-depleted promoter regions, confirming many of the previously proposed regulatory G4s (31,251). Thus, G-quadruplexes appear to be excellent targets for anti-cancer therapeutics (248).

Figure 41. G-quadruplex structure. (A) Orientation of guanines in a G-quadruplex quartet. (B) Monovalent cations often occupy the middle of two quartets, helping to stabilize the partial negative charge shared among the O6 oxygen of adjacent quartets. Phosphate backbone is shown as vertical black lines.


A
B
K+
K+

Currently there are greater than 1,000 characterized G-quadruplex stabilizing ligands that have been discovered through virtual screening (VS), traditional high-throughput screening (HTS), and plenty of serendipity (see: http://g4ldb.org/ for a listing of many verified G4 ligands). Although many of these compounds (TMPyP4, pyridostatin, telomestatin, BRACO-19, etc.) bind with high affinity to G4s, it is often by an end-pasting mechanism and, therefore, non-specific. Furthermore, these compounds commonly do not possess drug-like properties, e.g. they do not pass Lipinski's rule of five (252), nor do they have documented ADMET (absorption, distribution, metabolism, excretion, and toxicity) profiles (87). Extensive work has gone into modifying general end-pasting drug scaffolds such as porphyrins (253-256), phenanthrolines (257-260), anthracenes (261,262), naphthalenes (263), and quinolones (264,265) (**Figure 42A-E**). Only Quarfloxin (CX-3543) (**Figure 43A**), an end-paster, has progressed to clinical trials (15). Alternative G4 drug discovery strategies have focused on developing ligands that target the loops and grooves. An example groove-binding ligand is distamycin A (**Figure 43B**) which was shown by Randazzo et al. to interact with the grooves of the parallel tetramolecular quadruplex $[d(TGGGGT)]_4$ by ^{1}H-NMR (proton nuclear magnetic resonance spectroscopy) studies (266,267). Unfortunately, most G4 groove-binding ligands have poor selectivity over double-stranded DNA (dsDNA), which was the case for distamycin A and netropsin (268) (**Figure 43C**). To address this selectivity problem, many researchers have turned to VS drug discovery methods.

Figure 42. Common G-quadruplex “end-pasting” molecular scaffolds found in the literature. (A) porphyrin, (B) phenanthroline, (C) anthracene, (D) naphthalene, (E) quinoline.

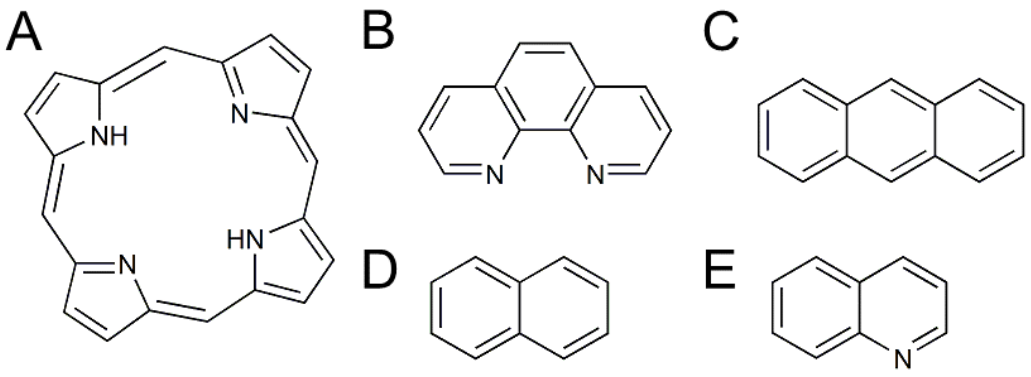
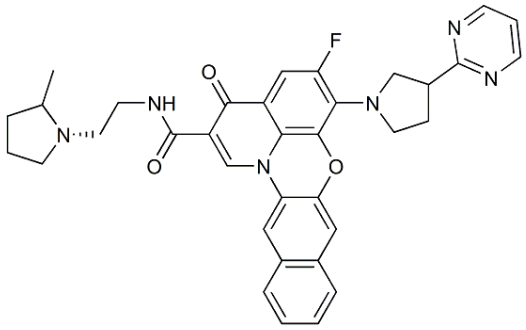
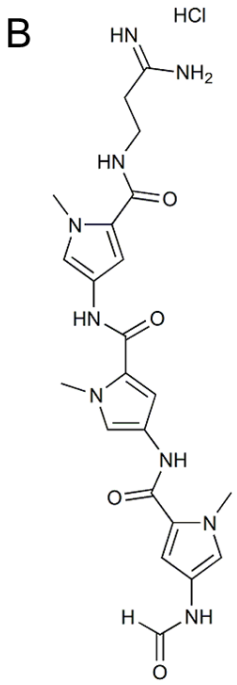

A
NH
N
HN
N
B
N
N
C
D
E
N

Figure 43. Structures of (A) Quarfloxin, (B) Distamycin A, and (C) Netropsin.

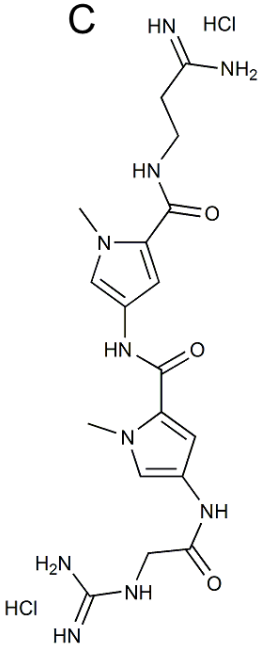
A



B



C



VS strategies have been building momentum in G4 drug discovery both as a low-cost enrichment step and as a lead development step in the discovery pipeline, which our laboratory has previously discussed (269). Whereas traditional HTS methods rely on obtaining and screening hundreds or thousands of compounds from curated libraries, VS simply requires knowledge of known ligand structures (for similarity and pharmacophore searches) or a receptor structure to which a library of virtual compounds can be docked. These methods are known as ligand-based or receptor-based drug discovery, respectively. Ligand-based methods use an identified set of known active ligands to search a database for compounds that have similar properties. These techniques operate under the assumption that ligands with a similar 2D or 3D structure will offer similar interactions with their targets. Conversely, receptor-based methods screen virtual libraries against a target structure, and so require an X-ray crystallographic, NMR, or homology-derived 3D atomistic model of the target. These coordinate files can be downloaded from databases such as the Protein Data Bank (PDB) (>133,000) or the Nucleic Acid Database (NAD) (>900), which are continuously being updated with new structures.

VS platforms have been extensively used in ligand discovery (270,271), however, until now, there has not been an assessment of strategies specifically targeting G4s. Here we briefly discuss some of the common screening strategies, such as docking and pharmacophore screening, as well as relevant aspects including: library preparation, scoring, and analysis. This is followed with a commentary on suggested best practices for *in silico* G4 drug discovery based on the authors' own experience and knowledge gleaned from successful campaigns.

Pharmacophore & Similarity Based Screening

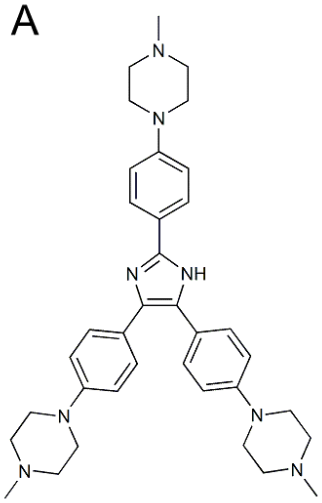
Ligand-based methods such as pharmacophore and similarity search platforms are widely used and often integrated into a VS docking campaign pre- and/or post-docking (see section on docking below). The term 'pharmacophore' as we use it refers to an abstract, 3D physio-chemical representation of the chemical moieties necessary for ligand-receptor interaction. Pharmacophore screens use multiple ligands of the same binding site to derive an ensemble of chemical features necessary for an ideal interaction (i.e. hydrogen bond donor/acceptor, aromatic ring elements,

cations, anions, etc.). The resulting model is known as a hypothesis. These hypotheses, which are 3D chemical descriptors, are then used to screen a virtual library to find “pharmacophore-similars” that satisfy the hypothesis (272). The result is a list of compounds that are ranked for their probability of favorable interactions based on their physical and chemical similarity to the initial query structure. Various pharmacophore search platforms are available such as Pharmer (ZINC) (273), Discovery Studio’s 3D-QSAR module (Accelrys: http://accelrys.com/products/discovery-studio), LigandScout (Inteligand) (274), MOE (Chemical Computing Group) (275), Phase (Schrödinger) (276), SYBYL-X2.1.1 (Certera) (http://tripos.com), and Pharao (Silicos) (277).

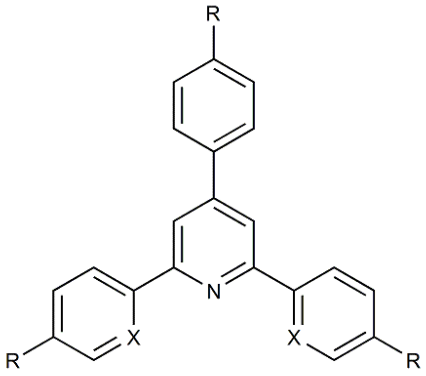
An example of a successful G4 pharmacophore screening campaign comes from Chen et al. (278) in which the authors used Discovery Studio’s 3D-QSAR pharmacophore generation module to construct a model based on acridine derivatives. By weighting hydrophobic interactions higher than aromatic interactions in the hypothesis the authors enriched for compounds with scaffolds unlike the acridines. This was achieved by screening their own in-house library. The resulting compound was a triaryl-substituted imidazole derivative (**Figure 44A**) that has a K_D of 0.5 μM against a human telomere G4 and displayed selectivity over dsDNA based on circular dichroism (CD) and fluorescence melting experiments. Interestingly, this compound is very similar to the triaryl-pyridines discovered previously (279) (**Figure 44B**).

Figure 44. Structures of (A) a triaryl imidazole and (B) a triaryl pyridine.

A



B

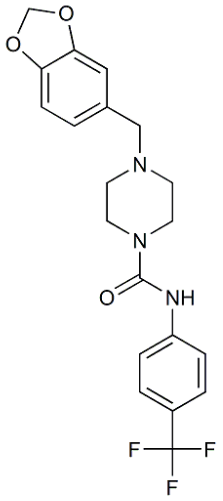


The second and most rapid ligand-based strategy is known as a structural similarity search. These platforms require only knowledge of active ligand's chemical composition (i.e. a chemical structure). In the past, this method utilized a rigid-body alignment approach using 2D (2-dimensional) and 3D chemical fingerprints to align and rank each molecule. This method was enhanced with the advent of semi-flexible and flexible superposition algorithms that allow for a more comprehensive search in 3D space by ranking each molecule based on the volume overlap within the query structure. See (280) for a more in-depth discussion.

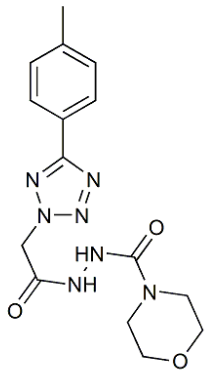
The structural similarity software vROCS (OpenEye) (https://docs.eyesopen.com/rocs/) has been utilized by Musumeci et al. (281) to screen the Maybridge (http://www.maybrdige.com) HitFinder database (~14,400 compounds) using Distamycin A (**Figure 43B**) as a query. Using the Tanimoto coefficient (Section 6.4) and vROCS's colour scoring (atom/feature similarity) criteria the authors discovered a set of novel G-quadruplex groove-binding ligands (**Figure 45A-C**). These ligands bound with higher affinity to the grooves of human telomeric quadruplexes over dsDNA [detected by UV-Vis, fluorescence, and oligo affinity support analysis (281)] but had no observable melting temperature (Tm) shift. It was also shown that 3 of the 7 compounds induced a DNA damage response at the telomeres, further confirming their G4 binding activity. While this isn't the first reported campaign using vROCS in G4 drug discovery (282) it is a proof-of-concept that this relatively straight forward lead-discovery approach can enrich for novel scaffolds which interact in a favorable manner.

Figure 45. Structures of the human telomere groove-binding ligands discovered by Musumeci et al.

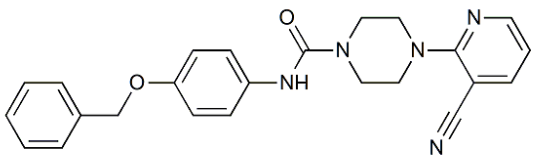
A



B



C



Libraries

Arguably the most important consideration in virtual screening methodologies is the selection of compound library. VS libraries contain hundreds to millions of virtual compounds that will inevitably dictate the scaffold diversity of resultant hits. The benefit of using large, diverse libraries is the expanded chemical search space. Fortunately, there are many large libraries available: MayBridge (http://www.maybrdige.com), AnalytiCon (https://ac-discovery.com/screening-libraries/), ZINC (http://zinc.docking.org/), ChemDiv (http://www.chemdiv.com/services-menu/screening-libraries/), SPECS (http://www.specs.net), Mcule (https://mcule.com/database/), eMolecules (https://www.emolecules.com), PubChem (https://pubchem.ncbi.nlm.nih.gov/), Life Chemicals (http://www.lifechemicals.com), ChemBridge (http://www.chembridge.com/screening_libraries/). Some databases, such as the ZINC database, offer sub-libraries for a more tailored search (e.g. lead-like, fragment-like, drug-like, and natural products), which often contain readily purchasable or synthesized compounds. Conversely, some researchers choose to develop their own curated libraries (278,283) that can be beneficial when the user has limited computing resources available.

The biggest challenge is finding the optimal balance among speed, accuracy, and library composition. A library of ~1 million compounds docked to a single receptor will take as little as weeks to as much as a year of computing time with a single workstation using a rigorous algorithm. Many researchers have circumvented this by reducing libraries to smaller, more manageable subsets that only contain, compounds that conform to a predefined criterion. Specifically, using a shape-based or pharmacophore search on a library, one can significantly reduce the size and enrich for chemical moieties that are well suited to the system of interest [see (278,281,282,284-286)]. However, limiting searches to a pre-defined chemical search space introduces significant bias and is bound to limit compound diversity.

Alternatively, increased computing power by use of a research cluster or computing grid can greatly reduce the computational time required for a screening campaign of >1 million compounds (287). The authors have had success using grid computing which can dock as many as ~25 million compounds in just a few days to a single receptor site (269). While grid computing

is becoming more commonplace in research institutes, not everyone has access to large-scale grids. Therefore, care must be taken if curating a library to be docked at a smaller scale. Hand picking small subsets of compounds can lead to significant bias (see discussion section below), or worse, no enrichment of meaningful hits (288).

Docking

Docking has been in use since the early 80's and has gained traction commensurate to the number of published protein and nucleic acid structures since (289). In general, docking seeks to use the physical and chemical information provided by an atomistic receptor to dock whole or fragmented molecules from a library and rank them using a scoring function. Each docking platform has its own algorithm as well as flavor of scoring function, which has made cross-platform comparisons difficult (290-292). The lack of convergence onto any one platform is likely due to the unique features inherent to each, such as: cost, speed, scoring terms, ease-of-use, scalability, receptor flexibility, ligand flexibility, and the option of implementing molecular dynamics (MD) force fields.

There are multiple docking platforms suitable for use with nucleic acid receptors. These include: DOCK v4-6 (UCSF) (293), AutoDock (Scripps) (294), AutoDock Vina (Scripps) (295), GOLD (Cambridge Crystallographic Data Centre) (296), Surflex-DOCK (BioPharmics) (297), Glide (Schrödinger) (298), and ICM (Molsoft) (299). Many of these programs have been compared elsewhere in the context of protein docking (300). The authors have also compared two of these platforms, Surflex-DOCK and AutoDock, in the context of nucleic acids (301). Both platforms performed equally well with Surflex being slightly faster and more easily scalable. DOCK, AutoDock, AutoDock Vina, and GOLD are all freely available to academic institutions. Each docking platform varies with respect to sampling algorithms and scoring functions.

A sampling algorithm is a systematic way to sample from a population of possible molecular conformations and binding modes without exhausting all possibilities. The primary hurdle in docking is the vast number of potential docked positions for a given set of molecules. Minimizing the computational time necessary for each docking run is of prime importance for high-throughput

screening. Strategies to minimize computational time include: library optimization (reduced size, generation of tautomers, protonation, filtering), robust computational infrastructure (computing grids), and selection of the appropriate sampling algorithm(s). Such sampling algorithms include: geometric matching algorithms (GM), incremental construction methods (IC), Monte Carlo (MC) searches, genetic algorithms (GA), and molecular dynamics (MD) [see ref. (302) for an overview]. **Table 8** lists the various algorithms employed by the docking platforms discussed here.

Table 8. Docking platforms and algorithms presented and discussed in this review.

Docking Platform	Algorithm
Glide	OPLS-AA force field optimization with Monte-Carlo refinement (FFMC/MD)
ICM	Flexible Monte-Carlo (MC)
AutoDock Vina	Flexible Monte-Carlo (MC)
DOCK V4-V6	Geometric matching and Incremental (GM/IC/MD)
AutoDock	Flexible Genetic Algorithm (GA)
GOLD	Flexible Genetic Algorithm (GA)
Surflex-DOCK	Hammerhead fragment-based algorithm and Genetic algorithm (IC/GA)

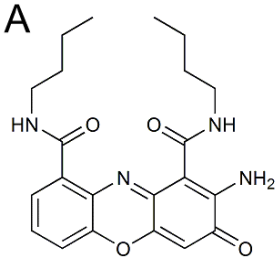
Docking: Geometric Matching (GM)

The first sampling algorithm, which is used in versions 4-6 of DOCK (303), shares characteristics with similarity or pharmacophore search algorithms. DOCK (v4-6) uses a geometric matching (GM) algorithm to place fragments into the receptor. In this algorithm the receptor is treated as a rigid object in which flexible ligands are docked. The receptor is defined by a set of overlapping spheres, while each ligand is defined by rigid segments whose conformation can be optimized within the user-defined binding site. The ligand 'flexibility' comes from an anchor-and-grow algorithm that uses the molecule's rotatable bonds to partition it into rigid segments. Initial docked segments are deemed 'anchors', and from these anchors the remainder of the molecule is appended, followed by optimization and scoring (293). By mapping each molecule into the active site of a receptor, GMs have the advantage of being very rapid techniques and well suited for large database screening (293,300). The recent release of DOCK v6 has added features that allow much more versatility to both the docking and scoring functions. Most importantly it has incorporated MD simulation capabilities (304), validated using a set of RNA-ligand complexes.

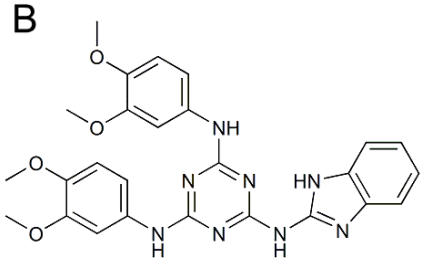
Park and Kang used DOCK (v5.4) in conjunction with the UNITY-3D pharmacophore platform to identify three compounds (**Figure 46A-C**) that stabilize the c-MYC G-quadruplex (284). The authors filtered 560,000 compounds from publicly accessible databases, ChemDiv and SPECS, based on a query generated in UNITY-3D. The resulting set of compounds were energy minimized and then docked into an NMR-derived c-myc quadruplex (PDB ID: 2A5R). The authors optimized this receptor by changing inosine bases back to their original wild type bases followed by short energy minimization. After docking and scoring, each compound was re-scored using a Generalized Born solvent accessible surface area (GBSA) scoring function to account for solvation. Interestingly, the top three compounds showed little or no thermal stabilization based on Förster resonance energy transfer (FRET) screening, but were diverse in structure, and had polymerase stalling ability as well as *in vivo* activity in Ramos, CA46, and HeLa cell lines.

Figure 46. Structures of the reported c-myc quadruplex stabilizing compounds from Kang et al.

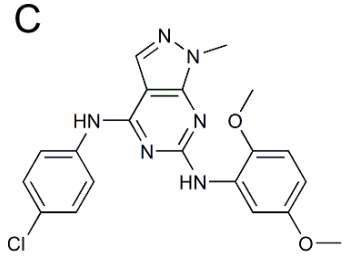
A



B



C



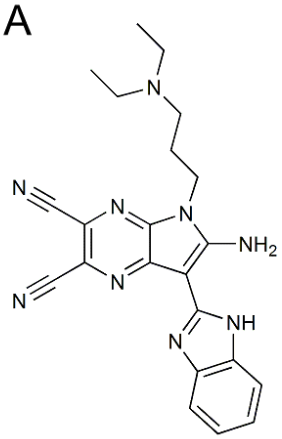
Docking: Incremental Construction (IC)

The incremental construction (IC) docking approach fragments the ligand where it has rotatable bonds and systematically docks each fragment. This allows for very rapid flexible ligand docking and is employed by DOCK (v4+) and Surflex-Dock. The Surflex-Dock approach, which is an adaptation of the Hammerhead docking procedure (305), places head fragments from each ligand into the receptor site and aligns them to 'probe' atoms. These probes are predefined idealized representations of favorable interactions. After placement, each head fragment is scored, and the top scoring fragments are retained. The algorithm next aligns the tail fragments to the head fragments and adjacent probes and scores them. In this way there is a drastic reduction in computation time by only following up with a small portion of possible conformations (301).

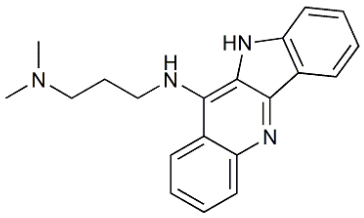
As an example of successful IC docking, Hou et al. identified a novel c-myc stabilizing ligand with the Surflex-Dock platform (306). Using the NMR derived quadruplex (PDB ID: 1XAV) the authors docked 28,530 compounds from the ChemBridge database, which was filtered for compounds containing ≥3 aromatic rings. These compounds were subsequently re-docked to a duplex DNA structure (PDB ID: 1Z3F) and scored based on intercalation. A third round of docking and scoring was performed on each compound, this time in the groove(s) of a duplex DNA (PDB ID: 1K2Z). Compounds with a score ratio of >1.0 (G4 score/ dsDNA score) and >1.1 (G4 score/ Groove score) were chosen. Although the resultant top hit, a pyrollopyrazine derivative (**Figure 47A**) was less effective than the control compound SYUIQ-5 (**Figure 47B**) in luciferase assays, it was much more selective for the G-quadruplex over dsDNA as determined by surface plasmon resonance assays. Furthermore, there are no similar reported scaffolds in the G4 ligand database, indicating that this is a novel quadruplex stabilizing ligand.

Figure 47. Structures of (A) a pyrollopyrazine compound which stabilizes the c-myc quadruplex discovered by Hou et al. and (B) SYUIQ-5.

A



B



Docking: Stochastic Sampling

ICM, AutoDock Vina, AutoDock, Glide, Surflex-Dock, and GOLD are platforms that incorporate the stochastic algorithms: Monte Carlo (MC) and genetic algorithms (GA). Stochastic sampling algorithms iteratively generate new molecular conformations to be placed and scored using random movements (MC) or 'mutations' and 'selection' (GA). Although stochastic algorithms can be computationally more expensive than GM or IC methods alone (301), they have traditionally out-performed in reproducing poses of ligands co-crystallized with their receptors (300).

Docking: Stochastic Sampling (MC)

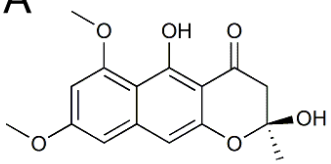
In MC algorithms, each ligand's initial conformation is altered through random steps of bond rotation, rigid-body translation, or rotation, and subsequently scored until a pre-defined number of steps have been reached. At each step, the score is assessed based on steric conflict that is followed by an empirical potential calculation (see Scoring section below). If this new step has improved the score sufficiently, then the molecule's configuration will be saved and used in another iteration of random conformational sampling (307). MC sampling is used in ICM, Glide, AutoDock Vina, and earlier versions of AutoDock.

ICM has previously been shown to perform exceedingly well at reproducing the correctly docked conformation of ligands to protein receptors over DOCK, AutoDock 3.0, and GOLD (308). In 2010, Lee et al. (309) used ICM-Pro to screen a natural products database (AnalytiCon) of 20,000 compounds against the *c-myc* nuclear hypersensitivity element III1 (NHE III1) G4, which was modified from a human telomeric quadruplex structure (PDB ID: 1KF1). Testing the top 5 scoring compounds in polymerase stop assays resulted in the discovery of fonsecin B (**Figure 48A**), a naphthopyrone pigment, which at the time of discovery was a novel scaffold. Later, Chan et al. (310), using the same modified receptor (PDB ID: 1KF1) and docking platform (ICM-Pro), screened 3,000 compounds from a library of FDA approved drugs. This screening campaign led to the identification of methylene blue (**Figure 48B**), a phenothiazinium derivative. Methylene blue is already known to be a dsDNA intercalator and is likely a G4 end-paster. Thus, the authors modified this scaffold with side chains to improve selectivity. Interestingly, one derivative (**Figure 48C**) had

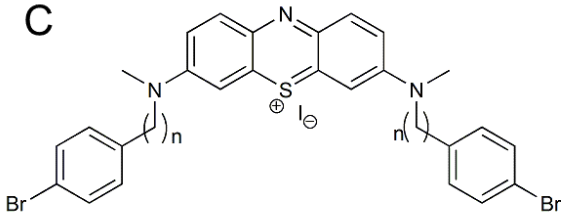
higher affinity for the c-myc quadruplex both *in vivo* (luciferase assays, MTT proliferation assays), as well as *in vitro* (fluorescent intercalator displacement assay, PCR stop assay, mass spec, UV-vis). The Ma group later applied the same approach (ICM-Pro targeting c-myc PDB ID: 1KF1) (311) to screen a natural product-like database of 20,000 compounds to identify potential groove-binding scaffolds by limiting their search space to the grooves. This screen resulted in a compound containing carbamide, diphenyl ether, and tetracyclic moieties (**Figure 48D**), which is a unique G4 scaffold. NMR titration and re-docking were then used to show that the ligand is a de facto groove-binder. Whether this ligand is specific for G-quadruplexes over dsDNA has yet to be determined.

Figure 48. Structures of (A) fonsecin B, (B) methylene blue, and the c-myc quadruplex stabilizing compounds (C) a methylene blue derivative discovered by Chan et al., and (D) a carbamide containing compound discovered by Ma et al.

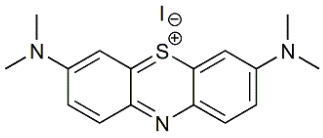
A



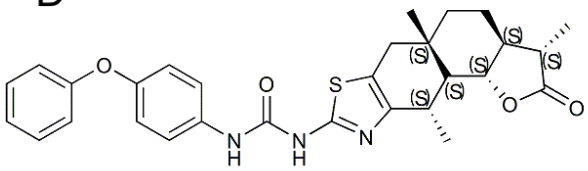
C



B

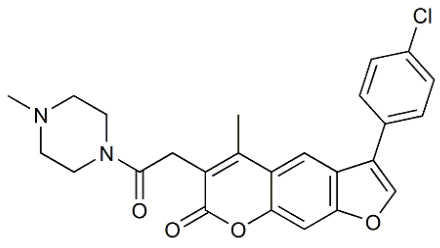


D



Another MC software which offers speed, a user-friendly interface, and great reproducibility of co-crystallized conformations (312) is AutoDock Vina. Vina was used by Alcaro et al. (285) in 2013 as a final step in their screening pipeline where they discovered a psoralen derivative (**Figure 49**). Psoralens have long been known as DNA intercalators; however, this is the first reported instance of psoralens as G-quadruplex stabilizers. Their initial screening began with the ZINC library of >2.7 million compounds, which were filtered down to ~4,000 compounds using 7 query structures in shape-based ROCS [Rapid Overlay of Chemical Structures (313)] and 2D fingerprint filter MACCS (Molecular ACCess System – MDL Information Systems inc.). This screen was followed by the removal of inorganic components, adjusting pH to 7.4, energy minimizing the structures, and finally, removing compounds that have a similarity of less than 0.7 Tanimoto coefficient (see Screening Analysis section below). Altogether, ~7,000 compounds were docked in AutoDock Vina using ensemble docking against the human telomere quadruplexes (PDB ID: 143D, 1KF1, 2HY9, and 2JPZ). This integrated VS approach resulted in 904 compounds that were clustering to obtain 28 compounds for testing, resulting in the psoralen.

Figure 49. The psoralen derivative discovered by Alcaro et al., which stabilized the human telomere quadruplexes.

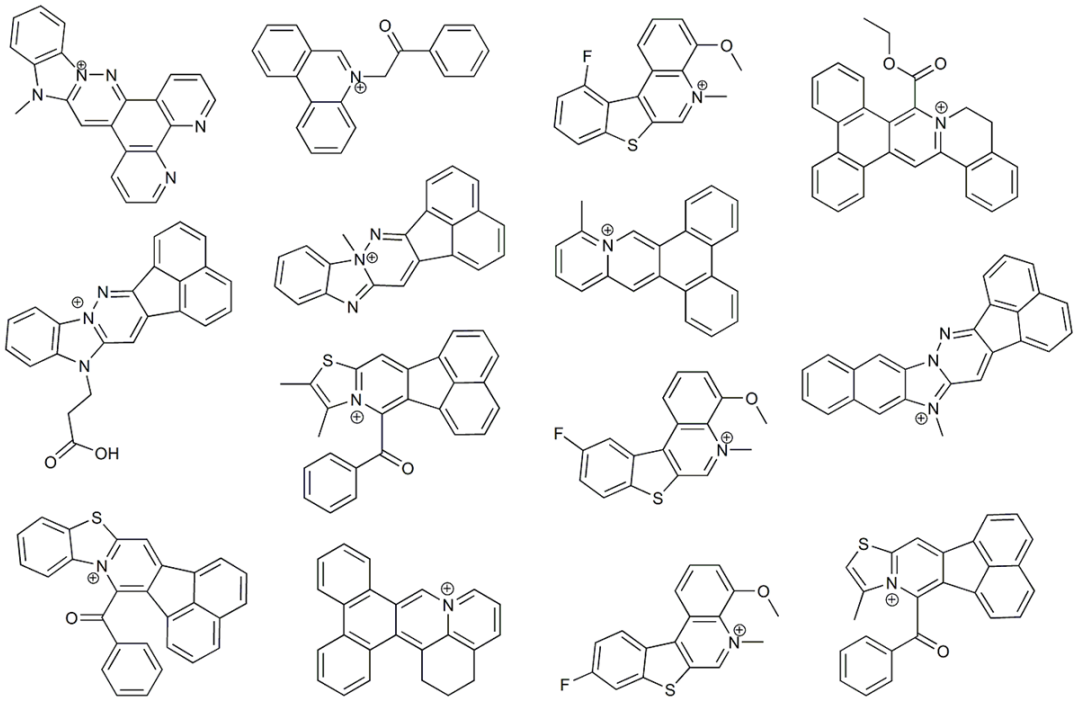

Cl
N
N
O
O
O
O

Docking: Stochastic Sampling (GA)

GA sampling uses a molecule's location, orientation, and conformation to specify the state of a random population of individuals. These individuals have a genotype (the ligand's states) and a phenotype (atomic coordinates) and the ligand's overall fitness is equivalent to its interaction energy with the receptor. Individuals from the initial top scoring population are iteratively "mated" and have offspring that gain random mutations (state changes) as well as inherit genes (states) from both parents, known as "crossover". Selection of each offspring for subsequent mating occurs based on the individual's fitness score (314). Algorithms such as these can be computationally expensive, but have traditionally performed well at reproducing known ligand orientations in active sites (300). GOLD and AutoDock (v3.0+) use this type of sampling method.

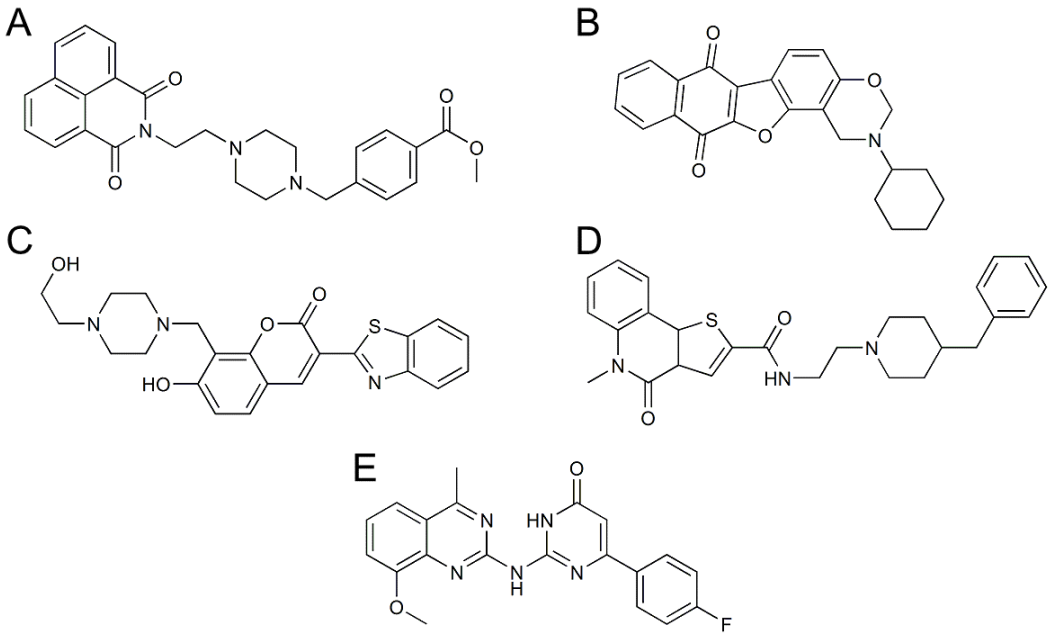
Kaserer and colleagues used GOLD (282) in parallel with the structural similarity search ROCS(313) and the pharmacophore search LigandScout (274) to find consensus hits between the three techniques. The pharmacophore models were generated based on the human telomere quadruplexes in complex with naphthalene diimide derivative BMSG-SH-3 (PDB ID: 3SC8), naphthalene diimide derivative MM41 (PDB ID: 3UYH), and berberine (PDB ID: 3R6R). Overall, they found 252 unique hits from the Specs.net database. Next, using vROCS, they selected the 9 best-performing shape-based models, which were based on queries derived from co-crystallized ligands (PDB IDs: 3UYH, 3SC8, 3R6R) or from the energy-minimized ligand. Using an Implicit Mills-Dean force field, with additional weighting for aromatic interactions, they found 2620 hits. Last, the authors selected the human telomere quadruplex (PDB ID: 3CE5) to directly dock the Specs library. From this screen, the top 10 ranked molecules were selected. In total from the three techniques, 5 consensus compounds and 30 other top scoring compounds were tested, plus some derivatives. Overall, they found 14 ligands (**Figure 50**) that were active and had affinities that compared well with other contemporary VS screening approaches (278,285,315,316). This tour de force campaign demonstrated that a combined approach with cross validation can significantly enrich for real hits, although it did not produce much scaffold diversity.

Figure 50. Structures of the 14 compounds discovered by Kaserer et al. using a multi-platform consensus approach to target the human telomere quadruplexes.



Autodock (v4.2) remains a powerful tool in identifying G4 groove-binding ligands. In 2009, Cosconati et al. (317) used Autodock to screen the Life Chemicals database of ~6,000 compounds against the tetramolecular, parallel G4 sequence $[d(TGGGGT)]_4$ (PDB ID: 1S45) using a grid enveloping just one of the identical grooves. The compounds were scored and selected based on visual inspection. Specifically, compounds unable to form H-bonds with guanine bases or establish electrostatic interactions with the backbone phosphates were removed. Thirty top-scoring compounds were selected and used in NMR titrations, which resulted in an impressive 6 out of 30 interacting as groove-binders. This study was followed up with a more in-depth investigation by Trotta et al. (318) showing that 3 of these compounds (**Figure 51A-C**) bind with higher affinity to the grooves of $[d(TGGGGT)]_4$ than distamycin A using isothermal titration calorimetry (ITC) and NMR. Similarly, Di Leva (319) used Autodock to screen ~19,000 compounds from the ChemDiv database against the 24 nt human telomere quadruplex (PDB ID: 2GKU). Out of the 18 compounds tested, one (**Figure 51D**) showed significant thermal stabilization and appeared to interact as a groove-binder based on NMR and re-docking experiments. The identified benzylpiperidine-containing compound was also shown to cause telomere damage in three cancer lines (HeLa, U2OS, HT29), but not a normal fibroblast line (BJ-hTERT). Subsequently, Amato et al. (320) used Autodock to screen ~59,000 compounds from the Mcule database against a G-triplex structure (PDB ID: 2MKM), an apparent intermediate state in the G-quadruplex folding pathway (321). 15 compounds were selected for purchase, but only one (**Figure 51E**) had significant stabilizing ability. Although this compound did not distinguish G-quadruplex from G-triplex, it did have selectivity for the higher order structures over dsDNA. Thus, these studies are undeniably a testament to Autodock's ability to successfully enrich for compounds which target the grooves of G-quadruplexes.

Figure 51. Structures of (A-C) the parallel groove-binders discovered by Trotta et al. and (D) the human telomere interacting groove-binder discovered by Di Leva. (E) The dual G-quadruplex/G-triplex stabilizing compound discovered by Amato et al.


A
B
C
D
E

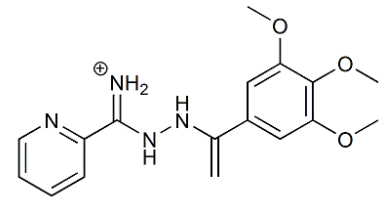
Docking: Stochastic Sampling and MD

MD is primarily associated with simulations of molecular and macromolecular systems but is also applied to other modeling techniques such as docking. MD has long been used as a method to simulate structural changes and molecular interactions at the resolution of atoms using force fields (322). Force fields are the equations that are solved to determine the potential of a given system and are necessary to determine the force acting on each atom. Once a force is determined, Newton's laws of motion can dictate the new atomic position (323). The forces, therefore, must consider each atom's charge, bond length, and angle relative to all other atoms in each system. Thus, docking that utilizes MD allows for the ultimate amount of flexibility of ligand and receptor, resulting in efficient local optimization of docked ligands (300,302). Unfortunately, this level of flexible sampling comes with a high computational cost (324), and so is typically only used as a post-docking refinement step or in estimations of binding free energy (325).

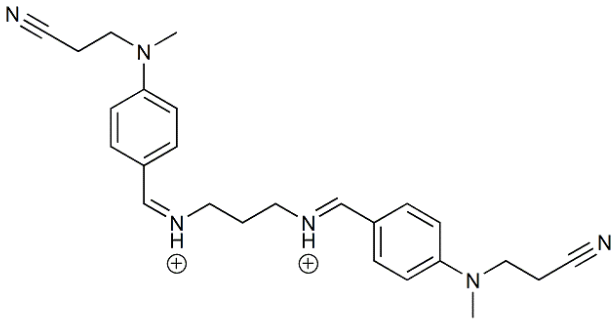
Glide incorporates both MC and MD in its algorithm and performs well relative to other flexible algorithms (GOLD, ICM) in protein docking (300). Glide (grid-based ligand docking with energetics) docks in essentially two stages: (1) each ligand is passed through hierarchical filters that evaluate spatial fit and complementarity of ligand-receptor interactions and, (2) poses that pass the initial screen are subjected to MD minimization based on the OPLS-AA force field (optimized potentials for liquid simulations – all atom force field) (324). Kar and colleagues (286) applied Glide (v5.7) SP (standard precision) mode to dock 14,400 molecules from the Maybridge database, followed by re-docking with the more extensive XP (extra precision) mode to the human telomere quadruplex (PDB ID: 2ld8). A docking site was not selected, rather, the authors constructed a grid encompassing the entire quadruplex. Two G4 ligands (**Figure 52A, B**) were selected from this screen and were shown to have moderately low affinities (K_D of 31 and 137 μM), as measured by fluorescence titrations, but had selectivity over GC-rich dsDNA.

Figure 52. Structures of the two telomere interacting compounds discovered by Kar et al. with moderate selectivity for G4s over dsDNA.

A

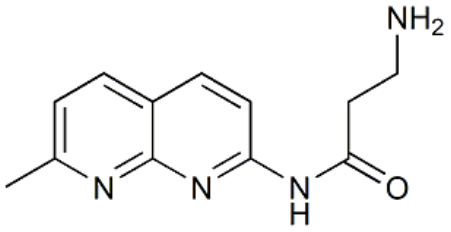


B



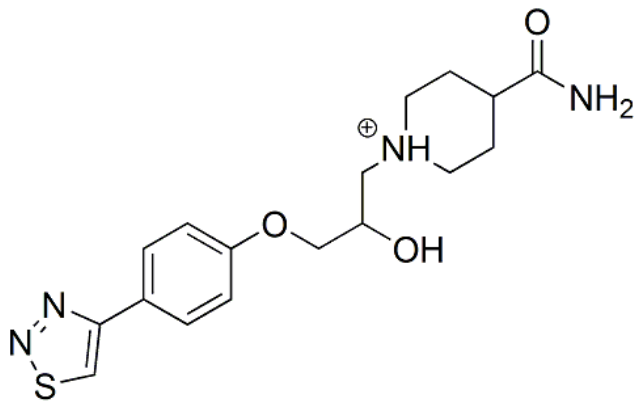
In a mixed pharmacophore/docking approach, Rocca et al. (326) used the pharmacophore screen LigandScout (274) to generate hypotheses based on 9 ligands known to bind DNA and RNA G4s. The ligand conformations for hypothesis derivation were extracted from top-ranked docked positions in the human telomere or TERRA (TElomeric Repeat-containing RNA) quadruplexes (PDB IDs: 3CE5, 2KBP). 257,000 natural product compounds from the ZINC database were minimized and ionized to pH 7.4 in Maestro's Ligprep module (Schrödinger) before being subjected to pharmacophore screening. The compounds were subsequently filtered based on Lipinski's rule of five. The resulting compounds (~12,000) were clustered and then subjected to Glide's ensemble docking and scoring. Testing of the top 20 scored compounds resulted in one ligand (**Figure 53**) that showed interaction *in vitro* as determined by CD, FRET melting, and mass spectrometry. However, the compound reported is a naphthyridine derivative, a class which has previously been reported to interact with telomeric G4s and inhibit telomerase (327).

Figure 53. Structure of the naphthyridine compound discovered by Rocca et al. and shown to stabilize both RNA and DNA G-quadruplexes.


NH2
N
N
N
H
O

Bhat and colleagues (328) have put forward a robust VS workflow to derive novel ligands targeting the *c-myc* NHEIII1 quadruplex (PDB ID: 2A5P). The steps are as follows: (1) the Maybridge database (~55,000 compounds) was imported into Maestro's Ligprep program, which generates all protonation states, conformations, and tautomeric structures for a given pH (~1.5 million compounds); (2) multiple stages of refinement for conformational restraints, conformational groups, and Lipinski's rule of five (~88,000 compounds); (3) Glide docking and re-docking to the 5' end of the quadruplex using all three modes: HTVS (high-throughput virtual screening), SP, and XP. This campaign resulted in three compounds which were chosen for testing, and one, a carbamoylpiperidinium-containing compound (**Figure 54**), stabilized the c-myc G4 by an end-pasting mechanism. A biological response was also observed in cells by luciferase expression assays as well as the induction of apoptosis selectively in T47D cancer cells, but not normal NKE cells.

Figure 54. Structure of the carbamoylpiperidinium containing compound discovered by Bhat et al. and shown to stabilize the c-myc G4 by an end-pasting mechanism.


O
NH2
NH
O
OH
N
N
S

Stand-alone MD simulations have also been used to study the interactions of known ligands with their receptors. These simulations have inherent advantage over traditional docking in that they can explicitly model solvent contributions. Not only does MD allow for calculation of relative binding free energies, but it can also estimate the k_{on} and k_{off} rate constants. The latter has been difficult to assess due to the long timescale simulations needed for the ligand to come back 'on' to the receptor. This has been addressed with biased force fields in what is known as funnel-metadynamics (329). This technique has been applied by Moraca et al. (325) to accurately calculate the free energy of binding of the ligand berberine to the human telomere sequence (PDB ID: 3R6R). Steady state fluorescence measurements were made to determine the actual free energy of ΔG = -9.8 kcal/mol, which compares well with the calculated ΔG = -10.3 kcal/mol. Techniques such as this will likely play a major role in virtual lead development in the future.

Scoring Functions

Docking algorithms attempt to find solutions to the orientation and ranking of ligand-receptor interactions. In doing so, the algorithms must have a way to order the thousands or millions of complexes. Ranking is achieved by scoring, which approximates the binding affinity (ΔG_{bind}). Relative binding free energies can be approximated by free energy perturbation methods using molecular dynamics simulations (324); however, these methods are far too computationally expensive for routine docking, and so more approximate solutions have been devised.

The first type of free energy approximation is the "empirical" (330) scoring function, which is an additive equation derived from each of the different modes of interaction of the system (324,331). As implied by the name, empirical score values are derived from a set of known ligand-receptor complexes. As an example (as adapted from (324)):

$$\Delta G_{bind} = \Delta G_{hb} + \Delta G_{ionic} + \Delta G_{rot} + \Delta G_{vdw} \qquad (1)$$

where ΔG_{bind} would be the total docking score based on the additive scores from H-bonds (*hb*), ionic interactions (*ionic*), rotational constraints of constituent groups (*rot*), and Van der Waals (*vdw*) interactions. These terms can also be modified by the user with weighting to favor or disfavor interactions depending on the system in question. Similarly, there are modifier (or "penalty") terms

which can be applied to disfavor improper H-bond angles, distance restraints, hydrophobic interactions, and torsions. Autodock 4, DOCK v4-6, GOLD, Surflex-Dock, and Autodock Vina use empirical scoring terms.

Another scoring method relies on force-field (FF) based scoring functions. These functions implement current molecular mechanics (MM) force fields (e.g. AMBER, CHARMM) to estimate enthalpy of binding from VDW and electrostatic interactions, strain energies, and solvation effects. The latter is typically estimated by calculating the desolvation energy using MM/PBSA (Poisson-Boltzmann surface area) or MM/GBSA (generalized Born surface area) methods. However, MM/PBSA and MM/GBSA are too computationally expensive to be used in high throughput screening (332). FF scoring is achieved by pair-wise evaluation of each non-bonded interaction, with the following general format (example taken from Autodock v4 manual(294)):

$$\Delta G = (V_{bound}^{L-L} - V_{unbound}^{L-L}) + (V_{bound}^{P-P} - V_{unbound}^{P-P}) + \left(V_{bound}^{P-L} - V_{unbound}^{P-L} + \Delta S_{conf}\right) \quad (2)$$

where *L* is the ligand, *P* is the receptor, and *V* is the calculated potential term from MD force fields. Eq. (2) shows the 6 pair-wise evaluations and entropy term to account for any changes in conformational entropy. The force field potentials used here are comparable to that used in the Amber, CHARMM, or GROMACS force fields but can be modified by the user if desired. Glide, ICM, and early versions of DOCK and Autodock use FF based scoring functions with empirical weighting.

Discussion

Virtual screening approaches in the discovery of new G-quadruplex ligands have clearly shown promise. Higher throughput computational screens are allowing for more comprehensive searches of chemical space. As workstation computing power increases and more researchers gain access to resources such as computing clusters, we expect to see the number of successful VS screening campaigns increase. While some campaigns described here have proven the utility of virtual drug discovery methods, the methodologies and pitfalls are worth discussing.

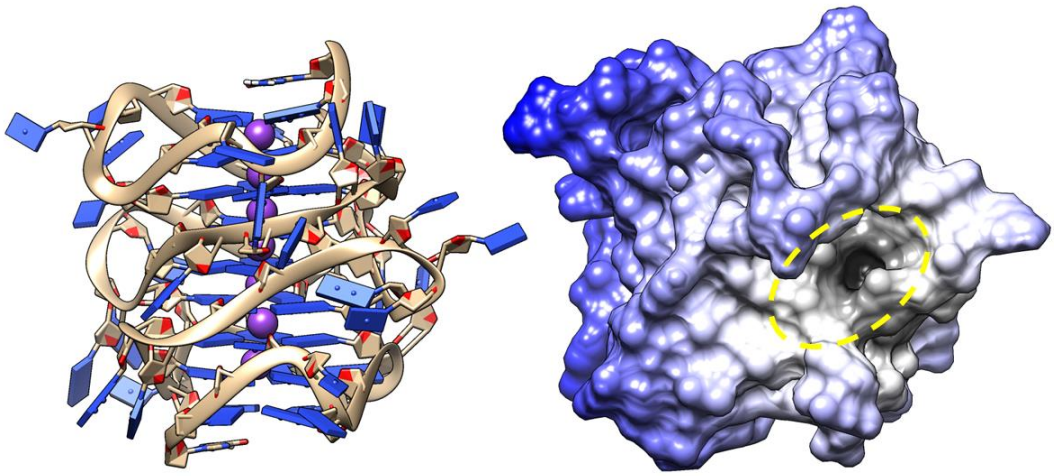
Receptors

G-quadruplex receptors can be downloaded from one of the various databases (Protein Data Bank, Nucleic Acid Data Bank, etc.) or modeled. NMR solution structures should be used if available or the quadruplex modeled based on similar NMR coordinates. X-ray crystallography derived structures are prone to conformational bias and may be artificial due to the packing environment (78). Thus, X-ray structures require pre-treatment with modeling techniques, such as short MD simulations with energy-minimizations (78). Often, modified bases such as inosine are used to select for a single conformation in NMR or X-ray crystallographic techniques. These should be modified, as described above (284), by swapping the inosine with their natural residue and carefully energy-minimizing the structure prior to docking. A second major concern is loop flexibility. Loops have inherently high mobility, and this is rarely accounted for in traditional rigid receptor docking. This can be addressed with short MD simulations to allow the loops to search conformational space. Multiple conformations from MD or NMR PDB files can then be used in ensemble docking (285,326), which is an excellent tactic for highly flexible receptors. These approaches are sufficient for most single G4 systems, such as the c-myc or 22-24 nucleotide long telomere G-quadruplexes. However, targeting non-canonical G4s (245), or large, multi-G4 complexes (145,162) with unknown structure can be challenging.

Multi-G4 systems have the interesting characteristic of large loop domains that span G4-G4 junctions. In theory, these loop-G4 pockets (**Figure 55**) can potentially serve as highly specific binding sites, much like that of enzyme substrate pockets. Unfortunately, determination of large nucleic acid secondary structures with traditional techniques is difficult. Nucleic acids rarely exist as a homogenous population in solution, making them difficult to characterize. Base substitution of non-tetrad guanine with inosine has often been used in NMR experiments to elucidate small G4 structures, but larger systems are not amenable due to spectral overlap and low proton density. Large DNA and RNA complexes are also difficult to crystallize. Even when crystallization is achieved, the packing environment can promote the formation of unrepresentative structures or features that are absent or only present in a small minority of the molecules in solution (78). Thus, low resolution techniques (small angle X-ray/neutron scattering, analytical ultracentrifugation,

dynamic light scattering, and CD spectroscopy) paired with MD simulations are now commonly used to develop structures currently unobtainable with traditional techniques (145,171,245,333).

Figure 55. (Left) hTERT core promoter G-quadruplex model created by Chaires and Trent et al.(145). Phosphate backbone is shown in tan, nucleotides in blue, and potassium in purple. (Right) Surface representation showing a large binding pocket (dark area inside of yellow dashed oval) at the junction between the first and second G4s of the hTERT G-quadruplex. Images were rendered in Chimera v1.12 (176).



Libraries

When docking any library, no matter the size, there will always be top scoring compounds. Thus, many campaigns often use theoretical validation, such as re-docking, complementary screening, or receiver operating characteristic (334) analyses for assessing the ability of their screen to detect real hits (282,284,286,310,335). This is particularly important when working with small databases and can help to minimize false positives. Conversely, docking known ligands as positive controls or simply increasing the library size can help to identify real high scoring compounds. Keep in mind that the use of in-house curated libraries, small drug libraries, or filtered libraries can impose serious limitations on the potential for identifying unique scaffolds. In fact, this appears to be the case for some of the campaigns mentioned here (278,285,306,310,317,326). Three of these reports limited their chemical search space by filtering larger databases and one searched a database with only 3,000 compounds. In every instance, the resulting hits were already known to bind nucleic acids. Conversely, focused libraries can be useful when searching for lead compounds based on validated hits, which was the case for Musumeci et al. (281) who used a similarity search to enrich for groove-binders based on a well characterized groove-binding ligand.

Before using a library, the compounds typically require optimization. Fortunately, many of the virtual screening databases have pre-optimized ligands for screening. Optimization ensures that each ligand has been desalted, neutralized, energy-minimized, and correctly protonated before docking (252,270,336). This can be achieved using programs such as Maestro's Ligprep module (177). Similarly, when ligands are optimized for a screen this should be reported, along with other relevant information such as the version of the database, the type of database, total number of compounds docked, purchased, tested, and validated as hits. Optimizations such as these will allow for a more comprehensive comparison of G4 docking techniques.

Screening

The choice of screening approach(s) used is highly dependent on the user's intentions. Lead optimization strategies should include pharmacophore or similarity screens of large databases (281,282), potentially followed with docking and/or MD minimizations and GBSA

scoring. This approach will minimize time by selecting for ligands that closely resemble the validated scaffolds while also allowing for solvated and flexible receptor-ligand interactions to determine improvement of binding scores (325). However, this approach may not be so useful in *de novo* drug discovery campaigns, where it is more advantageous to screen a large diverse library. As mentioned above, filtering and reducing your library places an inherent bias on the size of the chemical space that will be evaluated, leading to redundancy in scaffolds (see **Figures 44 & 50**). Ideally one should select as large a database as possible and screen with a rapid, flexible ligand screening platform, such as Surflex-dock or Autodock (which has now been surpassed in speed and user friendliness with the release of Autodock Vina), followed with extensive re-docking, consensus docking, or MD simulations with MM/PBSA or MM/GBSA calculations.

A second consideration is the definition of the site to be docked. Most docking platforms have features to allow for ligand-based docking site generation. Conversely, in *de novo* discovery (which is often the case of groove-binders), there is usually no defined site, and therefore a site must be chosen by the user. This is done by defining a 3D grid about the putative ligand binding site (Glide, Autodock, Autodock Vina, DOCK, and ICM), by generation of a space filling protomol [which is a pre-computed representation of an ideal ligand (337)] (Surflex-Dock), or simply by defining a bound ligand or set of residues (Surflex-dock, GOLD). Drawing on the authors' own experience, the docking site should be as small and focused as possible. Unfortunately, most ligands bound to G4s in crystal structures are cationic, polycyclic, and highly conjugated end-pasters bound to the 5' or 3' tetrad faces. Using complexes such as these as the basis for docking will undoubtedly result in top ranked compounds with similar features (282,284,309,310) and, thus, not useful in the discovery of groove-binders or loop-interacting ligands.

Screening Analysis

Regardless of screening strategy, the user will likely generate more compounds than can feasibly be tested. If only the highest-ranking molecules are to be purchased, then visual inspection is recommended. Although tedious, this process appears to increase enrichment in real groove-binders (317,320) by removing erroneous 'false-positives', high steric clashing compounds, and

molecules with poor hydrogen bonding interactions. This can also help rule out compounds docked into unintentional sites because of poorly defined docking sites or grids. Additionally, post-docking clustering based on molecular similarity criterion can reduce the redundancy in large screens and help inform purchasing decisions for diverse scaffolds (285,317,320,326). This is commonly done using similarity coefficients. Tanimoto, Dice, and Cosine similarity coefficients are numerical values computed from molecular attributes that are commonly used in clustering analyses (324). Selectivity can also be enriched for by re-docking the top-ranking compounds to potential off-targets, such as dsDNA (306), and others have reported enrichment from cross-platform consensus scoring (282).

Conclusion

We present here a comprehensive overview of G4 virtual screening methodologies, along with suggestions to help guide future campaigns. These reports have shown that proper receptor optimization, large screening libraries, and appropriate downstream analyses of hits can result in great enrichment for novel G4 ligands. Conversely, we find that filtered libraries impose a major limitation on ligand diversity. Furthermore, there is a fundamental deficiency in reporting relevant information regarding VS campaigns, such as: library sizes, library preparation (optimizing, filtering, tautomer generation), and contents (fragment-like, drug-like, natural products), total purchased vs. tested compounds, receptor preparation (protonation, modified bases, energy minimizations, MD), and downstream analysis (clustering, visual inspections, re-docking). This information is critical for evaluating G4 virtual drug discovery strategies.

There are potentially hundreds or thousands of G-quadruplexes that form within promoters, telomeres, RNA transcripts, and even LINEs and SINES (250,251,338,339). As articulated previously (15), G-quadruplexes are easily targetable with heterocyclic aromatic compounds because of the common tetrad face. Selectivity, then, must come from groove-interacting ligands or by end-pasting molecules with “built-in” selectivity for loops around the 5’ or 3’ interface. This selectivity is best achieved using massive, un-filtered libraries targeted at small pockets in and

around the loops and grooves (**Figure 55**). The authors have recently used Surflex-Dock version 2.1 to screen the ZINC drug-like libraries (versions 2014 and 2016) for a total of ~45 million compounds docked to multiple residue-defined loop/groove pockets of a modeled hTERT G-quadruplex (modeled using guanine stacks from the parallel c-myc G4, PDB ID: 1XAV) (145). The quadruplex was subjected to MD simulations and stripped of waters and ions before docking. Docking was carried out using a computing grid known as the DataseamGrid (www.kydataseam.com), which utilizes computers across schools in Kentucky. Ligands were used as-is from the ZINC drug-like database. Purchased compounds were chosen by hierarchical clustering of the top 6,000 molecules using Tanimoto similarity coefficients. From this analysis, 69 compounds were selected and screened using FRET, CD, ITC, fluorescent intercalator displacement assays, and analytical ultracentrifugation. The initial FRET screen resulted in ~33/69 G4 interacting compounds. The top 3 were further characterized, resulting in 2 potent groove or loop interacting ligands (unpublished, see Chapter VI), which are currently undergoing optimization and lead development.

Virtual screening of G-quadruplexes and other higher order nucleic acid structures is still in its infancy. As noted here, few VS platforms have been used in G4 drug discovery and even fewer have been used extensively enough with nucleic acids as to permit cross-platform comparisons. Furthermore, like protein systems, nucleic acids remain sensitive (if not more so) to the limitations of VS technologies. As mentioned previously (271,340), receptor flexibility remains difficult to address in a high-throughput manner, and so G4 loops remain a challenge to target. Similarly, while docking algorithms can be very reproducible and rapid, there remains a dire need for accurate, robust scoring approximations (340). Fortunately, the predictions (341) of hit enrichment from high performance computing and big libraries were correct. So while the world awaits breakthroughs in scoring, receptor flexibility, and machine learning (271,342), it might be wise to seek out your nearest computing cluster to carry out your G4 screening.

CHAPTER VI

TARGETING THE HIGHER-ORDER HTERT G-QUADRUPLEX: VIRTUAL DRUG DISCOVERY OF SELECTIVE HTERT REPRESSING SMALL MOLECULES

Non-canonical DNA structures known as G-quadruplexes are now widely accepted as viable targets in the pursuit of anticancer therapeutics. Unfortunately, few virtual or actual drug screening campaigns against monomolecular G-quadruplexes have resulted in selective and drug-like small molecules. This dearth of selectivity is likely due to an inadequacy of chemical space searched, as well as shortcomings in defining receptors. Herein, we show that by increasing the chemical search space to tens of millions of virtual compounds that it is possible to discover novel, small molecule scaffolds that are selective for G-quadruplexes over duplex DNA. Using *in vitro* screening techniques (fluorescent thermal shift assays and competition dialysis) and fundamental biophysical interaction techniques (isothermal titration calorimetry, analytical ultracentrifugation, CD melting), we demonstrate that we can enrich for selective small molecules which are specific for the loops and grooves of a multimer G-quadruplex formed in the core promoter of the *human telomerase reverse transcriptase* (*hTERT*) gene. Further, using exhaustive virtual docking and molecular dynamics simulations, we show that the lead molecule, a disubstituted 2-aminoethyl-quinazoline, binds in loop pockets and grooves, stabilizing G-quadruplex stacking junctions. Lastly, we provide evidence that this molecule downregulates *hTERT* transcription in breast cancer cells, making it a promising lead molecule for treatment of hTERT-reliant cancers.

Introduction

Nucleic acids have the capacity to form multiple types of secondary structure, such as duplex (B-, A-, and Z-DNA), triplex, and quadruplex. Genetic regions with high guanine content, specifically with multiple runs of consecutive guanines, can form highly stable structures known as G-quadruplexes (G4s) (40). G4s are composed of a stack of planar layers of guanine nucleotide tetrads held together through Hoogsteen hydrogen bonding. Analyses of the human genome have revealed between 376,000 and 716,310 potential G4 forming regions (11,31), and as much as 40% of these PQSs reside in gene promoters (12). With the advent of fluorescent G4-specific small molecules and antibodies there is now direct evidence of non-telomeric G4 formation in cells (27,28,343). Importantly, these promoter G4s are abundant in oncogenes (31) and impose proximal regulatory effects on transcription of adjacent genes (15,344). While this has energized research targeting promoter G4s of proteins that have previously been thought of as “undruggable” (345), or in general difficult to inhibit, there remains an outstanding issue of quadruplex drug specificity (15,346,347).

To date, most of the drug discovery and biological investigations of G4s have been limited to the small (<12 kDa) monomeric forms—likely owing to the ease at which they can be studied by NMR or X-ray crystallography. A variety of monomeric promoter G4s have been reported: c-MYC (16), KRAS (101), HRAS (102), HIF (103), and VEGF (104), which have been utilized in structure-based drug discovery (347). Numerous small molecules have now been unearthed that target monomeric G4s over duplex and triplex DNA, but few have demonstrated selectivity for a single G4 target (346). In some cases, small molecules can even convert certain monomeric G-quadruplex topologies to their “preferred” topology (348). A commonality shared among all G-quadruplexes are their planar 5’- and 3’- G-tetrad faces. These faces allow small molecules to end-paste, which is thought to maximize π-π stacking interactions between the ligand and the guanine bases. It follows that this type of interaction results in mostly indiscriminate binding (346). This non-specificity can be improved upon by adding or altering constitutive groups or sidechains of the core

scaffold to increase favorable interactions with distinctive loop or groove regions (349-351); although, this often comes at a price, as the molecules tend to deviate from “drug-likeness” the more they are modified (137).

An emerging alternative method of selective G-quadruplex drug discovery is through targeting higher-order G4 assemblies. G-quadruplexes can stack on top of one another through sandwiching of flanking nucleotides or direct stacking of terminal G-tetrads (54,243). While this phenomenon is well known and has important biological implications (20,352), few multimeric G4s have been structurally characterized. The allure of targeting multimer G4s is that they offer larger, potentially unique binding pockets for small molecules at G4 stacking junctions and loops, which could allow for the circumvention of non-specificity due to end-pasting that is commonly observed in monomeric G4 drug discovery (20,141,347).

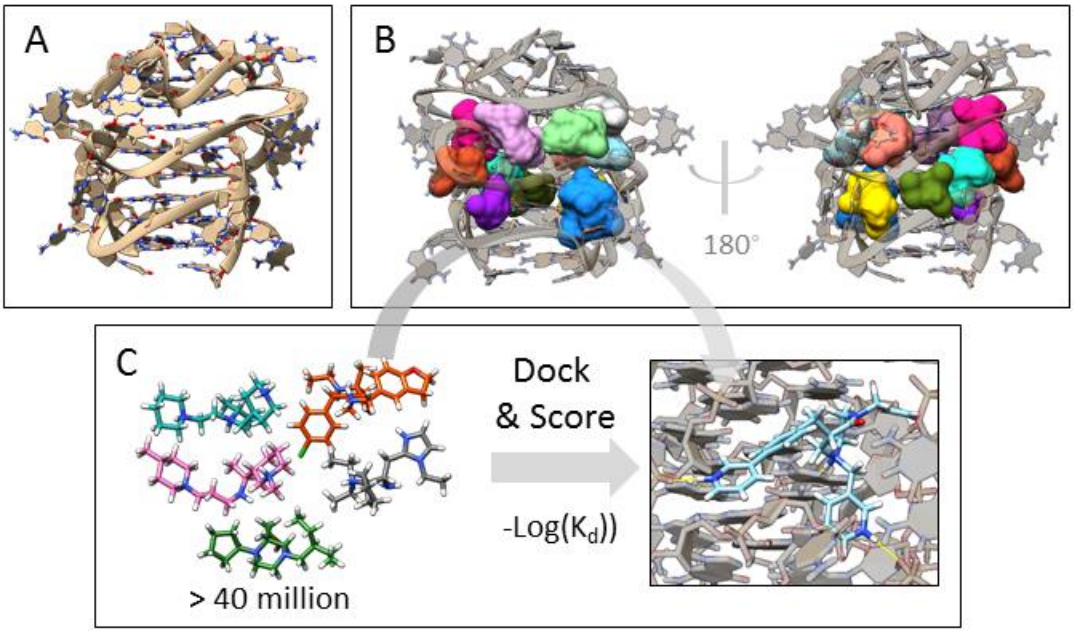
One such G4 multimer has been identified in the core promoter region the *human telomerase reverse transcriptase* (*hTERT*) gene (105,106,144). hTERT, and its cognate RNA (human telomerase RNA component, hTR), form the ribonucleoprotein complex that is responsible for maintenance of the telomeres. In normal somatic cells, hTERT activity is tightly regulated or entirely absent (353). Under these “normal” circumstances, dividing hTERT-negative cells will eventually experience telomere shortening, which elicits a DNA damage response that ultimately results in senescence or cell death (354). Interestingly, the forced re-expression of hTERT in hTERT-negative cell lines is all that’s necessary to extend cellular replication (355,356). Knockdown of hTERT in a variety of cancer cell lines and tumor models results in reduced telomere maintenance, sensitization of cancer cells to chemotherapeutics, and in some cases, the direct induction of apoptosis or senescence (357-359). It follows that hTERT’s nearly exclusive overexpression in malignant cell types has made it an ideal target for anti-cancer therapeutics. The majority of contemporary techniques targeted at telomerase inhibition, such as small molecules inhibitors, gene therapy, anti-sense oligonucleotides, and immunotherapies, have all shown that hTERT inhibition is a viable mechanism to treat cancer (214). Clinically, however, none have been successful (215). Of note, Imetelstat, a 13-nt antisense oligonucleotide, which appears to directly inhibit telomerase through base-pairing with hTR, has shown preclinical promise in a variety of

cancer models for its ability to block telomere extension (360,361). Unfortunately, like many other telomerase inhibitors, clinical trials have halted due to hematopoietic toxicity, potentially because of non-specific telomerase inhibition in stem cells (216,362). Currently, there are no FDA approved small molecule telomerase inhibitors. Hence, alternative hTERT inhibition strategies are warranted.

Overexpression of hTERT largely occurs through alterations in transcription and/or through increases in gene copy number (222,363-366). The hTERT core promoter region (-180 to +1 of TSS) has a low nucleosome occupancy, sensitivity to bovine pancreatic deoxyribonuclease (DNase I) treatment, and a high GC content (12,367,368). The open chromatin and multiple PQSs within the 12 G-tracts support the formation of DNA G-quadruplexes in the context of the nucleus. Further, functional genetic studies have found links between elevated *hTERT* promoter activity in cancers which contain G>A point mutations within these G-tracts, supporting a G4 regulatory mechanism (218,219). In fact, a recent investigation of urothelial carcinomas of the bladder found that 60-80% contain the *hTERT* mutations “G124A” or “G146A” (located within the core promoter region -124 and -146 from the TSS, respectively), and these mutations confer a selective advantage over non-mutated cells (369). These mutations lead to allele specific deposition of H3K4me3 marks, which promote transcriptional activity through a mechanism involving the swapping of specificity protein 1 (SP1) to E-twenty-six (ETS) transcription factors (222). However, this switch to transcription activating transcription factor binding only partly explains the changes in *hTERT* expression (223).

The secondary structure of the G-rich region spanning -168 to -100 has been extensively investigated by our lab and others (105,106,143-145,223,224,370). While all investigations are in agreement that non-B DNA secondary structure is formed by this sequence, there was some previous contention as to the exact nature of the major form (106,144,145). We have recently established that the best model for this region of DNA is a stack of three parallel G-quadruplexes (143,145), which is endowed with multiple unique loop and groove pockets that can be targeted in a rational drug discovery approach (**Figure 56**).

Figure 56. Overview of Surflex-dock virtual screening. (A) Structure of the all-parallel stacked hTERT G-quadruplex used here for *in silico* screening. (B) Twelve Surflex-dock protomols (colored space-filling blobs) superimposed on the hTERT G-quadruplex structure as in A. The hTERT structure has been made slightly transparent to emphasize the protomols within loop pockets. (C) Graphical representation of the overall docking procedure, where a pool of more than 40 million virtual small molecules are docked into each of the twelve protomol sites indicated in B, and subsequently scored based on interactions made, such as hydrogen bonding, physical clashes, and van der Waals interactions.


A
B
180°
C
Dock
& Score
-Log(Kd))
> 40 million

The initial investigations of the hTERT core promoter G4 have also involved efforts towards the development of hTERT G4-targeting small molecules. In the first report, Palumbo and colleagues demonstrated that both TMPyP4 and Telomestatin, two rather promiscuous G4 binding small molecules, could stabilize the hTERT quadruplex (106). In a second report, Micheli et al. used modified perylene diimide compounds to investigate Taq polymerase inhibition and stabilization of both monomeric and multimeric forms. While perylene diimide based small molecules are well known for their ability to bind G-quadruplexes of various topologies, this study provides evidence that targeting the multimeric selectively over the monomeric units is viable. Subsequently, Kang *et al.* identified an hTERT G4-stabilizing small molecule with an acridine scaffold that significantly reduced hTERT mRNA and protein levels in breast cancer cells (223). Although this study is rigorous in demonstrating the ability of this small molecule to reduce hTERT levels in cells, the authors do not provide significant evidence of selectivity for the hTERT G4 over other quadruplex topologies or duplex DNA. Moreover, this molecule [designated as GTC365 (223)] is an acridine derivative, and as such would be expected to show moderate to high affinity for duplex DNA via intercalation (371), and non-specific interactions with various G-quadruplex topologies through end-pasting (138,139), making it unlikely to be a truly selective small molecule. Altogether, the above investigations reveal a cancer specific epigenetic mechanism that exists for repressing *hTERT* through targeting its G-quadruplex secondary structure.

There are a variety of obstacles when attempting structure-based drug discovery of novel small molecules for a particular G-quadruplex target, which we have discussed previously (347). The biggest shortcoming is the dearth of small molecules in libraries and limited search of chemical space. Often for time's sake, one of these two elements will be neglected, resulting in the rediscovery of molecular scaffolds that are not unique or a general lack of "hits". Further, the lack of high-resolution information on G4 groove- and loop-interacting small molecules exacerbates this problem. Herein, we show that with large enough chemical diversity, search space, and rigorousness of sampling, we can successfully discover unique small molecules that selectively target the loop and groove pockets of a higher order G-quadruplex. Using an unparalleled G-quadruplex virtual drug discovery approach, we present here the discovery of novel, drug-like, and

selective small molecules targeting the hTERT core promoter G-quadruplex multimer. Further, we demonstrate that modifications of our lead molecule, a disubstituted 2-aminoethyl-quinazoline, result in differential stabilizing effects *in vitro* and in cells, demonstrating the potential for rational improvements as a lead candidate molecule in human telomerase repression.

Materials and Methods

Oligonucleotides

Oligos were purchased from IDT (Integrated DNA Technologies, Coralville, IA) and Sigma-Aldrich (St. Louis, MO) with standard desalting. FRET-labeled oligos used in FTSA experiments had 6-FAM (6-carboxyfluorescein) attached to their 3' end and TAMRA (5-carboxytetramethylrhodamine) attached to their 5' end. Upon receipt, stock oligos were dissolved in MilliQ ultrapure water (18.2 MΩ x cm at 25°C) at concentrations between 0.1 and 1 mM and stored at -20.0°C until use. Folding was achieved by diluting stock oligos into their respective buffer and heating to 99.9°C in a water bath for 20 minutes, followed by slow cooling overnight and subsequent storage at 4°C. Unlabeled oligonucleotide concentrations were determined from their extinction coefficients (ε_{260}) using the nearest-neighbor method. FRET-labeled oligo concentrations were determined using the extinction coefficient of 6-FAM (ε_{495} = 75,000 cm^{-1} M^{-1}).

Small Molecules

In the first round of screening, 69 small molecules were purchased as 1 or 5 mg quantities from the distributer Molport.com, diluted to 10 mM in DMSO, and stored at -80°C until use. Three of the compounds which were studied more extensively, 2R, 3B, 3B1, and 3B5 were either re-purchased (3B from ChemBridge.com and 3B1 from Molport.com) or synthesized (2R, 3B5) in-house. Compound concentrations were determined using molar extinction coefficients derived from UV-Visible spectroscopy measurements of stock compounds diluted into potassium phosphate buffer at pH 7.2.

Buffers

All buffer reagents, unless otherwise specified, were purchased from Sigma-Aldrich. Phosphate buffer (8 mM phosphate, 1 mM sodium EDTA, pH 7.2) was used throughout with supplementation of KCl as indicated. Buffers were filtering through 0.2 μm filter paper prior to use.

Preparative size-exclusion chromatography (SEC)

Oligonucleotide purification of the unlabeled TERT-FL, Tel48, and Tel72 sequences was achieved using SEC as detailed previously (168). Briefly, oligonucleotides were annealed at concentrations of 40-100 μM in their respective buffers, filtered through 0.2 μm filters, and injected onto an equilibrated Superdex 75 16/600 SEC column (GE Healthcare 28-9893-33) using a Waters 600 HPLC system. The flow rate was maintained at 0.5 mL/min. and sample fractions were collected every 2 minutes from 100 to 180 minutes run time. The molecular weights of fractionated species were estimated based on a regression analysis of elution time vs. log(MW) of protein standards (Sigma #69385), with elution profiles monitored at 260 nm and 280 nm. Purifications were carried out at room temperature and fractionated samples were stored at 4°C prior to concentration and downstream analysis.

In silico drug screening

Virtual screening was performed using Surflex-Dock 2.11 (372) on the KY Dataseam computing grid (http://www.kydataseam.com/) using over 40 million virtual ligands from the ZINC 2014 and 2016 drug-like libraries (373). The all-parallel G-quadruplex hTERT model created previously (143,145) was used as the receptor. Twelve docking sites were chosen by targeting the G4-G4 junctions, loops, and grooves, but not the terminal G-tetrad faces (**Figure 56**). Docking in Surflex-dock was carried out as previously described(301). Briefly, the command used in Surflex-Dock for each run was “surflexdock -pgeom +self_score +pflex dock_list <library> <protomol> <receptor> <output>”. Protomols were generated using residue selection, which allows for the manual selection of residues in or around putative binding pockets. The options “proto_thresh” and “proto_bloat” were left at their default settings. Protomols were visualized in UCSF Chimera v1.11

(176) to ensure adequate coverage of docking area (**Figure 56**). Scoring was based on the empirical -Log(K_D) term reported for each docked molecule (clash, polar, and strain values were not considered in scoring). From each docking site (24 in total, 12 for 2014 library and repeated for the 2016 library), the top scoring 500 poses were analyzed in Schrodinger's Canvas (374) application using a hierarchical clustering algorithm to cluster molecules based on binary fingerprints and Tanimoto similarity criteria. The highest ranked 100 centroid molecules (most representative scaffolds of the clade) were then chosen for purchasing. From this selection only 69 were available from Molport.com.

Fluorescence thermal shift assay (FTSA)

Small molecule screening by FTSA was performed on an Applied Biosystems StepOnePlus Real-Time PCR instrument in 96-well plates, adapted from previous work (208). Briefly, 10 mM compound stock solutions in DMSO were used to create 96-well stock solution plates by diluting each compound to 2x final concentration in potassium phosphate buffer. The same volume of DMSO was used as a control. FRET-labeled DNA, post-annealing, were quantified by UV-Vis and diluted to 2x final concentration. FTSA reaction mixes were made up in 96-well Applied Biosystems MicroAmp PCR plates by mixing 10 µL of 2x compound solution (or buffer/DMSO control) with 2x FRET-labeled DNA to yield 20 µL of 1x reaction mix. Final concentration of DNA was 0.25 µM in all experiments unless otherwise specified. Compound concentrations are as specified in each figure legend. After mixing, plates were then spun down at 1250 rpm for 2-3 minutes in a benchtop centrifuge to remove bubbles. Samples were denatured by ramping the temperature from 20.0°C to 99.8°C in 0.2°C increments at a rate of approximately 0.7°C/min. Fluorescence quenching of 6-FAM was monitored at each 0.2°C step using the instrument's onboard FAM filter over the entire reaction, providing a melting curve. Melting temperatures (T_m) were determined from the 1st derivative of the normalized melting curves (209), and differences in melting temperatures (ΔT_m) were determined by taking the difference of control and sample wells:

$$\Delta T_m = T_{m,sample} - T_{m,control}$$

Where $T_{m,sample}$ and $T_{m,control}$ are the melting temperatures of the sample and control, respectively. Measurements are averages of triplicate experiments repeated on 3 separate days unless otherwise specified.

Competition Dialysis

Competition dialysis assays were conducted as described (375) in either 96-well plate format or scaled up to beakers. Nucleic acids were annealed in a 185 mM K+ phosphate buffer, SEC purified, and concentrated to create stock solutions of 20-100 μM (strand concentration). From these stocks, nucleic acids were diluted to 75 μM working concentration based on monomeric unit, where each nucleotide, base pair, triplex, or tetrad is considered a single monomeric unit (e.g. the full-length hTERT is three G-quadruplexes and has 9 G-tetrads, whereas c-MYC only has 3 G-tetrads, so c-MYC is 3x the strand concentration as hTERT). Each dialysis membrane had 200 μL (96 well) or 500 μL (scale up) of nucleic acid sample or buffer only (control). The dialysis membranes were then submerged in at least 2 mL (96 well) or 200 mL (scale up) of a 4 or 5 μM solution of compound and allowed to incubate on a rocker (96 well) or with stir bar (scale up) overnight at room temperature. Approximately 24 hours later, the sample was removed and mixed with Triton X-100 to a final concentration of 1% v/v to disrupt ligand interactions with receptors. Ligand concentrations were determined from their extinction coefficients by measuring absorbance using a Tecan Safire II plate reader (Tecan, Männedorf, Switzerland). Total concentration of bound compound (C_b) was determined as follows:

$$C_b = C_t - C_f$$

Where C_t is the total concentration of ligand in sample membrane and C_f is the concentration of compound in the buffer membrane (which did not deviate from the 4 or 5 μM compound in dialysis buffer). Calculation of apparent binding affinities (K_{app}) was achieved using the following equation:

$$K_{app} = C_b / \{C_f \times ([NA]_{total} - C_b)\}$$

Where $[NA]_{total}$ is the total DNA concentration in the well (75 μM).

Thiazole orange (TO) displacement assay

The TO displacement assay, which is functionally similar to standard fluorescence intercalator displacement assays, was performed exactly as described previously (196). Briefly, annealed nucleic acids, in their respective annealing buffers, were mixed with Thiazole Orange (TO) and test compound at final concentrations of 2 μM DNA, 1 μM TO, and 5 μM compound in a total volume of 150 μL in a 96 well black flat-bottom polystyrene plate. Control wells with DNA and TO, and TO alone were also prepared in the respective buffers. After a brief incubation, TO fluorescence emission was measured at 1 nm intervals from 510 to 750 nm with an excitation wavelength of 500 nm. The percentage of TO displacement (%FID) was calculated from the intensity at 527 nm using the following equations:

$$\%\mathrm{FID} = 100 - (100 \times \frac{\mathrm{F}}{\mathrm{F_o}})$$

$$\mathrm{F} = \mathrm{F_{(ligand+DNA+TO)}} - \mathrm{F_{(buffer+TO)}} - \mathrm{F_{(DNA+ligand)}}$$

$$\mathrm{F_o} = \mathrm{F_{(DNA+TO)}} - \mathrm{F_{(buffer+TO)}}$$

where F is the fluorescence intensity reading from each well at 527 nm (λ_{ex} = 500 nm).

Circular dichroism spectroscopy (CD)

CD melting studies and spectra were collected on a Jasco-710 spectropolarimeter (Jasco Inc. Eason, MD) equipped with a Peltier thermostat regulated cell holder and magnetic stirrer. CD and melting spectra were collected using the following instrument parameters: 0.5 or 1cm path length quartz cuvette, 210 or 240 to 340 nm wavelength range, 1.0 nm step size, 200 nm/min scan rate, 1.0 nm bandwidth, 2 s integration time, and 4 scan accumulation. Spectra were recorded at 20.0°C and melting spectra were collected over a range of 4°C to 98°C with 4°C steps, 4°C/min ramp speed, and a 1-minute equilibration time at each temperature before acquisition. Spectra were corrected by subtracting a buffer blank. Spectra were normalized to molar circular dichroism (Δε) based on DNA strand concentration using the following equation:

$$\Delta\varepsilon = \theta/(32982cl)$$

where θ is ellipticity in millidegrees, c is molar DNA concentration in mol/L, and l is the path length of the cell in cm. In T_m shift experiments, the concentration of DNA was 1.1 μM, and compounds

25 μM (except for the control, BRACO-19, which was 2.5 μM). The same v/v DMSO was used as the control.

Analytical ultracentrifugation (AUC)

Sedimentation velocity (SV) experiments were performed in a Beckman Coulter ProteomeLab XL-A analytical ultracentrifuge (Beckman Coulter Inc., Brea, CA) at 20.0°C and 40,000 rpm in standard 2-sector cells using An50Ti or An60Ti rotors. 100 to 150 scans over an 8-hour period were collected and analyzed in Sedfit (170) using the continuous C(s) model with a partial specific volume of 0.55 mL/g for DNA. AUC drug binding experiments were carried out as detailed previously (210), with a final compound concentration of 100 μM and 10 μM DNA (10:1 [compound]:[DNA]). All compounds with reported stoichiometry from AUC experiments were monitored at a wavelength of 318 nm.

Molecular dynamics simulations

Starting coordinates for small molecule-G4 complexes were based on the output of flexible docking performed using Glide XP (376) with the Maestro (177) suite using the hTERT G4 model created previously as the receptor (143,145). Briefly, Sitemap (377) was used to generate multiple docking sites among the loops, grooves, and G-tetrad faces of the hTERT receptor (see **Figure 56** Surflex-Dock protomols), followed by flexible docking and scoring at each site. Molecular dynamics simulations were subsequently carried out on the highest scoring ligand-hTERT complexes for a total of 5 ns. The PDB structures were imported into the xleap module of AMBER18 (178), neutralized with K^+ ions, and solvated in a rectangular box of TIP3P water molecules with a 12 Å buffer distance. All simulations were equilibrated using sander at 300 K and 1 atm using the following steps: (1) minimization of water and ions with weak restraints of 10.0 kcal/mol/Å on all nucleic acid and ligand residues (2000 cycles of minimization, 500 steepest decent before switching to conjugate gradient) and 10.0 Å cutoff, (2) heating from 0 K to 100 K over 20 ps with 50 kcal/mol/Å restraints on all nucleic acid and ligand residues, (3) minimization of entire system without restraints (2500 cycles, 1000 steepest decent before switching to conjugate gradient) with 10 Å cutoff, (4)

heating from 100 K to 300 K over 20 ps with weak restraints of 10.0 kcal/mol/Å on all nucleic acid and ligand residues, and (5) equilibration at 1 atm for 100 ps with weak restraints of 10.0 kcal/mol/Å on nucleic acid and ligand residues. The output from equilibration was then used as the input (.rst) file for 100 ns of unrestrained MD simulations using pmemd with GPU acceleration in the isothermal isobaric ensemble (P = 1 atm, T = 300 K). Periodic boundary conditions and PME were used. 2.0 fs time steps were used with bonds involving hydrogen frozen using SHAKE (ntc = 2). Trajectories were analyzed using the CPPTRAJ module in the AmberTools18 package. Small molecules were parameterized using the Antechamber (378) package with general AMBER force field (GAFF) (379) and AM1-BCC atomic charges (380). Calculations of theoretical relative Gibb's free energy (ΔG) of ligand-receptor complexes was achieved using the single-trajectory MMPBSA method (381). Trajectory residue interaction network analysis was performed on each trajectory using the structureViz (382) and RINalyzer (383) modules of the program Cytoscape (384) and UCSF Chimera v1.11 (176).

Molecular visualizations

All molecular visualizations of MD trajectories and models were performed in UCSF Chimera v1.11 (176).

Cell culturing

All cell lines were maintained in 5% CO_2 at 37°C and 95% humidity in media supplemented with 10% heat inactivated FBS, penicillin (100 U final concentration), and streptomycin (100 ug final concentration). HEK293 (ATCC CRL-1573) cells were grown in DMEM media while MCF7 (ATCC HTB-22) and MDA-MB-231 (ATCC HTB-26) cells were grown in EMEM supplemented with 0.01 mg/mL human recombinant insulin. AlamarBlue assays were conducted as outlined in the manual (Thermofisher #DAL1100) in 96-well clear bottom plates. Cells were treated with compounds for the indicated time and concentration and results are presented relative to the control (DMSO) treated.

Quantitative real-time polymerase chain reaction (qRT-PCR)

MCF7 cells were seeded at 2 x 10^5 cells/well in 6-well plates. After overnight attachment, media was replaced and treated with compounds in DMSO (as indicated in text), or DMSO alone (control), followed with 2 minutes of gentle mixing before placing back in the incubator. Media replacement and compound treatment were repeated twice more in 24-hour intervals such that the end time point was 72 hours of total treatment time. After 72 hours, cells were aspirated and washed before harvesting of total RNA with PureLink RNA mini kit (Invitrogen, #12183018A), followed by reverse transcription into cDNA using a high-capacity reverse transcription kit (Applied Biosystems #4368813). Quantitative PCR was performed using a standard SYBR Green Master Mix (Applied Biosystems #4309155) in 96-well plates on an Applied Biosystems StepOnePlus RT-PCR system using the standard ΔΔCt method. Primers were from PrimerBank (https://pga.mgh.harvard.edu/primerbank/) and verified as specific based on monophasic transitions during thermal denaturation. Primers (5' to 3'): hTERT F-TCCACTCCCCACATAGGAATAGTC, R- TCCTTCTCAGGGTCTCCACCT, c-MYC F-CGTCTCCACACATCAGCACAA, R-CACTGTCCAACTTGACCCTCTTG, GAPDH F-TGCACCACCAACTGCTTAGC, R- GGCATGGACTGTGGTCATGAG.

Western blotting

Total protein extracts prepared from MDA-MB-231 cells treated with either DMSO or compound treatment were subjected to SDS-PAGE followed by wet transfer to nitrocellulose membranes. Blots were rinsed 3x with TBST supplemented with milk and subsequently incubated with primary anti-telomerase reverse transcriptase antibody (Abcam, ab230527) or GAPDH (Abcam, ab9485) at 4°C overnight. Visualization was achieved by incubation and visualization of an anti-rabbit Alexa Fluor 488 conjugated secondary antibody using a PharosFX imaging system.

Data analysis

All data fitting, statistical analysis, and graphing were performed using Origin (version 2020) (OriginLab Corporation, Northampton, MA, USA).

Results

Virtual screening of the ZINC 2014 and 2016 drug-like small molecule databases

To begin our drug discovery campaign, we docked over 40 million virtual small molecules into sites located in the loops and grooves of the all-parallel stacked hTERT G-quadruplex model (**Figure 56**) using Surflex-Dock 2.11. Due to the size of the Kentucky DataseamGrid (http://www.kydataseam.com/), the entire docking and scoring procedure was accomplished in just under a month. From this screen, the top 500 scoring small molecules from each protomol site and library (2 libraries * 12 docking sites * 500 molecules = 12,000 top scoring molecules) were pooled and clustered by a similarity criterion to group core scaffolds. The centroid molecules, which are the most representative of each clade, were visually inspected within their docked sites to ensure reasonable contacts were made. A final list of centroid molecules was then compiled and used as input for a search at Molport.com. In total, 69 out of 100 searched compounds were available and purchased for testing.

Experimental screening of first-generation compounds

Screening of the 69 first generation small molecules was achieved using a high-throughput fluorescent thermal shift assay (FTSA) which was described previously (269). Unfortunately, due to the thigh GC content of the full-length hTERT sequence (TERT-FL), we were not able to obtain a FRET labeled full-length oligonucleotides for screening purposes. Instead, we used truncated versions of the TERT-FL which have previously been characterized (143).

The first (and 5' most) putative quadruplex sequence, PQS1, which was solved previously (105), and two other sequences that lack either the 3' PQS (PQS12) or the 5' PQS (PQS23) G4 motif were used in place of the hTERT-FL for screening (**Table 9**). We also included an array of promoter and telomere PQSs with the intention of quickly vetting for hTERT selective molecules (**Table 10**). Initial screening of the first-generation compounds was carried out in a phosphate buffer with physiological (185 mM) K^+ concentration. Subtle shifts in melting temperature (ΔT_m) were

observed for the hTERT sequences with only 3/69 compounds, ranging from ~0.5-2.5 °C (**Figure 57**). The hTERT G-quadruplex is extremely stable in physiological K^+ concentrations [likely owing to the added stability through stacking (54,143,144)] and so we decided to re-screen at slightly lower (100 mM) K^+ concentrations. Reducing the potassium confirmed that compounds 1-3 bound to at least one of the larger hTERT constructs (PQS12 or PQS23), but not a control hairpin sequence (HP) (**Figure 57A**, compounds 1, 2, and 3). This assay allowed for both high-throughput assessment of binding as well as an indication of selectivity over an array of hairpin and G4 topologies.

Table 9. hTERT oligonucleotides used throughout this study.

NAME	OLIGONUCLEOTIDE SEQUENCE 5' TO 3'	LENGTH (NT)	MW	E260 ($M^{-1}CM^{-1}$)
TERT-FL	GGGGAGGGGCTGGGAGGGCCCGGAGGGGGCTGGGCCGGGGACCCGGGAGGGGTCGGGACGGGGCGGGG	68	21633	672671
PQS1	AGGGGAGGGGCTGGGAGGGC	20	6369	202900
PQS12	AGGGGAGGGGCTGGGAGGGCCCGGAGGGGGCTGGGCCGGGGACCCGGGA	49	15523	478700
PQS23	AGGGGGCTGGGCCGGGGACCCGGGAGGGGTCGGGACGGGGCGGGG	45	14278	436500

Table 10. Additional DNA oligonucleotides used in this study.

Name	Sequence (5' to 3')	Length (nt)	MW (g/mol)
143D	AGGGTTAGGGTTAGGGTTAGGG	22	6967
2GKU	TTGGGTTAGGGTTAGGGTTAGGGA	24	7575
2HY9	AAAGGGTTAGGGTTAGGGTTAGGGAA	26	8219
2JSM	TAGGGTTAGGGTTAGGGTTAGGG	23	7271
2JPZ	TTAGGGTTAGGGTTAGGGTTAGGGTT	26	8183
2JSL	TAGGGTTAGGGTTAGGGTTAGGGTT	25	7879
2KF8	GGGTTAGGGTTAGGGTTAGGGT	22	6958
2KKA	AGGGTTAGGGTTAGGGTTAGGGT	23	7271
Tel48	TTAGGGTAAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTAGGGTTA GGGTT	50	15829
2O3M	AGGGAGGGCGCTGGGAGGAGGG	22	7012
Rb	CGGGGGGTTTTGGGCGGC	18	5644
Her2	AGGAGAAGGAGGAGGTGGAGGAGGAGGGC	29	9276
Bcl-2	AGGGGCGGGCGCGGGAGGAAGGGGGCGGGAGCGGGGC	37	11799
VEGF	GGGCGGGCCGGGGGCGGGGTCCCGGCGGGGCGGGAG	36	11388
KRAS	GGGAAGAGGGAAGAGGGGGAGG	22	7069
Zeb1	GGGGTGGGGGGGAAGGGGGAGGGAGGGGG	29	9396
Hif1a	GCGCGGGGAGGGGAGAGGGGGCGGGAGCGCG	31	9879
c-myc	TGGGGAGGGTGGGGAGGGTGGGGAAGG	27	8688
1XAV	TGAGGGTGGGTAGGGTGGGTAA	22	6992
I-motif telo	CCCTAACCCTAACCCTAACCCT	22	6504
I-motif c-myc	TTCCCCACCCTCCCCACCCTCCCCTAA	27	7917
AT Hairpin	ATATATATATCCCCATATATATAT	24	7269
GC Hairpin	GCGCGCGCGCTTTTGCGCGCGCGC	24	7339

Figure 57. Generation 1 FTSA screen. (A) Heatmap of Tm shifts for the first generation compounds in 185 mM K^+ potassium buffer vs. a DNA panel. HP is a FRET labeled hairpin control. (B) Structures of compounds 1-3.

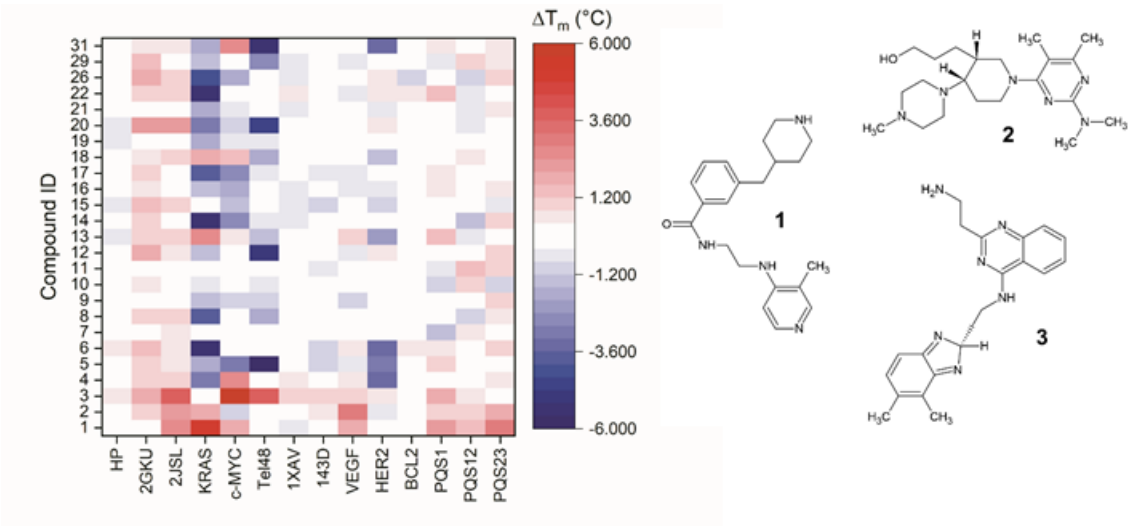
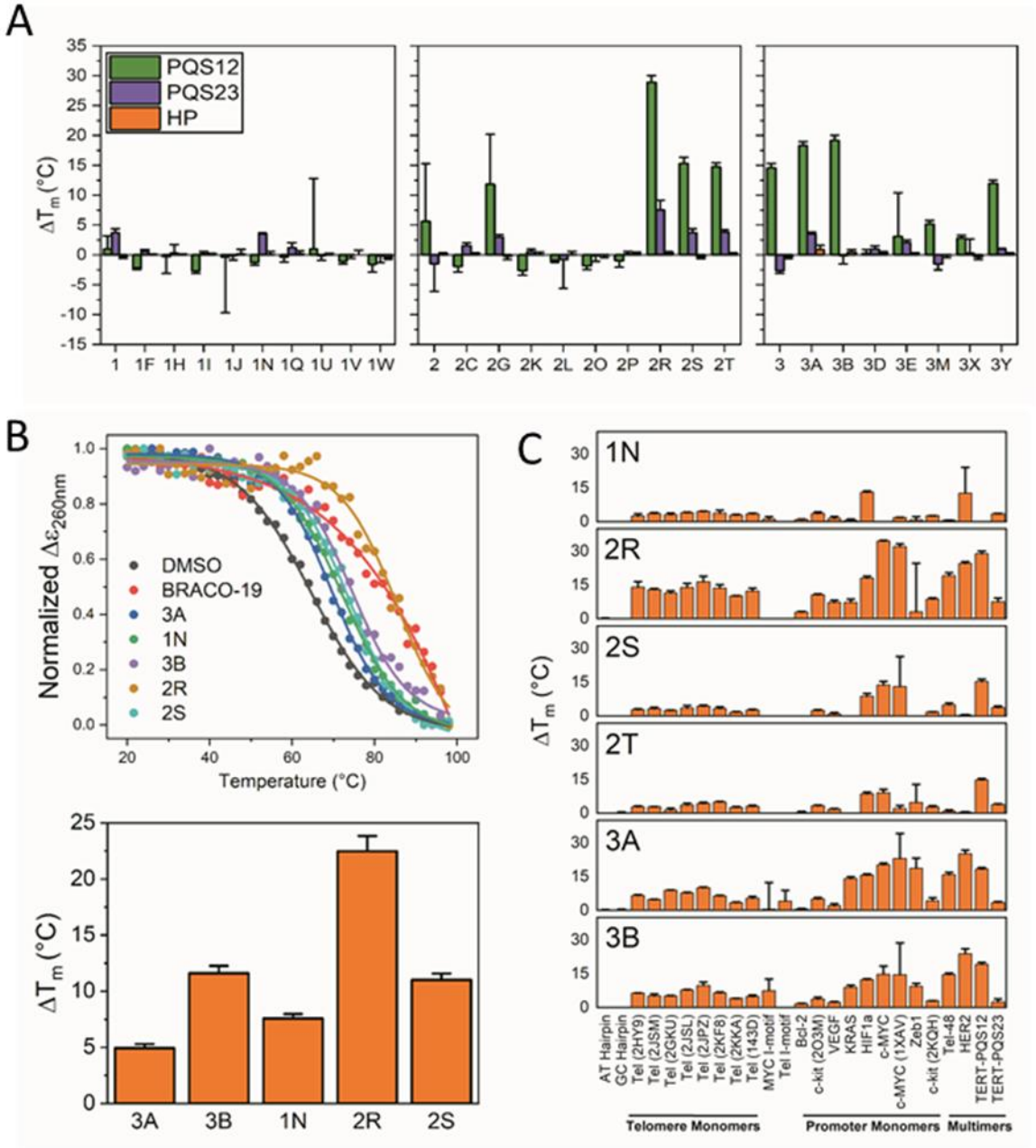

ΔTm (°C)
6.000
3.600
1.200
-1.200
-3.600
-6.000
Compound ID
HP
2GKU
2JSL
KRAS
c-MYC
Tel48
1XAV
143D
VEGF
HER2
BCL2
PQS1
PQS12
PQS23
1
2
3

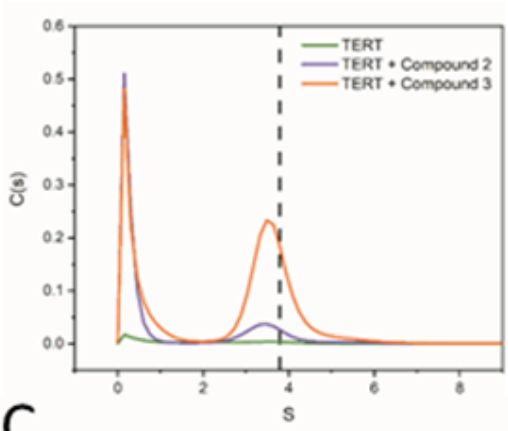
Figure 58. First and second generation FTSA screening. (A) FTSA results for screening compounds 1, 2, and 3, along with their derivatives (buffer consisted of 20 mM K^+). HP is a FRET labeled hairpin control. (B) Representative CD melting profiles and ΔT_m measurements for each of the indicated small molecules. (C) Results of FTSA DNA panel screening with indicated compounds. A few small (~3°C or less) negative shifts in melting temperature was observed in some cases but have been omitted here for the sake of clarity.


A
PQS12
PQS23
HP
ΔTm (°C)
1 1F 1H 1I 1J 1N 1Q 1U 1V 1W
2 2C 2G 2K 2L 2O 2P 2R 2S 2T
3 3A 3B 3D 3E 3M 3X 3Y
B
Normalized Δε260nm
DMSO
BRACO-19
3A
1N
3B
2R
2S
Temperature (°C)
ΔTm (°C)
3A 3B 1N 2R 2S
C
1N
2R
2S
2T
3A
3B
ΔTm (°C)
AT Hairpin
GC Hairpin
Tel (2HY9)
Tel (2JSM)
Tel (2GKU)
Tel (2JSL)
Tel (2JPZ)
Tel (2KF8)
Tel (2KKA)
Tel (143D)
MYC I-motif
Tel I-motif
Bcl-2
c-kit (2O3M)
VEGF
KRAS
HIF1a
c-MYC
c-MYC (1XAV)
Zeb1
c-kit (2KQH)
Tel-48
HER2
TERT-PQS12
TERT-PQS23
Telomere Monomers
Promoter Monomers
Multimers

We next sought to confirm the interaction of compounds 1-3 with both the truncated and full-length hTERT constructs by orthogonal methods. As a “medium-throughput” method for assessment of binding, as well as estimation of stoichiometry, we turned to analytical ultracentrifugation (AUC) (210). This technique allowed us to confirm binding of both compounds 2 and 3 to the truncated and full-length hTERT G4s (**Figure 59**), but not compound 1 as it had an inconvenient absorption spectrum. Instead, using isothermal titration calorimetry (ITC), compound 1 was confirmed as binding to PQS23, but not PQS12 (**Figure 59C**), consistent with FTSA results (**Figure 58**), and showed comparable binding affinity to that of compounds 2 and 3 (dissociation constants, K_D, ranged from 10-80 μM).

Figure 59. Isothermal titration calorimetry and AUC binding of compounds 1-3. (A) Representative AUC C(s) curves showing enrichment of compounds 2 and 3 at ~3.8 S, approximately where TERT-FL sediments (3.9-4.0 S dashed line). (B) Table of binding stoichiometries for various compounds determined from area under the curve in A. (C) ITC binding energetics profiles of compounds 1-3 when titrated into PSQ12 or PQS23 sequences.

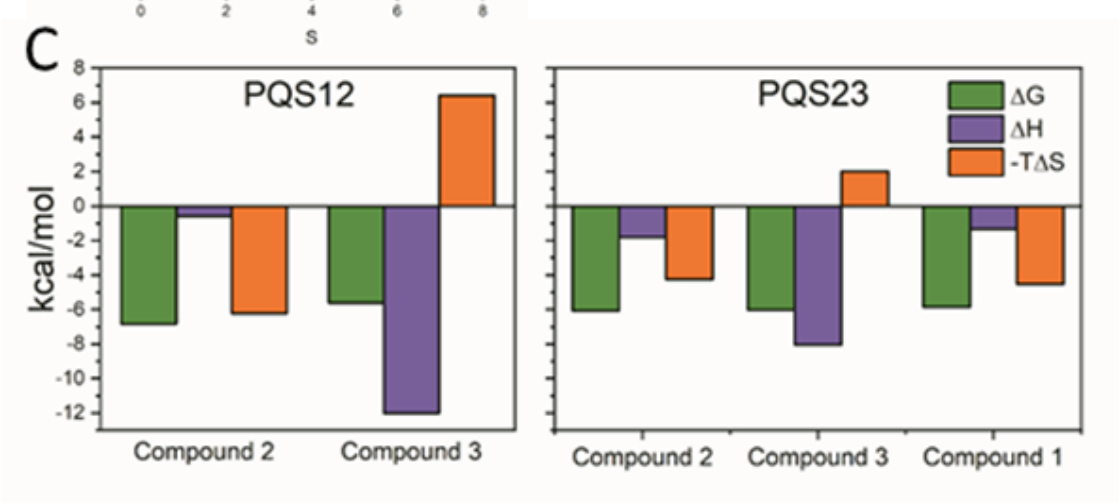
A



B

Compound ID	Binding Stoichiometry			
	PQS1	PQS12	PQS23	TERT-FL
2	1.5	1.4	1.7	2.6
3	3.3	3.1	3.8	8.6
3A	-	-	-	6.3
3B	-	-	-	7.9
2S	-	-	-	9.6

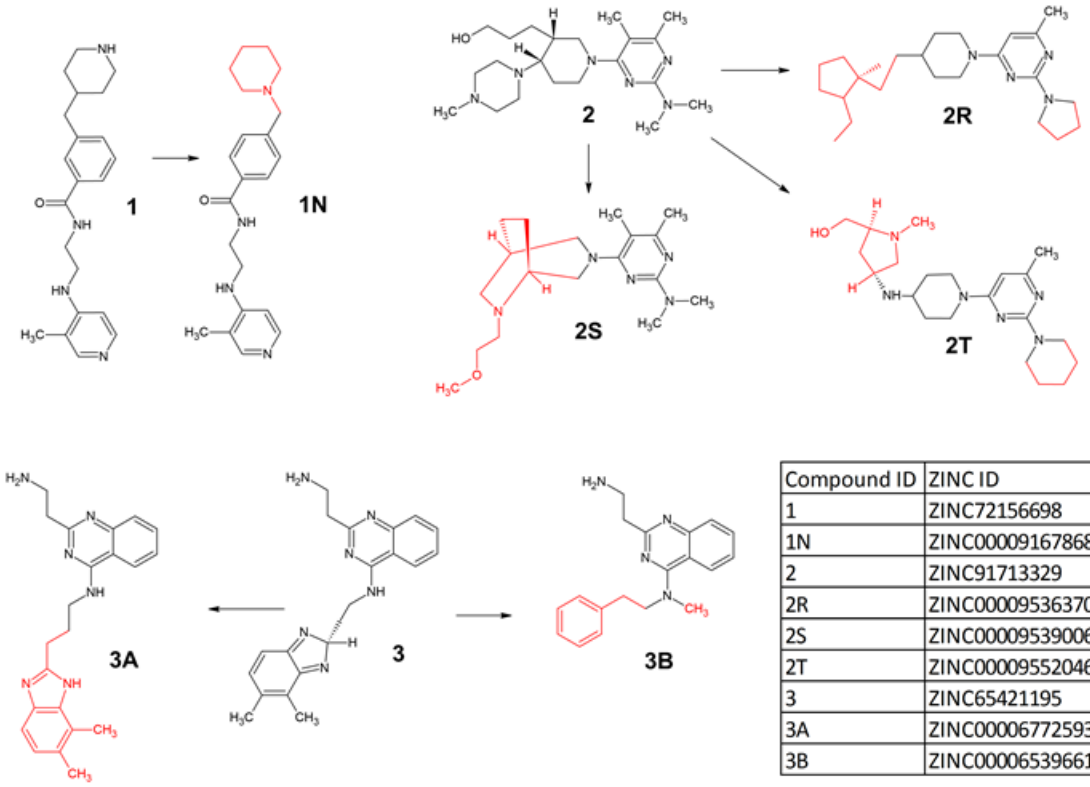
C



Selectivity assessment of second-generation compounds

The initial hits, compounds 1-3, had only moderate affinity, and so we proceeded to a structure activity relationship (SAR) type of approach. To this end we searched Molport.com for readily purchasable molecules with similar core scaffolds. In total, 25 molecules (arbitrarily designated with a letter, A-Y, preceded by the number of the compound they were derived from) were selected for purchasing and testing. **Figure 58A** shows the results of FTSA screening of compounds 1-3, along with their derivatives. Immediately it is evident that there is differential binding due to R-group modifications (which are shown in **Figure 60**). Compound 1 resulted in a single molecule with a comparable T_m shift (1N), whereas there were multiple derivatives of compounds 2 and 3 that exhibited enhanced stabilizing ability (2G, 2R, 2S, 2T and 3A, 3B, 3Y). These derivatives also exhibited differential binding to the hTERT truncated sequences PQS12 and PQS23. Only one molecule, 3A, had noticeable binding to the control hairpin duplex in FTSA screening experiments (**Figure 58A**). The compounds 2S, 3A, and 3B were subsequently confirmed to bind the full-length hTERT G4 (TERT-FL) by AUC binding experiments (**Figure 59B**) and compounds 1N, 2R, 2S, 3A, and 3B by CD melting studies (**Figure 58B**).

Figure 60. Structural variations of first-generation compounds (1-3) with their “SAR by catalogue” derivatives. Red is used to help visualize structural changes between parent molecules and derivatives.

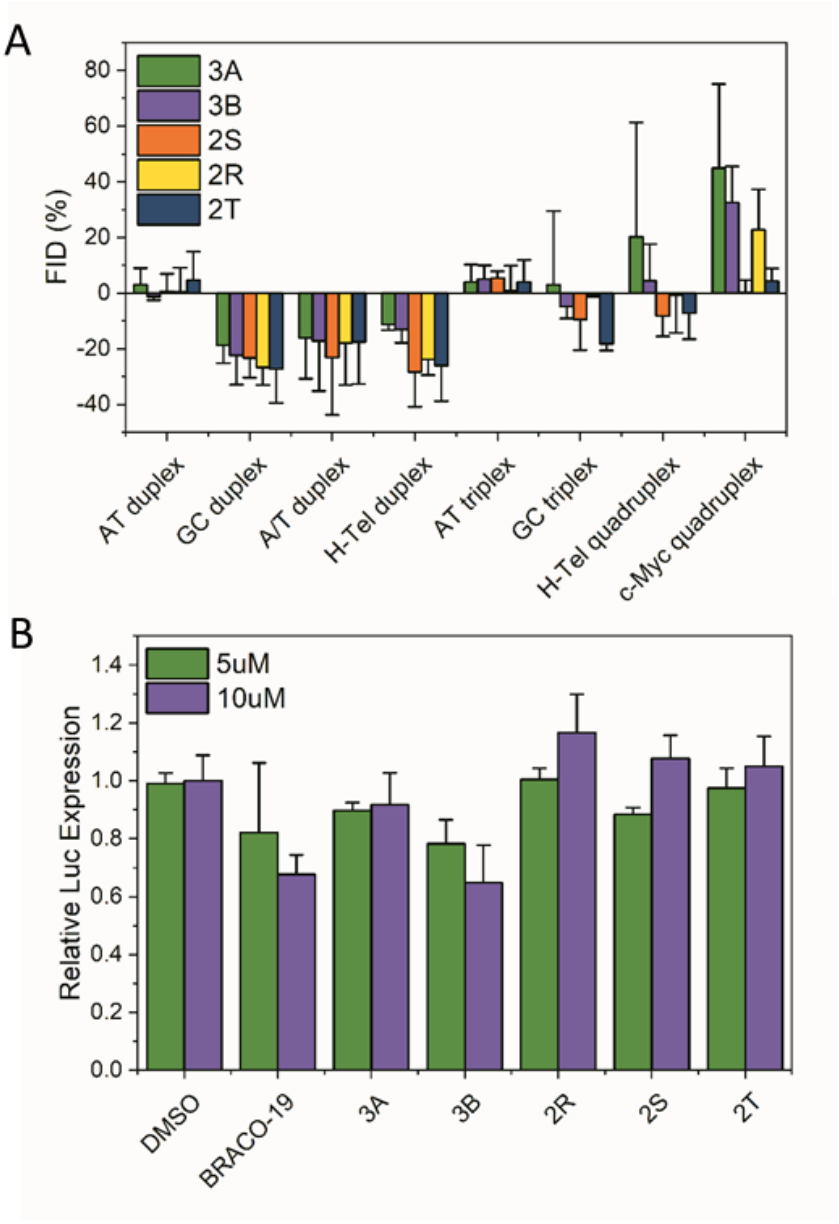


Compound ID	ZINC ID
1	ZINC72156698
1N	ZINC000091678681
2	ZINC91713329
2R	ZINC000095363705
2S	ZINC000095390066
2T	ZINC000095520468
3	ZINC65421195
3A	ZINC000067725939
3B	ZINC000065396611

As selectivity is of primary interest, we next wanted to investigate the DNA interaction profiles of the most stabilizing derivatives (3A, 3B, 2R, 1N, 2S, and 2T) against as many types of DNA topologies as possible (**Figure 58C**). Based on T_m shifts, we found that there was less discrimination among various G-quadruplex topologies than expected [it should be emphasized that T_m shifts are a function of affinity, stoichiometry, enthalpies of binding, and enthalpies of denaturation so direct comparisons are not necessarily reflective of differences in affinity (385)]. We did, however, observe differential interactions among the molecules, suggestive of distinct binding modes. For instance, although 2R, 2S, and 2T were all derived from compound 2, their binding profiles vary markedly. 2R has a profound thermal stabilizing effect (ΔT_m = +30 °C in some cases) and is indiscriminate among G-quadruplex DNA, whereas 2S and 2T exhibit some selectivity. Another differential interaction was observed with 3A and 3B. Although they have very similar profiles, only 3B stabilizes both the c-MYC derived G-quadruplexes (c-MYC, 1XAV) as well as the c-MYC I-motif. Further, 3A, 3B, 2R, and 1N all bind to the HER2 G-quadruplex, while 2S and 2T show no interaction. Last, we find that none of the derivatives tested bind the two duplexes (AT and GC duplex), which indicates their selectivity over duplex DNA.

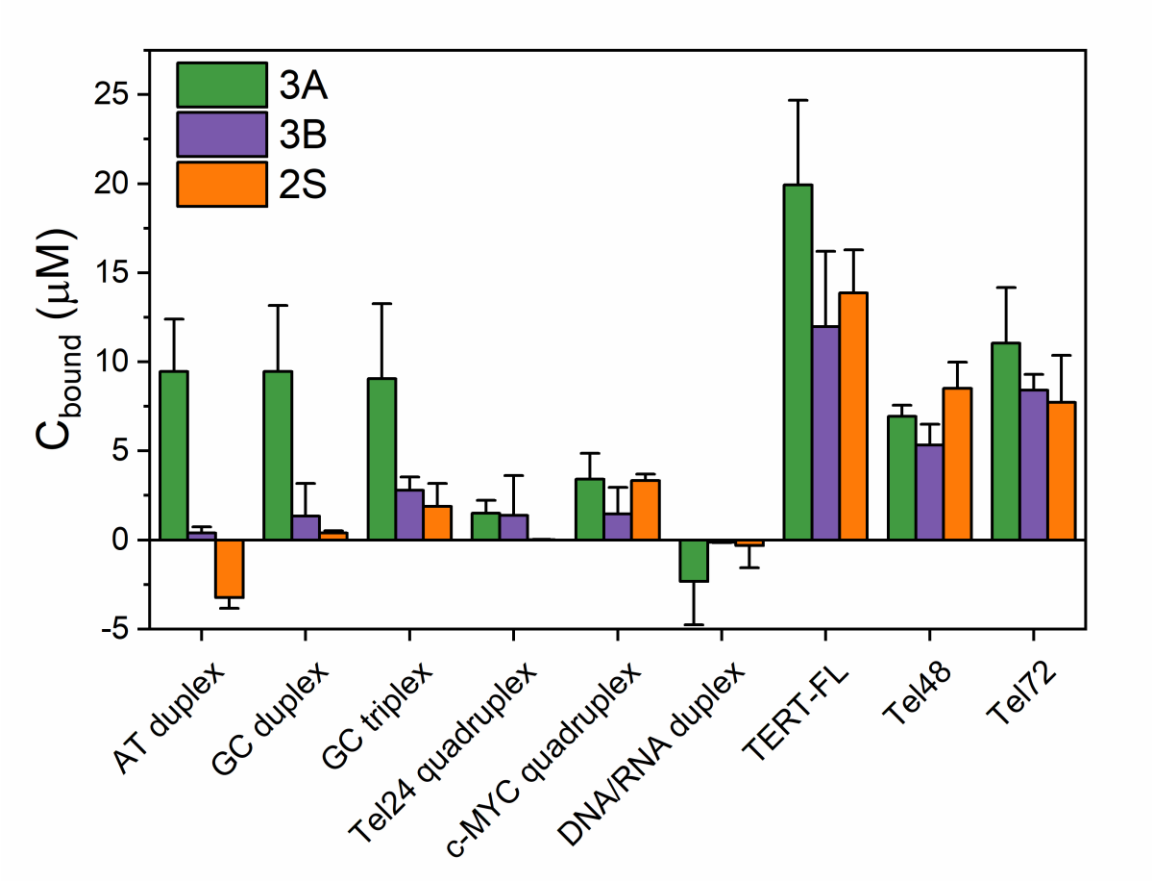
Many of the small molecules discovered through our FTSA screen had poor spectroscopic properties, with absorption spectra overlapping with DNA's intrinsic absorption. The poor spectroscopic properties being a limitation, we proceeded to a fluorescent indicator displacement assay (FID) (196,386), which can provide information on binding affinity as well as topological DNA preferences without relying on intrinsic spectroscopic properties of the test molecules. The results were somewhat ambiguous (**Figure 61**). First, compounds 2S and 2T appeared to be either too low in affinity to out compete the fluorescent indicator (TO) or were not cognate ligands for the DNA topologies tested. 3A, 3B, and 2R all showed only week preference for the parallel c-MYC G-quadruplex over other topologies, demonstrating that they have moderate affinities (K_D in the range of ~10-100 μM) for the parallel c-Myc structure. We also observed a large negative FIDs for all five compounds with respect to the GC duplex, A/T duplex, and H-Tel duplex. This phenomenon has been observed elsewhere (196) and is not readily explainable. Overall, this assay points to modest selection for the parallel type G4 structure with 3A, 3B, and 2R.

Figure 61. Results of FID and luciferase assays using second generation molecules. (A) FID assay showing % TO displaced based on measured fluorescence relative to controls for each indicated molecule against each given topology. (B) Relative reduction in luciferase expression based on a dual-reporter assay. In each case, samples are displayed as relative to DMSO control. Two different DMSO controls are used, as we found that DMSO had a small but significant dose-dependent effect in the ranges used in these assays.


A
3A
3B
2S
2R
2T
FID (%)
80
60
40
20
0
-20
-40
AT duplex
GC duplex
A/T duplex
H-Tel duplex
AT triplex
GC triplex
H-Tel quadruplex
c-Myc quadruplex
B
5uM
10uM
Relative Luc Expression
1.4
1.2
1.0
0.8
0.6
0.4
0.2
0.0
DMSO
BRACO-19
3A
3B
2R
2S
2T

Compounds 3A, 3B, and 2S have absorption spectra above 300 nm allowing for their use in absorption-based selectivity assays. One of the most powerful selectivity assays is the competition dialysis assay. In this assay, direct comparisons can be made between affinities of small molecules for any DNA structure based on the local enrichment of molecules inside the membrane due to strength of interaction (375). **Figure 62** displays the results of a competition dialysis experiment using 3A, 3B, and 2S. Compound 3A displays promiscuity with duplex, triplex, and G-quadruplex DNA (**Figure 58**). Conversely, compounds 3B and 2S exhibited a high selectivity for the larger, multimeric G-quadruplexes (TERT-FL, Tel48, and Tel72) over monomeric G4s, duplex or triplex DNAs. It is important to point out that the concentration of DNA used in the competition assays is the same in terms of monomeric unit (i.e. base pairs for duplex forms, triplets for triplex, and tetrads for quadruplexes). This allows for the direct comparisons in affinity across all receptor sizes and helps to account for multiple binding sites (375). Using the bound concentrations (C_{bound}), we are also able to calculate apparent dissociation constants (387), which are 24 μM for 3B and 22 μM for 2S with respect to TERT-FL, 45 μM or greater for Tel48 and Tel72, and ≥140 μM for all others. These affinity values are consistent with the prediction from the FID assay (**Figure 61**) and ITC data for their respective first-generation scaffolds (**Figure 59**). Overall, these data show that 3B and 2S have a strong preference for G-quadruplex DNA over duplex and triplex conformations and show preference for TERT-FL over the telomere G-quadruplex multimers (Tel48 and Tel72).

Figure 62. Results of competition dialysis. C_{bound} is the amount of compound in micromolar that was locally increased inside the dialysis membrane after background subtractions, directly reflecting the differential affinity of ligands with receptors. All DNA structures are at the same monomeric concentration, and so the higher local concentration of compound is interpreted directly as higher affinity.


3A
3B
2S
C_{bound} (μM)
25
20
15
10
5
0
-5
AT duplex
GC duplex
GC triplex
Tel24 quadruplex
c-MYC quadruplex
DNA/RNA duplex
TERT-FL
Tel48
Tel72

During the above *in vitro* discovery and screening processes we also worked toward generating an in-cell reporter method for screening. Others have previously demonstrated that luciferase-based G-quadruplex promoter assays allow for direct detection of transcriptional changes based on G4 formation and stabilization (223,388,389). Here, we used a previously validated (222) luciferase vector containing the full-length hTERT core promoter to screen for transcriptional down-regulation (**Figure 61B**). To our surprise, all derivatives of compound 2 (2R, 2S, and 2T) that were tested exhibited dose-dependent increases in luciferase expression. No changes were observed in cells treated with 3A. Only 3B and the positive control BRACO-19 (223) exhibited the anticipated dose-dependent reduction in luciferase expression. Thus, only compound 3B, a di-substituted quinazoline (**Figure 60**), was further pursued. Importantly, as of the time of writing, the core scaffold of 3B is unique and has not been reported as a DNA binding molecule.

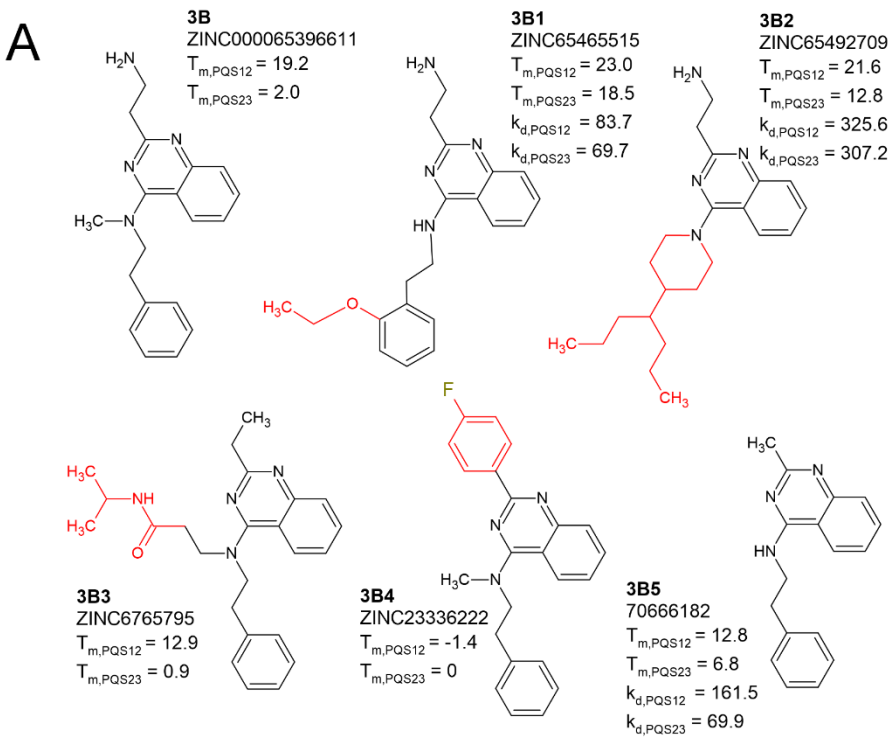
SAR of compound 3B derivatives

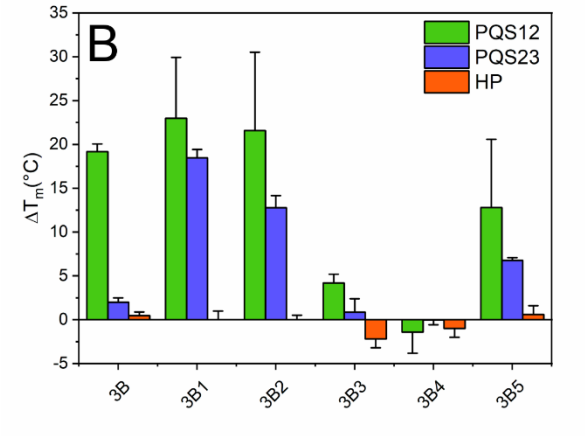
One of the drawbacks of virtual drug discovery approaches is the potential for unavailability of small molecules, whether that be by purchasing or synthesis. The most selective compound, 3B, was not able to be synthesized. Therefore, we proceeded to another round of "SAR by catalogue", with the intention of discovering a more selective and readily synthesizable small molecule. We purchased 5 derivatives of molecule 3B based on a structural similarity search (>80% similarity via Molport.com) and designated them as 3B1 through 3B5 (**Figure 63A**). **Figure 63B** shows the results of the FTSA thermal shift analysis, comparing compound 3B to each of its derivatives. We find that modifications made to the 2-phenylethyl moiety (such as in 3B1 and 3B2) resulted in minimal perturbations to T_m shifts with respect to PQS12, but large increases in T_m with respect to PQS23. 3B3, with the addition of the N-(propan-2-yl)acetamide side chain to the tertiary amine and removal of the terminal 2-aminoethyl group, had a major reduction in thermal stabilizing effect on PQS12 compared to the former. 3B3 also exhibited destabilization of the control hairpin (HP), which is undesirable as this indicates interaction with duplex DNA. The most substantial change in thermal stabilizing effect was observed with the swapping of the 2-aminoethyl group for a fluorophenyl moiety, which prevented binding entirely (3B4). The last compound, 3B5, showed a reduction in

thermal stabilization of PQS12 with a concomitant increase in stabilization of PQS23, likely indicating that the methyl group on the tertiary amine reduces interaction with PQS23.

Apparent binding affinities for 3B1, 3B2, and 3B5 were able to be estimated from FTSA titration experiments (**Figure 64**) (assuming similar binding stoichiometries). Their apparent K_D values are given in **Figure 63A** (3B3 and 3B4 Tm could not be fit to any binding models). These affinity estimations, while much higher than expected (which is reasonable when considering the dependency of K_D on temperature), indicate that compounds 3B1 and 3B5 bind most tightly to the truncated hTERT G-quadruplex sequences. These analyses also highlight the fact that T_m shift data from a single drug concentration may not necessarily approximate affinities (e.g. at saturation, 3B1 increases PQS23 melting temperature by 18.5 °C, with K_D = 69.7 μM, while 3B5 only increases the melting temperature by 6.8°C, yet yields a similar K_D = 69.9 μM). Subsequent ITC titration analysis of 3B1 to hTERT-FL showed a K_{D1} = ~3-4 uM, and that the binding is best fit by a three-site model (**Figure 63C**). Altogether, this data indicates that the 2-aminoethyl group plays a role in the interaction with PQS12, and that by removing the methyl group from the tertiary amine the affinity for PQS23 increases.

Figure 63. “Catalogue SAR” analysis of compound 3B derivatives. (A) Structures of 3B and derivatives 3B1-5. Red is intended to help visually emphasize the areas where molecular additions were made. The inset descriptions are (in order): compound ID, ZINC ID, FTSA T_m shifts in degrees Celsius, and dissociation constants, K_D as measured in FTSA titration experiments. (B) Plot of FTSA T_m shifts of 3B and its derivatives. (C) Representative ITC data of 3B1 titrated into hTERT-FL fit to a three-site model with fit and K_D values inset and residuals on bottom.


A
3B
ZINC000065396611
$T_{m,PQS12}$ = 19.2
$T_{m,PQS23}$ = 2.0
3B1
ZINC65465515
$T_{m,PQS12}$ = 23.0
$T_{m,PQS23}$ = 18.5
$k_{d,PQS12}$ = 83.7
$k_{d,PQS23}$ = 69.7
3B2
ZINC65492709
$T_{m,PQS12}$ = 21.6
$T_{m,PQS23}$ = 12.8
$k_{d,PQS12}$ = 325.6
$k_{d,PQS23}$ = 307.2
3B3
ZINC6765795
$T_{m,PQS12}$ = 12.9
$T_{m,PQS23}$ = 0.9
3B4
ZINC23336222
$T_{m,PQS12}$ = -1.4
$T_{m,PQS23}$ = 0
3B5
70666182
$T_{m,PQS12}$ = 12.8
$T_{m,PQS23}$ = 6.8
$k_{d,PQS12}$ = 161.5
$k_{d,PQS23}$ = 69.9


B
PQS12
PQS23
HP
ΔT_m(°C)
35
30
25
20
15
10
5
0
-5
3B
3B1
3B2
3B3
3B4
3B5

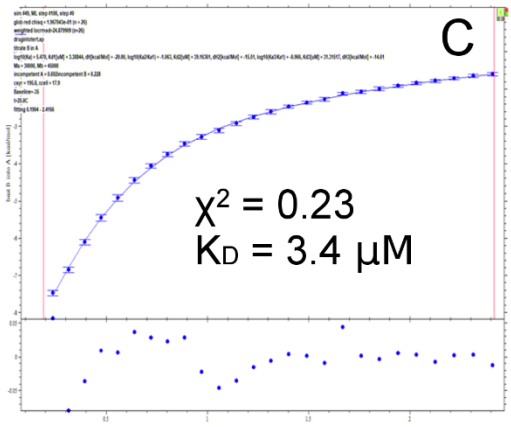
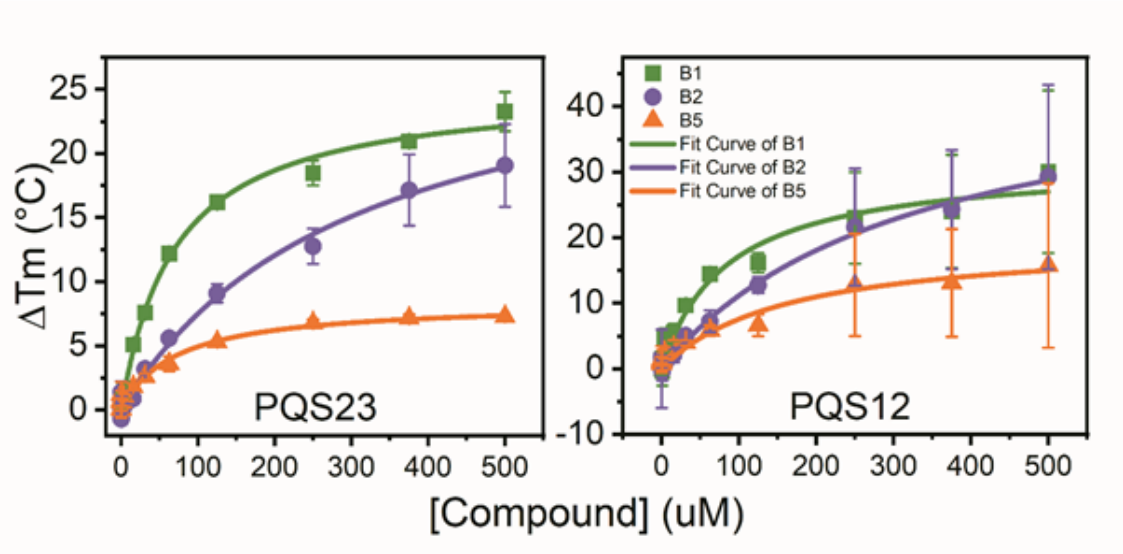
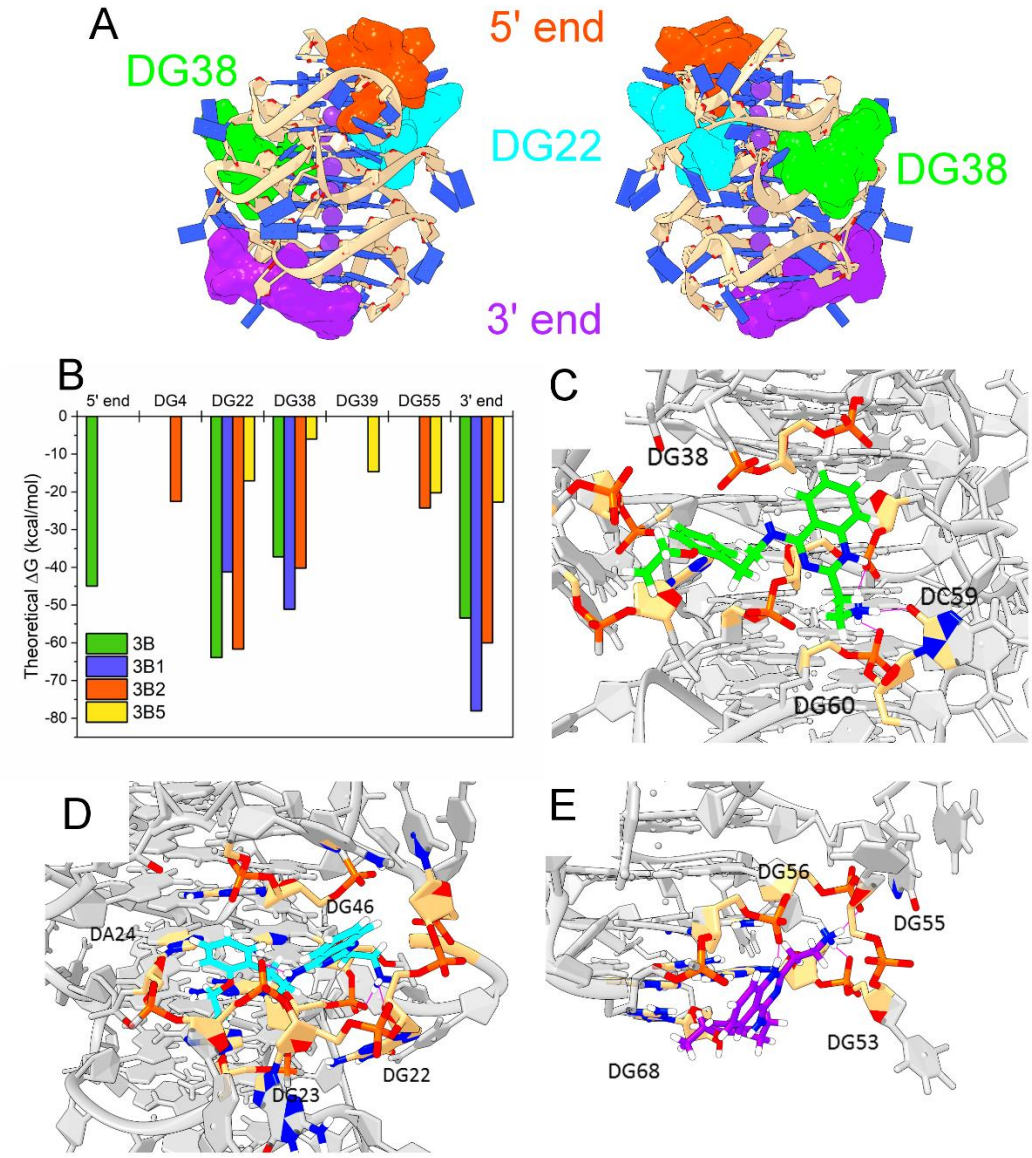

C
χ^2 = 0.23
K_D = 3.4 µM

Figure 64. FTSA dose-response curves for indicated compounds against both truncated TERT constructs: PQS12 and PQS23. Data were fit to a standard, single-site binding model. All fits had $R^2 \geq 0.97$.


ΔTm (°C)
PQS23
PQS12
[Compound] (uM)
B1
B2
B5
Fit Curve of B1
Fit Curve of B2
Fit Curve of B5

To gain some understanding of the differential binding of compound 3B and its derivatives, we conducted extensive virtual docking with post-docking MD simulations (**Figure 65**). Docking sites around loops, grooves, and terminal tetrad faces (3' and 5' ends) of the hTERT model from **Figure 56** were generated, followed by Glide XP (376) flexible docking of 3B and its derivatives, 3B1-5. Overall, seven sites had deep enough grooves identified for docking. Due to the high binding stoichiometry observed for compound 3B via AUC measurements (**Figure 58A-B**) and the ITC analysis indicating three sites bound by 3B1 (**Figure 61C**), multiple binding sites were investigated for each derivative. Using the top scoring docked positions for each molecule (as well as conformers and tautomers) we conducted 5 ns explicit solvent MD simulations (totaling 615 nanoseconds of simulation time). **Figure 65A** depicts the sites targeted in Glide docking. **Figure 65B** shows the post-MD theoretical Gibbs free energy values for the highest energy interaction poses calculated from evenly spaced frames of the trajectory. Surprisingly, the free energy values trend with the FTSA data in **Figure 64** (i.e. 3B, 3B1, and 3B2 trend toward higher affinity than 3B3, 3B4, and 3B5). Further, neither 3B1 or 3B2 were able to dock at the 5' end, and only 3B2 could bind to the site "DG4", indicating that the major stabilizing binding site for 3B1 and 3B2 resides in the loop pocket designated as "DG22". This pocket is formed by the connecting loop between the 5'-most and middle G4 units (**Figure 65D**), indicating that these molecules prefer loop/G4-junctions. Similarly, 3B1 and 3B2 interact strongly at site "DG38" and stabilize the phosphate backbone across the junction of the middle and 3'-most G4 units. 3B1 is predicted to have the highest affinity in a groove nearest the 3' tetrad face, with partial stacking on the terminal tetrad. Overall, this suggests that 3B1 has only three preferred sites, all of which involve groove and loop interactions, which is consistent with ITC results.

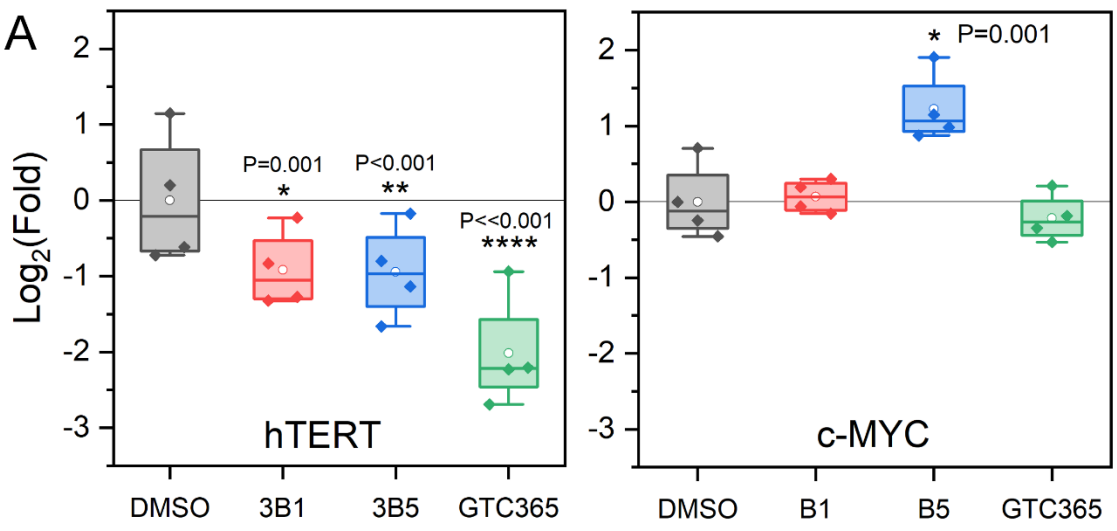
Figure 65. Glide XP docking and MD trajectory analysis of compound 3B and derivatives. (A) Schrodinger Sitemap docking sites (only the main 4 sites) shown as colored surfaces mapped onto the hTERT G-quadruplex structure. (B) Calculated theoretical Gibb's free energy of interaction for each compound-receptor combination from 5 nanoseconds of fully solvated MD. Scores are representative of only the largest free energy values for each ligand at each site. Missing values indicate that no docking solution was found. (C-E) Representative interactions of 3B1 docked in each of its three predicted binding sites (C – DG38, D – DG22, and E – 3' end). Ligand colors correspond to sites in A. Labeled residues in C-E indicate the residues essential to interaction.

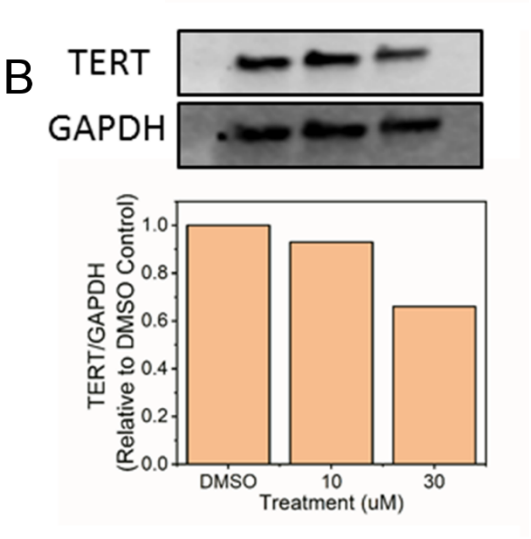

A
DG38
5' end
DG22
3' end
DG38
B
5' end
DG4
DG22
DG38
DG39
DG55
3' end
Theoretical ΔG (kcal/mol)
0
-10
-20
-30
-40
-50
-60
-70
-80
3B
3B1
3B2
3B5
C
DG38
DC59
DG60
D
DG46
DA24
DG22
DG23
E
DG56
DG55
DG53
DG68

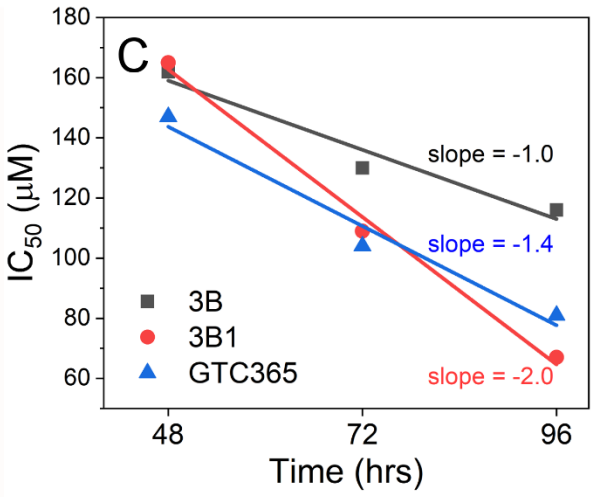
Biological assessment of compound 3B1

The above investigations demonstrate that a variety of these small molecules bind and stabilize the *hTERT* core promoter G-quadruplex multimer *in vitro* and have plausible binding modes in loop regions. We next wanted to see if a biological response would occur in cells known to have elevated hTERT expression with a wild-type core promoter (366). We began with quantitative PCR studies in MCF7 breast cancer cells (**Figure 66**). At 10 μM and 72 hours of treatment with compounds 3B1 and 3B5, we observed a statistically significant reduction in hTERT mRNA. Since *c-MYC* is the most thoroughly studied of all oncogene promoter G4s and its protein product can act directly on the *hTERT* promoter to modulate expression, we included it as a test for selectivity. We find that compound 3B5, but not 3B1 or GTC365, increases c-MYC levels significantly, indicating that 3B5 is not selective. The ability of 3B1 to reduce hTERT protein was subsequently confirmed by western blotting using the triple negative breast cancer cell line MDA-MB-231, which has a naturally higher level of hTERT protein expression (**Figure 66B**). Note, this seemingly subtle reduction in hTERT protein by western blotting is consistent with GTC365 (223), and studies using siRNA to directly knockdown hTERT (390,391).

Figure 66. Biological assessment of compounds 3B1 and 3B5. (A) Changes in hTERT and c-MYC mRNA levels as determined by RT-qPCR analysis with 10 μM treatment of compounds 3B1 and 3B5 for 72 hr. in MCF7 breast cancer cells. Measurements are normalized to GAPDH and displayed as relative to DMSO control. Statistical analysis was by two-way ANOVA and Tukey post-hoc test (alpha level of 0.05, n=4, each data point is from an independent experiment conducted on different days). (B) Immunoblot analysis and densitometry for hTERT protein after 10 and 30 μM treatment with compound 3B1 of MDA-MB-231 breast cancer cells for 72 hours. (C) IC_{50} measurements *verses* time for compounds 3B, 3B1, and GTC365 in MCF7 cells.

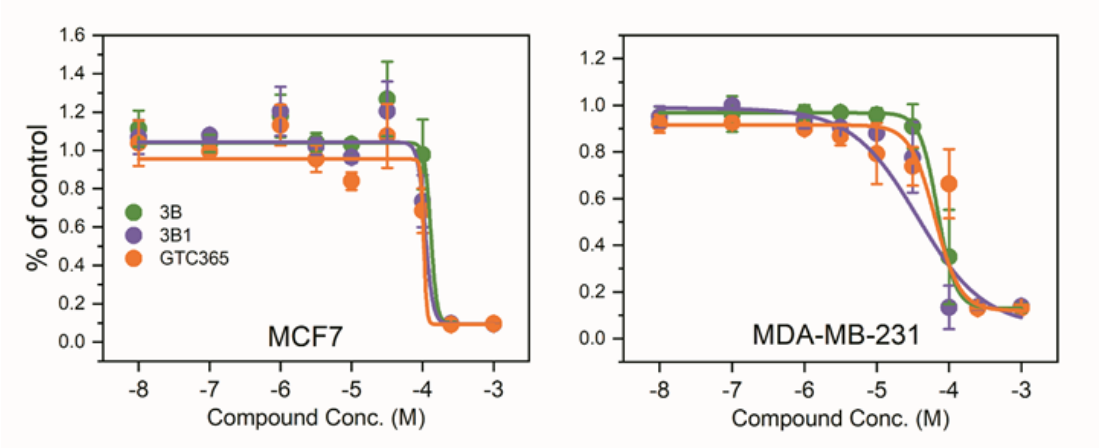

A
Log2(Fold)
2
1
0
-1
-2
-3
P=0.001
*
P<0.001
**
P<<0.001
****
hTERT
DMSO
3B1
3B5
GTC365
* P=0.001
c-MYC
DMSO
B1
B5
GTC365


B
TERT
GAPDH
TERT/GAPDH
(Relative to DMSO Control)
1.0
0.8
0.6
0.4
0.2
0.0
DMSO
10
30
Treatment (uM)


C
IC50 (μM)
180
160
140
120
100
80
60
slope = -1.0
slope = -1.4
slope = -2.0
3B
3B1
GTC365
48
72
96
Time (hrs)

We next investigated cell proliferation in the presence of 3B and 3B1 compared to GTC365 in breast cancer cells. Overall, IC_{50} values were approximately the same in both MCF7 and MDA-MB-231 cells (**Figure 67**). The values of IC_{50} measured between the two cells lines are consistent with an earlier report for GTC365, but differ slightly in magnitude (223). Because telomerase reduction can lead cells to a state of metabolically active senescence, rather than apoptosis (392), the IC_{50} values measured might not necessarily correlate with the effects of hTERT repression. Interestingly, by plotting the measured IC_{50} values in MCF7 cells versus time, we find that compound 3B1's IC_{50} has a much stronger dependency on time than GTC365. The dependency of IC_{50} on time is expected for agents that selectively repress hTERT expression (390,391). Altogether, these biological data support that 3B1 can repress human telomerase in breast cancer cells.

Figure 67. IC_{50} curves for compounds 3B, 3B1, and GTC365 in MCF7 and MDA-MB-231 breast cancer cells. Cells were treated for 72 hr. with compounds as indicated and analyzed by AlamarBlue assay. Data are reported relative to DMSO treated control cells. Data were fit to a standard dose-response model in Origin, from which IC_{20}, IC_{50}, and IC_{80} were obtained.



MCF7						
Compound	IC_{20} (μM)	Error	IC_{50} (μM)	Error	IC_{80} (μM)	Error
B	150	54	130	36	114	30
B1	129	67	109	23	92	19
H1	111	>1000	104	>1000	97	>1000

MDA-MB-231						
Compound	IC_{20} (μM)	Error	IC_{50} (μM)	Error	IC_{80} (μM)	Error
B	103	9	69	6	47	5
B1	158	146	37	22	9	3
H1	111	39	64	22	37	16

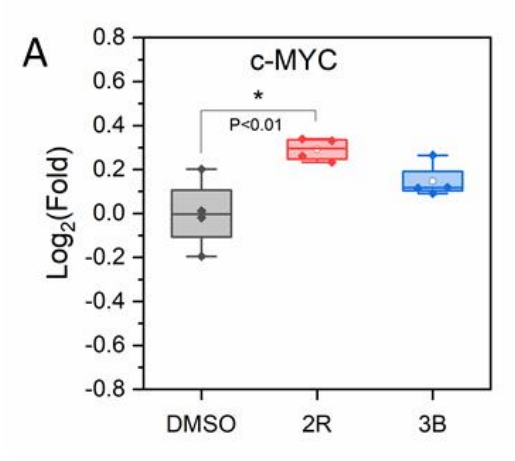
Discussion

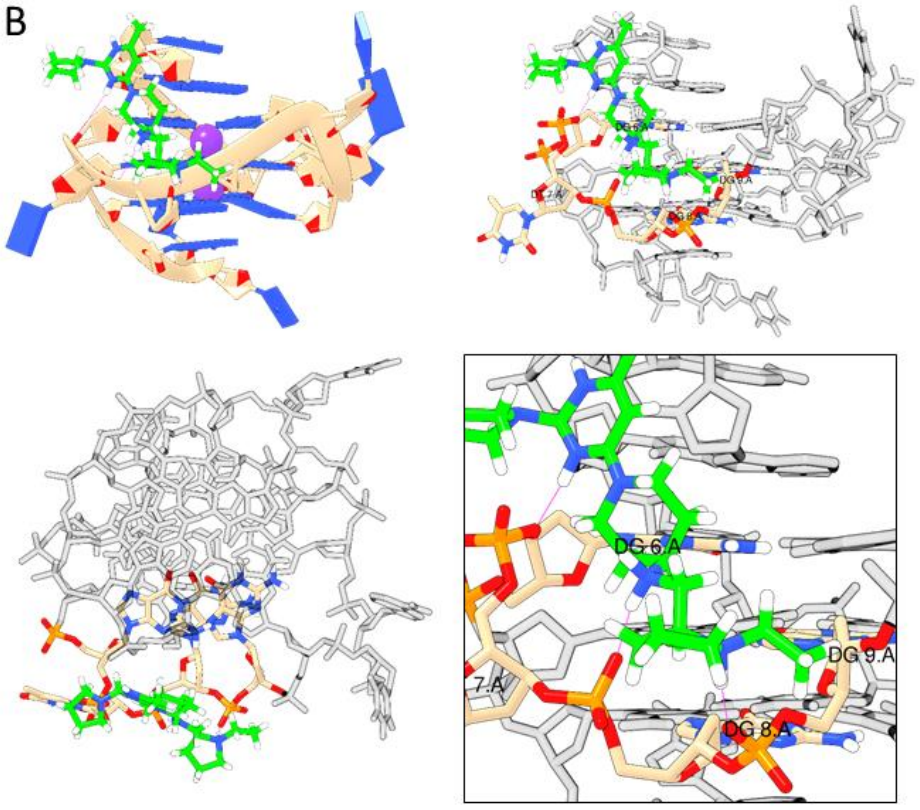
G-quadruplexes are now widely accepted as important drug targets for anticancer therapeutics. To date, most attempts at targeting promoter G4s with small molecules have been limited to their monomeric forms, and in some cases this has proven successful (16,103). However, this approach is limited in the following ways: (1) monomeric G-quadruplex units are small and often have shared characteristics, such as open G-tetrad faces, which encourage promiscuity of small molecules(18); (2) adding to, or expanding upon, core scaffolds to increase selection for monomeric G4s, leading to a decrease in drug-likeness (18) and potential bioavailability problems; (3) targeting a monomeric G4 fails to take into consideration its physiological context, as it may function in a larger, multimeric structure (20,144,352). Therefore, efforts towards revealing binding pockets among larger G4 multimers, by use of integrative structural biology approaches (143,165), are now needed to address this issue. Here, we have provided substantial evidence that this approach works by successfully discovering novel small molecules targeting the *hTERT* core promoter G-quadruplex multimer.

Discovery of novel small molecules for an intended target requires an adequate search of chemical space (271,347), which is achieved only through use of cheminformatics approaches [the only method that can approach the ~10^{33} possible drug-like small molecules (393)]. Here, we have utilized a G4 virtual screening approach of unprecedented in size (347) to targeting the loops and grooves of the all-parallel stacked model of the hTERT core promoter. From this, we have unearthed multiple new molecular scaffolds that show specificity towards G-quadruplex structures over duplex and other forms (**Figures 58, 62, and 63**). Although the focus of this work is selective targeting of the hTERT G4, we note that one such molecule, 2R (and to a lesser extent 2S and 2T), exhibited a substantial thermal stabilizing effect across nearly all G-quadruplexes tested in our FTSA panel, and no indications of interactions with other DNA topologies (**Figure 61**). 2R appeared to interact most favorably with the *c-MYC* G4s based on FTSA and FID (**Figures 58C and 61A**). Surprisingly, 2R caused a significant increase in c-MYC mRNA in MCF7 cells (10 μM treatment for 72 hr.) (**Figure 68A**). Molecular docking studies using the modified *c-MYC* NHEIII G-quadruplex

(PDB ID: 1XAV) reveal that 2R has a preferred interaction across the phosphate backbone of the first propeller loop (**Figure 68B**). This interaction seems to be facilitated by an ideal geometry of hydrogen bonding that spans three consecutive backbone residues. Thus, 2R's apparent exclusive interaction with G4s can be rationalized by its preference for the bent phosphate network found in parallel, single-nucleotide G4 loops. To the best of our knowledge, 2R has never been reported in any type of biological or biochemical investigation. The same holds true for compounds 2, 2S, and 2T. Thus, this somewhat serendipitous discovery is a testament to using massive virtual drug screening for enriching for new G4-interacting molecular scaffolds.

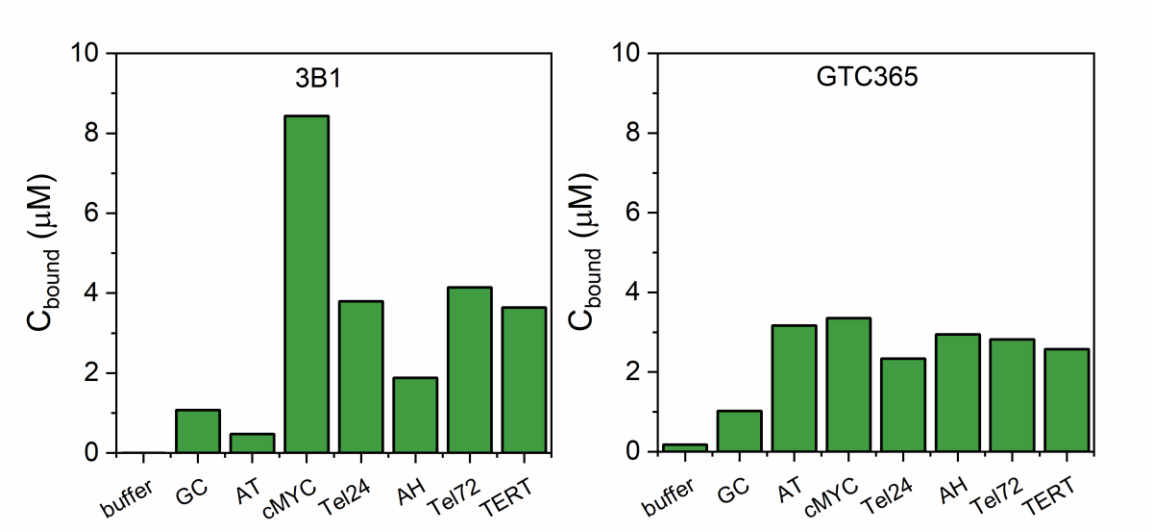
Figure 68. RT-qPCR and docking results of compound 2R. (A) RT-qPCR results of 2R 10 μM treatment of MCF7 breast cancer cells after 72 hours showing a significant increase in c-MYC mRNA relative to DMSO control. (B) Glide XP docking results of 2R with the modified all-parallel c-MYC G-quadruplex (PDB ID: 1XAV) showing an ideal hydrogen bonding network between H-bond donating amine groups of 2R with the first propeller loop's phosphate backbone. All four images are of the same docked position from different orientations.


A
c-MYC
*
P<0.01
Log2(Fold)
0.8
0.6
0.4
0.2
0.0
-0.2
-0.4
-0.6
-0.8
DMSO
2R
3B


B
DG 6.A
DG 9.A
7.A
DG 8.A

As for the primary goal of this work, we have discovered a unique disubstituted quinazoline-based small molecule, 3B1, that selectively targets the hTERT core promoter G-quadruplex. This discovery was accomplished using a SAR-by-catalogue approach combined with our G-quadruplex drug discovery funnel approach (269). In this process, we have taken top hits from over 40 million virtual docked small molecules and refined them using a robust high-throughput FTSA assay (**Figures 58, 59, 64**). At each generation, orthogonal assays have been employed that allowed us to define the regions of the scaffolds that contribute to potential selectivity (**Figures 58-60, 62, 63**). Scaffold 3B shows selectivity for the hTERT G4 over both duplex DNA and monomeric G4s, with a moderate selectivity over the higher-order telomere G4s. Further, we provide biological evidence that the modified scaffold 3B1 is able to reduce hTERT levels in breast cancer cells, similar to GTC365 (**Figures 65 and 67**) (223). However, in contrast to GTC365, 3B1 is a more drug-like small molecule [based on Lipinski's Rule of Five: MW = 336.4 g/mol, LogP = 3.184, < 5 H-bond donors, < 10 H-bond acceptors(137)]. Most importantly, 3B1 is more selective for G-quadruplex DNA over duplex DNA, whereas GTC365 shows a high degree of non-specificity based on competition dialysis (**Figure 69**). 3B1 is also unique, with no previous reports of disubstituted quinazolines with an aminoethyl group at the 2 position of the quinazoline ring system.

Figure 69. Competition dialysis results of 3B1 and GTC365 showing high selectivity of 3B1 for G4s over duplex DNA and the low selectivity of GTC365 for G4s over AT rich duplex DNA.


3B1
GTC365
C_{bound} (μM)
0
2
4
6
8
10
buffer
GC
AT
cMYC
Tel24
AH
Tel72
TERT

Based on extensive molecular docking and modeling studies, we find that 3B1 has three putative primary binding sites, the first of which is a pocket within an inter-G4 loop between PQS1 and PQS2, whereas the second is a shallow loop pocket spanning the PQS2 and PQS3 junction. A third site is found inside of a strand reversal loop pocket of the third (PQS3) G4 with partial stacking on the 3' tetrad face. Because there are 3 identified binding sites for 3B and 3B1, we can attempt to rationalize the unexpectedly large stoichiometry of ~8:1 observed for 3B (**Figure 59**). In each MD simulation of the molecules 3B, 3B1, and 3B2, there is a preferential stabilization of the phosphate backbone groups via the aminoethyl group. This interaction is slightly more favorable for 3B1, since there is also the secondary amine group at position 4 of the quinazoline ring. Since the AUC experiments are done in conditions of saturation (10:1 [Drug]:[DNA]), we reason that there are additional sites in which 3B is weakly associated through ionic interactions, leading to overestimations in stoichiometry from transport during sedimentation. However, we do not expect that this would impact competition dialysis experiments, as the ratio in these experiments is inverted (1:15 [Drug]:[DNA]). Consistently, three sites appear preferential based on ITC, MD, and AUC studies. This information will be beneficial moving forward as 3B1 is used in lead development, as we have shown that modifications to either the aminoethyl group or the 2-phenylethyl moiety reduce or improve its interactions with various hTERT sequence fragments (**Figure 63**).

Collectively, this study demonstrates that the discovery of novel, selective drug-like small molecules targeting multimeric G-quadruplexes can be achieved with an adequate search of chemical space. We show that a SAR approach, coupled with a robust, rapid FTSA assay, can yield novel small molecule scaffolds with moderate to high affinity. Further, we provide detailed insight into the binding mode of 3B1 in the binding pockets of the hTERT core promoter G-quadruplex that will benefit future campaigns and lead development strategies aimed at inhibiting the expression of human telomerase.

CHAPTER VII

CONCLUSION

In this dissertation, I provide an overview of an integrative structural biology approach for the medium- to high-resolution characterization of multimeric DNA G-quadruplexes. Importantly, this approach circumvents the need for perturbing the native sequences and allows for their study under biologically relevant conditions. Current approaches used to investigate the higher-order structure of DNA G-quadruplex multimers are limited. NMR techniques are hampered by the fact that proton resonances overlap significantly, making interpretation of higher-order G-quadruplexes difficult, if not impossible. X-ray crystallography, AFM, and EM-based methods require unnatural environments that can select out irrelevant structures or deform native structures through interactions with grids. Thus, my approach using a combination of robust solution-based biophysical techniques provides an excellent solution to this problem, as I have demonstrated with the higher-order telomere G4s and *hTERT* core promoter G-quadruplex. Further, this dissertation provides evidence that G-quadruplex multimers offer unique binding sites, allowing for development of selective ligands using *in silico* approaches.

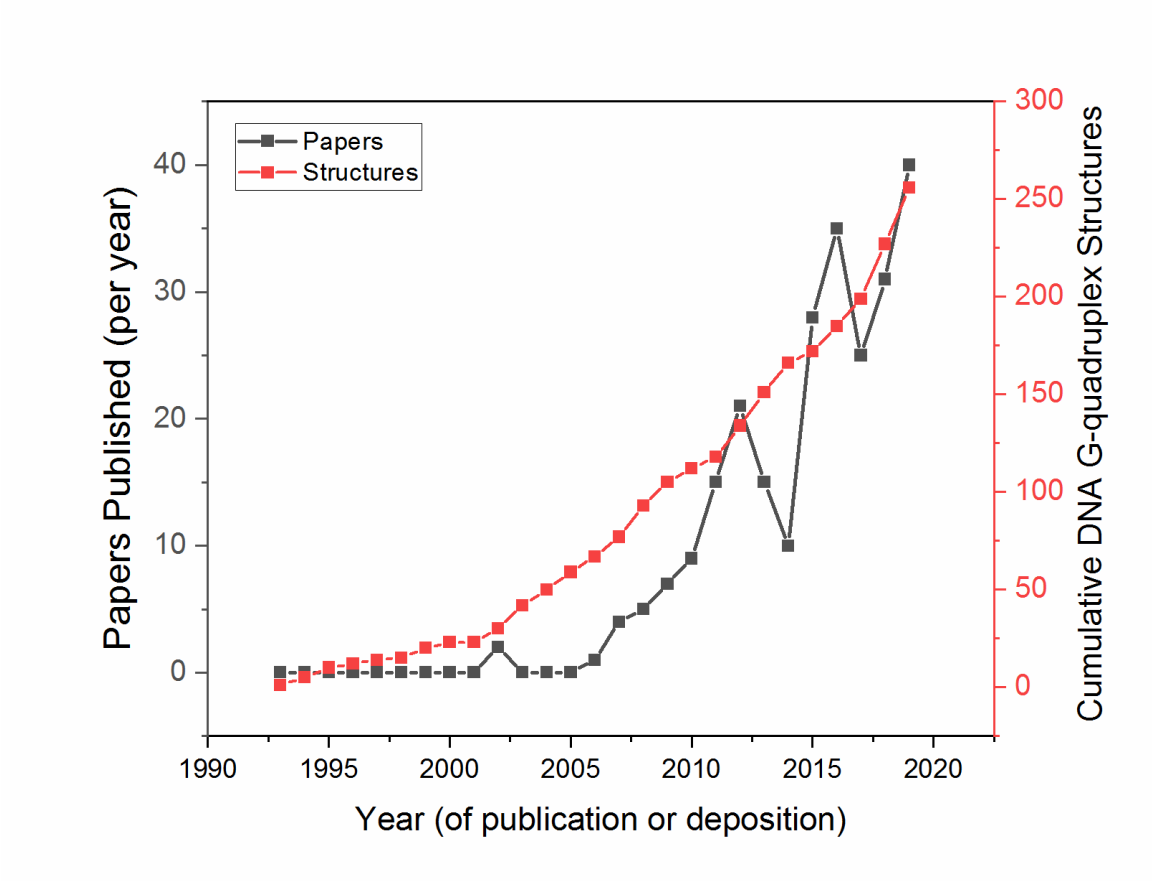
The structure formed in the higher-order human telomere sequence has been debated for almost a decade. Both *in vitro* and in-cell experimentation have shown that the two major topological forms of the monomeric telomere sequences are hybrid-1 and hybrid-2. However, in the context of longer sequences, information about the structure and assembly has yet to be firmly established. Low-resolution structural studies of long telomere repeats ($[TTAGGG]_n$, where $n >> 16$) by EM, AFM, and molecular tweezer studies conducted under harsh, non-physiological conditions have suggested that gaps exist between G-quadruplex monomers (i.e. that G-quadruplex formation is not maximized) (159,160). Studying the smaller sequences ($[TTAGGG]_n$, where $n \leq 16$) others have proposed that the maximum number of G-quadruplexes form, and that the most likely configuration is a compact, stacked combination of hybrid-1 and hybrid-2 topologies (141,162,394). Consistent with the latter, I've shown by SEC-SAXS and circular dichroism studies that the annealed telomere sequence, in a physiologically relevant buffer, maximizes its formation of G-quadruplex units, yet remains semi-flexible. The hybrid-1 and hybrid-2 topologies are most consistent with my analyses and prior intra-cellular studies (79). Further, MD simulations, SAXS modeling, and hydrodynamic studies suggest that the flexibility of the telomeres is due to stacking and unstacking transitions, with a significant proportion existing in the stacked state. G4 maximization and inter-G4 stacking offers an explanation to biological function, that is, the lack of gaps between G4s prevents access of proteins that recognize exposed single-stranded DNA, such as replication protein A (RPA), preventing aberrant DNA damage repair pathway activation. Altogether, this study adds significantly to our understanding of telomere biology insofar that it provides a detailed look at the structure and dynamics of the higher-order telomere at the highest resolution to date.

The atomistic models obtained from MD simulations of the higher-order telomer G4s reveal unique G4-junctions and loop sites which I have targeting using a high-throughput virtual screen. Using these models as receptors, I have identified a unique small molecule which binds with a stoichiometric ratio to the inter-G4 junctional regions. This molecule is an exciting lead for future development, as it shows selectivity for the higher-order G4 multimers over monomeric telomere G4s.

I next sought to determine the primary structure formed in the G-rich strand of the 68-nt long *hTERT* core promoter. This region of the *hTERT* core promoter is important as it is frequently mutated in a variety of cancers, leading to allele-specific increases in hTERT expression (222). These mutations appear to impinge upon secondary structure formation, such as G-quadruplexes (223). There has been an ongoing debate over the major structure formed within this sequence for over a decade, beginning with near simultaneous reports of a single parallel G-quadruplex in the first ~25 nucleotides (105), a novel parallel-antiparallel-hairpin structure (106), and an all-parallel stacked multimer (144,145). Determining the major structure formed is important from the perspective of both rational drug discovery and biological points of view (20), since targeting this structure with G-quadruplex ligands has been shown as a novel method of reducing hTERT expression in cancer cells (223). To this end, I have created two representative artificial DNA constructs that favor either an optimized all-parallel ("OP") or parallel-antiparallel-hairpin ("AH") for biophysical characterization and comparison to the wild-type ("WT"). Like my human telomere study, I have combined a suite of biophysical tools in an integrative fashion to compare the three oligonucleotide sequences. Spectroscopic studies with circular dichroism and DNase I cleavage assays indicated an absence of a hairpin moiety in both OP and WT and confirmed that both OP and WT had consistent CD spectra shape and magnitude—consistent with stacked parallel structures. Maximal parallel G-quadruplex formation in OP and WT (i.e. three parallel G-quadruplexes) was confirmed via ^{1}H-NMR studies. Lastly, I combined hydrodynamic techniques with SEC-SAXS and MD simulations to show, unequivocally, that the WT sequence is much more compact than the AH. Importantly, SEC-SAXS revealed that the AH structures contained multiple domains, whereas the WT sequence was very globular and is qualitatively akin to the all-parallel OP. From these results, we conclude that the major form of the *hTERT* promoter G-quadruplex is an all-parallel stacked G4 multimer. The immediate biological implications of this structure are uncertain. However, going forward there is now a model in which to test hypotheses, which is often accomplished in part by targeting with small molecules. Altogether, my combined results highlight the broad applicability of this integrative structural biology platform.

Nucleic acids have long remained in the shadow of proteins as biological targets in virtual screening approaches (269). As more *in vivo* evidence for the existence of novel non-B DNA structures is unearthed, more *in vitro* DNA structures will become *in silico* molecular targets. Indeed, G-quadruplexes, which are now widely accepted as important biological targets, are rapidly being used in virtual drug discovery approaches (**Figure 70**). Thus, it is prudent now to evaluate these initial virtual discovery campaigns and attempt to identify the aspects that make them successful. Therefore, I have created a compendium of the past decade of G-quadruplex virtual screening. In this comprehensive review, I have introduced the relevant screening methodologies, libraries, and scoring functions used in G4 virtual drug discovery, which provides a useful starting point for future investigators. Further, in comparing the contemporary methods, I have identified the two major limiting factors in successfully enriching for novel hits: library size and chemical search space. Overcoming these issues largely hinges on both advances in search and scoring algorithms, as well as increased accessibility to higher throughput computing infrastructure. I also show that most monomer G-quadruplexes that are targeted at their terminal tetrad faces enrich for heterocyclic molecules, known as “end-pasters”. End-pasters commonly lack selectivity due to their similar interactions with the common tetrad faces (like the specificity problem of tyrosine kinase inhibitors). Docking campaigns that focus on loops and grooves increase the likelihood of identifying selective hits. Thus, based on prior studies, we propose that virtual drug discovery should focus on the loop and groove regions of the higher-order G-quadruplex multimers. Lastly, we provide an overview of best practices that were gleaned from previous studies, which will undoubtably benefit G-quadruplex virtual campaigns moving forward.

Figure 70. Plot of publications per year in the Scifinder database based on the search term “G-quadruplex virtual screening” (left Y-axis, black), and the cumulative number of deposited atomic coordinate files for DNA G-quadruplexes in the Protein Databank (right Y-axis, red).


Papers
Structures
Papers Published (per year)
Cumulative DNA G-quadruplex Structures
Year (of publication or deposition)
0
10
20
30
40
0
50
100
150
200
250
300
1990
1995
2000
2005
2010
2015
2020

I next utilized a massive virtual and actual screening campaign targeting the *hTERT* core promoter G-quadruplex multimer. hTERT, and its cognate RNA hTR, form the ribonucleoprotein complex telomerase that is responsible for maintenance of the telomeres. hTERT's expression levels often correlate with its activity in cancer (395). Knockdown of hTERT is sufficient to cause telomere atrophy and, in some instances, direct induction of senescence (357-359). Due to its absence in most "normal" cells, hTERT has been a target of anti-cancer drug discovery for decades. To date, no small molecules inhibitors have been clinically successful (215), which has led many investigators to seek alternative strategies of inhibition. Recently, a new mechanism of hTERT repression was reported that involves targeting its core promoter with G-quadruplex interacting small molecules (223). Thus, I have used the all parallel hTERT G-quadruplex multimer receptor for *in silico* drug discovery. I docked over 40 million virtual small molecules to 12 different loop and groove locations across the hTERT G-quadruplex using Surflex-Dock v2.11. Based on docking scores, the top 500 molecules from each site were combined and clustered for scaffold similarity and to reduce the total top molecules. From the final group of unique scaffolds, 69 were purchased for screening. Using an iterative process of screening with thermal shift assays coupled with a catalogue-SAR approach, multiple small molecule scaffolds were identified that selectively stabilized G-quadruplexes over duplex and triplex DNA. Using competition dialysis and cell assays, I discovered that a single, di-substituted quinazoline molecule showed selectivity for hTERT over other G-quadruplex topologies. Additional catalogue-SAR investigations resulted in a disubstituted 2-aminoethyl-quinazoline molecule that significantly reduced hTERT mRNA levels in breast cancer cells and had no effect on *c-MYC* mRNA. Rigorous docking and molecular dynamics coupled with free energy calculations confirmed that this molecule preferentially stabilized loops and grooves. Importantly, the identified compounds are unique, and have not been reported in the literature to date. Thus, the resulting molecule represents an ideal candidate for future optimization as a novel, selective hTERT-repressing small molecule.

In conclusion, the integrative structural biology platform outlined in this work represents a robust means to answering questions about higher-order DNA G-quadruplex structures that previously could not be answered. It also emphasizes that selectivity can be achieved by targeting

the unique pockets created when G-quadruplex multimers form. However, I note that there is room for improvement. Refinement of G-quadruplex multimers against small-angle X-ray scattering is complex. While the interpretation of SAXS scattering is somewhat straight forward from a qualitative standpoint, the same cannot be said for the structural refinement, especially when heterogeneity, dynamics, and flexibility are taken into account (396). Therefore, while this approach allows us to integrate complexes of monomer subunits that have been solved by NMR or X-ray crystallography techniques previously, it does not allow us to identify G-quadruplex structures *ab initio*. To this end, future work will need to focus on the creation and implementation of G-quadruplex conformational search algorithms by which all plausible G-tetrad stacking and loop conformations can be generated for selection against SAXS scattering and prior experimental information. Similar algorithms are currently in use to some extent in the RNA community, although they are not applicable to G-quadruplex folds (397). Limitations aside, this dissertation is the first application of this robust integrative approach to characterizing DNA G-quadruplex multimers. This work represents the first steps towards characterizing, targeting, and understanding multimeric G-quadruplex structures as targets in human disease.

REFERENCES

1. Watson, J.D. and Crick, F.H. (1953) Molecular structure of nucleic acids; a structure for deoxyribose nucleic acid. *Nature*, **171**, 737-738.
2. Ghosh, A. and Bansal, M. (2003) A glossary of DNA structures from A to Z. *Acta Crystallogr D Biol Crystallogr*, **59**, 620-626.
3. Gellert, M., Lipsett, M.N. and Davies, D.R. (1962) Helix formation by guanylic acid. *Proc Natl Acad Sci U S A*, **48**, 2013-2018.
4. Zimmerman, S.B., Cohen, G.H. and Davies, D.R. (1975) X-ray fiber diffraction and model-building study of polyguanylic acid and polyinosinic acid. *J Mol Biol*, **92**, 181-192.
5. Webba da Silva, M. (2007) Geometric formalism for DNA quadruplex folding. *Chemistry*, **13**, 9738-9745.
6. Henderson, E., Hardin, C.C., Walk, S.K., Tinoco, I., Jr. and Blackburn, E.H. (1987) Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs. *Cell*, **51**, 899-908.
7. Zahler, A.M., Williamson, J.R., Cech, T.R. and Prescott, D.M. (1991) Inhibition of telomerase by G-quartet DNA structures. *Nature*, **350**, 718-720.
8. Blackburn, E.H. (1991) Structure and function of telomeres. *Nature*, **350**, 569-573.
9. Meyerson, M., Counter, C.M., Eaton, E.N., Ellisen, L.W., Steiner, P., Caddle, S.D., Ziaugra, L., Beijersbergen, R.L., Davidoff, M.J., Liu, Q. *et al.* (1997) hEST2, the putative human telomerase catalytic subunit gene, is up-regulated in tumor cells and during immortalization. *Cell*, **90**, 785-795.
10. Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C.B., West, M.D., Ho, P.L., Coviello, G.M., Wright, W.E., Weinrich, S.L. and Shay, J.W. (1994) Specific association of human telomerase activity with immortal cells and cancer. *Science*, **266**, 2011-2015.

11. Huppert, J.L. and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res*, **33**, 2908-2916.

12. Huppert, J.L. and Balasubramanian, S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res*, **35**, 406-413.

13. Ribeiro de Almeida, C., Dhir, S., Dhir, A., Moghaddam, A.E., Sattentau, Q., Meinhart, A. and Proudfoot, N.J. (2018) RNA Helicase DDX1 Converts RNA G-Quadruplex Structures into R-Loops to Promote IgH Class Switch Recombination. *Mol Cell*, **70**, 650-662 e658.

14. Duchler, M. (2012) G-quadruplexes: targets and tools in anticancer drug design. *J Drug Target*, **20**, 389-400.

15. Balasubramanian, S., Hurley, L.H. and Neidle, S. (2011) Targeting G-quadruplexes in gene promoters: a novel anticancer strategy? *Nat Rev Drug Discov*, **10**, 261-275.

16. Siddiqui-Jain, A., Grand, C.L., Bearss, D.J. and Hurley, L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription. *Proc Natl Acad Sci U S A*, **99**, 11593-11598.

17. Li, Q., Xiang, J.F., Yang, Q.F., Sun, H.X., Guan, A.J. and Tang, Y.L. (2013) G4LDB: a database for discovering and studying G-quadruplex ligands. *Nucleic Acids Res*, **41**, D1115-1123.

18. Asamitsu, S., Obata, S., Yu, Z., Bando, T. and Sugiyama, H. (2019) Recent Progress of Targeted G-Quadruplex-Preferred Ligands Toward Cancer Therapy. *Molecules*, **24**.

19. Kolesnikova, S., Hubalek, M., Bednarova, L., Cvacka, J. and Curtis, E.A. (2017) Multimerization rules for G-quadruplexes. *Nucleic Acids Res*, **45**, 8684-8696.

20. Kolesnikova, S. and Curtis, E.A. (2019) Structure and Function of Multimeric G-Quadruplexes. *Molecules*, **24**.

21. Potaman, V., Sinden, R. (2013), *Madame Curie Bioscience Database*. Landes Bioscience, Austin, TX, Vol. 2020 https://www.ncbi.nlm.nih.gov/books/NBK6545/.

22. Burge, S., Parkinson, G.N., Hazel, P., Todd, A.K. and Neidle, S. (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res*, **34**, 5402-5415.

23. Sen, D. and Gilbert, W. (1988) Formation of parallel four-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, **334**, 364-366.

24. Sundquist, W.I. and Klug, A. (1989) Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature*, **342**, 825-829.

25. Rhodes, D. and Lipps, H.J. (2015) G-quadruplexes and their regulatory roles in biology. *Nucleic Acids Res*, **43**, 8627-8637.

26. Schaffitzel, C., Berger, I., Postberg, J., Hanes, J., Lipps, H.J. and Pluckthun, A. (2001) In vitro generated antibodies specific for telomeric guanine-quadruplex DNA react with Stylonychia lemnae macronuclei. *Proc Natl Acad Sci U S A*, **98**, 8572-8577.

27. Biffi, G., Tannahill, D., McCafferty, J. and Balasubramanian, S. (2013) Quantitative visualization of DNA G-quadruplex structures in human cells. *Nat Chem*, **5**, 182-186.

28. Henderson, A., Wu, Y., Huang, Y.C., Chavez, E.A., Platt, J., Johnson, F.B., Brosh, R.M., Jr., Sen, D. and Lansdorp, P.M. (2014) Detection of G-quadruplex DNA in mammalian cells. *Nucleic Acids Res*, **42**, 860-869.

29. Liu, H.Y., Zhao, Q., Zhang, T.P., Wu, Y., Xiong, Y.X., Wang, S.K., Ge, Y.L., He, J.H., Lv, P., Ou, T.M. *et al.* (2016) Conformation Selective Antibody Enables Genome Profiling and Leads to Discovery of Parallel G-Quadruplex in Human Telomeres. *Cell Chem Biol*, **23**, 1261-1270.

30. Henderson, A., Wu, Y., Huang, Y.C., Chavez, E.A., Platt, J., Johnson, F.B., Brosh, R.M., Jr., Sen, D. and Lansdorp, P.M. (2017) Detection of G-quadruplex DNA in mammalian cells. *Nucleic Acids Res*, **45**, 6252.

31. Chambers, V.S., Marsico, G., Boutell, J.M., Di Antonio, M., Smith, G.P. and Balasubramanian, S. (2015) High-throughput sequencing of DNA G-quadruplex structures in the human genome. *Nat Biotechnol*, **33**, 877-881.

32. Moye, A.L., Porter, K.C., Cohen, S.B., Phan, T., Zyner, K.G., Sasaki, N., Lovrecz, G.O., Beck, J.L. and Bryan, T.M. (2015) Telomeric G-quadruplexes are a substrate and site of localization for human telomerase. *Nat Commun*, **6**, 7643.

33. Cogoi, S., Paramasivam, M., Spolaore, B. and Xodo, L.E. (2008) Structural polymorphism within a regulatory element of the human KRAS promoter: formation of G4-DNA recognized by nuclear proteins. *Nucleic Acids Res*, **36**, 3765-3780.

34. Gonzalez, V., Guo, K., Hurley, L. and Sun, D. (2009) Identification and characterization of nucleolin as a c-myc G-quadruplex-binding protein. *J Biol Chem*, **284**, 23622-23635.

35. Makowski, M.M., Grawe, C., Foster, B.M., Nguyen, N.V., Bartke, T. and Vermeulen, M. (2018) Global profiling of protein-DNA and protein-nucleosome binding affinities using quantitative mass spectrometry. *Nat Commun*, **9**, 1653.

36. Spiegel, J., Adhikari, S. and Balasubramanian, S. (2020) The Structure and Function of DNA G-Quadruplexes. *Trends in Chemistry*, **2**, 123-136.

37. Bang, I. (1910) Untersuchungen über die Guanylsäure. *Biochemistry*, **26**.

38. Ralph, R.K., Connors, W. J., Khorana, H. G. (1962) Secondary structure and aggregation in deoxyguanosine oligonucleotides. *Journal of American Chemical Society*, **84**.

39. Arnott, S., Chandrasekaran, R. and Marttila, C.M. (1974) Structures for polyinosinic acid and polyguanylic acid. *Biochem J*, **141**, 537-543.

40. Lane, A.N., Chaires, J.B., Gray, R.D. and Trent, J.O. (2008) Stability and kinetics of G-quadruplex structures. *Nucleic Acids Res*, **36**, 5482-5515.

41. Bhattacharyya, D., Mirihana Arachchilage, G. and Basu, S. (2016) Metal Cations in G-Quadruplex Folding and Stability. *Front Chem*, **4**, 38.

42. Adrian, M., Heddi, B. and Phan, A.T. (2012) NMR spectroscopy of G-quadruplexes. *Methods*, **57**, 11-24.

43. Webba da Silva, M. (2007) NMR methods for studying quadruplex nucleic acids. *Methods*, **43**, 264-277.

44. Kypr, J., Kejnovska, I., Renciuk, D. and Vorlickova, M. (2009) Circular dichroism and conformational polymorphism of DNA. *Nucleic Acids Res*, **37**, 1713-1725.

45. Lin, C. and Yang, D. (2017) Human Telomeric G-Quadruplex Structures and G-Quadruplex-Interactive Compounds. *Methods Mol Biol*, **1587**, 171-196.

46. Dai, J., Carver, M. and Yang, D. (2008) Polymorphism of human telomeric quadruplex structures. *Biochimie*, **90**, 1172-1183.

47. Zhang, Z., Dai, J., Veliath, E., Jones, R.A. and Yang, D. (2010) 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 Res*, **38**, 1009-1021.

48. Liu, C., Zhou, B., Geng, Y., Yan Tam, D., Feng, R., Miao, H., Xu, N., Shi, X., You, Y., Hong, Y. *et al.* (2019) A chair-type G-quadruplex structure formed by a human telomeric variant DNA in K(+) solution. *Chem Sci*, **10**, 218-226.

49. Karsisiotis, A.I., O'Kane, C. and Webba da Silva, M. (2013) DNA quadruplex folding formalism--a tutorial on quadruplex topologies. *Methods*, **64**, 28-35.

50. Chung, W.J., Heddi, B., Schmitt, E., Lim, K.W., Mechulam, Y. and Phan, A.T. (2015) Structure of a left-handed DNA G-quadruplex. *Proc Natl Acad Sci U S A*, **112**, 2729-2733.

51. Winnerdy, F.R., Bakalar, B., Maity, A., Vandana, J.J., Mechulam, Y., Schmitt, E. and Phan, A.T. (2019) NMR solution and X-ray crystal structures of a DNA molecule containing both right- and left-handed parallel-stranded G-quadruplexes. *Nucleic Acids Res*, **47**, 8272-8281.

52. Dvorkin, S.A., Karsisiotis, A.I. and Webba da Silva, M. (2018) Encoding canonical DNA quadruplex structure. *Sci Adv*, **4**, eaat3007.

53. Huppert, J.L. (2010) Structure, location and interactions of G-quadruplexes. *FEBS J*, **277**, 3452-3458.

54. Kogut, M., Kleist, C. and Czub, J. (2019) Why do G-quadruplexes dimerize through the 5'-ends? Driving forces for G4 DNA dimerization examined in atomic detail. *PLoS Comput Biol*, **15**, e1007383.

55. Yu, H.Q., Miyoshi, D. and Sugimoto, N. (2006) Characterization of structure and stability of long telomeric DNA G-quadruplexes. *J Am Chem Soc*, **128**, 15461-15468.

56. Matsugami, A., Okuizumi, T., Uesugi, S. and Katahira, M. (2003) Intramolecular higher order packing of parallel quadruplexes comprising a G:G:G:G tetrad and a G(:A):G(:A):G(:A):G heptad of GGA triplet repeat DNA. *J Biol Chem*, **278**, 28147-28153.

57. Do, N.Q., Chung, W.J., Truong, T.H.A., Heddi, B. and Phan, A.T. (2017) G-quadruplex structure of an anti-proliferative DNA sequence. *Nucleic Acids Res*, **45**, 7487-7493.

58. Moyzis, R.K., Buckingham, J.M., Cram, L.S., Dani, M., Deaven, L.L., Jones, M.D., Meyne, J., Ratliff, R.L. and Wu, J.R. (1988) A highly conserved repetitive DNA sequence, (TTAGGG)n, present at the telomeres of human chromosomes. *Proc Natl Acad Sci U S A*, **85**, 6622-6626.

59. Neidle, S. and Parkinson, G. (2002) Telomere maintenance as a target for anticancer drug discovery. *Nat Rev Drug Discov*, **1**, 383-393.

60. Townsley, D.M., Dumitriu, B. and Young, N.S. (2014) Bone marrow failure and the telomeropathies. *Blood*, **124**, 2775-2783.

61. Aubert, G. and Lansdorp, P.M. (2008) Telomeres and aging. *Physiol Rev*, **88**, 557-579.

62. O'Sullivan, R.J. and Karlseder, J. (2010) Telomeres: protecting chromosomes against genome instability. *Nat Rev Mol Cell Biol*, **11**, 171-181.

63. Wright, W.E., Tesmer, V.M., Huffman, K.E., Levene, S.D. and Shay, J.W. (1997) Normal human chromosomes have long G-rich telomeric overhangs at one end. *Genes Dev*, **11**, 2801-2809.

64. de Lange, T. (2005) Shelterin: the protein complex that shapes and safeguards human telomeres. *Genes Dev*, **19**, 2100-2110.

65. Stewart, J.A., Chaiken, M.F., Wang, F. and Price, C.M. (2012) Maintaining the end: roles of telomere proteins in end-protection, telomere replication and length regulation. *Mutat Res*, **730**, 12-19.

66. Levy, M.Z., Allsopp, R.C., Futcher, A.B., Greider, C.W. and Harley, C.B. (1992) Telomere end-replication problem and cell aging. *J Mol Biol*, **225**, 951-960.

67. Cong, Y.S., Wen, J. and Bacchetti, S. (1999) The human telomerase catalytic subunit hTERT: organization of the gene and characterization of the promoter. *Hum Mol Genet*, **8**, 137-142.

68. Shay, J.W. and Bacchetti, S. (1997) A survey of telomerase activity in human cancer. *Eur J Cancer*, **33**, 787-791.

69. Hiyama, E. and Hiyama, K. (2007) Telomere and telomerase in stem cells. *Br J Cancer*, **96**, 1020-1024.

70. Williamson, J.R., Raghuraman, M.K. and Cech, T.R. (1989) Monovalent cation-induced structure of telomeric DNA: the G-quartet model. *Cell*, **59**, 871-880.

71. Caruthers, M.H. (2011) A brief review of DNA and RNA chemical synthesis. *Biochem Soc Trans*, **39**, 575-580.

72. Kang, C., Zhang, X., Ratliff, R., Moyzis, R. and Rich, A. (1992) Crystal structure of four-stranded Oxytricha telomeric DNA. *Nature*, **356**, 126-131.

73. Schultze, P., Smith, F.W. and Feigon, J. (1994) Refined solution structure of the dimeric quadruplex formed from the Oxytricha telomeric oligonucleotide d(GGGGTTTTGGGG). *Structure*, **2**, 221-233.

74. Wang, Y. and Patel, D.J. (1993) Solution structure of the human telomeric repeat d[AG3(T2AG3)3] G-tetraplex. *Structure*, **1**, 263-282.

75. Parkinson, G.N., Lee, M.P. and Neidle, S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **417**, 876-880.

76. Dai, J., Punchihewa, C., Ambrus, A., Chen, D., Jones, R.A. and Yang, D. (2007) Structure of the intramolecular human telomeric G-quadruplex in potassium solution: a novel adenine triple formation. *Nucleic Acids Res*, **35**, 2440-2450.

77. Dai, J., Carver, M., Punchihewa, C., Jones, R.A. and Yang, D. (2007) 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 Res*, **35**, 4927-4940.

78. Li, J., Correia, J.J., Wang, L., Trent, J.O. and Chaires, J.B. (2005) Not so crystal clear: the structure of the human telomere G-quadruplex in solution differs from that present in a crystal. *Nucleic Acids Res*, **33**, 4649-4659.

79. Bao, H.L., Liu, H.S. and Xu, Y. (2019) Hybrid-type and two-tetrad antiparallel telomere DNA G-quadruplex structures in living human cells. *Nucleic Acids Res*, **47**, 4940-4947.

80. d'Adda di Fagagna, F., Reaper, P.M., Clay-Farrace, L., Fiegler, H., Carr, P., Von Zglinicki, T., Saretzki, G., Carter, N.P. and Jackson, S.P. (2003) A DNA damage checkpoint response in telomere-initiated senescence. *Nature*, **426**, 194-198.

81. Griffith, J., Bianchi, A. and de Lange, T. (1998) TRF1 promotes parallel pairing of telomeric tracts in vitro. *J Mol Biol*, **278**, 79-88.

82. Tahara, H., Shin-Ya, K., Seimiya, H., Yamada, H., Tsuruo, T. and Ide, T. (2006) G-Quadruplex stabilization by telomestatin induces TRF2 protein dissociation from telomeres and anaphase bridge formation accompanied by loss of the 3' telomeric overhang in cancer cells. *Oncogene*, **25**, 1955-1966.

83. Zhou, G., Liu, X., Li, Y., Xu, S., Ma, C., Wu, X., Cheng, Y., Yu, Z., Zhao, G. and Chen, Y. (2016) Telomere targeting with a novel G-quadruplex-interactive ligand BRACO-19 induces T-loop disassembly and telomerase displacement in human glioblastoma cells. *Oncotarget*, **7**, 14925-14939.

84. Zaug, A.J., Podell, E.R. and Cech, T.R. (2005) Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro. *Proc Natl Acad Sci U S A*, **102**, 10864-10869.

85. Chaires, J.B., Gray, R.D., Dean, W.L., Monsen, R., DeLeeuw, L.W., Stribinskis, V. and Trent, J.O. (2020) Human POT1 unfolds G-quadruplexes by conformational selection. *Nucleic Acids Res*, **48**, 4976-4991.

86. Altschuler, S.E., Croy, J.E. and Wuttke, D.S. (2012) A small molecule inhibitor of Pot1 binding to telomeric DNA. *Biochemistry*, **51**, 7833-7845.

87. Neidle, S. (2010) Human telomeric G-quadruplex: the current status of telomeric G-quadruplexes as therapeutic targets in human cancer. *FEBS J*, **277**, 1118-1125.

88. Bell, G.I., Selby, M.J. and Rutter, W.J. (1982) The highly polymorphic region near the human insulin gene is composed of simple tandemly repeating sequences. *Nature*, **295**, 31-35.

89. Sun, H., Karow, J.K., Hickson, I.D. and Maizels, N. (1998) The Bloom's syndrome helicase unwinds G4 DNA. *J Biol Chem*, **273**, 27587-27592.

90. van Wietmarschen, N., Merzouk, S., Halsema, N., Spierings, D.C.J., Guryev, V. and Lansdorp, P.M. (2018) BLM helicase suppresses recombination at G-quadruplex motifs in transcribed genes. *Nat Commun*, **9**, 271.

91. Woodford, K.J., Howell, R.M. and Usdin, K. (1994) A novel K(+)-dependent DNA synthesis arrest site in a commonly occurring sequence motif in eukaryotes. *J Biol Chem*, **269**, 27029-27035.

92. Dang, C.V. (2012) MYC on the path to cancer. *Cell*, **149**, 22-35.

93. Gabay, M., Li, Y. and Felsher, D.W. (2014) MYC activation is a hallmark of cancer initiation and maintenance. *Cold Spring Harb Perspect Med*, **4**.

94. Chen, H., Liu, H. and Qing, G. (2018) Targeting oncogenic Myc as a strategy for cancer treatment. *Signal Transduct Target Ther*, **3**, 5.

95. Siebenlist, U., Hennighausen, L., Battey, J. and Leder, P. (1985) Chromatin structure of the human c-myc oncogene: definition of regulatory regions and changes in Burkitt's lymphomas. *Haematol Blood Transfus*, **29**, 261-265.

96. Berberich, S.J. and Postel, E.H. (1995) PuF/NM23-H2/NDPK-B transactivates a human c-myc promoter-CAT gene via a functional nuclease hypersensitive element. *Oncogene*, **10**, 2343-2347.

97. Boles, T.C. and Hogan, M.E. (1987) DNA structure equilibria in the human c-myc gene. *Biochemistry*, **26**, 367-376.

98. Simonsson, T., Pecinka, P. and Kubista, M. (1998) DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Res*, **26**, 1167-1172.

99. Grand, C.L., Han, H., Munoz, R.M., Weitman, S., Von Hoff, D.D., Hurley, L.H. and Bearss, D.J. (2002) The cationic porphyrin TMPyP4 down-regulates c-MYC and human telomerase

reverse transcriptase expression and inhibits tumor growth in vivo. *Mol Cancer Ther*, **1**, 565-573.

100. Konig, S.L., Evans, A.C. and Huppert, J.L. (2010) Seven essential questions on G-quadruplexes. *Biomol Concepts*, **1**, 197-213.

101. Cogoi, S. and Xodo, L.E. (2006) G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res*, **34**, 2536-2549.

102. Cogoi, S., Shchekotikhin, A.E. and Xodo, L.E. (2014) HRAS is silenced by two neighboring G-quadruplexes and activated by MAZ, a zinc-finger transcription factor with DNA unfolding property. *Nucleic Acids Res*, **42**, 8379-8388.

103. Welsh, S.J., Dale, A.G., Lombardo, C.M., Valentine, H., de la Fuente, M., Schatzlein, A. and Neidle, S. (2013) Inhibition of the hypoxia-inducible factor pathway by a G-quadruplex binding small molecule. *Sci Rep*, **3**, 2799.

104. Wu, Y., Zan, L.P., Wang, X.D., Lu, Y.J., Ou, T.M., Lin, J., Huang, Z.S. and Gu, L.Q. (2014) Stabilization of VEGF G-quadruplex and inhibition of angiogenesis by quindoline derivatives. *Biochim Biophys Acta*, **1840**, 2970-2977.

105. Lim, K.W., Lacroix, L., Yue, D.J., Lim, J.K., Lim, J.M. and Phan, A.T. (2010) Coexistence of two distinct G-quadruplex conformations in the hTERT promoter. *J Am Chem Soc*, **132**, 12331-12342.

106. Palumbo, S.L., Ebbinghaus, S.W. and Hurley, L.H. (2009) Formation of a unique end-to-end stacked pair of G-quadruplexes in the hTERT core promoter with implications for inhibition of telomerase by G-quadruplex-interactive ligands. *J Am Chem Soc*, **131**, 10878-10891.

107. Fleming, A.M. and Burrows, C.J. (2017) 8-Oxo-7,8-dihydro-2'-deoxyguanosine and abasic site tandem lesions are oxidation prone yielding hydantoin products that strongly destabilize duplex DNA. *Org Biomol Chem*, **15**, 8341-8353.

108. Wilson, D.M., 3rd and Bohr, V.A. (2007) The mechanics of base excision repair, and its relationship to aging and disease. *DNA Repair (Amst)*, **6**, 544-559.

109. Roychoudhury, S., Pramanik, S., Harris, H.L., Tarpley, M., Sarkar, A., Spagnol, G., Sorgen, P.L., Chowdhury, D., Band, V., Klinkebiel, D. *et al.* (2020) Endogenous oxidized DNA bases and APE1 regulate the formation of G-quadruplex structures in the genome. *Proc Natl Acad Sci U S A*.

110. Bettin, N., Oss Pegorar, C. and Cusanelli, E. (2019) The Emerging Roles of TERRA in Telomere Maintenance and Genome Stability. *Cells*, **8**.

111. Collie, G.W., Parkinson, G.N., Neidle, S., Rosu, F., De Pauw, E. and Gabelica, V. (2010) Electrospray mass spectrometry of telomeric RNA (TERRA) reveals the formation of stable multimeric G-quadruplex structures. *J Am Chem Soc*, **132**, 9328-9334.

112. Mestre-Fos, S., Penev, P.I., Suttapitugsakul, S., Hu, M., Ito, C., Petrov, A.S., Wartell, R.M., Wu, R. and Williams, L.D. (2019) G-Quadruplexes in Human Ribosomal RNA. *J Mol Biol*, **431**, 1940-1955.

113. Hoffmann, R.F., Moshkin, Y.M., Mouton, S., Grzeschik, N.A., Kalicharan, R.D., Kuipers, J., Wolters, A.H., Nishida, K., Romashchenko, A.V., Postberg, J. *et al.* (2016) Guanine quadruplex structures localize to heterochromatin. *Nucleic Acids Res*, **44**, 152-163.

114. Yang, S.Y., Lejault, P., Chevrier, S., Boidot, R., Robertson, A.G., Wong, J.M.Y. and Monchaud, D. (2018) Transcriptome-wide identification of transient RNA G-quadruplexes in human cells. *Nat Commun*, **9**, 4730.

115. Christiansen, J., Kofod, M. and Nielsen, F.C. (1994) A guanosine quadruplex and two stable hairpins flank a major cleavage site in insulin-like growth factor II mRNA. *Nucleic Acids Res*, **22**, 5709-5716.

116. Bugaut, A. and Balasubramanian, S. (2012) 5'-UTR RNA G-quadruplexes: translation regulation and targeting. *Nucleic Acids Res*, **40**, 4727-4741.

117. Kumari, S., Bugaut, A., Huppert, J.L. and Balasubramanian, S. (2007) An RNA G-quadruplex in the 5' UTR of the NRAS proto-oncogene modulates translation. *Nat Chem Biol*, **3**, 218-221.

118. Yatsunyk, L.A., Mendoza, O. and Mergny, J.L. (2014) "Nano-oddities": unusual nucleic acid assemblies for DNA-based nanostructures and nanodevices. *Acc Chem Res*, **47**, 1836-1844.

119. Tucker, W.O., Shum, K.T. and Tanner, J.A. (2012) G-quadruplex DNA aptamers and their ligands: structure, function and application. *Curr Pharm Des*, **18**, 2014-2026.

120. Mariani, P., Spinozzi, F., Federiconi, F., Amenitsch, H., Spindler, L. and Drevensek-Olenik, I. (2009) Small angle X-ray scattering analysis of deoxyguanosine 5'-monophosphate self-assembing in solution: nucleation and growth of G-quadruplexes. *J Phys Chem B*, **113**, 7934-7944.

121. Calzolari, A., Felice, R.D., Molinari, E. and Garbesi, A. (2002) G-quartet biomolecular nanowires. *Applied Physics Letters*, **80**, 3331-3333.

122. Yang, X., Wang, X.B., Vorpagel, E.R. and Wang, L.S. (2004) Direct experimental observation of the low ionization potentials of guanine in free oligonucleotides by using photoelectron spectroscopy. *Proc Natl Acad Sci U S A*, **101**, 17588-17592.

123. Guo, Y., Yao, W., Xie, Y., Zhou, X., Hu, J. and Pei, R. (2016) Logic gates based on G-quadruplexes: principles and sensor applications. *Microchimica Acta*, **183**, 21-34.

124. Roxo, C., Kotkowiak, W. and Pasternak, A. (2019) G-Quadruplex-Forming Aptamers-Characteristics, Applications, and Perspectives. *Molecules*, **24**.

125. Bates, P.J., Kahlon, J.B., Thomas, S.D., Trent, J.O. and Miller, D.M. (1999) Antiproliferative activity of G-rich oligonucleotides correlates with protein binding. *J Biol Chem*, **274**, 26369-26377.

126. Bates, P.J., Laber, D.A., Miller, D.M., Thomas, S.D. and Trent, J.O. (2009) Discovery and development of the G-rich oligonucleotide AS1411 as a novel treatment for cancer. *Exp Mol Pathol*, **86**, 151-164.

127. Bates, P.J., Reyes-Reyes, E.M., Malik, M.T., Murphy, E.M., O'Toole, M.G. and Trent, J.O. (2017) G-quadruplex oligonucleotide AS1411 as a cancer-targeting agent: Uses and mechanisms. *Biochim Biophys Acta Gen Subj*, **1861**, 1414-1428.

128. Anantha, N.V., Azam, M. and Sheardy, R.D. (1998) Porphyrin binding to quadrupled T4G4. *Biochemistry*, **37**, 2709-2714.

129. Gowan, S.M., Harrison, J.R., Patterson, L., Valenti, M., Read, M.A., Neidle, S. and Kelland, L.R. (2002) A G-quadruplex-interactive potent small-molecule inhibitor of telomerase exhibiting in vitro and in vivo antitumor activity. *Mol Pharmacol*, **61**, 1154-1162.

130. Shin-ya, K., Wierzba, K., Matsuo, K., Ohtani, T., Yamada, Y., Furihata, K., Hayakawa, Y. and Seto, H. (2001) Telomestatin, a novel telomerase inhibitor from Streptomyces anulatus. *J Am Chem Soc*, **123**, 1262-1263.

131. Rodriguez, R., Muller, S., Yeoman, J.A., Trentesaux, C., Riou, J.F. and Balasubramanian, S. (2008) A novel small molecule that alters shelterin integrity and triggers a DNA-damage response at telomeres. *J Am Chem Soc*, **130**, 15758-15759.

132. De Cian, A., Delemos, E., Mergny, J.L., Teulade-Fichou, M.P. and Monchaud, D. (2007) Highly efficient G-quadruplex recognition by bisquinolinium compounds. *J Am Chem Soc*, **129**, 1856-1857.

133. Xu, H., Di Antonio, M., McKinney, S., Mathew, V., Ho, B., O'Neil, N.J., Santos, N.D., Silvester, J., Wei, V., Garcia, J. *et al.* (2017) CX-5461 is a DNA G-quadruplex stabilizer with selective lethality in BRCA1/2 deficient tumours. *Nat Commun*, **8**, 14432.

134. Drygin, D., Siddiqui-Jain, A., O'Brien, S., Schwaebe, M., Lin, A., Bliesath, J., Ho, C.B., Proffitt, C., Trent, K., Whitten, J.P. *et al.* (2009) Anticancer activity of CX-3543: a direct inhibitor of rRNA biogenesis. *Cancer Res*, **69**, 7653-7661.

135. Sun, Z.Y., Wang, X.N., Cheng, S.Q., Su, X.X. and Ou, T.M. (2019) Developing Novel G-Quadruplex Ligands: from Interaction with Nucleic Acids to Interfering with Nucleic Acid(-)Protein Interaction. *Molecules*, **24**.

136. Drygin, D., Lin, A., Bliesath, J., Ho, C.B., O'Brien, S.E., Proffitt, C., Omori, M., Haddach, M., Schwaebe, M.K., Siddiqui-Jain, A. *et al.* (2011) Targeting RNA polymerase I with an oral small molecule CX-5461 inhibits ribosomal RNA synthesis and solid tumor growth. *Cancer Res*, **71**, 1418-1430.

137. Lipinski, C.A. (2000) Drug-like properties and the causes of poor solubility and poor permeability. *J Pharmacol Toxicol Methods*, **44**, 235-249.

138. Campbell, N.H., Parkinson, G.N., Reszka, A.P. and Neidle, S. (2008) Structural basis of DNA quadruplex recognition by an acridine drug. *J Am Chem Soc*, **130**, 6722-6724.

139. Ferreira, R., Artali, R., Benoit, A., Gargallo, R., Eritja, R., Ferguson, D.M., Sham, Y.Y. and Mazzini, S. (2013) Structure and stability of human telomeric G-quadruplex with preclinical 9-amino acridines. *PLoS One*, **8**, e57701.

140. Taetz, S., Baldes, C., Murdter, T.E., Kleideiter, E., Piotrowska, K., Bock, U., Haltner-Ukomadu, E., Mueller, J., Huwer, H., Schaefer, U.F. *et al.* (2006) Biopharmaceutical characterization of the telomerase inhibitor BRACO19. *Pharm Res*, **23**, 1031-1037.

141. Petraccone, L., Trent, J.O. and Chaires, J.B. (2008) The tail of the telomere. *J Am Chem Soc*, **130**, 16530-16532.

142. Petraccone, L., Garbett, N.C., Chaires, J.B. and Trent, J.O. (2010) An integrated molecular dynamics (MD) and experimental study of higher order human telomeric quadruplexes. *Biopolymers*, **93**, 533-548.

143. Monsen, R.C., DeLeeuw, L., Dean, W.L., Gray, R.D., Sabo, T.M., Chakravarthy, S., Chaires, J.B. and Trent, J.O. (2020) The hTERT core promoter forms three parallel G-quadruplexes. *Nucleic Acids Res*.

144. Micheli, E., Martufi, M., Cacchione, S., De Santis, P. and Savino, M. (2010) Self-organization of G-quadruplex structures in the hTERT core promoter stabilized by polyaminic side chain perylene derivatives. *Biophys Chem*, **153**, 43-53.

145. Chaires, J.B., Trent, J.O., Gray, R.D., Dean, W.L., Buscaglia, R., Thomas, S.D. and Miller, D.M. (2014) An improved model for the hTERT promoter quadruplex. *PLoS One*, **9**, e115580.

146. Palumbo, S.L., Memmott, R.M., Uribe, D.J., Krotova-Khan, Y., Hurley, L.H. and Ebbinghaus, S.W. (2008) A novel G-quadruplex-forming GGA repeat region in the c-myb promoter is a critical regulator of promoter activity. *Nucleic Acids Res*, **36**, 1755-1769.

147. Morgan, R.K., Batra, H., Gaerig, V.C., Hockings, J. and Brooks, T.A. (2016) Identification and characterization of a new G-quadruplex forming region within the kRAS promoter as a transcriptional regulator. *Biochim Biophys Acta*, **1859**, 235-245.

148. Gonzalez, V. and Hurley, L.H. (2010) The c-MYC NHE III(1): function and regulation. *Annu Rev Pharmacol Toxicol*, **50**, 111-129.

149. Rigo, R. and Sissi, C. (2017) Characterization of G4-G4 Crosstalk in the c-KIT Promoter Region. *Biochemistry*, **56**, 4309-4312.

150. Zhao, C., Wu, L., Ren, J., Xu, Y. and Qu, X. (2013) Targeting human telomeric higher-order DNA: dimeric G-quadruplex units serve as preferred binding site. *J Am Chem Soc*, **135**, 18786-18789.

151. Ward, A.B., Sali, A. and Wilson, I.A. (2013) Biochemistry. Integrative structural biology. *Science*, **339**, 913-915.

152. Ambrus, A., Chen, D., Dai, J., Bialis, T., Jones, R.A. and Yang, D. (2006) Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res*, **34**, 2723-2735.

153. Luu, K.N., Phan, A.T., Kuryavyi, V., Lacroix, L. and Patel, D.J. (2006) Structure of the human telomere in K+ solution: an intramolecular (3 + 1) G-quadruplex scaffold. *J Am Chem Soc*, **128**, 9963-9970.

154. Phan, A.T., Kuryavyi, V., Luu, K.N. and Patel, D.J. (2007) Structure of two intramolecular G-quadruplexes formed by natural human telomere sequences in K+ solution. *Nucleic Acids Res*, **35**, 6517-6525.

155. Palm, W. and de Lange, T. (2008) How shelterin protects mammalian telomeres. *Annu Rev Genet*, **42**, 301-334.

156. Palm, W., Hockemeyer, D., Kibe, T. and de Lange, T. (2009) Functional dissection of human and mouse POT1 proteins. *Mol Cell Biol*, **29**, 471-482.

157. Renciuk, D., Kejnovska, I., Skolakova, P., Bednarova, K., Motlova, J. and Vorlickova, M. (2009) Arrangements of human telomere DNA quadruplex in physiologically relevant K+ solutions. *Nucleic Acids Res*, **37**, 6625-6634.

158. Xu, Y., Ishizuka, T., Kurabayashi, K. and Komiyama, M. (2009) Consecutive formation of G-quadruplexes in human telomeric-overhang DNA: a protective capping structure for telomere ends. *Angew Chem Int Ed Engl*, **48**, 7833-7836.

159. Wang, H., Nora, G.J., Ghodke, H. and Opresko, P.L. (2011) Single molecule studies of physiologically relevant telomeric tails reveal POT1 mechanism for promoting G-quadruplex unfolding. *J Biol Chem*, **286**, 7479-7489.

160. Kar, A., Jones, N., Arat, N.O., Fishel, R. and Griffith, J.D. (2018) Long repeating (TTAGGG) n single-stranded DNA self-condenses into compact beaded filaments stabilized by G-quadruplex formation. *J Biol Chem*, **293**, 9473-9485.

161. Abraham Punnoose, J., Ma, Y., Hoque, M.E., Cui, Y., Sasaki, S., Guo, A.H., Nagasawa, K. and Mao, H. (2018) Random Formation of G-Quadruplexes in the Full-Length Human Telomere Overhangs Leads to a Kinetic Folding Pattern with Targetable Vacant G-Tracts. *Biochemistry*, **57**, 6946-6955.

162. Petraccone, L., Spink, C., Trent, J.O., Garbett, N.C., Mekmaysy, C.S., Giancola, C. and Chaires, J.B. (2011) Structure and stability of higher-order human telomeric quadruplexes. *J Am Chem Soc*, **133**, 20951-20961.

163. Chaires, J.B., Dean, W.L., Le, H.T. and Trent, J.O. (2015) Hydrodynamic Models of G-Quadruplex Structures. *Methods Enzymol*, **562**, 287-304.

164. Schneidman-Duhovny, D., Kim, S.J. and Sali, A. (2012) Integrative structural modeling with small angle X-ray scattering profiles. *BMC Struct Biol*, **12**, 17.

165. Rout, M.P. and Sali, A. (2019) Principles for Integrative Structural Biology Studies. *Cell*, **177**, 1384-1403.

166. Chi, Q., Wang, G. and Jiang, J. (2013) The persistence length and length per base of single-stranded DNA obtained from fluorescence correlation spectroscopy measurements using mean field theory. *Physica A: Statistical Mechanics and its Applications*, **392**, 1072-1079.

167. Bloomfield, V.A., Crothers, D.M., Tinoco Jr., I. (2000) *Nucleic Acids: Structures, Properties, and Functions*. University Science Books, Sausalito.

168. Miller, M.C., Le, H.T., Dean, W.L., Holt, P.A., Chaires, J.B. and Trent, J.O. (2011) Polymorphism and resolution of oncogene promoter quadruplex-forming sequences. *Org Biomol Chem*, **9**, 7633-7637.

169. Irvine, G.B. (2001) Determination of molecular size by size-exclusion chromatography (gel filtration). *Curr Protoc Cell Biol*, **Chapter 5**, Unit 5 5.

170. Schuck, P. (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. *Biophys J*, **78**, 1606-1619.

171. Del Villar-Guerra, R., Gray, R.D. and Chaires, J.B. (2017) Characterization of Quadruplex DNA Structure by Circular Dichroism. *Curr Protoc Nucleic Acid Chem*, **68**, 17 18 11-17 18 16.

172. Kirby, N., Cowieson, N., Hawley, A.M., Mudie, S.T., McGillivray, D.J., Kusel, M., Samardzic-Boban, V. and Ryan, T.M. (2016) Improved radiation dose efficiency in solution SAXS using a sheath flow sample environment. *Acta Crystallogr D Struct Biol*, **72**, 1254-1266.

173. Hopkins, J.B., Gillilan, R.E. and Skou, S. (2017) BioXTAS RAW: improvements to a free open-source program for small-angle X-ray scattering data reduction and analysis. *J Appl Crystallogr*, **50**, 1545-1553.

174. Trewhella, J., Duff, A.P., Durand, D., Gabel, F., Guss, J.M., Hendrickson, W.A., Hura, G.L., Jacques, D.A., Kirby, N.M., Kwan, A.H. *et al.* (2017) 2017 publication guidelines for structural modelling of small-angle scattering data from biomolecules in solution: an update. *Acta Crystallogr D Struct Biol*, **73**, 710-728.

175. Franke, D., Petoukhov, M.V., Konarev, P.V., Panjkovich, A., Tuukkanen, A., Mertens, H.D.T., Kikhney, A.G., Hajizadeh, N.R., Franklin, J.M., Jeffries, C.M. *et al.* (2017) ATSAS 2.8: a comprehensive data analysis suite for small-angle scattering from macromolecular solutions. *J Appl Crystallogr*, **50**, 1212-1225.

176. Pettersen, E.F., Goddard, T.D., Huang, C.C., Couch, G.S., Greenblatt, D.M., Meng, E.C. and Ferrin, T.E. (2004) UCSF Chimera--a visualization system for exploratory research and analysis. *J Comput Chem*, **25**, 1605-1612.

177. Schrodinger. (2018). 11.8 ed. Schrodinger, LLC, New York, NY.

178. D.A. Case, R.M.B., D.S. Cerutti, T.E. Cheatham, III, T.A. Darden, R.E. Duke, T.J. Giese, H. Gohlke,, A.W. Goetz, N.H., S. Izadi, P. Janowski, J. Kaus, A. Kovalenko, T.S. Lee, S. LeGrand, P. Li, C., Lin, T.L., R. Luo, B. Madej, D. Mermelstein, K.M. Merz, G. Monard, H. Nguyen, H.T. Nguyen, I., Omelyan, A.O., D.R. Roe, A. Roitberg, C. Sagui, C.L. Simmerling, W.M. Botello-Smith, J. Swails, and R.C. Walker, J.W., R.M. Wolf, X. Wu, L. Xiao and P.A. Kollman. (2016). UCSF, University of California, San Francisco.

179. Ortega, A., Amoros, D. and Garcia de la Torre, J. (2011) Prediction of hydrodynamic and other solution properties of rigid proteins from atomic- and residue-level models. *Biophys J*, **101**, 892-898.

180. Le, H.T., Buscaglia, R., Dean, W.L., Chaires, J.B. and Trent, J.O. (2013) Calculation of hydrodynamic properties for G-quadruplex nucleic acid structures from in silico bead models. *Top Curr Chem*, **330**, 179-210.

181. Dolinsky, T.J., Nielsen, J.E., McCammon, J.A. and Baker, N.A. (2004) PDB2PQR: an automated pipeline for the setup of Poisson-Boltzmann electrostatics calculations. *Nucleic Acids Res*, **32**, W665-667.

182. Baker, N.A., Sept, D., Joseph, S., Holst, M.J. and McCammon, J.A. (2001) Electrostatics of nanosystems: application to microtubules and the ribosome. *Proc Natl Acad Sci U S A*, **98**, 10037-10041.

183. Tria, G., Mertens, H.D., Kachala, M. and Svergun, D.I. (2015) Advanced ensemble modelling of flexible macromolecules using X-ray solution scattering. *IUCrJ*, **2**, 207-217.

184. Bernado, P., Mylonas, E., Petoukhov, M.V., Blackledge, M. and Svergun, D.I. (2007) Structural characterization of flexible proteins using small-angle X-ray scattering. *J Am Chem Soc*, **129**, 5656-5664.

185. Svergun, D. (1992) Determination of the regularization parameter in indirect-transform methods using perceptual criteria. *Journal of Applied Crystallography*, **25**, 495-503.

186. Svergun, D., Barberato, C. and Koch, M.H.J. (1995) CRYSOL– a Program to Evaluate X-ray Solution Scattering of Biological Macromolecules from Atomic Coordinates. *Journal of Applied Crystallography*, **28**, 768-773.

187. Capp, J.A., Hagarman, A., Richardson, D.C. and Oas, T.G. (2014) The statistical conformation of a highly flexible protein: small-angle X-ray scattering of S. aureus protein A. *Structure*, **22**, 1184-1195.

188. Ortega, A. and García de la Torre, J. (2007) Equivalent Radii and Ratios of Radii from Solution Properties as Indicators of Macromolecular Conformation, Shape, and Flexibility. *Biomacromolecules*, **8**, 2464-2475.

189. Amorós, D., Ortega, A. and García de la Torre, J. (2011) Hydrodynamic Properties of Wormlike Macromolecules: Monte Carlo Simulation and Global Analysis of Experimental Data. *Macromolecules*, **44**, 5788-5797.

190. Garcia, H.G., Grayson, P., Han, L., Inamdar, M., Kondev, J., Nelson, P.C., Phillips, R., Widom, J. and Wiggins, P.A. (2007) Biological consequences of tightly bent DNA: the other life of a macromolecular celebrity. *Biopolymers*, **85**, 115-130.

191. Fink, H.-P. (1989) Structure analysis by small-angle X-ray and neutron scattering. *Acta Polymerica*, **40**, 224-224.

192. Kikhney, A.G. and Svergun, D.I. (2015) A practical guide to small angle X-ray scattering (SAXS) of flexible and intrinsically disordered proteins. *FEBS Lett*, **589**, 2570-2577.

193. Bhattacharjee, S.M., Giacometti, A. and Maritan, A. (2013) Flory theory for polymers. *J Phys Condens Matter*, **25**, 503101.

194. Murphy, M.C., Rasnik, I., Cheng, W., Lohman, T.M. and Ha, T. (2004) Probing single-stranded DNA conformational flexibility using fluorescence spectroscopy. *Biophys J*, **86**, 2530-2537.

195. Rambo, R.P. and Tainer, J.A. (2011) Characterizing flexible and intrinsically unstructured biological macromolecules by SAS using the Porod-Debye law. *Biopolymers*, **95**, 559-571.

196. Del Villar-Guerra, R., Trent, J.O. and Chaires, J.B. (2018) G-Quadruplex Secondary Structure Obtained from Circular Dichroism Spectroscopy. *Angew Chem Int Ed Engl*, **57**, 7171-7175.

197. Greve, J., Maestre, M.F. and Levin, A. (1977) Circular dichroism of adenine and thymine containing synthetic polynucleotides. *Biopolymers*, **16**, 1489-1504.

198. Lim, C.J., Zaug, A.J., Kim, H.J. and Cech, T.R. (2017) Reconstitution of human shelterin complexes reveals unexpected stoichiometry and dual pathways to enhance telomerase processivity. *Nat Commun*, **8**, 1075.

199. Lei, M., Podell, E.R. and Cech, T.R. (2004) Structure of human POT1 bound to telomeric single-stranded DNA provides a model for chromosome end-protection. *Nat Struct Mol Biol*, **11**, 1223-1229.

200. Taylor, D.J., Podell, E.R., Taatjes, D.J. and Cech, T.R. (2011) Multiple POT1-TPP1 proteins coat and compact long telomeric single-stranded DNA. *J Mol Biol*, **410**, 10-17.

201. Flynn, R.L., Chang, S. and Zou, L. (2012) RPA and POT1: friends or foes at telomeres? *Cell Cycle*, **11**, 652-657.

202. Flynn, R.L., Centore, R.C., O'Sullivan, R.J., Rai, R., Tse, A., Songyang, Z., Chang, S., Karlseder, J. and Zou, L. (2011) TERRA and hnRNPA1 orchestrate an RPA-to-POT1 switch on telomeric single-stranded DNA. *Nature*, **471**, 532-536.

203. Ray, S., Bandaria, J.N., Qureshi, M.H., Yildiz, A. and Balci, H. (2014) G-quadruplex formation in telomeres enhances POT1/TPP1 protection against RPA binding. *Proc Natl Acad Sci U S A*, **111**, 2990-2995.

204. Chen, Y. (2019) The structural biology of the shelterin complex. *Biol Chem*, **400**, 457-466.

205. McCauley, M.J. and Williams, M.C. (2007) Mechanisms of DNA binding determined in optical tweezers experiments. *Biopolymers*, **85**, 154-168.

206. Gao, C., Liu, Z., Hou, H., Ding, J., Chen, X., Xie, C., Song, Z., Hu, Z., Feng, M., Mohamed, H.I. *et al.* (2020) BMPQ-1 binds selectively to (3+1) hybrid topologies in human telomeric G-quadruplex multimers. *Nucleic Acids Res*.

207. Phatak, P., Cookson, J.C., Dai, F., Smith, V., Gartenhaus, R.B., Stevens, M.F. and Burger, A.M. (2007) Telomere uncapping by the G-quadruplex ligand RHPS4 inhibits clonogenic tumour cell growth in vitro and in vivo consistent with a cancer stem cell targeting mechanism. *Br J Cancer*, **96**, 1223-1233.

208. Darby, R.A., Sollogoub, M., McKeen, C., Brown, L., Risitano, A., Brown, N., Barton, C., Brown, T. and Fox, K.R. (2002) High throughput measurement of duplex, triplex and quadruplex melting curves using molecular beacons and a LightCycler. *Nucleic Acids Res*, **30**, e39.

209. Mergny, J.L. and Lacroix, L. (2003) Analysis of thermal melting curves. *Oligonucleotides*, **13**, 515-537.

210. Dean, W.L., Gray, R.D., DeLeeuw, L., Monsen, R.C. and Chaires, J.B. (2019) Putting a New Spin of G-Quadruplex Structure and Binding by Analytical Ultracentrifugation. *Methods Mol Biol*, **2035**, 87-103.

211. Wang, Q., Liu, J.Q., Chen, Z., Zheng, K.W., Chen, C.Y., Hao, Y.H. and Tan, Z. (2011) G-quadruplex formation at the 3' end of telomere DNA inhibits its extension by telomerase, polymerase and unwinding by helicase. *Nucleic Acids Res*, **39**, 6229-6237.

212. Jafri, M.A., Ansari, S.A., Alqahtani, M.H. and Shay, J.W. (2016) Roles of telomeres and telomerase in cancer, and advances in telomerase-targeted therapies. *Genome Med*, **8**, 69.

213. Hanahan, D. and Weinberg, R.A. (2011) Hallmarks of cancer: the next generation. *Cell*, **144**, 646-674.

214. Cunningham, A.P., Love, W.K., Zhang, R.W., Andrews, L.G. and Tollefsbol, T.O. (2006) Telomerase inhibition in cancer therapeutics: molecular-based approaches. *Curr Med Chem*, **13**, 2875-2888.

215. Gomez, D.L., Armando, R.G., Cerrudo, C.S., Ghiringhelli, P.D. and Gomez, D.E. (2016) Telomerase as a Cancer Target. Development of New Molecules. *Curr Top Med Chem*, **16**, 2432-2440.

216. Salloum, R., Hummel, T.R., Kumar, S.S., Dorris, K., Li, S., Lin, T., Daryani, V.M., Stewart, C.F., Miles, L., Poussaint, T.Y. *et al.* (2016) A molecular biology and phase II study of imetelstat (GRN163L) in children with recurrent or refractory central nervous system malignancies: a pediatric brain tumor consortium study. *J Neurooncol*, **129**, 443-451.

217. Takakura, M., Kyo, S., Kanaya, T., Hirano, H., Takeda, J., Yutsudo, M. and Inoue, M. (1999) Cloning of human telomerase catalytic subunit (hTERT) gene promoter and identification of proximal core promoter sequences essential for transcriptional activation in immortalized and cancer cells. *Cancer Res*, **59**, 551-557.

218. Vinagre, J., Almeida, A., Populo, H., Batista, R., Lyra, J., Pinto, V., Coelho, R., Celestino, R., Prazeres, H., Lima, L. *et al.* (2013) Frequency of TERT promoter mutations in human cancers. *Nat Commun*, **4**, 2185.

219. Borah, S., Xi, L., Zaug, A.J., Powell, N.M., Dancik, G.M., Cohen, S.B., Costello, J.C., Theodorescu, D. and Cech, T.R. (2015) Cancer. TERT promoter mutations and telomerase reactivation in urothelial cancer. *Science*, **347**, 1006-1010.

220. Huang, F.W., Hodis, E., Xu, M.J., Kryukov, G.V., Chin, L. and Garraway, L.A. (2013) Highly recurrent TERT promoter mutations in human melanoma. *Science*, **339**, 957-959.

221. Killela, P.J., Reitman, Z.J., Jiao, Y., Bettegowda, C., Agrawal, N., Diaz, L.A., Jr., Friedman, A.H., Friedman, H., Gallia, G.L., Giovanella, B.C. *et al.* (2013) TERT promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal. *Proc Natl Acad Sci U S A*, **110**, 6021-6026.

222. Bell, R.J., Rube, H.T., Kreig, A., Mancini, A., Fouse, S.D., Nagarajan, R.P., Choi, S., Hong, C., He, D., Pekmezci, M. *et al.* (2015) Cancer. The transcription factor GABP selectively binds and activates the mutant TERT promoter in cancer. *Science*, **348**, 1036-1039.

223. Kang, H.J., Cui, Y., Yin, H., Scheid, A., Hendricks, W.P., Schmidt, J., Sekulic, A., Kong, D., Trent, J.M., Gokhale, V. *et al.* (2016) A Pharmacological Chaperone Molecule Induces Cancer Cell Death by Restoring Tertiary DNA Structures in Mutant hTERT Promoters. *J Am Chem Soc*.

224. Song, J.H., Kang, H.J., Luevano, L.A., Gokhale, V., Wu, K., Pandey, R., Sherry Chow, H.H., Hurley, L.H. and Kraft, A.S. (2019) Small-Molecule-Targeting Hairpin Loop of hTERT Promoter G-Quadruplex Induces Cancer Cell Death. *Cell Chem Biol*, **26**, 1110-1121 e1114.

225. Saltzberg, D., Greenberg, C.H., Viswanath, S., Chemmama, I., Webb, B., Pellarin, R., Echeverria, I. and Sali, A. (2019) Modeling Biological Complexes Using Integrative Modeling Platform. *Methods Mol Biol*, **2022**, 353-377.

226. Franke, D. and Svergun, D.I. (2009) DAMMIF, a program for rapid ab-initio shape determination in small-angle scattering. *J Appl Crystallogr*, **42**, 342-346.

227. Harkness, R.W.t. and Mittermaier, A.K. (2016) G-register exchange dynamics in guanine quadruplexes. *Nucleic Acids Res*, **44**, 3481-3494.

228. Holm, A.I., Kohler, B., Hoffmann, S.V. and Brondsted Nielsen, S. (2010) Synchrotron radiation circular dichroism of various G-quadruplex structures. *Biopolymers*, **93**, 429-433.

229. Ambrus, A., Chen, D., Dai, J., Jones, R.A. and Yang, D. (2005) Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization. *Biochemistry*, **44**, 2048-2058.

230. Masiero, S., Trotta, R., Pieraccini, S., De Tito, S., Perone, R., Randazzo, A. and Spada, G.P. (2010) A non-empirical chromophoric interpretation of CD spectra of DNA G-quadruplex structures. *Org Biomol Chem*, **8**, 2683-2692.

231. Galas, D.J. and Schmitz, A. (1978) DNAse footprinting: a simple method for the detection of protein-DNA binding specificity. *Nucleic Acids Res*, **5**, 3157-3170.

232. Brenowitz, M., Senear, D.F., Shea, M.A. and Ackers, G.K. (1986) Quantitative DNase footprint titration: a method for studying protein-DNA interactions. *Methods Enzymol*, **130**, 132-181.

233. Herrera, J.E. and Chaires, J.B. (1994) Characterization of preferred deoxyribonuclease I cleavage sites. *J Mol Biol*, **236**, 405-411.

234. Weston, S.A., Lahm, A. and Suck, D. (1992) X-ray structure of the DNase I-d(GGTATACC)2 complex at 2.3 A resolution. *J Mol Biol*, **226**, 1237-1256.

235. Perrin, F. (1934) Brownian Motion of an Ellipsoid - I. Dielectric dispersion for ellipsoidal molecules. *J. Phys. Radium*, **5**, 497 - 511.

236. Garcia de la Torre, J. and Harding, S.E. (2013) Hydrodynamic modelling of protein conformation in solution: ELLIPS and HYDRO. *Biophys Rev*, **5**, 195-206.

237. Le, H.T., Dean, W.L., Buscaglia, R., Chaires, J.B. and Trent, J.O. (2014) An investigation of G-quadruplex structural polymorphism in the human telomere using a combined approach of hydrodynamic bead modeling and molecular dynamics simulation. *J Phys Chem B*, **118**, 5390-5405.

238. Miller, M.C. and Trent, J.O. (2011) Resolution of quadruplex polymorphism by size-exclusion chromatography. *Curr Protoc Nucleic Acid Chem*, **Chapter 17**, Unit17 13.

239. Serdyuk, I.N., Zaccai, N. R. & Zaccai, G. (2007) *Methods in Molecular Biophysics: Structure, Dynamics, Function*. Cambridge University Press.

240. Rambo, R.P. and Tainer, J.A. (2010) Improving small-angle X-ray scattering data for structural analyses of the RNA world. *RNA*, **16**, 638-646.

241. Grun, J.T., Hennecker, C., Klotzner, D.P., Harkness, R.W., Bessi, I., Heckel, A., Mittermaier, A.K. and Schwalbe, H. (2019) Conformational Dynamics of Strand Register Shifts in DNA G-Quadruplexes. *J Am Chem Soc*.

242. Lebowitz, J., Lewis, M.S. and Schuck, P. (2002) Modern analytical ultracentrifugation in protein science: a tutorial review. *Protein Sci*, **11**, 2067-2079.

243. Do, N.Q., Lim, K.W., Teo, M.H., Heddi, B. and Phan, A.T. (2011) Stacking of G-quadruplexes: NMR structure of a G-rich oligonucleotide with potential anti-HIV and anticancer activity. *Nucleic Acids Res*, **39**, 9448-9457.

244. Kato, Y., Ohyama, T., Mita, H. and Yamamoto, Y. (2005) Dynamics and thermodynamics of dimerization of parallel G-quadruplexed DNA formed from d(TTAGn) (n=3-5). *J Am Chem Soc*, **127**, 9980-9981.

245. Meier, M., Moya-Torres, A., Krahn, N.J., McDougall, M.D., Orriss, G.L., McRae, E.K.S., Booy, E.P., McEleney, K., Patel, T.R., McKenna, S.A. *et al.* (2018) Structure and hydrodynamics of a DNA G-quadruplex with a cytosine bulge. *Nucleic Acids Res*.

246. Schiavone, D., Guilbaud, G., Murat, P., Papadopoulou, C., Sarkies, P., Prioleau, M.N., Balasubramanian, S. and Sale, J.E. (2014) Determinants of G quadruplex-induced epigenetic instability in REV1-deficient cells. *EMBO J*, **33**, 2507-2520.

247. Takahama, K., Takada, A., Tada, S., Shimizu, M., Sayama, K., Kurokawa, R. and Oyoshi, T. (2013) Regulation of telomere length by G-quadruplex telomere DNA- and TERRA-binding protein TLS/FUS. *Chem Biol*, **20**, 341-350.

248. Hansel-Hertsch, R., Di Antonio, M. and Balasubramanian, S. (2017) DNA G-quadruplexes in the human genome: detection, functions and therapeutic potential. *Nat Rev Mol Cell Biol*, **18**, 279-284.

249. Guilbaud, G., Murat, P., Recolin, B., Campbell, B.C., Maiter, A., Sale, J.E. and Balasubramanian, S. (2017) Local epigenetic reprogramming induced by G-quadruplex ligands. *Nat Chem*, **9**, 1110-1117.

250. Bochman, M.L., Paeschke, K. and Zakian, V.A. (2012) DNA secondary structures: stability and function of G-quadruplex structures. *Nat Rev Genet*, **13**, 770-780.

251. Hansel-Hertsch, R., Beraldi, D., Lensing, S.V., Marsico, G., Zyner, K., Parry, A., Di Antonio, M., Pike, J., Kimura, H., Narita, M. *et al.* (2016) G-quadruplex structures mark human regulatory chromatin. *Nat Genet*, **48**, 1267-1272.

252. Lipinski, C.A., Lombardo, F., Dominy, B.W. and Feeney, P.J. (2001) Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Adv Drug Deliv Rev*, **46**, 3-26.

253. Shi, X., Barkigia, K.M., Fajer, J. and Drain, C.M. (2001) Design and synthesis of porphyrins bearing rigid hydrogen bonding motifs: highly versatile building blocks for self-assembly of polymers and discrete arrays. *J Org Chem*, **66**, 6513-6522.

254. Izbicka, E., Wheelhouse, R.T., Raymond, E., Davidson, K.K., Lawrence, R.A., Sun, D., Windle, B.E., Hurley, L.H. and Von Hoff, D.D. (1999) Effects of cationic porphyrins as G-quadruplex interactive agents in human tumor cells. *Cancer Res*, **59**, 639-644.

255. Wei, C., Han, G., Jia, G., Zhou, J. and Li, C. (2008) Study on the interaction of porphyrin with G-quadruplex DNAs. *Biophys Chem*, **137**, 19-23.

256. Maraval, A., Franco, S., Vialas, C., Pratviel, G., Blasco, M.A. and Meunier, B. (2003) Porphyrin-aminoquinoline conjugates as telomerase inhibitors. *Org Biomol Chem*, **1**, 921-927.

257. Kieltyka, R., Fakhoury, J., Moitessier, N. and Sleiman, H.F. (2008) Platinum phenanthroimidazole complexes as G-quadruplex DNA selective binders. *Chemistry*, **14**, 1145-1154.

258. Pierce, S.E., Kieltyka, R., Sleiman, H.F. and Brodbelt, J.S. (2009) Evaluation of binding selectivities and affinities of platinum-based quadruplex interactive complexes by electrospray ionization mass spectrometry. *Biopolymers*, **91**, 233-243.

259. Zhang, H., Xiang, J., Hu, H., Liu, Y., Yang, F., Shen, G., Tang, Y. and Chen, C. (2015) Selective recognition of specific G-quadruplex vs. duplex DNA by a phenanthroline derivative. *Int J Biol Macromol*, **78**, 149-156.

260. Mergny, J.L., Lacroix, L., Teulade-Fichou, M.P., Hounsou, C., Guittat, L., Hoarau, M., Arimondo, P.B., Vigneron, J.P., Lehn, J.M., Riou, J.F. *et al.* (2001) Telomerase inhibitors based on quadruplex ligands selected by a fluorescence assay. *Proc Natl Acad Sci U S A*, **98**, 3062-3067.

261. Gama, S., Rodrigues, I., Mendes, F., Santos, I.C., Gabano, E., Klejevskaja, B., Gonzalez-Garcia, J., Ravera, M., Vilar, R. and Paulo, A. (2016) Anthracene-terpyridine metal complexes as new G-quadruplex DNA binders. *J Inorg Biochem*, **160**, 275-286.

262. Perry, P.J., Gowan, S.M., Reszka, A.P., Polucci, P., Jenkins, T.C., Kelland, L.R. and Neidle, S. (1998) 1,4- and 2,6-disubstituted amidoanthracene-9,10-dione derivatives as inhibitors of human telomerase. *J Med Chem*, **41**, 3253-3260.

263. Hampel, S.M., Sidibe, A., Gunaratnam, M., Riou, J.F. and Neidle, S. (2010) Tetrasubstituted naphthalene diimide ligands with selectivity for telomeric G-quadruplexes and cancer cells. *Bioorg Med Chem Lett*, **20**, 6459-6463.

264. Paritala, H. and Firestine, S.M. (2009) Benzo(h)quinoline derivatives as G-quadruplex binding agents. *Bioorg Med Chem Lett*, **19**, 1584-1587.

265. Zeng, D.Y., Kuang, G.T., Wang, S.K., Peng, W., Lin, S.L., Zhang, Q., Su, X.X., Hu, M.H., Wang, H., Tan, J.H. *et al.* (2017) Discovery of Novel 11-Triazole Substituted Benzofuro[3,2-b]quinolone Derivatives as c-myc G-Quadruplex Specific Stabilizers via Click Chemistry. *J Med Chem*, **60**, 5407-5423.

266. Martino, L., Virno, A., Pagano, B., Virgilio, A., Di Micco, S., Galeone, A., Giancola, C., Bifulco, G., Mayol, L. and Randazzo, A. (2007) Structural and thermodynamic studies of the interaction of distamycin A with the parallel quadruplex structure [d(TGGGGT)]4. *J Am Chem Soc*, **129**, 16048-16056.

267. Pagano, B., Fotticchia, I., De Tito, S., Mattia, C.A., Mayol, L., Novellino, E., Randazzo, A. and Giancola, C. (2010) Selective Binding of Distamycin A Derivative to G-Quadruplex Structure [d(TGGGGT)](4). *J Nucleic Acids*, **2010**.

268. Randazzo, A., Galeone, A., Esposito, V., Varra, M. and Mayol, L. (2002) Interaction of distamycin A and netropsin with quadruplex and duplex structures: a comparative 1H-NMR study. *Nucleosides Nucleotides Nucleic Acids*, **21**, 535-545.

269. Holt, P.A., Buscaglia, R., Trent, J.O. and Chaires, J.B. (2011) A Discovery Funnel for Nucleic Acid Binding Drug Candidates. *Drug Dev Res*, **72**, 178-186.

270. Shoichet, B.K. (2004) Virtual screening of chemical libraries. *Nature*, **432**, 862-865.

271. Lionta, E., Spyrou, G., Vassilatis, D.K. and Cournia, Z. (2014) Structure-based virtual screening for drug discovery: principles, applications and recent advances. *Curr Top Med Chem*, **14**, 1923-1938.

272. Horvath, D. (2011) Pharmacophore-based virtual screening. *Methods Mol Biol*, **672**, 261-298.

273. Koes, D.R. and Camacho, C.J. (2011) Pharmer: efficient and exact pharmacophore search. *J Chem Inf Model*, **51**, 1307-1314.

274. Wolber, G. and Langer, T. (2005) LigandScout: 3-D pharmacophores derived from protein-bound ligands and their use as virtual screening filters. *J Chem Inf Model*, **45**, 160-169.

275. Labute, P., Williams, C., Feher, M., Sourial, E. and Schmidt, J.M. (2001) Flexible alignment of small molecules. *J Med Chem*, **44**, 1483-1490.

276. Dixon, S.L., Smondyrev, A.M. and Rao, S.N. (2006) PHASE: a novel approach to pharmacophore modeling and 3D database searching. *Chem Biol Drug Des*, **67**, 370-372.

277. Taminau, J., Thijs, G. and De Winter, H. (2008) Pharao: pharmacophore alignment and optimization. *J Mol Graph Model*, **27**, 161-169.

278. Chen, S.B., Tan, J.H., Ou, T.M., Huang, S.L., An, L.K., Luo, H.B., Li, D., Gu, L.Q. and Huang, Z.S. (2011) Pharmacophore-based discovery of triaryl-substituted imidazole as new telomeric G-quadruplex ligand. *Bioorg Med Chem Lett*, **21**, 1004-1009.

279. Waller, Z.A., Shirude, P.S., Rodriguez, R. and Balasubramanian, S. (2008) Triarylpyridines: a versatile small molecule scaffold for G-quadruplex recognition. *Chem Commun (Camb)*, 1467-1469.

280. Lemmen, C. and Lengauer, T. (2000) Computational methods for the structural alignment of molecules. *J Comput Aided Mol Des*, **14**, 215-232.

281. Musumeci, D., Amato, J., Zizza, P., Platella, C., Cosconati, S., Cingolani, C., Biroccio, A., Novellino, E., Randazzo, A., Giancola, C. *et al.* (2017) Tandem application of ligand-based virtual screening and G4-OAS assay to identify novel G-quadruplex-targeting chemotypes. *Biochim Biophys Acta Gen Subj*, **1861**, 1341-1352.

282. Kaserer, T., Rigo, R., Schuster, P., Alcaro, S., Sissi, C. and Schuster, D. (2016) Optimized Virtual Screening Workflow for the Identification of Novel G-Quadruplex Ligands. *J Chem Inf Model*, **56**, 484-500.

283. Shan, C., Lin, J., Hou, J.Q., Liu, H.Y., Chen, S.B., Chen, A.C., Ou, T.M., Tan, J.H., Li, D., Gu, L.Q. *et al.* (2015) Chemical intervention of the NM23-H2 transcriptional programme on c-MYC via a novel small molecule. *Nucleic Acids Res*, **43**, 6677-6691.

284. Kang, H.J. and Park, H.J. (2015) In silico identification of novel ligands for G-quadruplex in the c-MYC promoter. *J Comput Aided Mol Des*, **29**, 339-348.

285. Alcaro, S., Musetti, C., Distinto, S., Casatti, M., Zagotto, G., Artese, A., Parrotta, L., Moraca, F., Costa, G., Ortuso, F. *et al.* (2013) Identification and characterization of new DNA G-quadruplex binders selected by a combination of ligand and structure-based virtual screening approaches. *J Med Chem*, **56**, 843-855.

286. Kar, R.K., Suryadevara, P., Jana, J., Bhunia, A. and Chatterjee, S. (2013) Novel G-quadruplex stabilizing agents: in-silico approach and dynamics. *J Biomol Struct Dyn*, **31**, 1497-1518.

287. Levesque, M.J., Ichikawa, K., Date, S. and Haga, J.H. (2009) Design of a grid service-based platform for in silico protein-ligand screenings. *Comput Methods Programs Biomed*, **93**, 73-82.

288. Schneider, G. (2010) Virtual screening: an endless staircase? *Nat Rev Drug Discov*, **9**, 273-276.

289. Kuntz, I.D., Blaney, J.M., Oatley, S.J., Langridge, R. and Ferrin, T.E. (1982) A geometric approach to macromolecule-ligand interactions. *J Mol Biol*, **161**, 269-288.

290. Hevener, K.E., Zhao, W., Ball, D.M., Babaoglu, K., Qi, J., White, S.W. and Lee, R.E. (2009) Validation of molecular docking programs for virtual screening against dihydropteroate synthase. *J Chem Inf Model*, **49**, 444-460.

291. Warren, G.L., Andrews, C.W., Capelli, A.M., Clarke, B., LaLonde, J., Lambert, M.H., Lindvall, M., Nevins, N., Semus, S.F., Senger, S. *et al.* (2006) A critical assessment of docking programs and scoring functions. *J Med Chem*, **49**, 5912-5931.

292. Kontoyianni, M., McClellan, L.M. and Sokol, G.S. (2004) Evaluation of docking performance: comparative data on docking algorithms. *J Med Chem*, **47**, 558-565.

293. Ewing, T.J., Makino, S., Skillman, A.G. and Kuntz, I.D. (2001) DOCK 4.0: search strategies for automated molecular docking of flexible molecule databases. *J Comput Aided Mol Des*, **15**, 411-428.

294. Morris, G.M., Huey, R., Lindstrom, W., Sanner, M.F., Belew, R.K., Goodsell, D.S. and Olson, A.J. (2009) AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem*, **30**, 2785-2791.

295. Trott, O. and Olson, A.J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem*, **31**, 455-461.

296. Jones, G., Willett, P., Glen, R.C., Leach, A.R. and Taylor, R. (1997) Development and validation of a genetic algorithm for flexible docking. *J Mol Biol*, **267**, 727-748.

297. Jain, A.N. (2003) Surflex: fully automatic flexible molecular docking using a molecular similarity-based search engine. *J Med Chem*, **46**, 499-511.

298. Friesner, R.A., Banks, J.L., Murphy, R.B., Halgren, T.A., Klicic, J.J., Mainz, D.T., Repasky, M.P., Knoll, E.H., Shelley, M., Perry, J.K. *et al.* (2004) Glide: a new approach for rapid, accurate docking and scoring. 1. Method and assessment of docking accuracy. *J Med Chem*, **47**, 1739-1749.

299. Neves, M.A., Totrov, M. and Abagyan, R. (2012) Docking and scoring with ICM: the benchmarking results and strategies for improvement. *J Comput Aided Mol Des*, **26**, 675-686.

300. Pagadala, N.S., Syed, K. and Tuszynski, J. (2017) Software for molecular docking: a review. *Biophys Rev*, **9**, 91-102.

301. Holt, P.A., Chaires, J.B. and Trent, J.O. (2008) Molecular docking of intercalators and groove-binders to nucleic acids using Autodock and Surflex. *J Chem Inf Model*, **48**, 1602-1615.

302. Meng, X.Y., Zhang, H.X., Mezei, M. and Cui, M. (2011) Molecular docking: a powerful approach for structure-based drug discovery. *Curr Comput Aided Drug Des*, **7**, 146-157.

303. Lorber, D.M. and Shoichet, B.K. (2005) Hierarchical docking of databases of multiple ligand conformations. *Curr Top Med Chem*, **5**, 739-749.

304. Lang, P.T., Brozell, S.R., Mukherjee, S., Pettersen, E.F., Meng, E.C., Thomas, V., Rizzo, R.C., Case, D.A., James, T.L. and Kuntz, I.D. (2009) DOCK 6: combining techniques to model RNA-small molecule complexes. *RNA*, **15**, 1219-1230.

305. Welch, W., Ruppert, J. and Jain, A.N. (1996) Hammerhead: fast, fully automated docking of flexible ligands to protein binding sites. *Chem Biol*, **3**, 449-462.

306. Hou, J.Q., Chen, S.B., Zan, L.P., Ou, T.M., Tan, J.H., Luyt, L.G. and Huang, Z.S. (2015) Identification of a selective G-quadruplex DNA binder using a multistep virtual screening approach. *Chem Commun (Camb)*, **51**, 198-201.

307. Hart, T.N. and Read, R.J. (1992) A multiple-start Monte Carlo docking method. *Proteins*, **13**, 206-222.

308. Bursulaya, B.D., Totrov, M., Abagyan, R. and Brooks, C.L., 3rd. (2003) Comparative study of several algorithms for flexible ligand docking. *J Comput Aided Mol Des*, **17**, 755-763.

309. Lee, H.M., Chan, D.S., Yang, F., Lam, H.Y., Yan, S.C., Che, C.M., Ma, D.L. and Leung, C.H. (2010) Identification of natural product fonsecin B as a stabilizing ligand of c-myc G-quadruplex DNA by high-throughput virtual screening. *Chem Commun (Camb)*, **46**, 4680-4682.

310. Chan, D.S., Yang, H., Kwan, M.H., Cheng, Z., Lee, P., Bai, L.P., Jiang, Z.H., Wong, C.Y., Fong, W.F., Leung, C.H. *et al.* (2011) Structure-based optimization of FDA-approved drug methylene blue as a c-myc G-quadruplex DNA stabilizer. *Biochimie*, **93**, 1055-1064.

311. Ma, D.L., Chan, D.S., Fu, W.C., He, H.Z., Yang, H., Yan, S.C. and Leung, C.H. (2012) Discovery of a natural product-like c-myc G-quadruplex DNA groove-binder by molecular docking. *PLoS One*, **7**, e43278.

312. Castro-Alvarez, A., Costa, A.M. and Vilarrasa, J. (2017) The Performance of Several Docking Programs at Reproducing Protein-Macrolide-Like Crystal Structures. *Molecules*, **22**.

313. Hawkins, P.C., Skillman, A.G. and Nicholls, A. (2007) Comparison of shape-matching and docking as virtual screening tools. *J Med Chem*, **50**, 74-82.

314. Morris, G.M., Goodsell, D.S., Halliday, R.S., Huey, R., Hart, W.E., Belew, R.K. and Olson, A.J. (1998) Automated docking using a Lamarckian genetic algorithm and an empirical binding free energy function. *Journal of Computational Chemistry*, **19**, 1639-1662.

315. Castillo-Gonzalez, D., Mergny, J.L., De Rache, A., Perez-Machado, G., Cabrera-Perez, M.A., Nicolotti, O., Introcaso, A., Mangiatordi, G.F., Guedin, A., Bourdoncle, A. *et al.* (2015) Harmonization of QSAR Best Practices and Molecular Docking Provides an Efficient Virtual Screening Tool for Discovering New G-Quadruplex Ligands. *J Chem Inf Model*, **55**, 2094-2110.

316. Ma, D.L., Lai, T.S., Chan, F.Y., Chung, W.H., Abagyan, R., Leung, Y.C. and Wong, K.Y. (2008) Discovery of a drug-like G-quadruplex binding ligand by high-throughput docking. *ChemMedChem*, **3**, 881-884.

317. Cosconati, S., Marinelli, L., Trotta, R., Virno, A., Mayol, L., Novellino, E., Olson, A.J. and Randazzo, A. (2009) Tandem application of virtual screening and NMR experiments in the discovery of brand new DNA quadruplex groove binders. *J Am Chem Soc*, **131**, 16336-16337.

318. Trotta, R., De Tito, S., Lauri, I., La Pietra, V., Marinelli, L., Cosconati, S., Martino, L., Conte, M.R., Mayol, L., Novellino, E. *et al.* (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**, 1280-1287.

319. Di Leva, F.S., Zizza, P., Cingolani, C., D'Angelo, C., Pagano, B., Amato, J., Salvati, E., Sissi, C., Pinato, O., Marinelli, L. *et al.* (2013) Exploring the chemical space of G-quadruplex binders: discovery of a novel chemotype targeting the human telomeric sequence. *J Med Chem*, **56**, 9646-9654.

320. Amato, J., Pagano, A., Cosconati, S., Amendola, G., Fotticchia, I., Iaccarino, N., Marinello, J., De Magis, A., Capranico, G., Novellino, E. *et al.* (2017) Discovery of the first dual G-triplex/G-quadruplex stabilizing compound: a new opportunity in the targeting of G-rich DNA structures? *Biochim Biophys Acta Gen Subj*, **1861**, 1271-1280.

321. Gray, R.D., Trent, J.O. and Chaires, J.B. (2014) Folding and unfolding pathways of the human telomeric G-quadruplex. *J Mol Biol*, **426**, 1629-1650.

322. Weiner, S.J., Kollman, P.A., Case, D.A., Singh, U.C., Ghio, C., Alagona, G., Profeta, S. and Weiner, P. (1984) A new force field for molecular mechanical simulation of nucleic acids and proteins. *Journal of the American Chemical Society*, **106**, 765-784.

323. Hospital, A., Goni, J.R., Orozco, M. and Gelpi, J.L. (2015) Molecular dynamics simulations: advances and applications. *Adv Appl Bioinform Chem*, **8**, 37-47.

324. Leach, A.R. (2001) *Molecular Dynamics Simulation Methods, Molecular Modelling: Principles and Applications*. Pearson/Prentice Hall.

325. Moraca, F., Amato, J., Ortuso, F., Artese, A., Pagano, B., Novellino, E., Alcaro, S., Parrinello, M. and Limongelli, V. (2017) Ligand binding to telomeric G-quadruplex DNA investigated by funnel-metadynamics simulations. *Proc Natl Acad Sci U S A*, **114**, E2136-E2145.

326. Rocca, R., Moraca, F., Costa, G., Nadai, M., Scalabrin, M., Talarico, C., Distinto, S., Maccioni, E., Ortuso, F., Artese, A. *et al.* (2017) Identification of G-quadruplex DNA/RNA binders: Structure-based virtual screening and biophysical characterization. *Biochim Biophys Acta Gen Subj*, **1861**, 1329-1340.

327. Nakatani, K., Hagihara, S., Sando, S., Sakamoto, S., Yamaguchi, K., Maesawa, C. and Saito, I. (2003) Induction of a remarkable conformational change in a human telomeric sequence by the binding of naphthyridine dimer: inhibition of the elongation of a telomeric repeat by telomerase. *J Am Chem Soc*, **125**, 662-666.

328. Bhat, J., Mondal, S., Sengupta, P. and Chatterjee, S. (2017) In Silico Screening and Binding Characterization of Small Molecules toward a G-Quadruplex Structure Formed in the Promoter Region of c-MYC Oncogene. *ACS Omega*, **2**, 4382-4397.

329. Limongelli, V., Bonomi, M. and Parrinello, M. (2013) Funnel metadynamics as accurate binding free-energy method. *Proc Natl Acad Sci U S A*, **110**, 6358-6363.

330. Liu, J. and Wang, R. (2015) Classification of current scoring functions. *J Chem Inf Model*, **55**, 475-482.

331. Fenu, L.A., Lewis, R.A., Good A.C., Bodkin, M., Essex J.W. (2007) *Structure-Based Drug Discovery*. Dordrecht: Springer.

332. Sotriffer, C.A. (2011) Accounting for induced-fit effects in docking: what is possible and what is not? *Curr Top Med Chem*, **11**, 179-191.

333. Garbett, N.C., Mekmaysy, C.S. and Chaires, J.B. (2010) Sedimentation velocity ultracentrifugation analysis for hydrodynamic characterization of G-quadruplex structures. *Methods Mol Biol*, **608**, 97-120.

334. Triballeau, N., Acher, F., Brabet, I., Pin, J.P. and Bertrand, H.O. (2005) Virtual screening workflow development guided by the "receiver operating characteristic" curve approach.

Application to high-throughput docking on metabotropic glutamate receptor subtype 4. *J Med Chem*, **48**, 2534-2547.

335. Rocca, R., Moraca, F., Costa, G., Alcaro, S., Distinto, S., Maccioni, E., Ortuso, F., Artese, A. and Parrotta, L. (2014) Structure-based virtual screening of novel natural alkaloid derivatives as potential binders of h-telo and c-myc DNA G-quadruplex conformations. *Molecules*, **20**, 206-223.

336. Polgar, T. (2007) [The role of structure based virtual screening in the early phase of drug discovery]. *Acta Pharm Hung*, **77**, 223-234.

337. Jain, A.N. (2007) Surflex-Dock 2.1: robust performance from ligand energetic modeling, ring flexibility, and knowledge-based search. *J Comput Aided Mol Des*, **21**, 281-306.

338. Rigo, R., Palumbo, M. and Sissi, C. (2017) G-quadruplexes in human promoters: A challenge for therapeutic applications. *Biochim Biophys Acta Gen Subj*, **1861**, 1399-1413.

339. Sahakyan, A.B., Murat, P., Mayer, C. and Balasubramanian, S. (2017) G-quadruplex structures within the 3' UTR of LINE-1 elements stimulate retrotransposition. *Nat Struct Mol Biol*, **24**, 243-247.

340. Heikamp, K. and Bajorath, J. (2013) The future of virtual compound screening. *Chem Biol Drug Des*, **81**, 33-40.

341. Yuriev, E. (2014) Challenges and advances in structure-based virtual screening. *Future Med Chem*, **6**, 5-7.

342. Chen, H., Engkvist, O., Wang, Y., Olivecrona, M. and Blaschke, T. (2018) The rise of deep learning in drug discovery. *Drug Discov Today*, **23**, 1241-1250.

343. Zhang, X., Zhao, B., Yan, T., Hao, A., Gao, Y., Li, D. and Sui, G. (2018) G-quadruplex structures at the promoter of HOXC10 regulate its expression. *Biochim Biophys Acta Gene Regul Mech*, **1861**, 1018-1028.

344. Brooks, T.A., Kendrick, S. and Hurley, L. (2010) Making sense of G-quadruplex and i-motif functions in oncogene promoters. *FEBS J*, **277**, 3459-3469.

345. Carabet, L.A., Rennie, P.S. and Cherkasov, A. (2018) Therapeutic Inhibition of Myc in Cancer. Structural Bases and Computer-Aided Drug Discovery Approaches. *Int J Mol Sci*, **20**.

346. Dhamodharan, V. and Pradeepkumar, P.I. (2019) Specific Recognition of Promoter G-Quadruplex DNAs by Small Molecule Ligands and Light-up Probes. *ACS Chem Biol*, **14**, 2102-2114.

347. Monsen, R.C. and Trent, J.O. (2018) G-quadruplex virtual drug screening: A review. *Biochimie*, **152**, 134-148.

348. Ma, Y., Iida, K. and Nagasawa, K. (2020) Topologies of G-quadruplex: Biological functions and regulation by ligands. *Biochem Biophys Res Commun*.

349. Kataoka, Y., Fujita, H., Kasahara, Y., Yoshihara, T., Tobita, S. and Kuwahara, M. (2014) Minimal thioflavin T modifications improve visual discrimination of guanine-quadruplex topologies and alter compound-induced topological structures. *Anal Chem*, **86**, 12078-12084.

350. Rasadean, D.M., Sheng, B., Dash, J. and Pantos, G.D. (2017) Amino-Acid-Derived Naphthalenediimides as Versatile G-Quadruplex Binders. *Chemistry*, **23**, 8491-8499.

351. Dhamodharan, V., Harikrishna, S., Bhasikuttan, A.C. and Pradeepkumar, P.I. (2015) Topology specific stabilization of promoter over telomeric G-quadruplex DNAs by bisbenzimidazole carboxamide derivatives. *ACS Chem Biol*, **10**, 821-833.

352. Ducani, C., Bernardinelli, G., Hogberg, B., Keppler, B.K. and Terenzi, A. (2019) Interplay of Three G-Quadruplex Units in the KIT Promoter. *J Am Chem Soc*, **141**, 10205-10213.

353. Cong, Y.S., Wright, W.E. and Shay, J.W. (2002) Human telomerase and its regulation. *Microbiol Mol Biol Rev*, **66**, 407-425, table of contents.

354. Hayflick, L. (1998) A brief history of the mortality and immortality of cultured cells. *Keio J Med*, **47**, 174-182.

355. Bodnar, A.G., Ouellette, M., Frolkis, M., Holt, S.E., Chiu, C.P., Morin, G.B., Harley, C.B., Shay, J.W., Lichtsteiner, S. and Wright, W.E. (1998) Extension of life-span by introduction of telomerase into normal human cells. *Science*, **279**, 349-352.

356. Jiang, X.R., Jimenez, G., Chang, E., Frolkis, M., Kusler, B., Sage, M., Beeche, M., Bodnar, A.G., Wahl, G.M., Tlsty, T.D. *et al.* (1999) Telomerase expression in human somatic cells does not induce changes associated with a transformed phenotype. *Nat Genet*, **21**, 111-114.

357. Uziel, O., Beery, E., Dronichev, V., Samocha, K., Gryaznov, S., Weiss, L., Slavin, S., Kushnir, M., Nordenberg, Y., Rabinowitz, C. *et al.* (2010) Telomere shortening sensitizes cancer cells to selected cytotoxic agents: in vitro and in vivo studies and putative mechanisms. *PLoS One*, **5**, e9132.

358. Guo, K., Gokhale, V., Hurley, L.H. and Sun, D. (2008) Intramolecularly folded G-quadruplex and i-motif structures in the proximal promoter of the vascular endothelial growth factor gene. *Nucleic Acids Res*, **36**, 4598-4608.

359. Dong, X., Liu, A., Zer, C., Feng, J., Zhen, Z., Yang, M. and Zhong, L. (2009) siRNA inhibition of telomerase enhances the anti-cancer effect of doxorubicin in breast cancer cells. *BMC Cancer*, **9**, 133.

360. Marian, C.O., Wright, W.E. and Shay, J.W. (2010) The effects of telomerase inhibition on prostate tumor-initiating cells. *Int J Cancer*, **127**, 321-331.

361. Joseph, I., Tressler, R., Bassett, E., Harley, C., Buseman, C.M., Pattamatta, P., Wright, W.E., Shay, J.W. and Go, N.F. (2010) The telomerase inhibitor imetelstat depletes cancer stem cells in breast and pancreatic cancer cell lines. *Cancer Res*, **70**, 9494-9504.

362. Chiappori, A.A., Kolevska, T., Spigel, D.R., Hager, S., Rarick, M., Gadgeel, S., Blais, N., Von Pawel, J., Hart, L., Reck, M. *et al.* (2015) A randomized phase II study of the telomerase inhibitor imetelstat as maintenance therapy for advanced non-small-cell lung cancer. *Ann Oncol*, **26**, 354-362.

363. Gunes, C., Lichtsteiner, S., Vasserot, A.P. and Englert, C. (2000) Expression of the hTERT gene is regulated at the level of transcriptional initiation and repressed by Mad1. *Cancer Res*, **60**, 2116-2121.

364. Ducrest, A.L., Amacker, M., Mathieu, Y.D., Cuthbert, A.P., Trott, D.A., Newbold, R.F., Nabholz, M. and Lingner, J. (2001) Regulation of human telomerase activity: repression by

normal chromosome 3 abolishes nuclear telomerase reverse transcriptase transcripts but does not affect c-Myc activity. *Cancer Res*, **61**, 7594-7602.

365. Zhang, A., Zheng, C., Lindvall, C., Hou, M., Ekedahl, J., Lewensohn, R., Yan, Z., Yang, X., Henriksson, M., Blennow, E. *et al.* (2000) Frequent amplification of the telomerase reverse transcriptase gene in human tumors. *Cancer Res*, **60**, 6230-6235.

366. Yuan, X., Larsson, C. and Xu, D. (2019) Mechanisms underlying the activation of TERT transcription and telomerase activity in human cancer: old actors and new players. *Oncogene*, **38**, 6172-6183.

367. Wang, Y., Broderick, P., Matakidou, A., Eisen, T. and Houlston, R.S. (2010) Role of 5p15.33 (TERT-CLPTM1L), 6p21.33 and 15q25.1 (CHRNA5-CHRNA3) variation and lung cancer risk in never-smokers. *Carcinogenesis*, **31**, 234-238.

368. Cheng, K.A., Kurtis, B., Babayeva, S., Zhuge, J., Tantchou, I., Cai, D., Lafaro, R.J., Fallon, J.T. and Zhong, M. (2015) Heterogeneity of TERT promoter mutations status in squamous cell carcinomas of different anatomical sites. *Ann Diagn Pathol*, **19**, 146-148.

369. Gunes, C., Wezel, F., Southgate, J. and Bolenz, C. (2018) Implications of TERT promoter mutations and telomerase activity in urothelial carcinogenesis. *Nat Rev Urol*, **15**, 386-393.

370. Yu, Z., Gaerig, V., Cui, Y., Kang, H., Gokhale, V., Zhao, Y., Hurley, L.H. and Mao, H. (2012) Tertiary DNA structure in the single-stranded hTERT promoter fragment unfolds and refolds by parallel pathways via cooperative or sequential events. *J Am Chem Soc*, **134**, 5157-5164.

371. Hurwitz, J., Furth, J.J., Anders, M. and Evans, A. (1962) The role of deoxyribonucleic acid in ribonucleic acid synthesis. II. The influence of deoxyribonucleic acid on the reaction. *J Biol Chem*, **237**, 3752-3759.

372. Spitzer, R. and Jain, A.N. (2012) Surflex-Dock: Docking benchmarks and real-world application. *J Comput Aided Mol Des*, **26**, 687-699.

373. Sterling, T. and Irwin, J.J. (2015) ZINC 15--Ligand Discovery for Everyone. *J Chem Inf Model*, **55**, 2324-2337.

374. Sastry, M., Lowrie, J.F., Dixon, S.L. and Sherman, W. (2010) Large-scale systematic analysis of 2D fingerprint methods and parameters to improve virtual screening enrichments. *J Chem Inf Model*, **50**, 771-784.

375. Ragazzon, P. and Chaires, J.B. (2007) Use of competition dialysis in the discovery of G-quadruplex selective ligands. *Methods*, **43**, 313-323.

376. Repasky, M.P., Shelley, M. and Friesner, R.A. (2007) Flexible ligand docking with Glide. *Curr Protoc Bioinformatics*, **Chapter 8**, Unit 8 12.

377. Halgren, T.A. (2009) Identifying and characterizing binding sites and assessing druggability. *J Chem Inf Model*, **49**, 377-389.

378. Wang, J., Wang, W., Kollman, P.A. and Case, D.A. (2006) Automatic atom type and bond type perception in molecular mechanical calculations. *J Mol Graph Model*, **25**, 247-260.

379. Wang, J., Wolf, R.M., Caldwell, J.W., Kollman, P.A. and Case, D.A. (2004) Development and testing of a general amber force field. *J Comput Chem*, **25**, 1157-1174.

380. Jakalian, A., Jack, D.B. and Bayly, C.I. (2002) Fast, efficient generation of high-quality atomic charges. AM1-BCC model: II. Parameterization and validation. *J Comput Chem*, **23**, 1623-1641.

381. Miller, B.R., 3rd, McGee, T.D., Jr., Swails, J.M., Homeyer, N., Gohlke, H. and Roitberg, A.E. (2012) MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *J Chem Theory Comput*, **8**, 3314-3321.

382. Morris, J.H., Huang, C.C., Babbitt, P.C. and Ferrin, T.E. (2007) structureViz: linking Cytoscape and UCSF Chimera. *Bioinformatics*, **23**, 2345-2347.

383. Doncheva, N.T., Klein, K., Domingues, F.S. and Albrecht, M. (2011) Analyzing and visualizing residue networks of protein structures. *Trends Biochem Sci*, **36**, 179-182.

384. Shannon, P., Markiel, A., Ozier, O., Baliga, N.S., Wang, J.T., Ramage, D., Amin, N., Schwikowski, B. and Ideker, T. (2003) Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res*, **13**, 2498-2504.

385. Matulis, D., Kranz, J.K., Salemme, F.R. and Todd, M.J. (2005) Thermodynamic stability of carbonic anhydrase: measurements of binding affinity and stoichiometry using ThermoFluor. *Biochemistry*, **44**, 5258-5266.

386. Monchaud, D., Allain, C. and Teulade-Fichou, M.P. (2006) Development of a fluorescent intercalator displacement assay (G4-FID) for establishing quadruplex-DNA affinity and selectivity of putative ligands. *Bioorg Med Chem Lett*, **16**, 4842-4845.

387. Ragazzon, P.A., Garbett, N.C. and Chaires, J.B. (2007) Competition dialysis: a method for the study of structural selective nucleic acid binding. *Methods*, **42**, 173-182.

388. Fleming, A.M., Zhu, J., Ding, Y. and Burrows, C.J. (2019) Location dependence of the transcriptional response of a potential G-quadruplex in gene promoters under oxidative stress. *Nucleic Acids Res*, **47**, 5049-5060.

389. McLuckie, K.I., Waller, Z.A., Sanders, D.A., Alves, D., Rodriguez, R., Dash, J., McKenzie, G.J., Venkitaraman, A.R. and Balasubramanian, S. (2011) G-quadruplex-binding benzo[a]phenoxazines down-regulate c-KIT expression in human gastric carcinoma cells. *J Am Chem Soc*, **133**, 2658-2663.

390. Rubis, B., Holysz, H., Gladych, M., Toton, E., Paszel, A., Lisiak, N., Kaczmarek, M., Hofmann, J. and Rybczynska, M. (2013) Telomerase downregulation induces proapoptotic genes expression and initializes breast cancer cells apoptosis followed by DNA fragmentation in a cell type dependent manner. *Mol Biol Rep*, **40**, 4995-5004.

391. Shi, Y.A., Zhao, Q., Zhang, L.H., Du, W., Wang, X.Y., He, X., Wu, S. and Li, Y.L. (2014) Knockdown of hTERT by siRNA inhibits cervical cancer cell growth in vitro and in vivo. *Int J Oncol*, **45**, 1216-1224.

392. Shammas, M.A., Koley, H., Batchu, R.B., Bertheau, R.C., Protopopov, A., Munshi, N.C. and Goyal, R.K. (2005) Telomerase inhibition by siRNA causes senescence and apoptosis in Barrett's adenocarcinoma cells: mechanism and therapeutic potential. *Mol Cancer*, **4**, 24.

393. Polishchuk, P.G., Madzhidov, T.I. and Varnek, A. (2013) Estimation of the size of drug-like chemical space based on GDB-17 data. *J Comput Aided Mol Des*, **27**, 675-679.

394. Yang, D. and Okamoto, K. (2010) Structural insights into G-quadruplexes: towards new anticancer drugs. *Future Med Chem*, **2**, 619-646.

395. Kirkpatrick, K.L., Clark, G., Ghilchick, M., Newbold, R.F. and Mokbel, K. (2003) hTERT mRNA expression correlates with telomerase activity in human breast cancer. *Eur J Surg Oncol*, **29**, 321-326.

396. Powers, K.T., Gildenberg, M.S. and Washington, M.T. (2019) Modeling Conformationally Flexible Proteins With X-ray Scattering and Molecular Simulations. *Comput Struct Biotechnol J*, **17**, 570-578.

397. Chen, Y. and Pollack, L. (2016) SAXS studies of RNA: structures, dynamics, and interactions with partners. *Wiley Interdiscip Rev RNA*, **7**, 512-526.

CURRICULUM VITA

Robert C. Monsen

Biochemistry & Molecular Genetics

804 E. Madison Street

Louisville, KY 40204

EDUCATION

2010-2014 **B.S., Biochemistry, University of Southern Indiana, Evansville, IN**
Undergraduate research involved the characterization of actin filamentation in the slime mold *stemonitis flavogenita* under the supervision of Jeannie Collins, Ph.D.

2016-2018 **M.S., Biochemistry & Molecular Genetics, Louisville School of Medicine, Louisville, KY**
Research focus: characterizing and targeting higher-order DNA G-quadruplex systems with selective stabilizing ligands. (See next section for skills acquired)
Thesis Advisor: John Trent, Ph.D.

2018-2020 **Ph.D., Biochemistry & Molecular Genetics, Louisville School of Medicine, Louisville, KY**
Thesis: Structural characterization and selective drug targeting of higher-order DNA G-quadruplex systems
Research techniques include:
Molecular characterization techniques: molecular dynamics simulations (AMBER), hydro- and thermo-dynamic calculations of DNA, protein, and receptor-ligand complexes, analytical ultracentrifugation (AUC) (Beckman Coulter ProteomeLab XL-A), size exclusion chromatography (SEC) (Waters system), SEC-resolved small-angle X-ray scattering of DNA G-quadruplexes (Argonne BioCAT beamline), circular dichroism (CD) spectroscopy, ^{1}H-NMR (Bruker) with some experience in analyzing 2D NOESY, TOCSY, and COSY spectra of DNA.
Purifications techniques: Size-exclusion chromatography (Waters), anion-exchange (AKTA), and nickel column His-tag purification (Bio-Rad Profinia system).
Drug discovery techniques: *in silico* screening (Surflex-Dock and Glide), *in vitro* screening: fluorescence thermal shift assays (96-well, Applied biosystems PCR), AUC drug binding analysis, competition dialysis (96-well), fluorescence intercalator displacement (96-well), *in cellula* screening: dual luciferase assays using G-quadruplex promoter expression system (Promega, Perkin Elmer 96-well luminometer). Drug binding validation: Isothermal titration calorimetry (ITC), CD and UV-Vis melting analysis.
Other relevant techniques: SDS-PAGE, western blotting, molecular cloning and site-directed mutagenesis, luciferase assays, human cell culture, real-time quantitative PCR (RT-PCR) (Applied Biosystems StepOnePlus system), and cell proliferation assays (MTT/Alamar Blue).
Thesis Advisor: John Trent, Ph.D.

ACADEMIC APPOINTMENTS

2017-2018 **Teaching Assistant**
Assistant in biochemistry lab course for 1st year students.
Assistant in biochemistry lecture (metabolism) for 1st year students
Course Director: Brian Clem, Ph.D.

2018-2019 **Graduate Student Council**
Represented the Biochemistry and Molecular Genetics department as a member of the Graduate Student Council.

PAST EMPLOYMENT

2013, 2014 **Internship, Nylene plastics, Henderson, KY**
Operated small scale (~10 gallon) Parr reactors to produce novel copolymer nylons for specialized customer applications.
Primary research techniques: polymerization reactions, tensile strength testing, impact testing, notched Izod and Charpy testing, differential scanning calorimetry (DSC), and polymer viscosity measurements.

2015-2016 **R&D Technician, SABIC Innovative Plastics, Mt Vernon, IN**
Worked with Sr. Scientists in thin film and coatings to troubleshoot and/or develop new products based on the needs of customers.
Primary research techniques: UV-Vis spectroscopy, densitometry, radical polymerization reactions for coating applications, polymer viscosity measurements.

PUBLICATIONS

1. **Monsen, R.C.** and Trent, J.O. (2018) G-quadruplex virtual drug screening: A review. Biochimie, 152, 134-148.
2. Dean, W.L., Gray, R.D., DeLeeuw, L., **Monsen, R.C.** and Chaires, J.B. (2019) Putting a New Spin of G-Quadruplex Structure and Binding by Analytical Ultracentrifugation. Methods Mol Biol, 2035, 87-103.
3. Chaires, J.B., Gray, R.D., Dean, W.L., **Monsen, R.**, DeLeeuw, L.W., Stribinskis, V. and Trent, J.O. (2020) Human POT1 unfolds G-quadruplexes by conformational selection. Nucleic Acids Res, 48, 4976-4991.
4. **Monsen, R.C.**, DeLeeuw, L., Dean, W.L., Gray, R.D., Sabo, T.M., Chakravarthy, S., Chaires, J.B. and Trent, J.O. (2020) The hTERT core promoter forms three parallel G-quadruplexes. Nucleic Acids Res, 48, 5720-5734.
5. **Monsen, R.C.**, Chakravarthy, S., Dean, W.L., Chaires, J.B., Trent, J.O. (2020) The solution structures of higher-order human telomere G-quadruplex multimers. BioRxiv, 2020.2011.2013.382036.

AWARDS/SCHOLARSHIPS

2016 **Fellowship, School of Interdisciplinary and Graduate Studies (SIGS), University of Louisville School of Medicine**
Ph.D. fellowship in the Biochemistry and Molecular Genetics program ($28,000/year stipend).

2017 **Beckman Coulter AUC Abstract Scholarship Winner**
Abstract contest winner for innovative research using analytical ultracentrifugation which included an all-expense paid trip to Glasgow,

Scotland to present research at the 2017 AUC conference (Estimated worth $4-5,000)

2017 **Graduate Student Council (GSC) Research Grant**
University of Louisville research grant award for equipment/reagents used during dissertation research ($500).

2018 **Graduate Student Council (GSC) Research Travel Grant**
University of Louisville grant award for research conference travel and poster/oral presentations ($350).

2019 **Fellowship, Arno Spatola Endowment Graduate Research Fellowship, Institute for Molecular Diversity & Drug Design, University of Louisville**
Research Fellowship supporting drug discovery, development, and collaborations ($15,000).

2020 **Argonne National Laboratory APS Beam Time Allocation**
X-ray beam time allotted for proposal to analyze various DNA G-quadruplex promoter and telomere systems for the Fall of 2020 (Estimated worth $20,000).

2020 **John M. Houchens Prize**
University of Louisville award to the doctoral student whose dissertation has potential for significant impact on a field.

HONORS

1. **Science Fair Judge** – Hoosier Science and Engineering Fair, Evansville IN (2015)
2. **Mentor** – Undergraduate summer rotation student – Poster "Small Molecule Inhibitors of hTERT" presented at R!L (Research! Louisville), Louisville KY (2018)
3. **Science Fair Judge** – Meyzeek Middle School Science Fair, Louisville KY (2019)
4. **Science Fair Judge** – Louisville Regional Science and Engineering Fair, Louisville KY (2019)
5. **Science Fair Judge** – DuPont Manual Regional Science Fair, Louisville KY (2019)
6. **Mentor** – High school summer rotation student – Poster "Automation of DNA-ligand MD Simulations with Free Energy Calculations for Enrichment of High Affinity Ligands in Virtual Screening" presented at R!L, Louisville KY – Won 2nd place among HS students (2019)

MEMBERSHIPS

G4 Society – Global community of nucleic acids researchers with the common goal of providing a framework in which the nucleic acids disciplines can collaborate and integrate ideas with a primary focus on G-quadruplex DNA (Since 2020)

ABSTRACTS AND PRESENTATIONS

7. Undergraduate Research Conference – Evansville, In (Fall, 2013)
   Poster Presentation: **Monsen, R. C.,** Collins, J. Capillary Electrophoretic Analysis of Actin Filaments of the Slime Mold *Stemonitis Flavogenita.*

8. AUC 2017 Conference – Glasgow, Scotland (Aug, 2017)

Poster Presentation: **Robert C. Monsen**, Lynn Deleeuw, William L. Dean, Jonathan B. Chaires, John O. Trent. Elucidation of the hTERT Core Promoter G-Quadruplex as a Target for Telomerase Inhibition.
Oral Presentation: **Robert C. Monsen**, Lynn Deleeuw, William L. Dean, Jonathan B. Chaires, John O. Trent. Elucidation of the hTERT Core Promoter G-Quadruplex as a Target for Telomerase Inhibition.

9. 1st Annual Commonwealth Computational Summit – Lexington, KY (Oct, 2017)
   Poster Presentation: **Robert C. Monsen**¸ Lynn Deleeuw, Jon Maguire, William L. Dean, Jonathan B. Chaires, John O. Trent. Structure-based Drug Discovery: Computational Virtual Screening.

10. Graduate Student Regional Research Conference – Louisville, KY (March, 2018)
    Poster Presentation: **Robert C. Monsen**¸ Lynn Deleeuw, Jon Maguire, William L. Dean, Jonathan B. Chaires, John O. Trent. The hTERT Core Promoter Sequence Forms Three Parallel G-quadruplexes.

11. 32nd Gibbs Biothermodynamics Conference – Carbondale, IL (Oct, 2018)
    Poster Presentation: **Robert C. Monsen**¸ Lynn Deleeuw, Jon Maguire, William L. Dean, Jonathan B. Chaires, John O. Trent. Structure-Based Design of Selective hTERT Promoter G-Quadruplex Ligands.

12. Research! Louisville – Louisville, KY (Oct, 2018)
    Poster Presentation: **Robert C. Monsen**¸ Lynn Deleeuw, Jon Maguire, William L. Dean, Jonathan B. Chaires, John O. Trent. Structure-Based Design of Selective hTERT Promoter G-Quadruplex Ligands.

13. 33rd Gibbs Biothermodynamics Conference – Carbondale, IL (Oct, 2019)
    Poster Presentation: **Robert C. Monsen**, Lynn Deleeuw, William L. Dean, Jonathan B. Chaires, John O. Trent. Biophysical Characterization of a Self-Organizing G-quadruplex in the hTERT Core Promoter.

14. 34th Gibbs Biothermodynamics Conference – Virtual (Zoom) (Oct, 2020)