Structural Dynamics and Electrostatic Properties of the *VEGF* and *PIM-1* Oncogenic Promoter G-Quadruplexes from Polarizable Molecular Dynamics Simulations

Rebekah J. Fogarty

Thesis submitted to the faculty of the Virginia Polytechnic Institute and State University
in partial fulfillment of the requirements for the degree of

Master of Science in Life Sciences
In
Biochemistry

Justin A. Lemkul, Chair
Anne M. Brown
Clement Vinauger

December 4th, 2024
Blacksburg, VA

Keywords: G-Quadruplex, Molecular Dynamics, Computer-Aided Drug Design

Structural Dynamics and Electrostatic Properties of the *VEGF* and *PIM-1* Oncogenic Promoter G-Quadruplexes from Polarizable Molecular Dynamics Simulations

Rebekah J. Fogarty

ACADEMIC ABSTRACT

G-Quadruplexes (GQs) are higher ordered nucleic acid structures that form within regions of DNA and RNA that are enriched with guanine nucleobases. These structures are highly stable and have been shown to function in genomic maintenance and regulating key biological processes. Due to their role in regulating gene expression, GQs also contribute to a wide variety of human diseases including neurodegenerative conditions, premature aging disorders, and various cancers. Therefore, these structures have gained growing interest as the subjects of various research investigations to explore potential methods for targeting and disease management on transcriptional and translational levels. However, targeting efforts have been relatively unsuccessful due to the conserved GQ core structure, leading to compounds that cannot bind to their targets with sufficient specificity. Here, we employed conventional and enhanced sampling molecular dynamics simulations on two oncogenic GQ structures with the Drude polarizable force field to gain crucial insights into structural and electrostatic properties contributing to overall GQ stability and potential small-molecule binding sites. In addition to these simulations, we also subjected these structures to the Site Identification by Ligand Competitive Saturation workflow to determine the favorability of various functional groups and gain insights into preferential binding of these GQ structures.

Structural Dynamics and Electrostatic Properties of the *VEGF* and *PIM-1* Oncogenic Promoter G-Quadruplexes from Polarizable Molecular Dynamics Simulations

Rebekah J. Fogarty

GENERAL AUDIENCE ABSTRACT

Genetic mutations can contribute to diseases in individuals, including many cancers. Some genes are more likely to result in abnormal activity in cancerous cells; these genes are known as oncogenes. These genes provide a unique targeting opportunity for cancer therapeutic approaches by targeting cancers on the genomic level. G-Quadruplexes (GQs) form in DNA and RNA, using complex stacking and hydrogen bonding interactions to stabilize these structures. GQ-forming sequences have been identified in many oncogenes, in regulatory regions that dictate the expression of the genes. An emerging treatment strategy is to stabilize these GQs, thus reducing oncogene expression. Due to the similar structures adopted by the majority of these GQs, targeting them effectively has been a challenge in the field. Therefore, it is necessary to characterize the structural properties of GQs that govern their behaviors, to better understand how targeting approaches can be improved in future drug design efforts. In this study, we utilized computational techniques such as molecular dynamics (MD) simulations to gain insight into these properties of two oncogenic GQ structures. Additionally, we conducted simulations with common drug fragments to help further improve and inform drug design for these GQ structures.

Acknowledgements

Many people have helped me throughout my academic career. I want to take this space to thank those who have supported me throughout this journey and encouraged me in the times that I needed it, as these past months have challenged me beyond measure.

To my mentors. Dr. Lemkul, I cannot thank you enough for taking a chance on the undergraduate student who showed up to a computational lab with goggles and a lab coat on the first day of my summer research program. You have supported me in and out of the laboratory and have shown me the importance of a mentor who appreciates and respects you as a person above all else. Navigating this difficult season of life would have been nearly impossible without your support. To my committee members, past and present, Dr. Brown, Dr. Vinauger, and Dr. Slade, thank you for encouraging me to grow into a better scientist.

To my friends. To Lemkul Lab, what a wild ride we have been on together. Thank you for understanding the struggles and helping me grow, personally and professionally, all these years. Thank you for showing me how important it is to surround myself with those who make work fun. Haley, thank you for all you have done for me over the last three years. Thank you for helping me navigate graduate school and pushing me to be a better Bekah in all aspects. I will be forever grateful for all the laughs (and tears) we have shared. Carter, thank you for being my biggest cheerleader. You are one of the few who understand what these past few months have been like, and you never left my corner. Thank you for making sure I was prioritizing myself, while also encouraging me to work hard.

To my family. Mom and dad, thank you and I love you. Thank you for always pushing me to be the best version of myself and showing me the importance of working hard throughout all my pursuits. I will never be able to express how grateful I am to have had you by my side and to cheer me on when I needed it all those years. I wish more than anything that both of you could stand beside me while this journey comes to a close, but I hope I have made you proud. Hannah and Logan, thank you for being the support I have needed in this final stretch of pursuing this degree. I am not sure if I would’ve pulled this off without you and the tough love that I needed at times.

Table of Contents

List of Figures

Chapter 1. Biological Roles of G-Quadruplexes and Potential for Drug Targeting

1.1 Overview of G-Quadruplex Biology

G-quadruplexes (GQs) are higher ordered nucleic acid structures that occur within guanine-rich nucleic acid sequences.[1] These structures are primarily found in genomic sequences with regulatory function including, but not limited to, promoter regions, origins of replication, telomeres, and untranslated mRNA regions.[2,3] Various bioinformatics studies have proposed that hundreds of thousands of putative GQ-forming sequences can be found within the human genome, as well as the genomes of various organisms across many domains of life.[2,4–6] Due to their conserved presence within the described regions, it is believed that GQs likely contribute to genetic regulation on both transcriptional and translational levels. Recent studies have shown that GQs are often implicated in disease states, including neurodegeneration, hematological disorders, and cancers.[7,8] Much is still unknown regarding the full biological roles of these GQs, however due to their implications in various diseases, they have become increasingly attractive therapeutic targets in recent years. Therefore, the characterization of their structural dynamics and electrostatic properties is necessary to inform these future drug targeting efforts.

1.2 History and Background

Research involving the characterization of DNA structures dates back to the 1950s with the foundational structural model of helical B-form DNA, through the work of Franklin, Watson, and Crick.[9] Following this discovery, additional investigations into secondary structures of nucleic acids were launched, typically including biophysical techniques such as circular dichroism spectroscopy. In 1962, Gellert et al. observed that guanylic acid could form hydrogen-bonded helical structures.[8] These arrangements were found to enhance stability through the hydrogen bond donor and acceptor interactions involving each guanine base.[1,10] Further, it was also hypothesized that the ability for these bases to form stacked-layers of four hydrogen-bonded guanine bases (known as G-tetrads) was driven by van der Waals forces.[10] Building on these observations of guanine and its associated nucleotides, the first GQ structure was resolved via NMR spectroscopy in 1975, confirming various hypotheses regarding these structural characteristics.[11] This finding inspired other investigations into GQ structures, resulting in the resolution of several GQ structures through NMR and X-ray crystallography techniques throughout the 1990s and into the early 2000s.[12] The advancements into computational techniques and analysis sparked interest in further investigation and characterization of GQs beyond the tetrad core configurations, leading to additional studies regarding potential biological roles and implications of these higher ordered structures including bioinformatic searches[2,4,5] and molecular dynamics investigations.[13]

1.3 Structural Characteristics

1.3.1 Primary Structure

Despite occurring within conserved guanine-containing nucleic acid sequences, GQs adopt many distinct topologies. GQs are composed of distinct primary, secondary,

and tertiary structures, all of which contribute to GQ dynamics, electronic properties, and their ability to interact with other biomolecules. The primary structure of GQs refers to the sequence that corresponds to the nucleobases that are present within that region of DNA or RNA. The consensus sequence of putative GQ forming motifs is $G_3XG_3YG_3ZG_3$, with X, Y, and Z corresponding to additional nucleotides that comprise loop region, etc. within the structure (Figure 1.1A).[14] GQ structures can form outside of this consensus sequence with some variation in sequence length and composition.[15,16] The consecutive guanine nucleotides are known as G-tracts and are separated by short sequences of other nucleotides. These additional nucleotide stretches make up loops or bulges in the GQ fold and help to determine the overall topological folding pattern and dynamic properties of the GQs. Additionally, these regions help to enhance specificity for binding targets to these GQs, an ongoing challenge due to the conserved nature of the G-tract sequences in GQs.

1.3.2 Secondary Structure

The most significant structural characteristic of GQs is the secondary structure which comprises the tetrad formation of the GQ core. The G-tracts in the nucleic acid sequences interact to adopt planar quartets, called tetrads, which are able to stack and give rise to the conserved tetrad core of folded GQ structures (Figure 1.1C). These tetrads consist of four guanine bases that interact through Hoogsteen hydrogen-bonding networks (Figure 1.1B). The orientation of these Hoogsteen hydrogen bonds results in an electronegative core due to the orientation of carbonyl oxygen (O6) atoms toward the GQ core. This electronegative core is then occupied by coordinated cations, most commonly K^+, that provide electrostatic interactions to stabilize the GQ structure, an attribute which will be discussed below. Typically, GQs contain at least two stacked tetrads with one cation coordinated between each pair of the tetrads, however it is it possible for GQs to form with stacked triads, or non-guanine base containing tetrads.[17,18]

1.3.3 GQ Tetrad Core Cations

Previous research has demonstrated that the electronegative tetrad core must be stabilized by the coordination of cations to mitigate the repulsion in the core.[19] These cations are typically monovalent, with K^+ being the dominant cation found in GQ structures. This stabilization has been found to be vital to proper GQ folding, as the presence of cations within the GQ core provides more of a stabilizing force than hydrogen bonding throughout the GQ.[20] Additionally, it has been demonstrated that the cellular salt concentrations and ions present strongly influence GQ folding. Notably, GQs preferentially fold in the presence of K^+ ions relative to Na^+ and Li^+, respectively,[19,21] and understanding the GQ-ion interactions have been a focus in the field. Experimental and computational studies have shown that due to ion size and electrostatic interactions, K^+ ions are believed to be the optimal occupant for the GQ core as K^+ ions are found to adopt a bipyramidal antiprismatic configuration contrasting with the planar coordination of other cation types.[21–23] Additionally, K^+ coordination is believed to be more advantageous due to the higher cellular concentration (~150 mM) relative to other ions, e.g., Na^+ (~10 mM).[24,25]

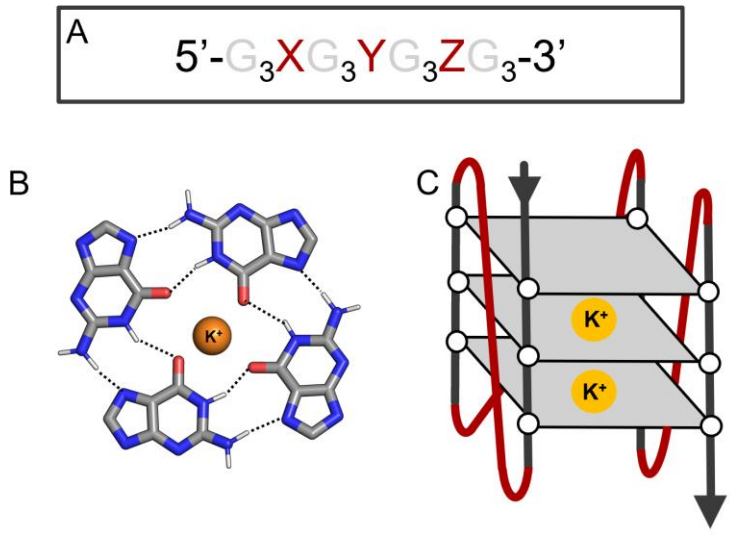

Figure 1.1 Overview of GQ structural hierarchy. (A) Putative GQ-forming consensus primary structure with G-tracts shown in grey and loop/linker regions (X, Y, Z) shown in red. (B) Secondary structure arises due to Hoogsteen hydrogen bonding to form G-tetrads with a core K^+ shown in orange. (C) Schematic of a tertiary folded GQ structure, with guanine bases shown as circles, tetrads as grey squares, and loop and linker regions shown in red. Core K^+ present between tetrads are shown in orange.

1.4 Biological Roles of GQs and Involvement in Cancer

GQs form *in vivo* during biological processes that result in transiently single-stranded DNA, such as during replication, transcription, and repair, and in RNA that is typically single-stranded to regulate translation. Therefore, GQs have been identified in nucleotide sequences responsible for modulating gene expression including promoter regions, telomeres, origins of replication, and untranslated regions in mRNA. The formation of GQs can have positive or negative effects on transcription.[26] In the context of promoters, GQs that form may inhibit the binding of necessary proteins, such as transcription factors, thereby impeding transcription of the gene controlled by that promoter.[27] In contrast, some promoter GQs serve as recognizable binding partners for proteins identified to increase expression, such as the SP1 transcription factor.[28] Therefore, by altering the stability of GQs *in vivo* with small molecules, the effects of the GQs may be exploited as a method of regulating gene expression to achieve a desired biological outcome. Additionally, when GQs are present downstream of the transcription start site, a GQ present on the coding strand will contribute to the re-initiation of transcription, whereas GQs present on the template strand in this region will hinder the progression of RNA polymerase, resulting in stalled transcription.[26,29] Similar positive and negative effects of GQs can be found in other locations as well corresponding with transcription factor recruitment, etc.

Recent bioinformatics efforts have concluded that putative GQ-forming sequences are six times as likely to be present within a gene promoter sequence than elsewhere in the human genome.[2] This percentage increases when investigating the prevalence of putative GQ sequences within oncogenes. When these oncogenes are expressed, various pro-

tumorigenic downstream effects may occur that contribute to tumor proliferation, survival, and overall growth and metastasis. However, oftentimes when GQs are present within the promoter regions of proto-oncogenes, these structures have been found to stall transcription, thereby hindering gene expression. GQs have been identified in the promoters of several major proto-oncogenes including, but not limited to, *BCL-2*,[30] *VEG-F*,[31] *PIM-1*,[32] *MYC*,[33] *KRAS*,[34] *c-KIT*,[35] and *WNT-1*.[36] Therefore, targeting oncogenes on a genomic level by designing GQ-binding ligands has emerged as a new chemotherapeutic strategy.

1.4.1 G-Quadruplexes as Drug Targets

Due to improved understanding of the regulatory roles of GQs and their proximity to genes often implicated in various human diseases, the interest in determining structures of GQs and GQ:ligand complexes has increased in recent years. Given the variable impact of GQ formation on gene expression, the development of small-molecule drugs may aid in modulating biological processes on the nucleic-acid level. As previously mentioned, GQ structures can result in altered gene expression by means of increased or decreased transcription and translation. Therefore, it is important to determine the appropriate goal of ligand design, including whether to stabilize or destabilize the folded GQ structure or to block the binding of proteins to these GQs. While GQs have served as the subjects for recent drug design efforts, including multiple GQ-targeting ligands progressing to clinical trials, most identified ligands lack specificity towards the GQ target. For example, highly aromatic compounds like Quarfloxin[37] and APTO-253[38] (Figure 1.2) both demonstrated promise during pre-clinical investigations by inducing a pro-apoptotic effect. However, both drugs failed in clinical trials due to bioavailability concerns and insufficient effectiveness, respectively. As anticipated, the highly conserved G-tetrad core of GQs presents a specificity challenge when carrying out drug design as off-target binding of candidate ligands can occur through binding to G-tetrads of other GQs. As such, it is important to exploit distinct features in GQ structures including loop regions, quadruplex-duplex junction sites, bases stacked on tetrads, and other defining features.

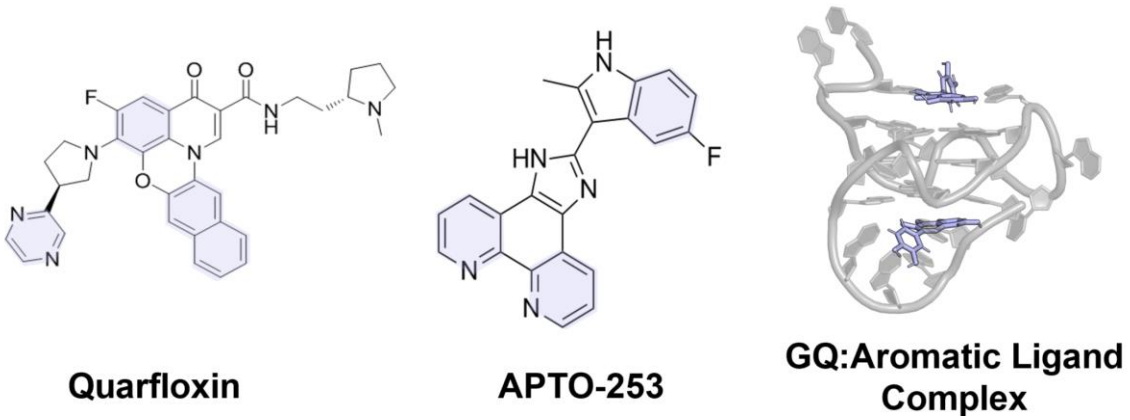

Figure 1.2 Recent GQ ligand clinical trial candidates, Quarfloxin and APTO-253. Aromatic rings are depicted in light blue. Non-specific GQ ligands typically bind to the solvent-accessible faces of the outermost tetrads in the core, as shown with a *c-MYC* GQ ligand complex (PDB: 2N6C).[39]

1.5 Molecular Dynamics Simulations and Advancements in GQ Research

Computational tools have been increasingly useful in investigating biological processes, particularly due to recent improvements in hardware performance, software optimization, and adaptation to graphical processing unit (GPU) platforms, and in the empirical energy functions describing the systems. Molecular dynamics (MD) simulations have served as a vital research technique in gaining a deeper understanding of atomistic contributions to biomolecular dynamics and their influence on biological functions. Specifically, MD simulations have proven to be especially useful in exploring GQ folding and identifying stabilizing forces that may be further exploited when designing small-molecule drugs to target GQs.[40–42] Despite providing details necessary for advancing the knowledge of GQ structures, these previous simulations have typically utilized non-polarizable, or additive, force fields (FFs). These FFs lack the ability to robustly model electrostatic interactions, as electronic polarization responses are approximated due to averaged point charges on atoms within the system. Due to the highly electronegative core present within GQ core, modeling GQ-ion interactions is crucial for gaining biologically relevant insights from these simulations. In many additive simulations conducted with GQs, ion coordination within the core is incorrectly modeled when compared to experimental data. For example, despite experimental work demonstrating the coordination of cations between multiple tetrads within the GQ core, many additive simulations are unable to maintain this characteristic due to ion-ion repulsion forces. Recent studies have demonstrated the need for more robust polarization changes to adequately model the distinct electronic properties and electrostatic interactions that are present in complex biological systems.[41,43]

Given the importance of induced polarization effects in stabilizing GQ-ion interactions,[41] use of additive FFs has only limited accuracy and GQs represent biomolecules in which electronic polarization is an important consideration. Recent progress in FF development has led to the emergence of non-additive force fields for use with nucleic acids, including the Drude FF based on the classical Drude oscillator model[44–46] and the AMOEBA multipole-induced dipole model.[47] The Drude FF has demonstrated success with simulating GQs in previous studies, as it preserves key interactions, such as core cation stability,[48] by allowing for induced polarization through the addition of negatively charged "Drude oscillator" particles to all heavy atoms within the system. The Drude oscillators are connected to their parent atoms via harmonic springs, thus their dynamics are treated classically. Throughout the simulation, the displacement of these negatively charged can be calculated, and therefore polarization responses to intrinsic electric fields can be calculated.[49] Additionally, the Drude polarizable model accounts for explicit dipole-dipole interactions between first and second neighbors through a Thole screening factor,[50] as well as lone pairs that yield better quadrupole moments and contribute to accounting for anisotropic polarizability.[51] A full description of the Drude parameterization process is can be found elsewhere.[49] However, it is important to note that the preliminary success of the Drude FF with nucleic acids including RNA and DNA GQs,[22,52,53] duplex DNA,[45] and the increased accuracy of DNA-ion interactions,[48] justifies the use of this FF for further GQ studies.

1.6 Computer-Aided Drug Design

Computational approaches have been developed to help streamline the drug discovery process. The traditional wet-lab approach to drug discovery has proven to be time-intensive and prohibitively expensive as most new drugs cost approximately $1 billion to develop.[54] However, with recent improvements in computational capabilities, computer modeling can serve as a key complement to experimental techniques in the evaluation of drug candidates.[55–57] Computational approaches have been specifically useful in identifying potential drug interactions with protein targets.[58] With growing interest in targeting gene expression at transcriptional and translational levels, novel tools are being developed to utilize nucleic acids as the primary drug targets. Notably, several recent studies targeting RNA have proven successful and have contributed to the to increased attention towards these efforts.[59,60] However, nucleic acids are challenging targets, particularly RNA, in part due to the intrinsic structural dynamics, as well as there being considerably resolved structures compared to proteins. Recent advancements have been made to expand protein-targeting methods to account for these characteristics of nucleic acids. Most commonly, Molecular docking is frequently employed in the drug discovery process, using programs such as AutoDock Vina.[61] While these methods are useful and provide insight into the nature of interactions with the target, rigid docking methods are not the most beneficial when exploring the targetability of flexible structures, including nucleic acids. Recent progress has been made towards computational methods that can be applied to dynamic systems to identify binding regions and structural properties for consideration during the drug discovery pipeline.

1.6.1 Site Identification by Ligand Competitive Saturation (SILCS)

One of the techniques that accounts for the dynamics in a system is the Site Identification by Ligand Competitive Saturation (SILCS) method.[62] This method has been validated for use with proteins, membrane systems, and has recently been extended for use with nucleic acids.[63–65] Notably, in recent years, SILCS has also been adapted for protein systems to support explicit polarization with the Drude polarizable FF.[66] In developing that method, it was found that the inclusion of explicit polarization led to differences in probe molecule occupancy maps, suggesting that a polarizable model may give more detailed information for drug design. As described above, the incorporation of polarization is especially valuable in the computational investigations of nucleic acids, including those of GQs. Therefore, as SILCS accounts for conformational dynamics and polarizability within simulations, it is a relevant drug-design technique for use with GQ systems.

SILCS works by introducing common small molecule drug functional groups as probe molecules surrounding the biomolecular target of interest (Figure 1.3). MD simulations are performed to allow these probe molecules to sample the target surface, and periodic shuffling of the molecules by a Grand Canonical Monte Carlo method[67] enhances sampling. Following the simulations, the areas of affinity for these specific probe molecules can then be mapped onto the structure and utilized for HotSpot generation[68] and pharmacophore modeling,[69] both of which are useful tools for subsequent steps in the computer-aided drug design pipeline.

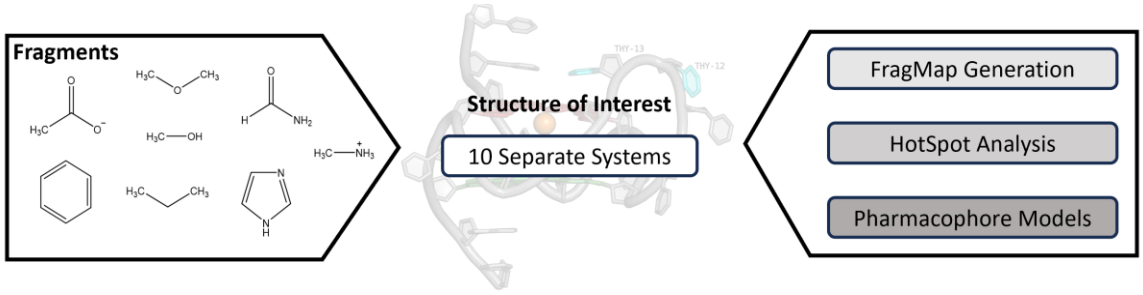

Figure 1.3 Overview of SILCS workflow.

1.7 Organization of Thesis

This introductory chapter serves to provide a foundational basis on GQ research, primarily focusing on the computer-aided drug design and targeting of GQs. In the following chapters, we will explore multiple applications of computational techniques that serve as preliminary steps within the drug design pipeline for oncogenic promoter GQs. In Chapter 2, we discuss the MD simulations performed on the *VEGF* promoter GQ, including both wild-type and experimentally mutated sequences. Through analysis of both conventional MD simulations, and the virtual fragment-based drug design simulations, we then discuss the implications and effectiveness of experimental mutants throughout the drug design process. In Chapter 3, we investigate the promoter GQ present in the *PIM-1* promoter. The applications of conventional MD and enhanced sampling methods allow us to explore a wider conformational landscape for the distinct GQs that form in this region. These GQs contain distinct topological characteristics that we analyze to gain further insight into their role in targeting through the SILCS analysis, allowing us to identify features of importance for preferential GQ binding. In Chapter 4, we share our overarching conclusions and recommendations for future investigations of oncogenic GQs.

1.8 Attribution

The work included within this thesis is my own with specific contributions from my advisor, Justin A. Lemkul (Ph.D., Virginia Tech Department of Biochemistry) and colleagues highlighted with the authors' initials at the beginning of each chapter, as follows:

Rebekah J. Fogarty (R. J. F.)
Justin A. Lemkul (J. A. L.)
Haley M. Michel (H. M. M.)

Chapter 2. Human Vascular Endothelial Growth Factor Oncogene G-Quadruplex Dynamics and Electrostatics

Attribution: R.J.F. and J.A.L. designed the research, R.J.F. performed the simulations and analyzed the data, R.J.F. and J.A.L. wrote the chapter.

2.1 Introduction

The human vascular endothelial growth factor (VEGF) is a primary mediator of angiogenesis, the process of generating new blood vessels.[70] Angiogenesis is vital in human embryonic development and wound healing.[71] However, within tumor cells, abnormally high levels of VEGF expression is responsible for the over-induction of angiogenesis, resulting in increased growth and proliferation and overall poor prognosis.[72–75] Elevated VEGF levels lead to the increased ability of tumor cells to surpass the "angiogenic switch," a necessary precursor for tumor growth and metastasis (Figure 2.1).[76,77] This angiogenic switch refers to the process by which tumors can rapidly proliferate following the release of activated VEGF. This increased proliferation occurs because of the higher concentration of VEGF and larger size of blood vessels within the tumor allowing for access to vital nutrients, without which tumors are unlikely able to grow beyond 2 millimeters in diameter due to the hypoxic environment.

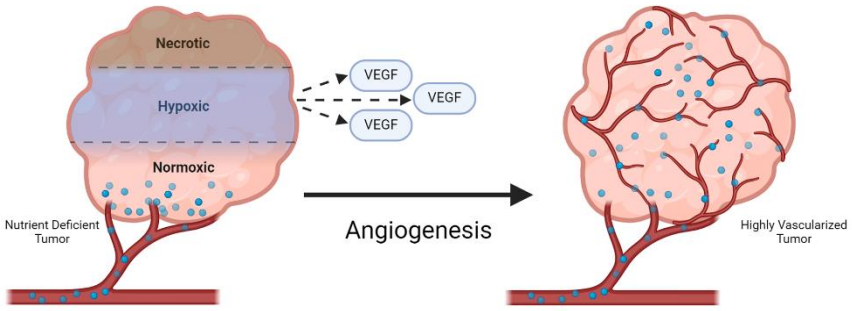

Figure 2.1 Mechanism of human VEGF triggering the angiogenic switch and resulting in tumor growth through nutrient absorption.

The promoter sequence of human *VEGF* gene is enriched in guanine and contains a 22-nucleotide region that forms a G-quadruplex (GQ). A mutated *VEGF* promoter GQ structure was resolved by Agrawal et al. using nuclear magnetic resonance (NMR) spectroscopy.[31] This mutated *VEGF* GQ structure contains two guanine to thymine mutations at positions 12 and 13 residing in the loop region of the GQ. This mutant structure is denoted *VEGF*-T12T13. These mutations were necessary to resolve the NMR ensemble of the GQ and are not believed to cause any disruption to the native topology and characteristic behaviors of the wild-type (WT) *VEGF* GQ. In this study, we test this assumption by employing computational tools to generate two distinct starting structures with both guanine and thymine bases at these positions, known throughout this study as *VEGF*-WT and *VEGF*-T12T13, respectively (Figure 2.2). The results of this study provide crucial insights into the potential structural and dynamic consequences of experimentally mutated structures being utilized in computational and future drug-targeting studies. We investigated the possibility that artificial mutations changed structural and electrostatic

behaviors of these GQs via conventional MD simulations using the polarizable Drude-2017 nucleic acid FF.[44,45] Additionally, both GQ structures served as the initial structures for two separate applications of the Site Identification by Ligand Competitive Saturation (SILCS) workflow[62,65], and provided a proof-of-concept for the implementation of polarizable SILCS with the Drude-2017 FF.

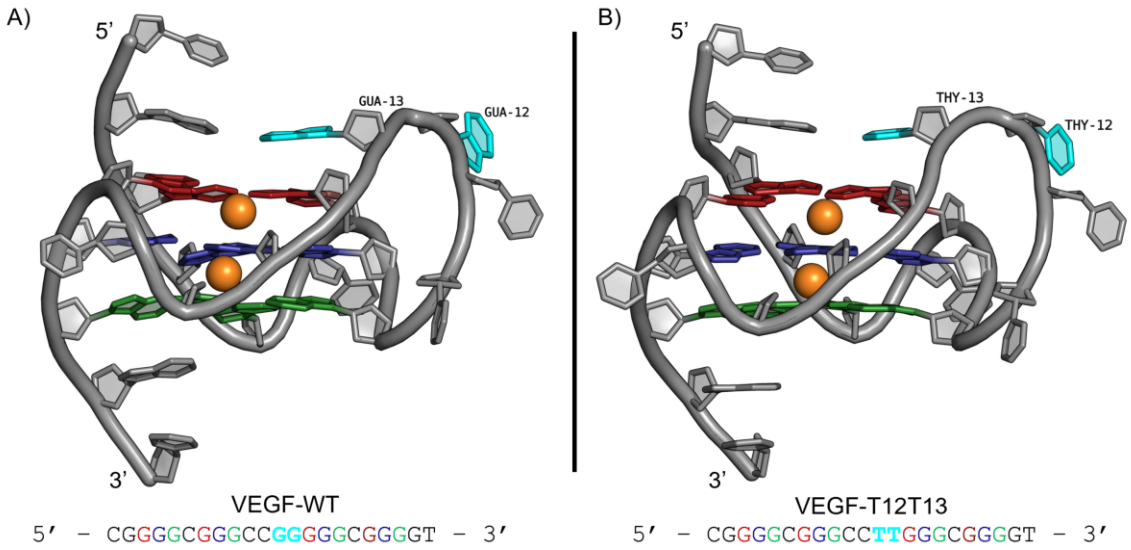

Figure 2.2 *VEGF* promoter GQ structures and sequences. Bases involved in forming the GQ core are colored by tetrad (1 – red, 2 – blue, 3 – green). Core K^+ ions shown as orange spheres. Sites of mutations are shown in cyan. (A) *VEGF*-WT GQ with WT bases built in CHARMM, (B) *VEGF*-T12T13 GQ from resolved NMR ensemble (PDB: 2M27).

2.2 Methods

System Construction

The starting coordinates for our MD simulations were taken from Protein Data Bank (PDB) entry 2M27.[31] This NMR ensemble corresponds to the mutated form of the GQ-forming sequence in the promoter region of the *VEGF* gene. The GQ structure is characterized by the three guanine-containing tetrads and the 4-nucleotide loop. Tetrad 1 is composed of guanines 3, 7, 14, and 18; tetrad 2 of guanines 4, 8, 15, and 19; and tetrad 3 of guanines 5, 9, 16, and 20 (Figure 2.2B). The 4-nucleotide loop consists of nucleotide positions 10-13. It is also known that the 22-nucleotide WT sequence adopts a GQ topology, however researchers relied on further mutation to formally characterize the folded structure via NMR spectroscopy.[31,78] Therefore, the resolved structure contained in PDB:2M27 contains mutated bases at positions 12 and 13, where the guanine nucleotides have been replaced with thymine. To generate a starting structure with the WT sequence, we deleted the existing thymine bases and rebuilt the coordinates of guanine nucleobases via the internal coordinate builder in the CHARMM program (Figure 2.2A).[79]

Previous studies have demonstrated the need for cations to exist within the tetrad core, specifically K^+ ions. Therefore, as the initial NMR structures lacked these essential ions, we further employed the CHARMM program to add two symmetric K^+ ions between the guanine-tetrads utilizing the coordinates of the inward-facing guanine carbonyl oxygen (O6) atoms. This cation addition within the electronegative core of the GQ allows for more

robust simulation of the GQ structural and electrostatic properties, as previously described.[48]

Equilibration and Unrestrained MD Simulations

All preparation steps for both the *VEGF*-WT and *VEGF*-T12T13 systems were conducted using the additive CHARMM36 nucleic acid FF.[80] Each GQ was placed in the center of a cubic box, maintaining a box-solute distance of 10 Å. The systems then underwent solvation through the addition of TIP3P water molecules and 150 mM KCl, including neutralizing counterions. This KCl concentration mimics an environment representative of eukaryotic cellular conditions that encourages proper GQ folding and characteristic behaviors.[25] The solvated *VEGF* GQ systems were then relaxed through energy minimization in CHARMM, consisting of 500 steps of steepest descent minimization and 500-steps of adopted-basis Newton-Raphson minimization.

The minimized systems were then subjected to an equilibration procedure in NAMD.[81] Throughout the duration of the 1-ns equilibration, position restraints were placed on all non-hydrogen atoms within the GQ, including the K^+ ions within the core. The surrounding water molecules and mobile KCl ions remained unrestrained during the equilibration, allowing for their motion within the simulation box. NPT conditions were maintained by applying a Langevin thermostat at 298 K with a force constant of 5 ps^{-1}. The system pressure remained 1 atm by employing a Langevin piston method with isotropic box scaling.[82] Periodic boundary conditions were applied throughout all system dimensions. Short-range van der Waals forces were smoothly switched to zero from 10 – 12 Å and neighbor lists were maintained within 14 Å. Electrostatic forces were calculated with the particle mesh Ewald (PME) method,[83] with a real-space cutoff of 12 Å and a Fourier grid spacing of ~1 Å. Four replicate simulations were performed for both systems, starting from different, random initial velocities during equilibration.

Prior to beginning production runs, these systems were made compatible with the Drude polarizable FF. By utilizing CHARMM, this preparation included the addition of lone pairs and Drude oscillator atoms to all non-hydrogens contained within the system. Additionally, the water molecules present within the simulation box were converted to the polarizable SWM4-NDP water model.[84] Drude-2017 parameters were then applied to all components within the simulation box, and the Drude oscillators underwent further minimization via CHARMM prior to production MD runs using steepest-descent and adopted-basis Newton-Raphson minimization. 1 ns of NPT equilibration was then employed at 1 atm pressure and 298 K, maintained by the Langevin thermostat and barostat methods. The same harmonic restraints described above were applied in this NPT equilibration process. The integration time step was limited to 1 fs due to the high-frequency vibrations of Drude-atom bonds. Additionally, polarization catastrophe was avoided through a "hard-wall" constraint of a 0.2 Å maximum Drude-atom bond length. The nonbonded scheme was the same except for the use of a potential switch for van der Waals forces rather than a force switch.

Each replicate system was then simulated in the absence of any restraints in OpenMM.[85] Each replicate was run for a simulation time of 1 μs, with similar NPT ensemble conditions as those described above, despite the application of a Monte Carlo barostat with box scaling every 25 integration steps.

SILCS Simulations

Here, the established SILCS-RNA protocol[65] was extended to be compatible with DNA GQ structures and Virginia Tech Advanced Research Computing architecture. As in the unrestrained MD simulations, both the *VEGF*-WT and *VEGF*-T12T13 GQ systems were generated using a minimized structure file containing only the GQ and the core cations as the starting structures. Twenty independent SILCS systems were then generated for both *VEGF*-WT and *VEGF*-T12T13; ten with the neutral SILCS solutes (benzene, propane, methanol, formamide, and imidazole) and ten with the charged SILCS solutes (acetate and methylammonium). The GQs were centered in a simulation box 15 Å from the edge, and the solutes were placed randomly around the GQ at an initial concentration of ~0.25 M. Water was also present within the box at ~55 M. Following the initial setup, Grand Canonical Monte Carlo (GCMC) refilling was employed to first remove all solutes and water from around the structure, and then re-insert them, allowing for larger fragments to fill in buried pockets on the GQ surface prior to supplementing the box with the smaller solute fragments and water. Once desired concentrations were reached for all solutes, steepest descent minimization of 5000-steps was followed by 100 ps of equilibration prior to beginning production runs for all systems. Production runs consist of 100 cycles of GCMC-MD with GCMC sampling, minimization, equilibration, and 1 ns of production MD for each system. These systems employed weak restraints of 50.208 kJ/mol/nm^2 on C1' and N3 (pyrimidines) or N1 (purines) atoms to prevent large conformational changes from occurring, as well as restraints of 67873.8 kJ/mol/nm^2 on the K^+ ions to maintain their occupancy of the GQ core, as this force constant overcomes the repulsion between the core K^+ ions and allows for distributions of ion positions similar to unrestrained simulations with the Drude-2017 FF that require no such restraints. All simulations were conducted with the CHARMM36 FF[80] and performed using GROMACS 2022.5[86] and the SilcsBio package, version 2023.1.

Following the completion of production runs, solute probability density maps indicating the interacting locations for each of the solutes for each of the 10 systems were generated. These maps are known as FragMaps. The calculated FragMaps were then combined in a two-step process to obtain a single map file that represents the regions of interaction for each solute across all SILCS systems. The process combines the maps generated for runs 1-5, as well as those generated for runs 6-10 for each solute. By comparing the resulting maps, overlap coefficients were calculated to assess for convergence between the runs, an overlap coefficient ≥ 0.75 indicates well-converged solute activity across the runs. The maps were then all combined to obtain a single map file representative of regions of affinity for each solute. These maps were then visualized in PyMOL using a specified Grid Free Energy (GFE) threshold (see below), which serves as an estimate of ligand binding affinity.

Polarizable SILCS Simulations with Drude FF

Following the completion of the additive SILCS production runs with the CHARMM36 FF,[80] the newly developed polarizable SILCS workflow was employed on the systems by using OpenMM 7.7.0[85] with the Drude-2017 FF.[46] Frames selected at 10-ns intervals from each of the 10 equilibrated, nonpolarizable systems were used to serve as the starting points for the polarizable systems, resulting in 100 systems. The neutral and charged solutes remained separate with 100 systems for each set, resulting in a combined

of 200 systems for both *VEGF*-WT and *VEGF*-T12T13. Then, using the CHARMM program, each of the starting systems was prepared for the polarizable production simulations by the addition of lone pairs to hydrogen bond acceptors and Drude oscillator atoms to all heavy atoms in the system. A brief minimization process was also then performed, followed by an equilibration of 1 ns prior to the initiation of the production runs. Each production run consisted of 20 ns of production MD, resulting in a total simulation time of 2 μs for all systems. The same weak harmonic restraints on the nucleic acid described above were applied in the polarizable production runs. Notably, no restraints were required on the K^+ ions present in the GQ core due to the improved modeling of these interactions by the Drude FF. This approach provides sufficient sampling time in relation to the additive SILCS systems given the 1-fs timestep and slower kinetics associated with the Drude FF. Following the completion of the production runs, FragMaps were then generated the same manner described above.

2.3 Results and Discussion

2.3.1 Conventional Molecular Dynamics Simulations

We first set out to determine whether common experimental mutations would influence the drug discovery pipeline by conducting simulations on the wild-type and mutated forms of the *VEGF* GQ. Further, we sought to compare the results of nonpolarizable and polarizable simulations in generating affinity maps for different probe molecules around the *VEGF* GQ. Due to the innate characteristics associated with various nucleotides, we hypothesized that we would observe distinct structural and electrostatic properties when comparing the two *VEGF* GQ systems, and that these differences would have implications within drug design processes. As discussed above, the experimentally resolved structure of the *VEGF* GQ contains two G→T mutations at nucleotides 12 and 13. As thymine takes up a smaller surface area and has an intrinsically different dipole moment compared to guanine, we hypothesized that structural and electrostatic variation would be observed between the two *VEGF* GQs.

Structural Implications of Mutations

We observed no major conformational changes in either the *VEGF*-WT or the *VEGF*-T12T13 GQs. Despite the hypothesis that the mutated bases may impact the overall stability and structural behaviors of the GQs, our findings did not support this hypothesis. To quantify the overall structural integrity of the GQ systems, we analyzed several aspects of the GQ structure and their flexibility to arrive at these conclusions. Due to minimal conformational changes within the system, the root-mean-squared deviation (RMSD) of both the *VEGF*-WT and *VEGF*-T12T13 GQ systems remained below 3.5 Å throughout the entirety of the simulation (Supporting Information, Figure S2.1). In these systems, the loop regions and non-tetrad bases exhibited the largest RMSD values, as the GQ core exhibited an RMSD of approximately 1-1.5 Å across all replicates for both the *VEGF* GQs, this is anticipated due to the stabilizing effects of the GQ core interactions (Hoogsteen hydrogen bonding and tetrad formation). This finding indicates that the main contribution to the RMSD arises from these regions, despite having minimal consequence on the overall conformation of the structure.

While RMSD confirmed that no large conformational changes occurred in these *VEGF* GQ systems, we also aimed to characterize any changes in structural flexibility that may be a result of the mutations. To do so, we calculated the root-mean-squared fluctuation (RMSF) of each nucleotide. As anticipated, we found that nucleotides involved in the Hoogsteen hydrogen bonding had the lowest RMSF values (Figure 2.3). Additionally, all nucleotides in the loop region and bulges in the GQs had large RMSF values, reflecting their intrinsic flexibility as a result of being unpaired and exposed to solvent. Interestingly, the nucleotide located at position 13, either guanine (*VEGF*-WT) or thymine (*VEGF*-T12T13), had an RMSF value nearly as low as those engaged in the GQ tetrad core, despite flanking the tetrads and being more solvent-exposed. This finding indicates that it is likely that this nucleotide is engaged in some additional stabilizing interactions, which are discussed in the next section. Overall, these observations confirm that the G→T mutations do not perturb the native structure or dynamic behaviors of the *VEGF* GQ.

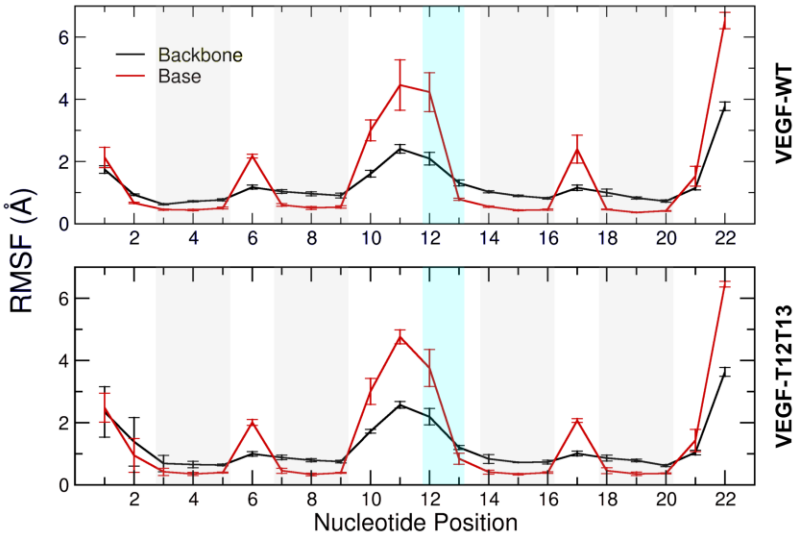

Figure 2.3 Average RMSF values per nucleotide across four replicate simulations for *VEGF* GQ systems. Error bars represent the standard deviation of the averages. Nucleotides in the GQ tetrad core are indicated by the grey shaded regions. Cyan shaded regions represent sites of mutation.

Tetrad Accessibility and Capping Base Pairs

When determining possible targets for drug targeting, accessibility of the structure is a key consideration and MD simulations can provide insights into the structural motifs of GQs that are most accessible to solvent and candidate drug molecules. Additionally, the solvent-accessible regions above and below the GQ core are common sites for drug targeting. As such, characterizing these sites in terms of dynamics, base pairing, and ion interactions is important for informing future drug design efforts for the wild-type GQ. Here, we discuss aspects of base-pairing. Ion interactions will be discussed in the next section.

As noted above, the Gua13 and Thy13 bases in the *VEGF*-WT and *VEGF*-T12T13 GQs exhibited low RMSF values. Experimental investigations of *VEGF*-T12T13 GQ determined that tetrad 1 is blocked by a base pair formed between Gua2 and Thy13, therefore sterically hindering that site.[31] Therefore, we analyzed whether these base pairs

were present throughout our MD simulations, finding that Gua2:Gua13 (in *VEGF*-WT) and Gua2:Thy13 (in *VEGF*-T12T13) formed base pairs. In both cases, Gua2 formed two hydrogen bonds with the nucleotide present at position 13, regardless of whether the base-pairing partner was thymine or guanine (Figure 2.4). We calculated the prevalence of these base pairs during the 4 μs of simulation time by calculating the distances between the H1 and O6 atoms of Gua2 and the H1 and O6 atoms or H3 and O4 atoms of the corresponding guanine and thymine bases present at position 13. Using an acceptor-hydrogen distance cutoff of 2.5 Å, we determined that these hydrogen bonds were present for over 98% of the simulation time for both Gua2:Gua13 and Gua2:Thy13. This finding demonstrates the strength of these interactions explains the rigidity of the bases as noted in the RMSF analysis (Figure 2.3). The persistence of the Gua2:Thy13 base pair validates that the simulations reproduce the experimentally observed properties of the *VEGF*-T12T13 GQ.

The presence of these base pairs reduces the solvent accessibility of tetrad 1, meaning any putative ligand would have to disrupt the Gua2:Gua13 or Gua2:Thy13 base pairs. Therefore, these interactions are key considerations in experimental efforts for developing drugs that target the *VEGF* GQ and whether or not the sequence used in the study is the mutant or wild-type.

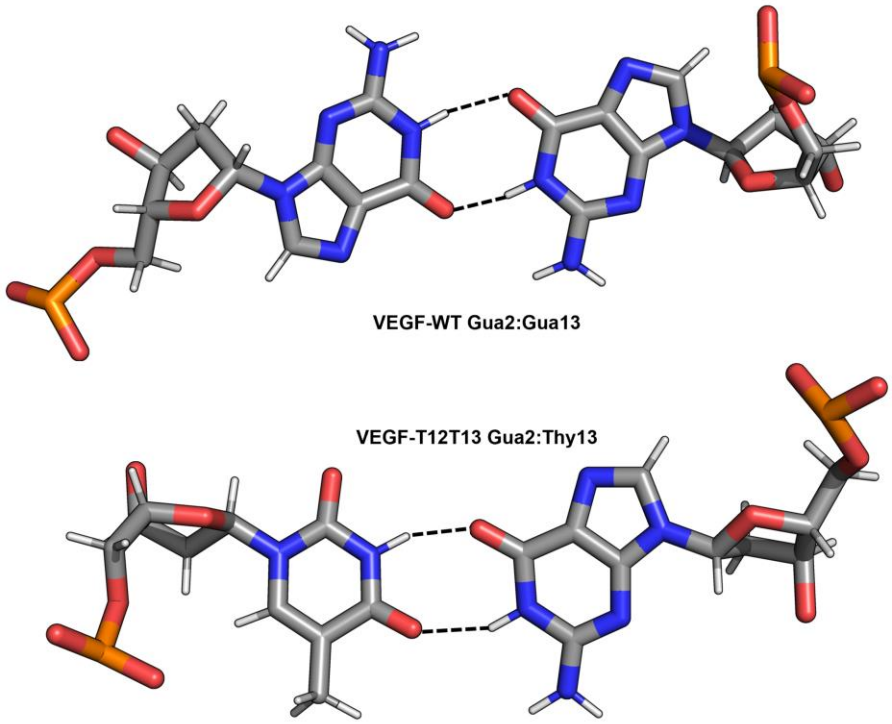

Figure 2.4 Noncanonical base pairs present within the *VEGF* GQs. The base at position 13 forms hydrogen bonds with Gua2, as indicated by the dashed lines.

Ion Dynamics and Interactions

To characterize potential interactions in which the nucleotides in our system participate in and to further contextualize the shifts in RMSD and RMSF values, we investigated the ion occupancy throughout our simulations. The K^+ ions from the bulk KCl in the solvent are known to interact with GQ structures, and by identifying regions that the ions commonly occupy, we sought to further inform drug design efforts with details

regarding ion binding and any influence these events had on the *VEGF* GQ structure.

We observed that both *VEGF* GQ systems exhibited regions of high ion affinity, specifically in the volumes above tetrad 1 (Figure 2.5), as these solvent-exposed regions are well-known ion binding sites in other GQ systems.[87] Additionally, both the *VEGF*-WT and *VEGF*-T12T13 GQs maintained their core K^+ ions within the tetrads. For our analysis, we will focus on auxiliary binding sites aside from these core cations. For both the *VEGF*-WT and *VEGF*-T12T13 GQs, ions interacted with the bases of Cyt1 and Gua2, respectively. For the *VEGF*-T12T13 GQ, K^+ ions also interacted with the mutant Thy13. This observation is reminiscent of a similar observation in the *c-kit1* GQ system, in which Thy12 above tetrad 1 frequently coordinated a K^+ ion.[48]

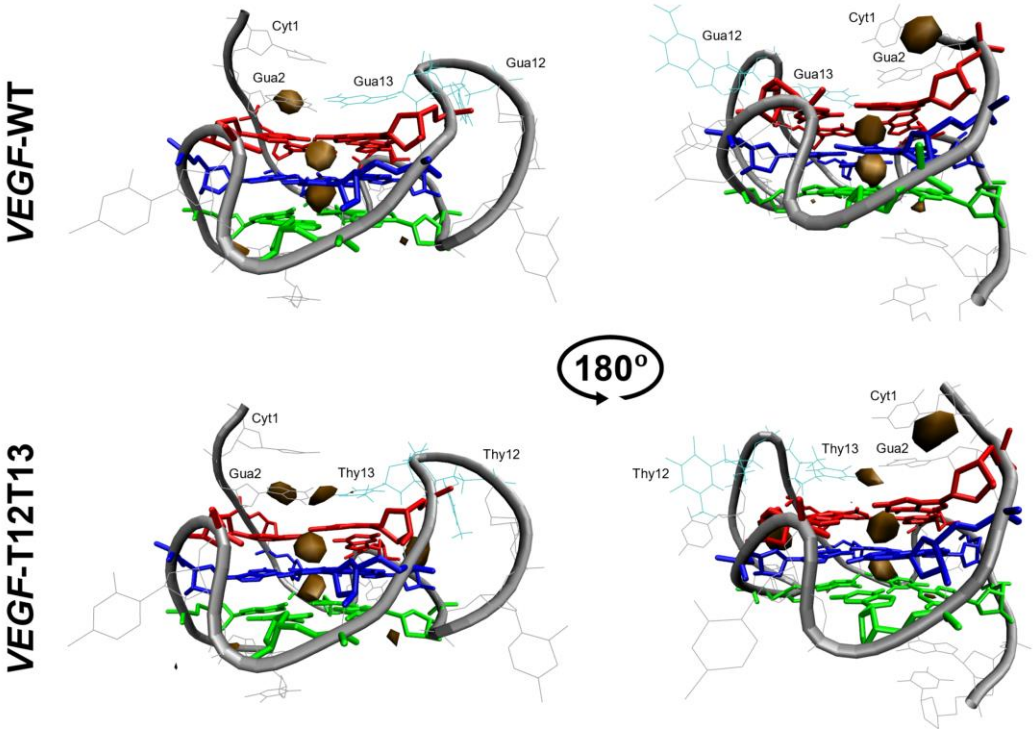

Figure 2.5 Ion occupancy maps for *VEGF* GQ systems. Brown regions indicate positions occupied by K^+ ions for at least 1% of total simulation time across all four replicates per system (4 μs total). Tetrads are colored (1 – red, 2 – blue, 3 – green) and the mutation locations are shown in cyan and labeled.

After observing this additional binding site for K^+ ions in the *VEGF*-T12T13 GQ, we sought to characterize this interaction. We determined that in addition to a steric explanation (thymine being smaller than guanine) for this difference, we can also attribute some of the ion coordination in this location to the electrostatic properties of the bases themselves. For instance, we observed that in the instances of ion coordination above tetrad 1 in the *VEGF*-T12T13 GQ, a K^+ ion was coordinated via direct interaction with the O2 atom of the Thy13 base (Figure 2.6). During each replicate simulation, the thymine base remained oriented such that the O2 atom was accessible to the aqueous solution. Notably, coordination of K^+ ions within 3.5 Å was observed across all simulation replicates, some for extended durations, such as replicate 4 for approximately 600 ns (Supporting Information, Figure S2.2) This accessible electronegative atom presents a favorable ligand

for the K^+ ions that is not present in the *VEGF*-WT GQ, explaining the lack of ion occupancy above tetrad 1 in the *VEGF*-WT GQ. Combined with the observation that Gua2 can form a persistent base pair with Thy13 in previous experimental work and simulations presented here, likely the properties of this site are a key consideration in drug design against the *VEGF* GQ. Subtle differences may exist in this regard as the position of the electronegative O2 atom in the *VEGF*-T12T13 GQ is replaced by the N2 amino group in Gua13 in the wild-type GQ (Figure 2.4), which will not favorably interact with a K^+ ion. Therefore, studies using the *VEGF*-T12T13 GQ may show behaviors that differ from the *VEGF*-WT GQ due to the additional K^+ occupancy near the Gua2:Thy13 noncanonical base pair, which itself is intrinsically different from the wild-type Gua2:Gua13 base pair.

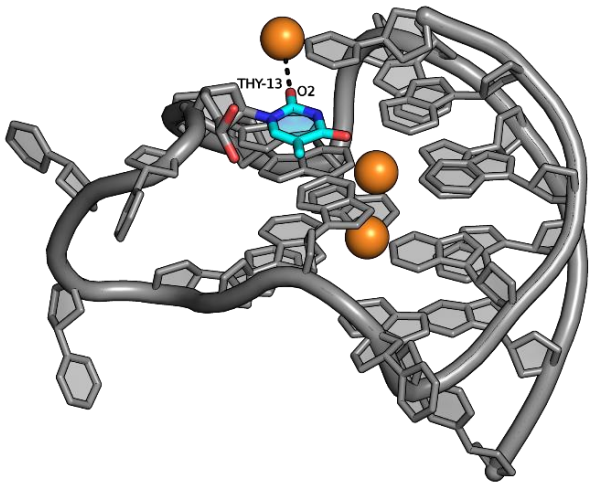

Figure 2.6 Ion interaction with Thy13 in the *VEGF*-T12T13 GQ. The O2 atom of the Thy13 base is oriented such that it can coordinate a K^+ ion.

2.3.2 Mixed Solvent Simulations with SILCS

Although our MD simulations of the *VEGF*-WT and *VEGF*-T12T13 GQs differed very little, only with respect to K^+ binding near Thy13, we hypothesized that the mutations may have some direct impacts on the types of molecules that would bind to the *VEGF* GQs. Here, we performed and analyzed some of the first GQ simulations applying the SILCS workflow with the Drude polarizable force field. In addition to determining the overall implications of the *VEGF* GQ mutations on drug design, we also sought to use the findings of this study to investigate the importance of modeling explicit polarization in computer-aided drug design efforts.

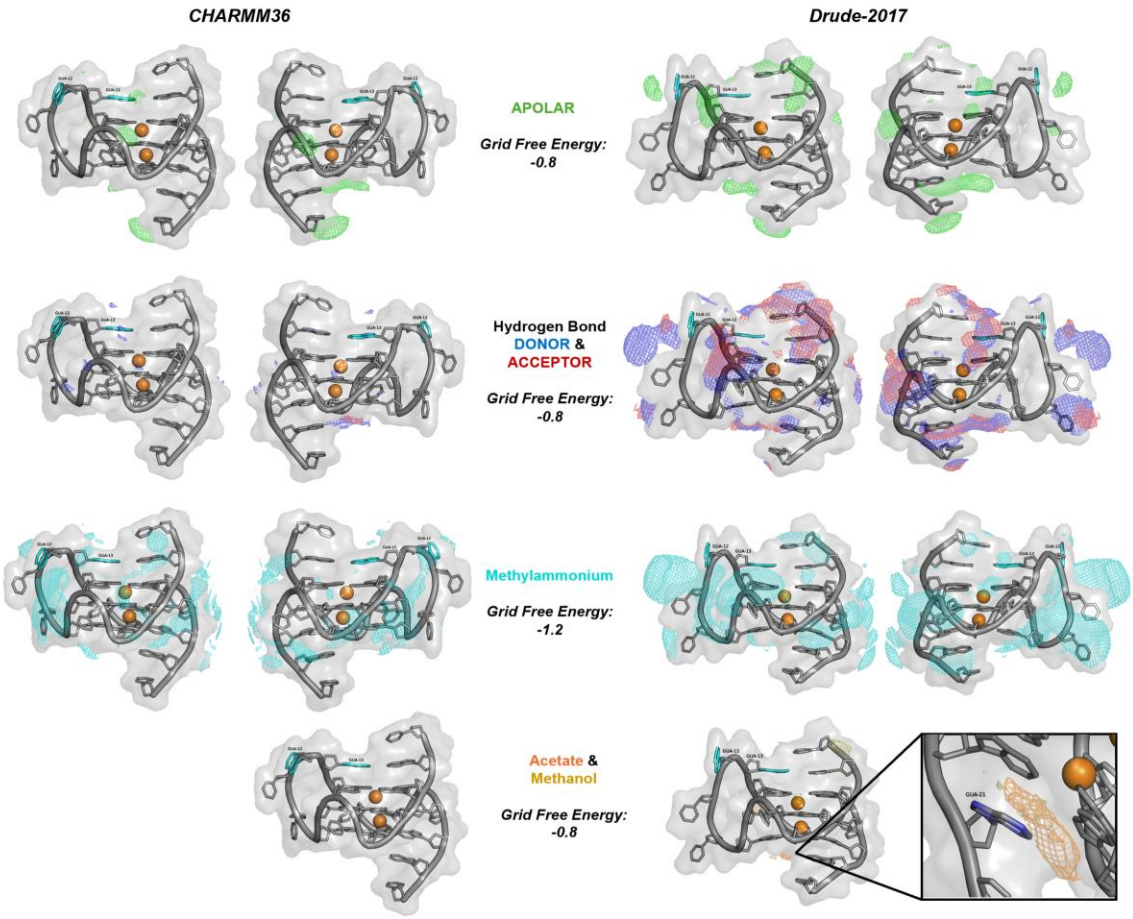

Figure 2.7 Comparison of additive and polarizable SILCS FragMaps for *VEGF*-WT GQ. The maps were generated from specific atoms of the SILCS solutes. Apolar maps represent benzene (C), propane (C), and imidazole (HC); hydrogen bond donor maps include formamide (N) and imidazole (NH); hydrogen bond acceptor maps include formamide (O), dimethylether (O), and imidazole (N). The GFE values used as a threshold to render the maps are provided, and the WT Gua12 and Gua13 bases are shown in cyan.

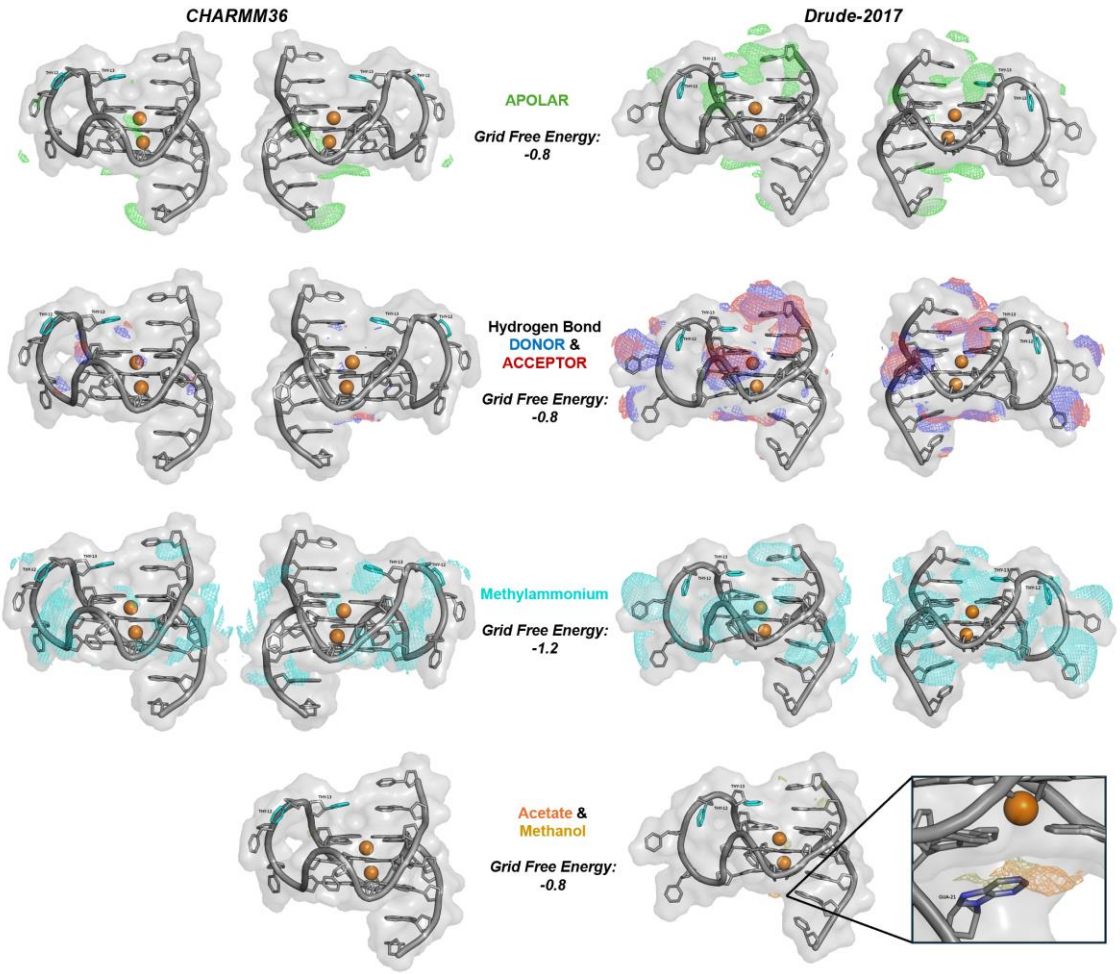

Figure 2.8 Comparison of additive and polarizable SILCS FragMaps for *VEGF*-T12T13 GQ. The maps were generated from specific atoms of the SILCS solutes. Apolar maps represent benzene (C), propane (C), and imidazole (HC); hydrogen bond donor maps include formamide (N) and imidazole (NH); hydrogen bond acceptor maps include formamide (O), dimethylether (O), and imidazole (N). The GFE values used as a threshold to render the maps are provided, and the mutated bases are shown in cyan.

In comparing the results of the additive and polarizable SILCS simulations, the Drude polarizable FF led to greater solute accumulation around the GQs. The FragMaps generated represent the overall pattern of affinity for general features of the specific SILCS solutes. That is, the apolar map is generated from the coordinates of benzene (C), propane (C), and imidazole (HC). The hydrogen bond donor map includes contributions from formamide (N) and imidazole (NH). The hydrogen bond acceptor map includes formamide (O), dimethylether (O), and imidazole (N). The solute occupancies are all converted into a Grid Free Energy (GFE) score, which serves as a measure of the affinity of the SILCS functional groups for the volumes of space that they illustrate. As is clear from Figures 2.7 and 2.8, when visualized at the same GFE threshold, the use of the Drude polarizable force field led to greater and more favorable solute sampling around the surfaces of both GQs.

We observed that favorable sampling of apolar and hydrogen bond donor and acceptor groups occurred across the entire surface of the *VEGF* GQs. Additionally, the additive SILCS systems lacked prominent sampling regions in distinct areas of the GQ, such as above tetrad 1 or surrounding the loop region. As previously discussed, any regions of enhanced specificity may lead to improved targeting for GQ-ligand design, therefore the lack of sampling in these regions in the additive SILCS systems suggests the need for polarizability in the drug design process in GQ systems. Overall, the Drude SILCS simulations exhibited sampling above and below the GQ tetrads, as well as around the loop bases. Additionally, as anticipated, positively charged methylammonium sampling was present at highly favorable GFE values across both additive and polarizable SILCS simulations, due to the intrinsically polyanionic nature of the DNA backbone. Interestingly, in both *VEGF* GQ systems, favorable sampling of the negatively charged acetate was observed below the GQ core and above the Gua21 base, but only with the Drude force field (Figures 2.7 and 2.8). The presence of anionic sampling in this region is likely a result of the K^+ coordinated between tetrads 2 and 3, combined with the greater solvent accessibility of this region relative to tetrad 1, which is occluded by base pairs, discussed above. Additionally, some experimental studies have suggested that acetate ions interact with guanine bases, therefore further characterization of these interactions may be necessary in the drug design process.[88] Overall, Drude SILCS systems exhibit improved favorability of solute sampling and provided more detailed insights into regions of affinity for the drug design process.

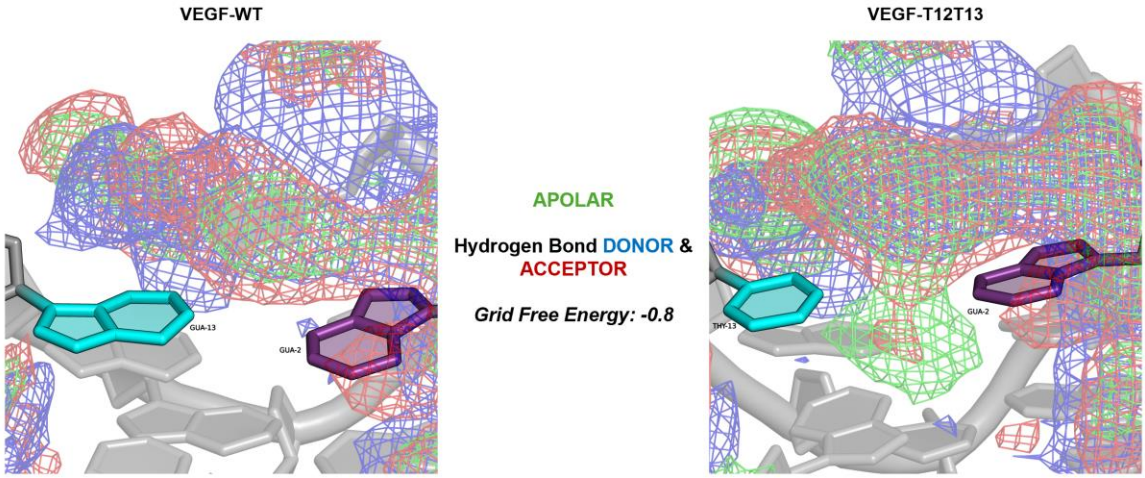

Figure 2.9 Polarizable SILCS FragMaps of the mutation site for *VEGF* GQ systems. Generated and rendered in the same manner as previous FragMaps. GFE values for generation provided.

Finally, we sought to explore differences in the SILCS FragMaps specifically near the mutation sites of the *VEGF*-T12T13 GQ relative to the *VEGF*-WT GQ. As the results of our conventional MD simulations demonstrated that Thy13 could coordinate a K^+ ion and that Gua2 could form base pairs with either Gua13 or Thy13, we hypothesized that the resulting FragMaps would differ. As expected, very little solute sampling was present beneath the Gau2:Gua13 and Gua2:Thy13 base pairs (Figure 2.9). Interestingly, in the region surrounding the mutated thymine bases, additional sampling of apolar and hydrogen bond acceptor functional groups was observed. This sampling was likely a result of the

disruption of the Gua2:Thy13 base pair, therefore allowing the SILCS solutes to interact directly with the Thy13 base. Additionally, as suggested from the results of our conventional MD analysis, the overall size of the nucleotide base present at position 13 (the smaller thymine vs. larger guanine) has important implications for the overall sampling surrounding this region.

2.4 Conclusions

Our investigation of the *VEGF* GQs revealed several properties that may be beneficial for future drug development efforts targeting this structure. *VEGF* often serves as a model GQ system and has been the subject of various chemotherapeutic investigations due to its role in tumor proliferation. Efforts to target the VEGF protein remain clinically ineffective and lead to many negative side effects, and it is hypothesized that by targeting its production at the transcriptional level, alternative treatment methods may be identified. Here, we explored the NMR-resolved *VEGF*-T12T13 and the computationally regenerated *VEGF*-WT GQs to explore any implications of the mutations on the structure, dynamics, and small-molecule binding properties. Overall, the structural stability and behaviors of the GQs remained largely unaltered due to the mutations. Additional work, such as a covariance analysis, should be carried out to investigate more subtle structural implications of the mutations. However, once we began exploring features more specific to drug design efforts, the impacts of the G→T mutations became more apparent. In our conventional MD and SILCS simulations, we observed that, due to the smaller nucleotide size and the orientation of the electronegative O2 atom, the *VEGF*-T12T13 GQ exhibited regions of ion affinity and fragment accessibility that were not present in the *VEGF*-WT GQ. The findings discussed here suggest that, despite the global similarities in the overall structural dynamics and behaviors of the *VEGF* GQs, some consequential differences exist in these systems. While the resolved *VEGF*-T12T13 GQ commonly serves as a model system for analyses on GQ and even drug design efforts, it is important to note the observable differences in properties of these two GQs. It is unlikely that, if utilizing the *VEGF*-T12T13 GQ for targeting investigations, specifically at the common targeting site above tetrad 1, one would generate a result that would be directly applicable to the wild-type sequence. Therefore, this study provides relevant details on the inherent differences in these two systems to further improve the drug design efforts against VEGF and inform proper selection of the system of interest in these future studies.

2.5 Acknowledgements

Computing time and resources were provided by Virginia Tech Advanced Research Computing and Expanse Cluster at the San Diego Supercomputer Center (SDSC) through allocation BIO230117 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program, which is supported by National Science Foundation grants #2138259, #2138286, #2138307, #2137603, and #2138296. This work was supported by the NIH (grant R35GM133754 to JAL) and USDA-NIFA (project VA-160092) and the Thomas F. and Kate Miller Jeffress Memorial Trust.

2.6 Supporting Information

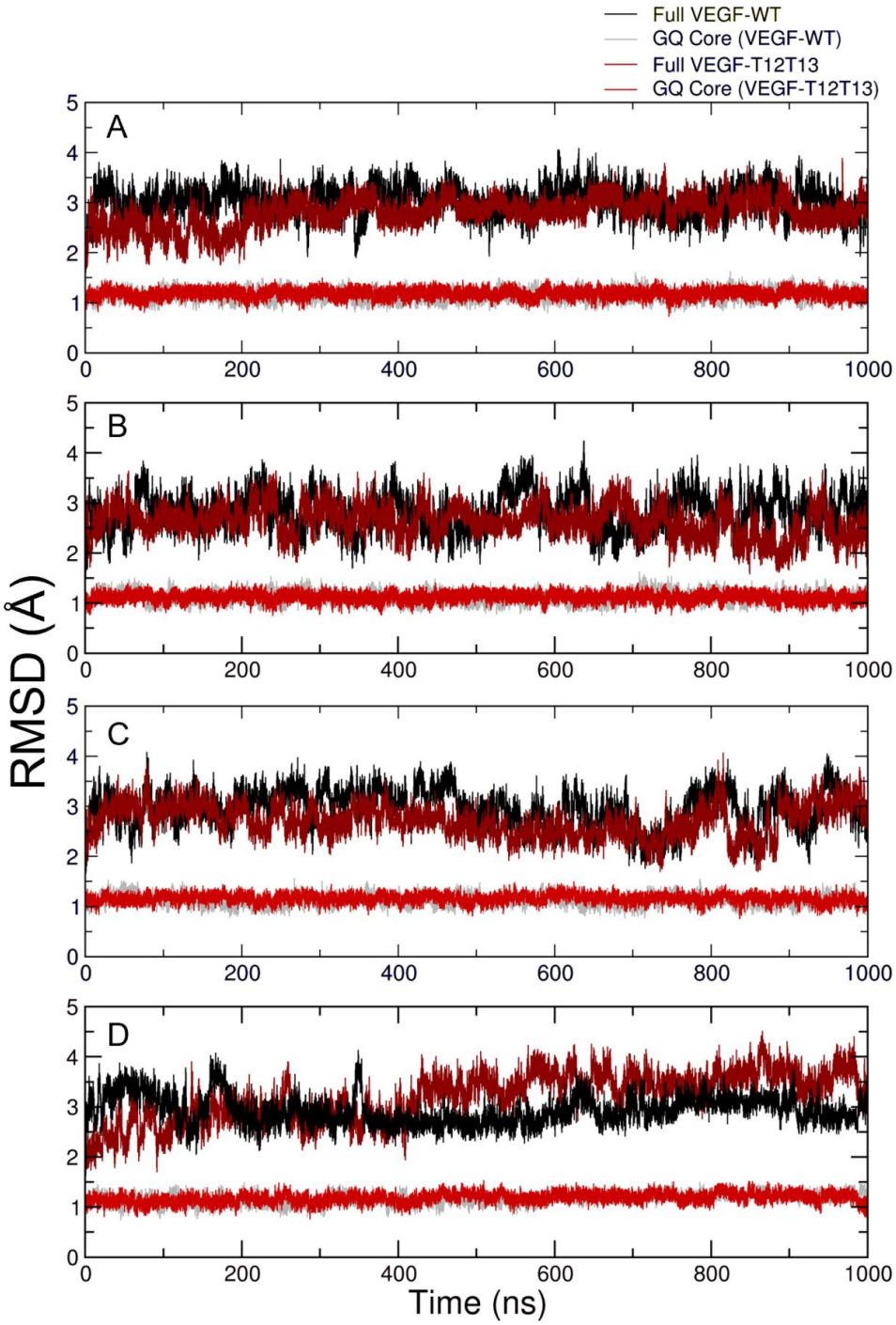

Figure S2.1 RMSD of whole structure and GQ core for the *VEGF*-T12T13 and *VEGF*-WT GQs for replicates 1 (A), 2 (B), 3 (C), and 4 (D) throughout simulation time.

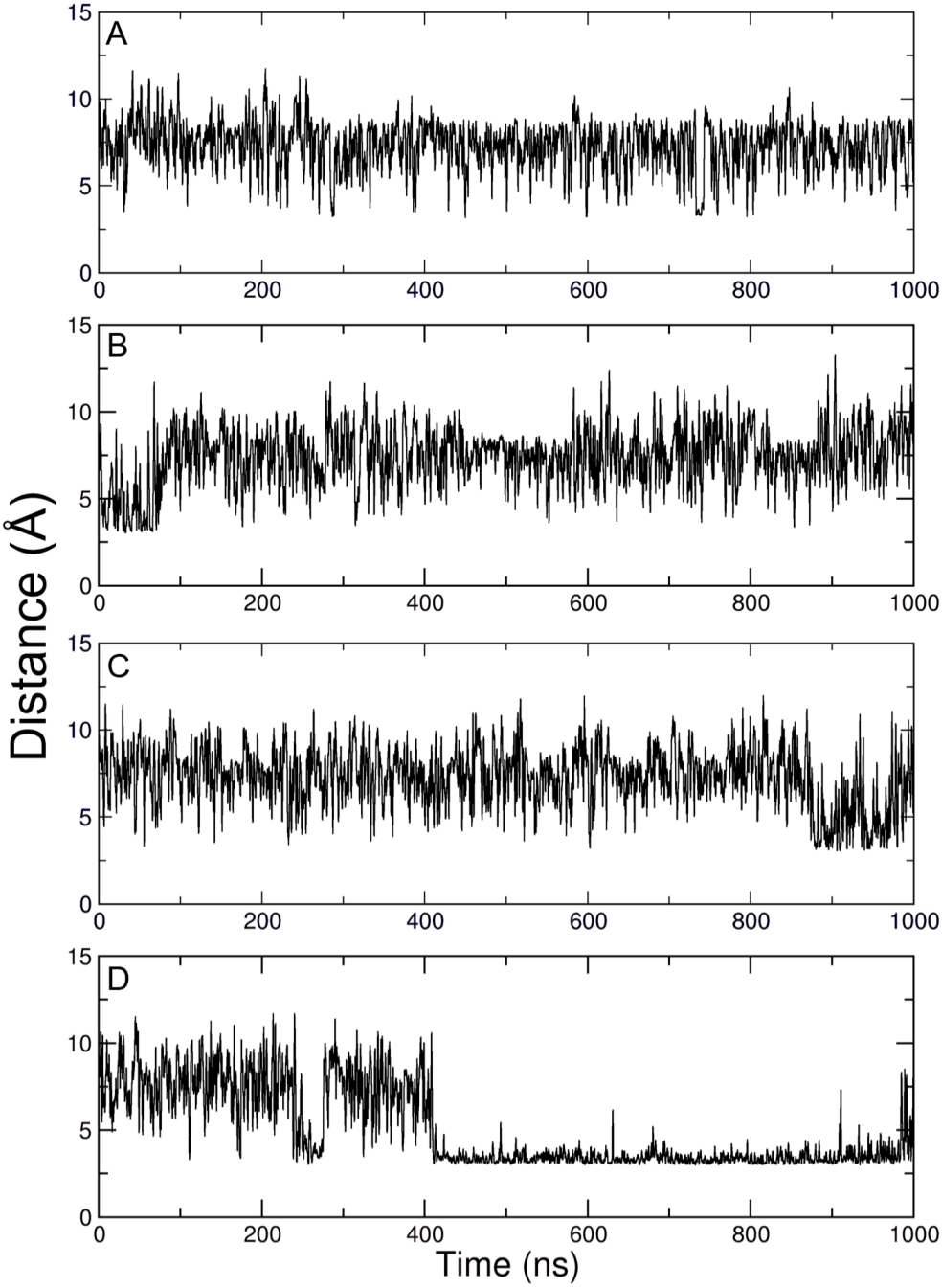

Figure S2.2 Distance between Thy13 O2 atom and nearest bulk K^+ ion for *VEGF*-T12T13 replicates 1 (A), 2 (B), 3 (C), and 4 (D) throughout simulation time. Data shown as 100-point (1-ns) running averages.

Chapter 3. Dynamics, Electrostatics, and Fragment Screening with Quadruplex-Duplex Hybrids and Atypical Tetrads

Attribution: R.J.F., H.M.M., and J.A.L. designed the research, H.M.M. developed the methodology, R.J.F. performed the simulations and analyzed the data, J.A.L. performed and analyzed the QM calculations, R.J.F. and J.A.L. wrote the chapter.

3.1 Introduction

The proviral integration site for Moloney murine leukemia virus-1 (*PIM-1*) gene is a proto-oncogene that encodes for the PIM-1 serine/threonine kinase. This kinase has a pro-tumorigenic role in a wide variety of biological processes including cellular proliferation and survival through the phosphorylation of various other oncogenes involved in transcription, cell cycle, and apoptosis regulation.[89,90] PIM-1 expression is known to be dysregulated in a variety of hematopoietic and solid tumors including breast, prostate, cervical, and lung cancers.[89,91–95] Most notably, this kinase is largely overexpressed in the highly aggressive and challenging-to-treat triple negative breast cancer (TNBC).[94,95] TNBC accounts for 10-20% of breast cancer diagnoses, however it disproportionately contributes to breast cancer metastatic cases and deaths due to the distinct disease profile.[96] TNBC tumors are characterized by the lack of the three common hormonal receptors present in other breast cancer subtypes; the estrogen, progesterone, and HER-2 receptors.

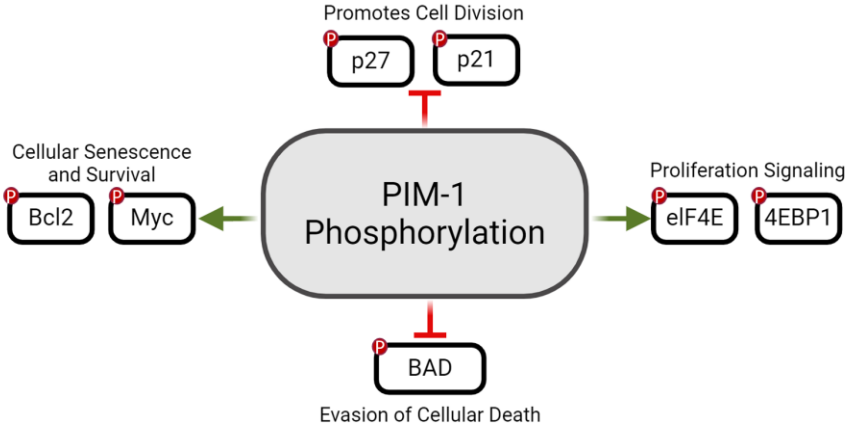

Figure 3.1 Downstream implications of PIM-1 kinase activity and phosphorylation with various other oncogenes that contribute to tumorigenesis, proliferation, and survival. Inactivation of PIM-1 activity is hypothesized to serve as one method to mitigate these downstream effects.

Due to the key biological implications of the PIM-1 kinase (Figure 3.1), and the known downstream response to the inhibition of the kinase activity, determining an appropriate targeting method may serve as a novel chemotherapeutic approach.[90,97–100] However, previous studies targeting the kinase directly have unsuccessful in clinical trials due to ineffectiveness and off-target effects.[101] As such, therapeutic intervention on the nucleic acid level emerges as a viable strategy for impairing aberrant PIM-1 function. The *PIM-1* promoter region has been found to form two distinct GQ structures that exist in equilibrium (Figure 3.2).[32] As discussed in Chapter 2, efforts to target GQs have been

ineffective in clinical trials due to the lack of specificity in ligand-binding sites. The two GQs that form in the promoter region of the *PIM-1* gene contain distinct structural and topological features that we hypothesize may serve as regions of enhanced specificity for designing selective ligands for preferential binding to the two forms. As these two forms exist in equilibrium, the dominant form 1 (PDB:7CV3) is nearly three times as prevalent *in vitro* studies when compared to form 2 (PDB:7CV4).[32] Form 1 folds into a three G-tetrad containing GQ, with a loop region composed of duplex DNA and a capping structure of a four-nucleotide hairpin loop. The *PIM-1* promoter sequence also adopts a distinct topology in Form 2; this form contains two G-tetrads stacked upon a mixed tetrad containing guanine and cytosine bases, in addition to a similar duplex and hairpin region described within form 1. Mixed tetrad configurations have been identified in telomeric GQ structures and are hypothesized to contribute to enhanced ligand binding specificity when targeting GQs.[18,102]

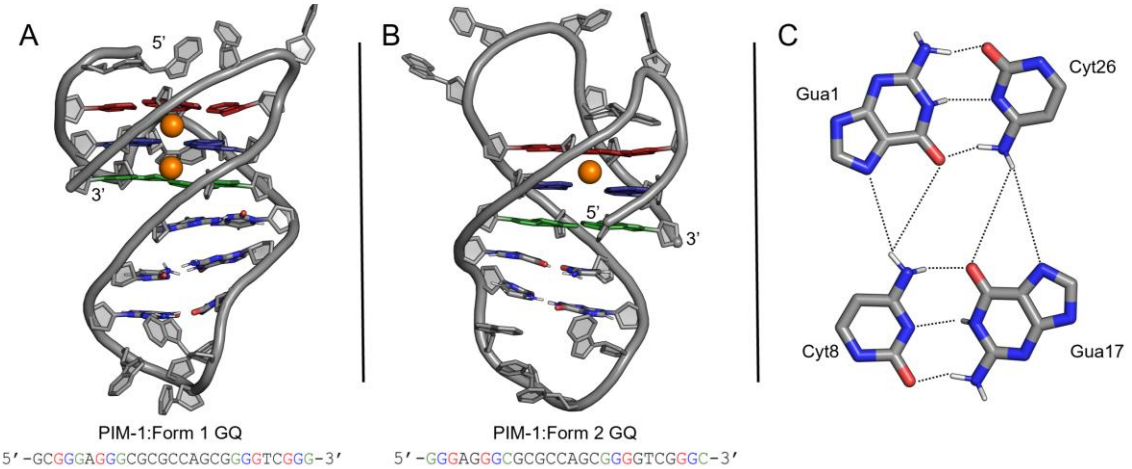

Figure 3.2 Structure of the *PIM-1* GQs. (A) *PIM-1:*Form 1 (major) GQ. (B) *PIM-1:*Form 2 (minor) GQ. Bases involved in forming the GQ core are colored by tetrad (1 – red, 2 – blue, 3 – green). Duplex region and Watson-Crick base pairs shown in stick representation colored by atom. Core K^+ ions shown as orange spheres. (C) Mixed tetrad composed of two Gua:Cyt Watson-Crick base pairs found to form tetrad 3 in the *PIM-1*:Form 2 GQ. Bases are labeled, and the proposed hydrogen bonding network is indicated by dotted lines.

The nature of tetrad formation has been a topic of great interest for multiple decades. While all-guanine tetrads are the most prevalent across DNA and RNA GQs, there has been experimental identification of tetrad configurations comprised of various other nucleobases and located in other higher-ordered nucleic acid structures, such as i-motifs.[103] These tetrads can form through the interaction of identical nucleobases, known as homo-tetrads, or the arrangement of proximal base pairs to form mixed-base tetrad structures like the one investigated here.[18,104] The biological role of these mixed tetrad configurations is not yet fully understood, however it has been shown that this base arrangement can aid in the interaction between multiple GQ structures, coordinate cations, and impact nucleic acid folding pathways.[105] As such, we hypothesize that the presence of a mixed tetrad configuration will have implications on structural dynamics and characteristic behaviors important for informed drug design.

Here, we aimed to further characterize electrostatic and structural properties of a mixed tetrad GQ through conventional and enhanced sampling MD simulations and conduct a comparative analysis of the coexistent *PIM-1* GQs. Additionally, through computational fragment-based drug design techniques, we sought to identify potential structural characteristics that may contribute to preferential ligand binding to modulate PIM-1 expression.

3.2 Methods

System Construction

In this study, MD simulations were performed on resolved NMR structures for the two coexisting *PIM-1* promoter GQs: the primary GQ (form 1) and the minor GQ population (form 2). Coordinates from the NMR ensemble of each isolated GQ population is contained in the Protein Data Bank (PDB) entries 7CV3 and 7CV4 and these resolved structures provided the initial starting coordinates for form 1 and form 2 simulations, respectively.[32]

The published structure of the primary GQ-duplex hybrid structure (PDB:7CV3), denoted here as *PIM-1:*Form 1, contains a conventional GQ topology containing three stacked guanine tetrads, and a loop region comprised of three additional Watson-Crick duplex base pairs (Figure 3.2A). Tetrad 1 is comprised of guanines 3, 7, 22, and 25; tetrad 2 of guanines 4, 8, 21, and 26; and tetrad 3 of guanines 5, 9, 20, and 27. Despite being observed within the same primary sequence, the secondary form (PDB:7CV4), known here as *PIM-1:*Form 2, exists *in vitro* as approximately 20% of the population and adopts a distinct topology (Figure 3.2B). This GQ contains a similar stacked tetrad core; guanines 3, 6, 19, and 24 comprise tetrad 1; guanines 2, 7, 18, and 25 comprise tetrad 2. However, tetrad 3 contains both guanine and cytosine bases and is arranged through additional interactions between adjacent Watson-Crick base pairs, G1:C26 and G17:C8 (Figure 3.2C). The secondary form also contains a similar loop region with two additional Watson-Crick base pairs.

The published structures lack essential, bound K^+ ions that are expected to be present within the GQ tetrad core, therefore these ions were added to ensure accurate electrostatic modeling. This task was completed using the CHARMM program.[79] K^+ ions were positioned symmetrically between stacked guanine tetrads by using the averaged coordinates of guanine carbonyl oxygen (O6) atoms in the consecutive tetrads. This method was applied to form 1 with two bound K+ ions added between the guanine containing tetrad core prior to simulation. This method is a standardized practice within the system construction of all-guanine tetrad containing GQs; however, core cation occupancy is not yet well characterized for mixed tetrad configurations. As a result, simulation systems were prepared with only one K^+ ion positioned between the all-guanine tetrads 1 and 2 in *PIM-1:*Form 2 (see below).

Equilibration and Unrestrained MD Simulations

All systems were initially prepared using the additive CHARMM36 nucleic acid FF.[80] The fully constructed *PIM-1* GQ systems, including the manually constructed core ions, were centered in cubic unit cells. Each unit cell was constructed to maintain a box-solute distance of 10 Å and was solvated with CHARMM-compatible TIP3P water molecules and KCl in 150 mM concentration, including neutralizing counterions. This

concentration was utilized as it has previously been demonstrated to simulate proper natural GQ folding and behaviors and is representative of the standard eukaryotic cellular KCl concentration. Following this construction process, the solvated systems were then subjected to relaxation through an energy minimization process in CHARMM. This process includes 500 steps of the steepest descent minimization, as well as 500 steps of adopted-basis Newton-Raphson minimization.

Following minimization, the systems underwent equilibration in NAMD.[81] The equilibration was carried out for 1 ns under NPT conditions while position restraints were applied to all heavy, non-hydrogen atoms within the GQ, including the core cations. The surrounding solvent molecules, water and KCl ions, were unrestrained and free to move throughout the simulation box. The NPT conditions were upheld through use of the Langevin thermostat at 298 K with a friction coefficient of 5 ps^{-1}. Additionally, a Langevin piston method was utilized to maintain a pressure of 1 atm.[82] Periodic boundary conditions were applied throughout all system dimensions. Short-range van der Waals forces were smoothly switched to zero from 10 – 12 Å and neighbor lists were maintained within 14 Å. Electrostatic forces were calculated with the particle mesh Ewald (PME) method,[83] with a real-space cutoff of 12 Å and a Fourier grid spacing of ~1 Å. Four replicate simulations were performed for both *PIM-1* GQs, starting from different, random initial velocities during equilibration.

To prepare for production, each equilibrated additive system was converted to a Drude polarizable model in CHARMM. This conversion step applies additional Drude oscillators to all non-hydrogen atoms. Additionally, for full compatibility with the Drude polarizable force field, lone pairs must be added, as well as Drude oscillators to all non-hydrogen atoms within the system. The additive water model was also then converted to the polarizable SWM4-NDP model.[84] The Drude-2017 FF parameters[44,45] were applied to the nucleic acid structure, as well as the ions within the simulation box. Further preparation for production simulations included the relaxation of the Drude oscillator positions by steepest descent minimization in CHARMM. The same harmonic restraints described above were applied in this NPT equilibration process. For the Drude systems, an integration time step was limited to 1-fs due to the high-frequency vibrations and a "hard-wall" constraint of a 0.2 Å maximum Drude-atom bond length was applied to avoid polarization catastrophe. Each replicate system was then subjected to unrestrained MD simulations in OpenMM.[85] Each replicate system was run until a simulation time of 1.5 μs was reached, unless otherwise specified. Some replicates were extended due to lack of convergence (described below). The same NPT ensemble described above was applied throughout the production runs, with the application of a Monte Carlo barostat instead of the Langevin method in NAMD.

QM Core Ion Interactions

To better understand the energetics of ion coordination behaviors in the GQ core in the presence of various mixed tetrad configurations, symmetry-adapted perturbation theory (SAPT) calculations[106,107] were performed using the Psi4 software package.[108] This method determines the noncovalent interaction energy between two molecules and can be further analyzed to gather information regarding the various contributing components. We employed SAPT0 calculations, which omit intramolecular electron correlation effects but allow for a decomposition of interaction energies in terms of electrostatic, exchange,

induction, and dispersion components. The SAPT0 calculations applied the jun-cc-pVDZ basis set to all elements except the K^+ ions, which used the 6-311G(2df,2pd) basis set.

To calculate the interaction energies between a positively charged K^+ ion and the core GQ bases, we considered the passage of one K^+ ion through both "slipped" and "non-slipped" mixed tetrad configurations of the *PIM-1* GQ core, with or without a second K^+ ion present. Thus, we considered three energy surfaces for each of the two tetrad geometries. In the "1 ion" system, we considered a single K^+ ion translating along the entire GQ tetrad core. In the "2 ions, top" system, one K^+ ion enters from above tetrad 1 and coordinates between tetrads 1 and 2 with a second ion fixed between tetrad 2 and the mixed tetrad. In the "2 ions, bottom" system, one K^+ translated from below the mixed tetrad to coordinate between tetrad 2 and the mixed tetrad, with a second ion fixed between tetrads 1 and 2. The coordinates for the GQ core with a non-slipped mixed tetrad were taken from PDB:7CV4 and the slipped mixed tetrad coordinates were taken from a snapshot from the MD simulations described below (Replicate 1, ~710 ns). The tetrad coordinates were aligned such that the GQ core axis was coincident with the z-axis and the K^+ ion was translated along the axis in 0.5-Å intervals using the CHARMM program.[79]

GaMD Enhanced Sampling Simulations

The process by which the PIM-1 GQ forms interconvert is unknown. Therefore, capturing conformational changes and quantifying the free energies associated with the different conformations of these GQs are of particular importance when considering GQs as potential drug targets. The enhanced sampling method, Gaussian-accelerated MD (GaMD)[109] was applied to both *PIM-1* GQs, form 1 and 2. GaMD is an enhanced sampling method that works by applying an overall boost to increase the potential energy of a system to overcome free energy barriers. When a system has insufficient energy to overcome these barriers, it may become energetically trapped and may be unable to sample all relevant conformations in a practical amount of simulation time.

The "dual-boost" GaMD approach, in which boost potentials are applied to the dihedral and total potential energy terms, was applied to unequilibrated, Drude-converted starting structures. Five GaMD simulations were conducted, each combining 2 ns conventional MD, 50 ns GaMD equilibration, and 500 ns GaMD production in NAMD, with the first 10 ns of the GaMD production period discarded as additional equilibration time. The threshold energy, E, was set to serve as the lower bound for the simulations. Therefore, $E=V_{max}$, or the maximum of both potential and dihedral energies. All other default simulation parameters were maintained, unless otherwise stated. Statistical averaging of potential energies and boost recalculations were set to occur every 400 ps throughout both equilibration and production simulation time. Prior to beginning production, GaMD equilibration was assessed for convergence through the evaluation of k_0 and the threshold energy, E, during equilibration. Once convergence was confirmed, production simulations were started from the exact state (including velocities) calculated in the final step of equilibration. The GaMD production simulations were reweighted to recover the potential mean of force (PMF) profiles by specifying collective variables and applying the cumulant expansion to the second order within the PyReweighting toolkit.[109] For the highly dynamic *PIM-1:*Form 2 GQ, eRMSD and a pseudodihedral defined by several key bases (described below) were selected to serve as collective variables for the

reweighting process. Due to minimal conformational sampling, the reweighting process was not conducted for *PIM*-1:Form 1 at this time.

SILCS Simulations

Here, the established SILCS-RNA protocol[65] was extended to be compatible with DNA GQ structures and Virginia Tech Advanced Research Computing architecture. As in the unrestrained MD simulations, both form 1 and form 2 of the *PIM-1* GQ were generated using a minimized structure file containing only the GQ and the core cations as the starting structures. Twenty independent SILCS systems were then generated for both systems: ten with the neutral SILCS solutes (benzene, propane, methanol, formamide, and imidazole) and ten with the charged SILCS solutes (acetate and methylammonium). The GQs were centered in a simulation box 15 Å from the edge, and the solutes were placed randomly around the GQ at an initial concentration of ~0.25 M. Water was also present within the box at ~55 M. Following the initial setup, Grand Canonical Monte Carlo (GCMC) refilling was employed to first remove all solutes and water from around the structure, and then re-insert them. This refilling process allows for larger fragments to fill in buried pockets on the GQ surface prior to supplementing the box with the smaller solute fragments and water. Once desired concentrations were reached for all solutes, steepest descent minimization of 5000-steps was followed by 100 ps of equilibration prior to beginning production runs for all systems. Production runs consist of 100 cycles of GCMC-MD with GCMC sampling, minimization, equilibration, and 1 ns of production MD for each system. These systems employed weak restraints of 50.208 kJ/mol/nm^2 on C1' and N3 (pyrimidines) or N1 (purines) atoms to prevent large conformational changes from occurring, as well as restraints of 67873.8 kJ/mol/nm^2 on the K^+ ions to maintain their occupancy of the GQ core, as this force constant overcomes the core cation repulsion in the GQ. All simulations were conducted with the CHARMM36 FF[80] and performed using GROMACS 2022.5[86] and the SilcsBio package, version 2023.1.

Following the completion of production runs, a solute probability density map indicating the interacting locations for each of the solutes for each of the 10 systems was generated, known as a FragMap. These generated FragMaps were then combined in a two-step process to obtain a single map file that represents the regions of interaction for each solute across all SILCS systems. The process combines the maps generated for runs 1-5, as well as those generated for runs 6-10 for each solute. By comparing the resulting maps, overlap coefficients were calculated to assess convergence between the runs, an overlap coefficient ≥ 0.75 indicates well-converged solute activity across the runs. The maps were then all combined to obtain a single map file representative of regions of affinity for each solute. These maps were then visualized in PyMOL using a specified Grid Free Energy (GFE) threshold, as this value serves as an estimate of ligand binding affinity.

Polarizable SILCS Simulations with Drude FF

Following the completion of the *PIM*-1 GQ additive SILCS production runs with the CHARMM36 FF[80], the newly developed polarizable SILCS workflow was employed on the systems by using OpenMM 7.7.0[85] with the Drude-2017 FF.[46] One in every 10 frames from each of the 10 equilibrated, nonpolarizable systems was selected to serve as the starting point for the polarizable systems, resulting in 100 systems. The neutral solutes and charged solutes remained separate with 100 systems for each set, resulting in a

combined 200 systems for both *PIM-1* GQs. Then, using the CHARMM program, each of the starting systems were prepared for the polarizable production by the addition of lone pairs to hydrogen bond acceptors and Drude oscillator atoms to all heavy atoms in the system. A brief minimization process was also then performed, followed by an equilibration of 1 ns prior to the initiation of the production runs. Each production run consists of 20 ns of production MD, resulting in a combined total simulation time of 2 μs for all systems. The same weak harmonic restraints for the nucleobases used in the additive SILCS systems described above were applied in the polarizable production runs. However, no restraints were necessary on the core K^+ ions when implementing Drude. This provides sufficient sampling time in relation to the additive SILCS systems given the 1 fs timestep and slower kinetics associated with the Drude FF. Following the completion of the production runs, FragMaps were then generated in the same manner as the additive systems.

3.3 Results and Discussion

Here, we discuss various structural and electronic implications of a mixed tetrad presence in a GQ-duplex hybrid structure. As described previously, the sequence of the *PIM-1* promoter region has been shown to form two distinct GQs, and by investigating their dynamics and interactions we sought to observe details that could provide insight into the potential pathways of interconversion. This information can also be used for computer-aided drug design efforts to selectively target each of the two forms.

3.3.1 Conventional Molecular Dynamics Simulations of the *PIM-1*:Form 1 GQ

As *PIM-1*:Form 1 is the major conformation adopted by this GQ-forming promoter sequence, we first sought to characterize any structural features and behaviors that may inform future drug development efforts. Additionally, as this structure has many features of a typical GQ structure, the analysis performed here provides us with details that are necessary to adequately describe the implications of the mixed tetrad in the secondary GQ conformation that we will investigate later in the chapter.

Structural Dynamics

First, we explored the overall structural dynamics of the *PIM-1*:Form1 GQ and sought to confirm that this GQ-duplex hybrid behaves in a manner similar to what has been previously observed in other GQs. Therefore, through root-mean-squared deviation (RMSD) and root-mean-squared fluctuation (RMSF) analysis, we confirmed our initial hypothesis that due to the topology of the structure, various regions demonstrated distinct structural characteristics. For example, we observed that across all four replicates, the GQ tetrad core in the *PIM-1*:Form 1 system exhibited an average RMSD value of approximately 1 Å, while the solvent-exposed duplex region displayed a higher RMSD value, likely due to the contribution of bases in the intervening loop that are not involved in the base pairing (Figure 3.3).

Our RMSF analysis showed that, as anticipated, the bases involved in the GQ core exhibited less flexibility than bases located in solvent-accessible regions throughout the GQ structure (Figure 3.3). Additionally, despite being present in the loop region, the bases

engaged in Watson-Crick base pairs in the duplex region below the GQ core exhibited low levels of flexibility as well, with minimal deviation across replicates. This observation confirms that the hydrogen bonds that these bases participate in serve as stabilizing factors for these nucleotides. The 5'-terminal Gua1 base was free to interact with solvent throughout the simulation time, and therefore manifested an RMSF value of approximately 2 Å, while the RMSF value calculated for the 3'-terminal Gua27 base was approximately 0.5 Å across the replicates. This low RMSF value is due to the stabilizing nature of the Hoogsteen hydrogen bonds in which Gua27 engaged, as this base remained an integral component of tetrad 3 throughout the duration of the simulations. We observed high levels of flexibility among various other bases in the GQ that are not involved in additional stabilizing interactions, such as bulges and loop bases at the top and bottom of the structure (Figure 3.2). Overall, these findings suggest that the *PIM-1:*Form 1 GQ behaves in a manner similar to other GQ systems that have been characterized.

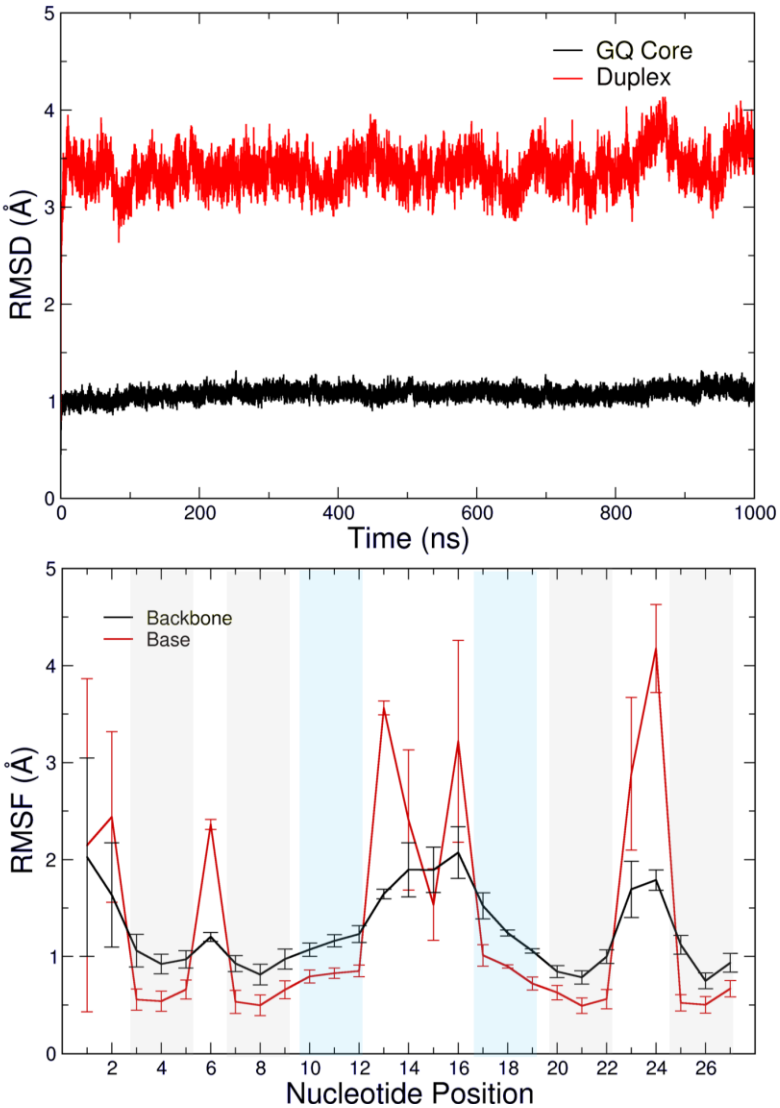

Figure 3.3 Average RMSD and RMSF values across four replicate simulations for *PIM-1:*Form 1 GQ. Error bars represent the standard deviation of the averages. Nucleotides in

the GQ tetrad core are indicated by the grey shaded regions. Bases involved in base-pairs in the DNA duplex loop are shown in the blue shaded regions.

PIM-1:Form 1 Base Electrostatic Properties

Previous work has shown that the electronic properties of bases present in GQs differ from that of conventional double-stranded B-form DNA due to the distinct hydrogen-bonding and stacking interactions of the GQs.[48,110] Therefore, we aimed to characterize the electronic properties of the *PIM-1* GQs, and provide a comparison between the two forms. Unique electronic properties contribute to the overall stability and structural properties of these GQs, as well as their interactions with ions and potential ligands. We explored our hypothesis that there would be observable differences between the *PIM-1:*Form 1 GQ and the *PIM-1:*Form 2 GQ by calculating the base dipole moment of all bases across the structure and calculating the average dipole moment of the bases based on their position within the structure. In the *PIM-1:*Form 1 GQ, we observed increased guanine dipole moments that are distinct from those found in canonical B-form DNA (Figure 3.4). Additionally, due to the interaction of tetrad 2 with the two core K^+ ions in this GQ, we observed the largest increase in dipole moment within these bases.

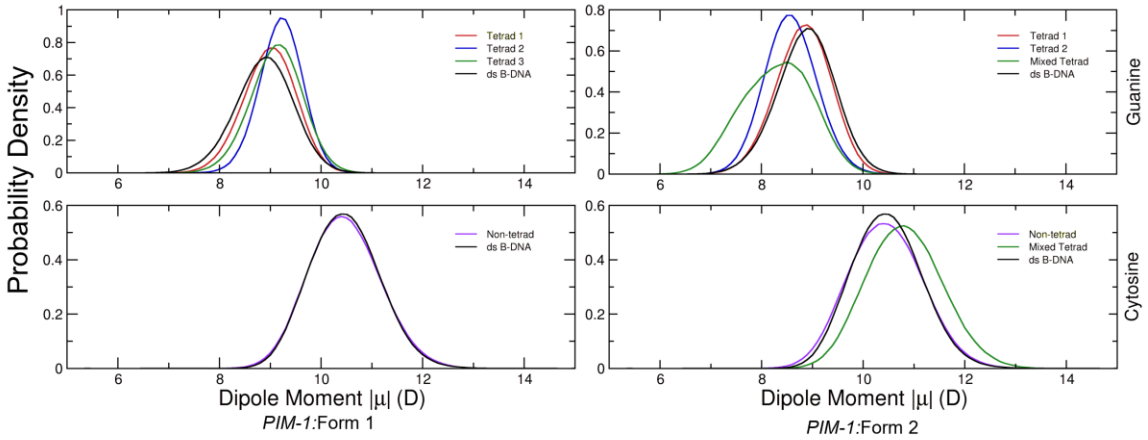

Figure 3.4 Base dipole moment distributions in the *PIM-1* GQs. Guanine and cytosine bases involved in various regions of GQ structure are plotted and compared to canonical double-stranded B-form DNA.

Ion Dynamics and Coordination

Investigations into GQ ion dynamics typically identify ion binding sites in the solvent-exposed regions above and below the tetrad core. In our simulations of the *PIM-1*:Form 1 GQ, we observed a region with affinity for an ion directly below tetrad 3, with some replicates maintaining an K^+ ion in that location for extended periods of simulation time. We performed an ion occupancy analysis to identify regions along the GQ structure that ions are present for ≥1% of the total 4 μs of simulation time. These ion occupancy maps revealed that, as hypothesized, the *PIM-1*:Form 1 GQ has ion binding sites at the solvent-accessible faces of the tetrad core (Supporting Information, Figure S3.1). This finding is particularly useful for informing drug design investigations, as it is predicted that drugs may have to compete with ions in these accessible regions to bind to the GQ. This property is discussed in more detail below.

Our analysis of the *PIM-1*:Form 1 GQ confirms that this GQ behaves in a manner expected by typical GQs in the overall patterns of dynamics and ion maintenance in the GQ core. Therefore, we will refer to these analyses as we discuss differences in the two *PIM-1* GQs throughout the chapter, especially as we further investigate the implications of a mixed tetrad present in the form 2 GQ.

3.3.2 Conventional Molecular Dynamics Simulations of the *PIM-1*:Form 2 GQ with a Mixed Guanine-Cytosine Tetrad

While ion coordination of conventional GQ structures has been widely investigated and characterized, ion interactions in structures containing mixed tetrads are comparatively under-studied. Previous investigations have shown that mixed tetrads in GQs can adopt multiple geometries that are stabilized by different interactions, such as a direct and a slipped configuration, thus differing in hydrogen bonding and cation coordination.[18,32] Therefore, to observe and analyze how these mixed tetrads interconvert between the two geometries, we performed four replicate simulations of the *PIM*-1:Form 2 GQ, which contains a direct mixed tetrad and only one core cation coordinated by all-guanine tetrads 1 and 2 as described above.

Given the uncertainty regarding the positioning of an ion between all-guanine tetrad 2 and mixed tetrad 3, we monitored the diffusion of K^+ ions during the simulation to determine if any ions would partition into the GQ core between these two tetrads. Among the four replicate simulations, we observed such ion partitioning in two replicates, whereas two replicates manifested no core ion coordination involving the mixed tetrad. Due to this variation in ion behavior (described in detail below), we extended the simulation time of all four replicates. We anticipated that additional time might allow the systems to overcome any energy barriers to gain more robust insights into the GQ dynamics and ion coordination of these systems. We extended all replicates to a total length of 1.5 μs, with an additional extension to 2 μs for replicate 3, for reasons discussed below. At the completion of our extended simulation time, replicate 1 exhibited bulk ion coordination of K^+ ions present in solution and replicate 3 demonstrated an unusual mechanism of ion partitioning within the electronegative core arising from the displacement of the ion initially placed between tetrads 1 and 2. Replicates 2 and 4 maintained the initially placed core K^+ ion and manifested no further ion coordination or ion dynamics of the GQ core (Supporting Information, Figure S3.2). As such, the remainder of the analysis on our conventional MD simulations will primarily focus on replicates 1 and 3 and characterizing the observations in the mixed tetrad.

PIM-1:Form 2 Base Electrostatic Properties

As with the *PIM-1:*Form 1 GQ, we investigated the base dipole moment properties, with an additional emphasis in the *PIM-1*:Form 2 GQ on cytosine bases given the presence of the mixed tetrad 3 in this structure. We observed that when involved in the mixed tetrad, the cytosine bases exhibited an increased dipole moment when compared to the cytosines present throughout other regions of the structure, as well as cytosines present within the canonical *PIM-1:*Form 1 GQ (Figure 3.4). Additionally, we observed an overall decrease in the dipole moment of the guanine bases in the mixed tetrad compared to those present within the GQ tetrad core of both *PIM-1* GQs. We observed that in the *PIM-1:*Form 2 GQ,

guanine bases in tetrad 2 were also depolarized when compared to tetrad 2 of the *PIM-1:*Form 1 GQ. This outcome is likely due to the variable core cation activity across the replicates of *PIM-1:*Form 2 GQ that will be discussed later.

Bulk Ion Partitioning and Tetrad Slippage in Replicate 1

The favorability of a K^+ ion to be coordinated by the mixed tetrad is largely unknown. We hypothesized that in our simulations of the *PIM-1:*Form 2 GQ with only one core cation placed in the GQ core, some ion partitioning from bulk solvent would take place, as has been observed in many prior GQ simulations.[48,111] In replicate 1, the *PIM-1*:Form 2 GQ behaved as hypothesized, retaining the core K^+ ion coordinated by all-guanine tetrads 1 and 2, as well as attracting a separate K^+ ion from bulk solvent at approximately 705 ns (Figure 3.5). Coordination of these two K^+ ions was maintained for the remainder of the simulation.

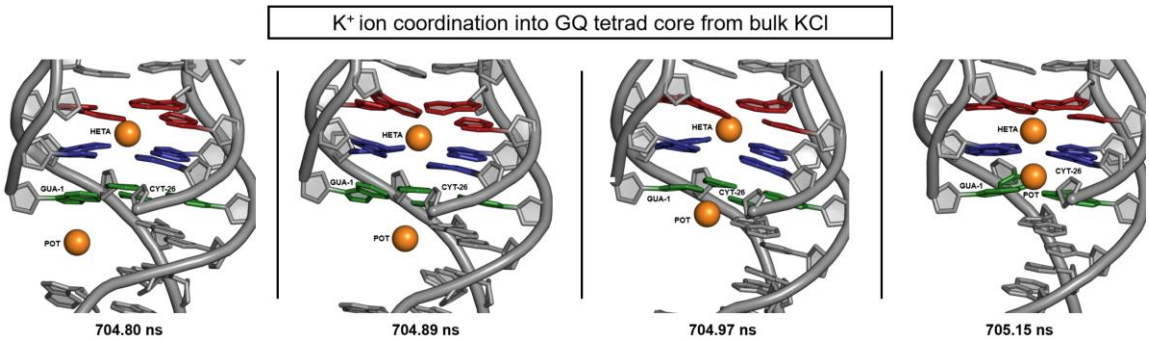

Figure 3.5 Mechanism of K^+ ion partitioning in replicate 1. Time stamps are located below each image. "HETA" label corresponds to the core K^+ ion placed within the GQ core, and "POT" labels correspond to K^+ ions from bulk solvent.

Mixed tetrads adopt distinct conformations depending on the hydrogen-bonding pattern in the tetrad.[18] As we observed this motif in the GQ to be dynamic and contribute to ion coordination, we sought to quantify the interactions that the mixed tetrad was involved in throughout the simulations. First, we calculated the RMSD values of various regions throughout the structure to determine the regions that contribute the most to conformational change. We calculated the RMSD of all heavy atoms of the nucleotides in various regions of the *PIM-1*:Form 2 GQ including the duplex region, the all-guanine tetrads 1 and 2, as well as the mixed tetrad guanine and cytosine bases (Figure 3.6A).

The duplex loop region had the highest RMSD values due to the frequency of solvent interactions resulting in base dynamics observed in the nucleotides not involved in Watson-Crick base-pairs. Additionally, as hypothesized, the ion behaviors of replicate 1 appeared to rely on large structural changes in the mixed tetrad, which can be observed here as well, as the mixed tetrad exhibited an average RMSD of 1.6 ± 0.4 Å in replicate 1. These findings confirm our hypothesis that some degree of conformational change must occur to allow ion partitioning across the mixed tetrad, confirming that it is likely that a conversion from a direct tetrad configuration to a slipped tetrad configuration may be observed within these replicates that coordinate ions due to interactions with the mixed tetrad.

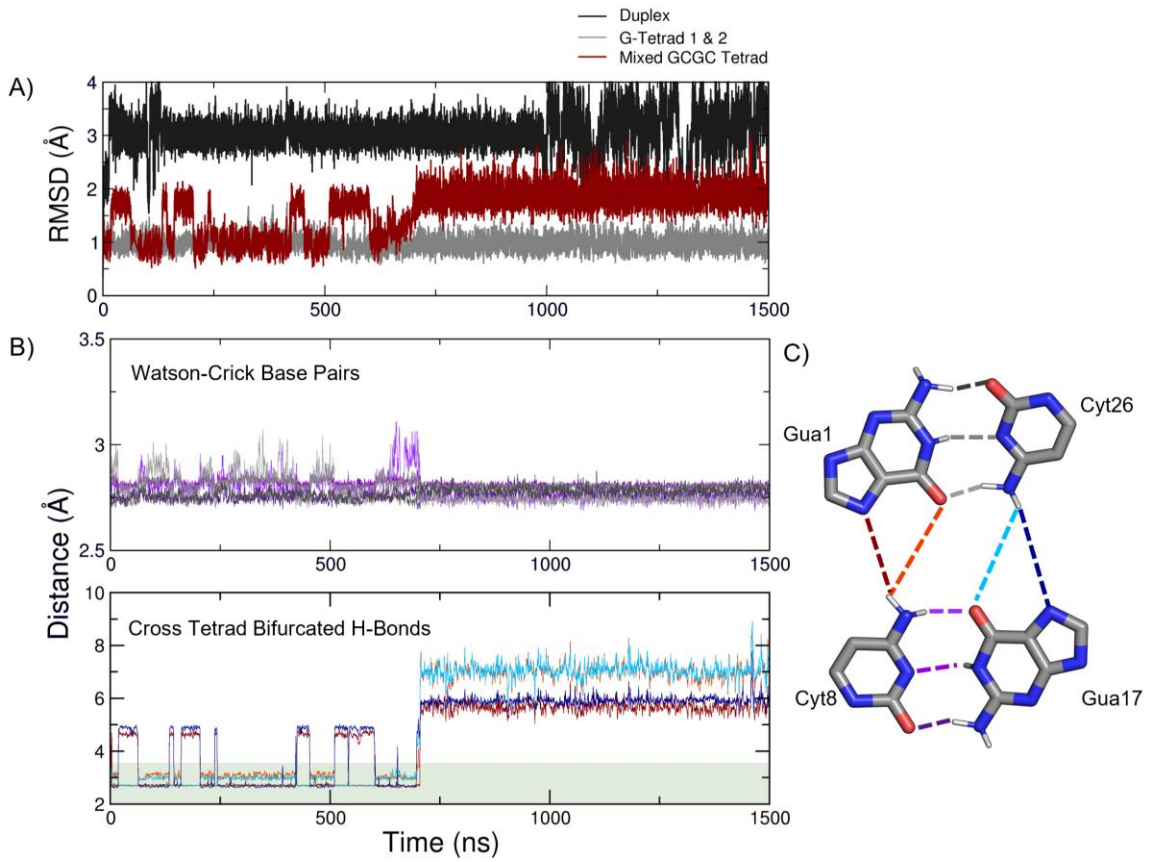

Figure 3.6 Characterization of the *PIM*-1:Form 2 mixed tetrad in replicate 1. (A) RMSD over time for all heavy atoms contained in distinct regions of the overall GQ structure. (B) Mixed tetrad hydrogen bonding distances. An approximate hydrogen bonding distance is below 3.5 Å (green shaded region). The distance data series in panels (A) and (B) are shown as 100-point (1-ns) running averages. (C) Hydrogen bonding patterns in the mixed tetrad. Watson-Crick Gua1:Cyt26 and Gua17:Cyt8 hydrogen bonds are shown in various shades of grey and purple, respectively. Bifurcated hydrogen bonds between Cyt8 and Cyt26 N4 and Gua O6 or N7 atoms are shown in shades of red and blue, respectively.

To characterize if such slippage was occurring within the mixed tetrad, we characterized the hydrogen-bonding patterns involving those bases, as these interactions distinguish between the two geometries. As described above, the mixed tetrad forms due to additional interactions occurring between two Watson-Crick base pairs. When additional, bifurcated hydrogen bonds form between these two base pairs, the mixed tetrad is engaged in a "direct" conformation (Figure 3.6C). Therefore, we calculated the distance between the guanine O6, N1, and N2 atoms and the cytosine N4, N3, and O2 atoms, respectively. Using a hydrogen bond distance threshold of 3.5 Å between the heavy atoms, we confirmed that, as hypothesized, the Watson-Crick base pairs remained intact throughout the entirety of the simulation time (Figure 3.6B).

Next, to test our hypothesis that the loss of the cross-tetrad bifurcated bonds can be observed at points of ion partitioning into the GQ core resulting from conformational changes, we quantified these hydrogen bonds, as well. For this analysis, the distances between the cytosine N4 atoms and the guanine O6 and N7 atoms were calculated. We

confirmed the initial presence of the direct mixed tetrad configuration, as all bifurcated hydrogen bonds were initially present at the start of the simulations (Figure 3.6B). Replicates 2 and 4 also maintained all four bifurcated bonds from the beginning of the simulation (Supporting Information, Figure S3.3). As hypothesized, the results of this analysis indicate the loss of the bifurcated hydrogen bonds in replicate 1 at approximately 705 ns. At this time, a K^+ ion from bulk solvent interacted with the guanine O6 atoms at a distance of approximately 2-3 Å and remained there for the rest of the simulation (Figure 3.7). These results suggest that a conformational transition from a direct to a slipped mixed tetrad configuration in replicate 1 occurred following bulk K^+ ion coordination. The destabilization of these hydrogen bonds corresponded to the coordination of a K^+ ion from bulk solvent (Figure 3.7). The increase of all four bifurcated hydrogen-bond distances to approximately 6-8 Å, values that persisted for the remainder of the simulation, further suggest that the mixed tetrad adopted a slipped conformation following the coordination of a second K^+ ion into the GQ core.

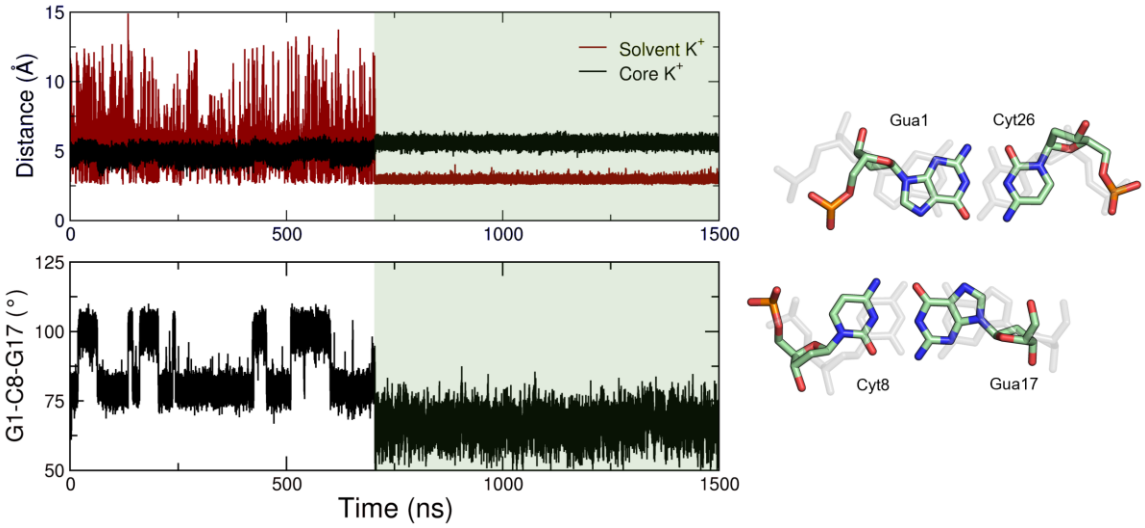

Figure 3.7 Ion coordination and mixed tetrad configuration throughout replicate 1. The top plot shows the minimum distances between a K^+ ion and Gua1 and Gua17 O6 atoms, and the bottom graph shows the Gua1-Cyt8-Gua17 base COM angle. Colored regions of the graphs correspond to distinct simulation snapshots and mixed tetrad configurations shown here, with the starting mixed tetrad arrangement shown in grey.

The conformational shift from the direct tetrad configuration to the slipped tetrad configuration is primarily characterized by the coordination of a stabilizing cation between the mixed tetrad and the all-guanine tetrad 2. Additionally, as described above, when this cation was present, the bifurcated bonds in the GQ tetrad core were broken as that space became occupied by the ion. The Watson-Crick base pairs further shifted to orient the guanine bases in the most favorable position to interact with the cation directly. As we have shown, it is reasonable to hypothesize that the conformational conversion to the slipped tetrad configuration occurred in replicate 1 due to the position of the coordinating K^+ ion. To further characterize the transition from the direct to the slipped conformations within the mixed tetrad, we investigated both the distance of the K^+ ion from Gua1 and Gua17 O6 atoms and the angle formed by the centers-of-mass of the Gua1, Cyt8, and Gua17 bases to

better visualize the slippage of the Watson-Crick pairs. Monitoring the angle is a more direct metric of the slipping behavior, as in a square-planar arrangement, this value should be ~90° but would become acute upon slipping. In addition to visualizing features of the conformational interconversion of the mixed tetrad, this analysis also provides details on the dynamics of both bulk and the core cations. The ion partitioning event corresponded to a sudden decrease in the angle of the mixed tetrad bases. As expected, the angle formed by the bases shifted from ~90° to ~60° upon conversion to the slipped conformation (Figure 3.7). Given that the slippage of the mixed tetrad results from the re-orientation of guanine bases to adopt more favorable interactions, this transition will result in a direct interaction between the guanine O6 atom and the K^+ ion, as seen in the all-guanine tetrads of GQs.

Ion and Mixed Tetrad Dynamics in Replicate 3

While the results of the simulation from replicate 1 supported our initial hypothesis of ion coordination between tetrad 2 and the mixed tetrad (discussed above), in replicate 3 we observed ion dynamics that were unprecedented in previous studies of GQ structures. The K^+ ion initially placed between tetrads 1 and 2 fluctuated within the GQ core, sampling the locations between all three tetrads, including the mixed tetrad. For the first 900 ns of this simulation, the core K^+ ion remained between tetrads 1 and 2 (Figure 3.8). However, from 900ns – 1 μs, the core K^+ ion shifted positions throughout the GQ and sampled locations among all tetrads, favoring a co-planar coordination with tetrad 2.

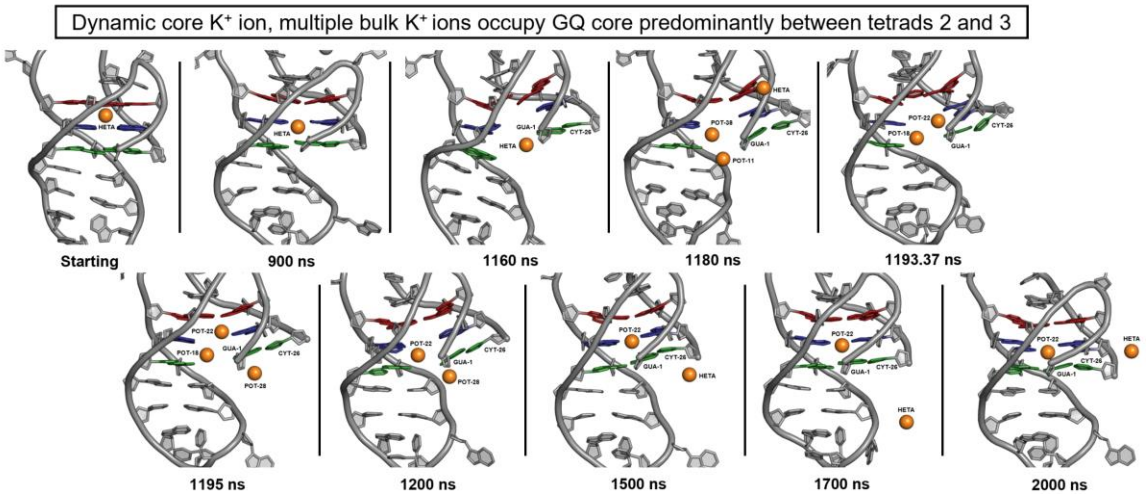

Figure 3.8 Mechanism of K^+ ion partitioning in replicate 3. Time stamps are located below each image. "HETA" label corresponds to the core K^+ ion placed within the GQ core, and "POT" labels correspond to K^+ ions from bulk solvent.

Upon the extension of the simulation, we observed unusual behaviors of the core ion and the tetrad core itself. Following the core cation engagement with the entire GQ core, we observed that the GQ core became increasingly dynamic as noted by increases in RMSD for all tetrads (Figure 3.9A), allowing for the core K^+ ion to diffuse into the bulk solvent, leaving the GQ core unoccupied from 1160.59 ns to 1180 ns. According to previous studies, the coordination of core cations is vital to overall GQ stability, therefore this finding indicates that the *PIM-1*:Form 2 GQ may have somewhat more plasticity than other GQs. At 1180 ns, the *PIM-1:*Form 2 GQ coordinated a K^+ ion from bulk solvent that

interacted with various positions in the entire core. When a second bulk K^+ ion partitioned into the core, the two ions briefly interacted with the mixed tetrad bases due to the solvent-accessible state of the tetrads, prior to one ion shifting upwards in the GQ core between tetrads 1 and 2, shown at approximately 1195 ns (Figure 3.8). The GQ then underwent a conformation change, resulting in the retention of only one cation in the GQ core, with this ion sampling various locations prior to establishing a preference for the coordination between tetrad 2 and the mixed tetrad. These findings indicate that the *PIM-1:*Form 2 GQ contains additional features that contribute to the stability of the overall fold, as well as demonstrating that there may be a strong affinity for the core K^+ ions within the mixed tetrad, but only after some degree of opening by disrupting hydrogen bonds.

For replicate 3, we performed the same analysis of the dynamics in the mixed tetrad that we described above for replicate 1 to fully characterize the behaviors of the mixed tetrad. Throughout replicate 3, the *PIM-1:*Form 2 GQ underwent large conformation changes that influenced ion dynamics, so our analysis of this system aimed to better capture the features of this GQ that contribute to these large changes. Starting with our RMSD analysis, we determined that the overall trends found in replicate 1 persisted with replicate 3, including the higher RMSD that we can observed in the freely solvent-accessible duplex loop, as well as the lower RMSD values in the remainder of the GQ core (Figure 3.9A). In this replicate, the mixed tetrad had an average RMSD of 1.7 ± 0.6 Å. This higher RMSD value corresponds to the more prominent dynamical behavior present in this motif relative to replicate 1. Additionally, we observed that at approximately 900 ns, a large increase in the RMSD values occurred for both the mixed tetrad as well as the other two all-guanine tetrads in the GQ core. This increase corresponds to the largely distorted state that this GQ adopts throughout much of the rest of the simulation due to this GQ remaining highly dynamic through the core and mixed tetrad specifically.

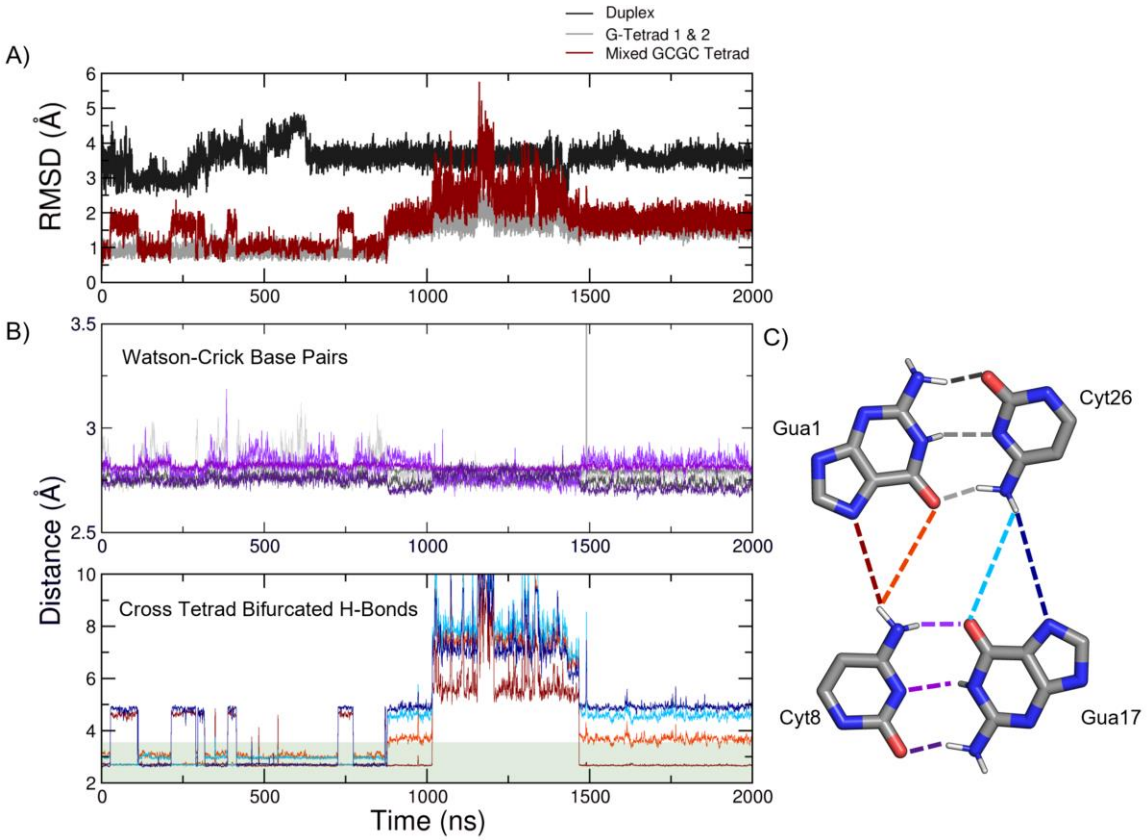

Figure 3.9 Characterization of the *PIM-1:*Form 2 mixed tetrad in replicate 3. (A) RMSD over time for all heavy atoms contained in distinct regions of the overall GQ structure. (B) Mixed tetrad hydrogen bonding distances. As in Figure 3.6, an approximate hydrogen bonding distance threshold of 3.5 Å is shown in the green shaded region. The distance data series in panels (A) and (B) are shown as 100-point (1-ns) running averages. (C) Hydrogen bonding patterns in the mixed tetrad using the same color scheme as in Figure 3.6.

We then investigated the hydrogen bonding network of the mixed tetrad of replicate 3. We confirmed that the Watson-Crick base pairs were maintained within hydrogen bonding distance, aside from a small fluctuation of Cyt26 out of plane at approximately 1489 ns and lasting for only ~2 ns. When investigating the pattern of the formation of the cross-tetrad bifurcated bonds within replicate 3, we observed similar shifts in the distances between the atoms as described in replicate 1. At the beginning of the simulation, we observed that all four of the hydrogen bonds were within hydrogen-bonding distance, with slight variations attributable to thermal fluctuation (Figure 3.9B,C). In replicate 3, the *PIM-1*:Form 2 GQ maintained these bifurcated hydrogen bonds until approximately 900 ns, corresponding to the time during which the core K^+ ion was sampling the entire GQ core. These ion dynamics likely resulted in the loss of some of the bifurcated hydrogen bonds, as the ion interacted with the mixed tetrad bases directly at this time. Additionally, at approximately 1 μs, we observed that the bifurcated hydrogen-bonding network was disrupted, with all interatomic distances increasing to 5 Å or higher, indicating a complete loss of this bifurcation network. The increased distances among the bases correspond to the increase in the RMSD of the mixed tetrad at this point in the simulation (Figure 3.9A). As shown in the snapshots provided in Figure 3.8, the *PIM-1:*Form 2 GQ existed in a

distorted state, such that these bifurcated hydrogen bonds were broken, contributing to the high RMSD. Additionally, the distances between the bases decreased and some bifurcated hydrogen bonds were reformed at approximately 1.4 μs, which corresponds to the collapse of the GQ tetrad core (Figures 3.8 and 3.9B). These findings confirm that the dynamics we observed within the mixed tetrad, supported by RMSD analysis, correspond to considerable conformational variability in this region. However, we observed that the GQ in this replicate did not adopt the proposed slipped conformation observed in replicate 1, as no ion was coordinated by the mixed tetrad directly.

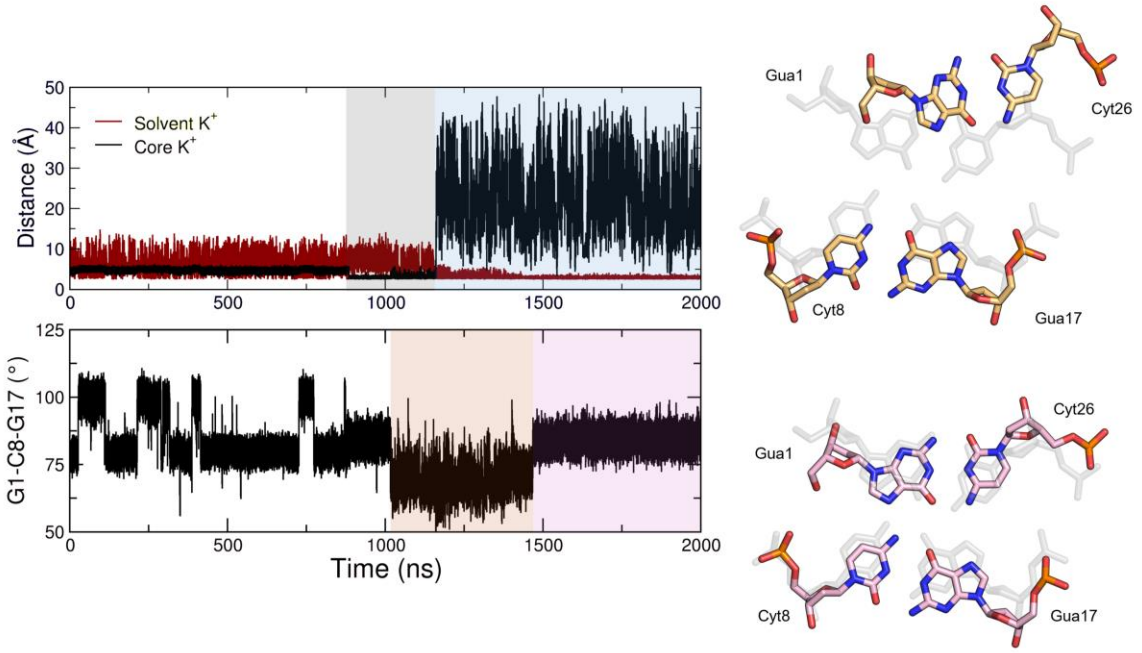

Figure 3.10 Ion coordination and mixed tetrad configuration throughout replicate 3. The top plot shows the minimum distances between a K^+ ion and Gua1 and Gua17 O6 atoms, and the bottom graphs shows the Gua1-Cyt8-Gua17 base COM angle. Colored regions of the graphs correspond to distinct simulation snapshots and mixed tetrad configurations shown here, with the starting mixed tetrad arrangement shown in grey.

We conducted the same geometric analysis as in the case of replicate 1 regarding K^+ distance to Gua1 and Gua17 O6 atoms, and the mixed tetrad center-of-mass angle to further characterize the conformational dynamics observed in replicate 3 and their implications on ion dynamics. First, the ion minimum distance analysis clearly demonstrates the distinct behaviors of the core K^+ ion we observed throughout the simulation (Figure 3.10). At approximately 900 ns, the core cation interacted with the bases in the mixed tetrad, demonstrating the downward shift from coordination by tetrads 1 and 2 to coordination by tetrad 2 and the mixed tetrad 3. As discussed above, this behavior indicates some degree of affinity for this region of the GQ core. Additionally, the distance analysis shows that this ion was expelled from the GQ core at approximately 1175 ns, following the distortion of the tetrads and the diffusion of the core K^+ ion into the bulk solvent, shown in the snapshots provided in Figure 3.8. Simultaneously, the distance between these guanine O6 atoms and bulk K^+ ions decreased, confirming the ion coordination within the GQ core from a bulk ion partitioning in. The core cation remained

solvated for the duration of the simulation and never returned to the GQ core, as indicated by the large distance from the GQ core. Additionally, following the initial ion coordination, the GQ core remained occupied by at least one K^+ ion from bulk solvent. The "distorted" GQ state was further characterized by observing the base angle of the mixed tetrad from 1100 - 1400 ns. The decrease in the angle, although similar to that of the slipped tetrad geometry, does not correspond to this state. Rather, as the GQ core "opened," the terminal Watson-Crick base pairs adopted a more acute angle. Additionally, following ion coordination from bulk solvent and the "reformation" of the overall GQ core, the angle returned to a value similar to that of the native structure. Overall, these results confirm that replicate 3 yielded distinct conformations of the GQ as well, despite the mixed tetrad never fully adopting the slipped geometry. These findings demonstrate different regions of affinity for cations throughout the GQ core, and a favorability for the GQ core to remain occupied by only one cation, at least on the time scale of these simulations.

Across the conventional MD simulations, our replicates exhibited a wide variety in the behaviors and structural dynamics of the GQ core. Two replicates (2 and 4) displayed minimal structural or ion dynamics within the GQ core, demonstrating the preference for one core K^+ ion to remain between tetrads 1 and 2. However, replicate 1 exhibited the partitioning of a K^+ ion through the slippage of the bottom mixed tetrad. This finding indicates that this GQ maintains some preference for coordinating two core cations, as long as a slipped configuration is adopted by the mixed tetrad. Replicate 3 displayed the highest level of structural and ion dynamics of the GQ core. This replicate displayed that ion modulation associated with various states of the GQ. Overall, these findings suggest that the dynamics and broad conformational ensemble associated with the *PIM-1*:Form 2 GQ may present new insights into the potential challenges with targeting this form *in vivo*, and may also explain why it is the minor form of the promoter GQ given the comparative rigidity of Form 1.

Ion Interaction Energy Calculations

The SAPT0 calculations of the non-slipped mixed tetrad systems revealed the expected energy minima between tetrads 1 and 2, and tetrad 2 and the mixed tetrad. Minima existed at equidistant points between the coordinating tetrads at $\Delta z \pm 1.8$ Å (positions relative to tetrad 2). In the "1 ion" system, these minima were of roughly equal magnitude, suggesting that a K^+ ion could be stably bound between tetrad 2 and a non-slipped mixed tetrad. However, there is a large energy barrier coincident with the location of the mixed tetrad that indicates it is essentially impossible for a K^+ ion to adopt this coordination if partitioning from the bulk solvent (Figure 3.11). The positioning of the mixed tetrad directly below tetrad 2 in a non-slipped configuration leads to steric repulsion as the ion moves along the tetrad axis. Even if the bases in the mixed tetrad were shifted relative to the tetrad axis, the ion would encounter similar repulsion with tetrad 2. Thus, any partitioning from below the mixed tetrad would require major rearrangement of the atoms in this configuration. In both "2 ion" systems, the positions of the minima are the same as in the "1 ion" system but shifted upward, reflecting some repulsion between the ions, but that both ions are favorably accommodated in the expected coordination geometries. The barrier to partitioning from the bulk solvent remains in the same location. Together, these quantum mechanical results indicate that a K^+ ion can be accommodated between tetrad 2 and the mixed tetrad in a non-slipped geometry, but this ion would have to partition into

the GQ core from outside of tetrad 1, prior to the entry of a K^+ ion that would be coordinated between tetrads 1 and 2.

In the slipped tetrad systems, the SAPT0 energy surface changes dramatically. The energy minimum between tetrads 1 and 2 remains at the same location and magnitude in the "1 ion" and "2 ions, top" systems, but the energy minimum between tetrad 2 and the mixed tetrad shifts slightly in the direction of the mixed tetrad (Figure 3.11). This result implies that the reconfiguration of the hydrogen bonding within the mixed tetrad leads to a favorable coordination geometry that is slightly closer to the mixed tetrad than in the case of the non-slipped mixed tetrad system. Further, the barrier to partitioning into the GQ core from below the mixed tetrad is abolished in this system. In fact, in the slipped mixed tetrad system, ion partitioning is essentially barrierless in either direction, suggesting that the rearrangement of hydrogen bonding interactions makes K^+ partitioning into the GQ core more favorable across the entire structure. Therefore, slipping of the mixed tetrad fundamentally alters the electrostatic properties of the GQ core. As in the non-slipped mixed tetrad system, the energy profiles of the "2 ions, top" and "2 ions, bottom" systems are systematically shifted upward relative to the "1 ion" system, but the barriers to partitioning and the locations and magnitudes of the energy minima are essentially unchanged.

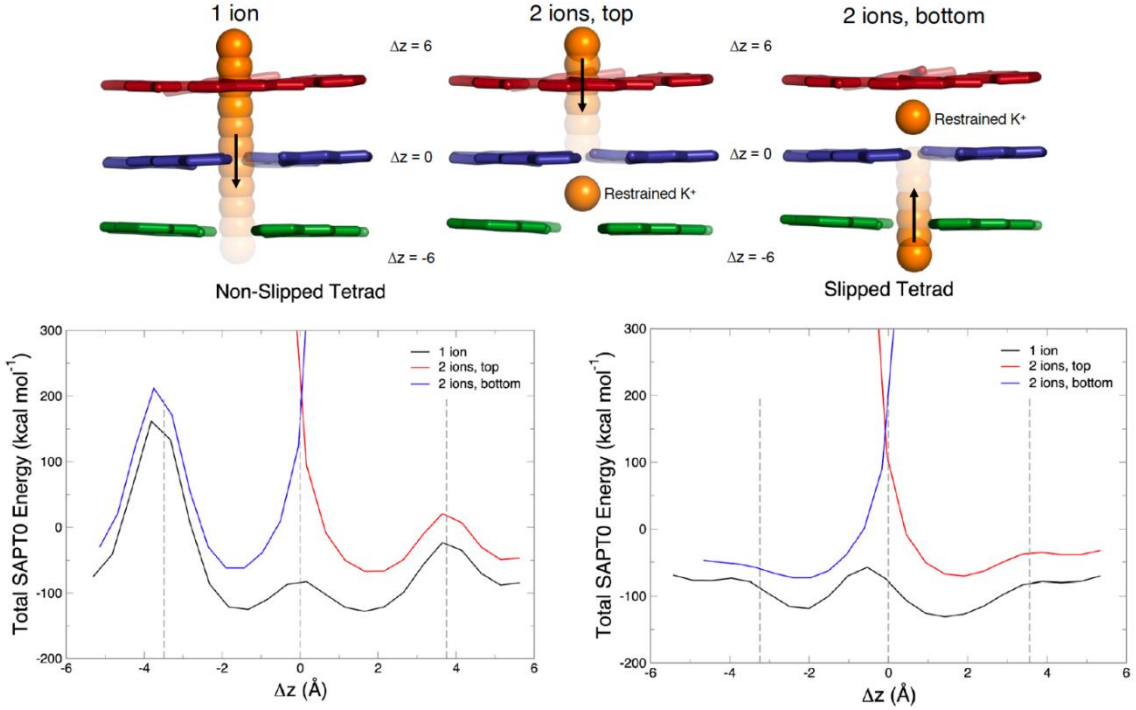

Figure 3.11 Energetics of K^+ partitioning through the GQ cores with non-slipped and slipped mixed tetrads, for each of the three ion coordination states indicated. In the images along the top, tetrad 1 is shown in red, tetrad 2 in blue, and the mixed tetrad in green. K^+ ions are shown as orange spheres with the directions of movement along the tetrad axis shown as arrows.

These findings support our observations from the conventional MD simulations, as we observed the coordinated K^+ ion placed between tetrads 1 and 2 to appear to prefer a position between tetrad 2 and the mixed tetrad in replicate 3, shown through the movement

throughout the GQ tetrad core. While this replicate adopted a non-slipped configuration of the mixed tetrad, this observation validates the findings here demonstrating that this partitioning would need to result from a downward movement from tetrad 1 due to steric occlusion. Additionally, as we observed the slipped tetrad conformation in replicate 1, we can confirm that the partitioning of a K^+ from below the GQ core was accessible from bulk solution, as this behavior is likely due to the low energy barriers associated with ion partitioning when this geometry is adopted.

3.3.3 Enhanced Sampling of *PIM*-1:Form 2 with GaMD

Due to the variable ion coordination behaviors and distinct differences across the replicates of our conventional MD simulations for the *PIM-1*:Form 2 GQ, we suspected that the GQs might be kinetically trapped in several different conformations across the free energy landscape. That is, that each of the replicate simulations may be representing a low-energy state, without the ability to sample additional conformations that may be separated by large energy barriers. This behavior is made even more likely from the inherently slower, but more realistic, kinetics of Drude polarizable simulations compared to nonpolarizable simulations. Therefore, we performed Gaussian-accelerated molecular dynamics (GaMD) simulations,[109] to more exhaustively sample the conformational landscape of these PIM-1 GQs. We performed these enhanced sampling simulations starting from four different structures arising from our conventional MD simulations. For example, as we previously mentioned, the terminal base pairs involved in the mixed tetrad of *PIM-1*:Form 2 were highly dynamic, resulting in the overall instability of the GQ tetrad core. Therefore, we identified various frames of the simulations that were most representative of the "distorted" and "reformed" conformations of the GQ following these conformational changes observed in replicate 3 of the conventional MD simulations. In addition to these representative starting points obtained from replicate 3 of conventional MD, we also simulated the *PIM-1*: Form 2 GQ, with both one and two core K^+ ions, by using the unequilibrated Drude structures from our conventional simulation setup (Figure 3.12).

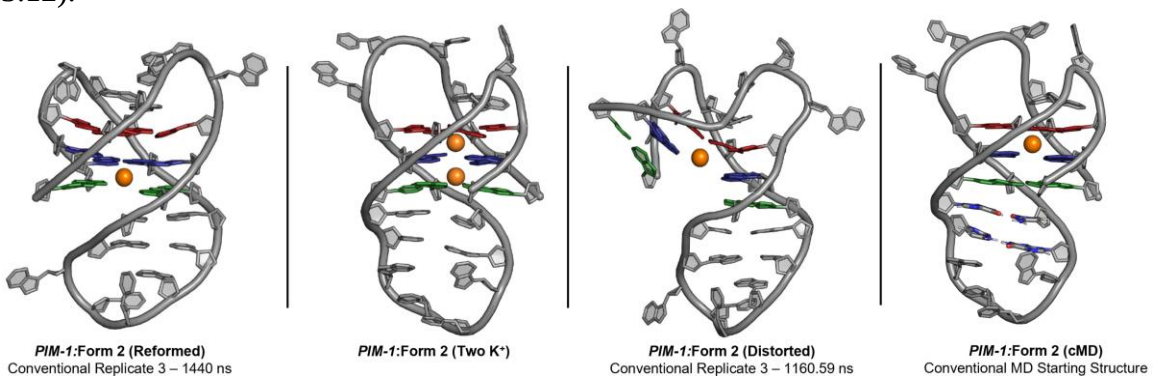

Figure 3.12 Starting structures for GaMD simulations of the *PIM-1*:Form 2 GQ. Four *PIM-1*:Form 2 GaMD systems were simulated in total.

As this study is one of the first applications of GaMD with the Drude polarizable FF, we first assessed our systems for convergence following the equilibration process. To do so, we analyzed the k_0 and E values that define the overall boost applied to the system. We observed that all the GaMD systems achieved convergence during the initial 50 ns of

equilibration time (Supporting Information, Figure S3.4). Having confirmed that the systems had converged, we began 500 ns of GaMD production simulation time in each system. For analysis, the first 10 ns of production time was discarded as additional equilibration time for the system to obtain a relaxed starting point for simulation time, therefore for GaMD production, we will be discussing our observations from the results of 490 ns of simulation time.

The rationale for performing enhanced sampling simulations rests on the highly dynamic behaviors that we observed in our conventional MD simulations. We hypothesized that the conformational changes in the *PIM-1:*Form 2 GQ correspond to different thermodynamically accessible states and by boosting potential energy in GaMD simulations, we would be able to capture more of the conformational ensemble of these *PIM-1* GQ systems. First, we observed the structural dynamics of the GaMD systems and aimed to characterize any distinguishable features that differ from the observations made in conventional MD simulations.

Across all GaMD replicates of the *PIM-1:*Form 2 GQ, we observed similar dynamics in the Watson-Crick base pairs that comprise the mixed tetrad as those observed in the conventional MD. Specifically, in all simulations in which there was no K^+ ion initially placed between tetrad 2 and the mixed tetrad, we observed that the terminal bases, Gua1 and Cyt26, interacted with the bulk solvent and thus resulted in an "open" conformation that allowed for the GQ core to remain accessible for additional cation coordination. This outcome aligns with the observations from replicate 3 of the conventional MD simulations. To characterize this behavior, we conducted both an eRMSD and a pseudo-dihedral analysis to better understand this swinging motion that the GQs are engaged in.

eRMSD is a metric that reflects the variation in the nucleobase orientations throughout a structure.[112] Thus, while providing a similar comparison from distinct time points in the MD trajectory to the initial starting structure as standard RMSD calculations, this measurement aims to provide a more specific analysis associated with the internal geometries of nucleic acid structures. The analysis of eRMSD values across the GaMD simulations revealed some distinction between replicates. Overall, all eRMSD values across all GaMD trajectories fell within the range of 0.6 to 1.6 (Figure 3.13A). eRMSD values ≤1.0 serve as a threshold, below which the structures are typically considered very similar. Therefore, we observed both the *PIM-1:*Form 2 (cMD) and *PIM-1:*Form 2 (Two K^+) GQs remained above this threshold for the entirety of the simulation time. It is likely that this result is due to a conformational change in the mixed tetrad as it adopts the "slipped" configuration over time, as we observed in both of these trajectories. Additionally, the *PIM-1:*Form 2 (Reformed) and *PIM-1:*Form 2 (Distorted) GQs exhibited low eRMSD values for approximately half of the simulation time prior to reaching steady, larger eRMSD values. At the beginning of the simulation, the *PIM*-1:Form 2 (Distorted) GQ remains highly dynamic with an unoccupied core and flexible terminal bases, properties that largely aligns with the observations made in the conventional MD, strongly resembling with the overall structure provided as the reference for this replicate. Additionally, the same behavior was observed in the *PIM*-1:Form 2 (Reformed) GQ system as this structure remained in that conformation prior to the opening of the mixed tetrad to coordinate an additional core K^+ ion. Therefore, as these replicates coordinated those

additional ions into the GQ core, the nucleobase interaction networks become changed, resulting in the higher observed eRMSD values.

In addition to the eRMSD analysis, we also calculated a pseudodihedral to characterize the overall structural dynamics of the systems (Figure 3.13B). The pseudo dihedral angle was calculated from the centers-of-mass of the Gua1, Cyt8, Gua7, and Gua3 bases (Figure 3.13C), allowing us to characterize the overall "swinging" motion of the terminal bases that facilitates ion coordination observed across the replicates. We observed that these bases adopted a smaller pseudodihedral angle ($\leq 45^{o}$) concomitant with the "opening" or "distortion" of the GQ core, whereas a larger angle was sampled when the GQ core adopted the more stable, "closed" conformation (Figure 3.13B). In particular, in the *PIM-1*:Form2 (Distorted) GQ, we observed fluctuations of smaller angles when this replicate remained highly dynamic and distorted, with a large increase in dihedral angle when this replicate coordinated bulk K^+ ions and adopted a more conventional GQ structure. The same pattern was observed in the *PIM*-1:Form 2 (Reformed) GQ, as this system exhibited a larger pseudodihedral angle prior to the "opening" of the GQ core, a behavior that corresponds to the large decrease in the angle at approximately 150 and 250 ns, prior to the coordination of a second core K^+. Additionally, in the *PIM-1*:Form 2 (cMD) and *PIM-1*:Form 2 (Two K^+) replicates, we observed only small variation in the pseudodihedral angle. Despite maintaining the value that is more indicative of the "open conformation," it is likely this configuration is due to the slipped-tetrad conformational change that we have ascribed to the binding of a second K^+.

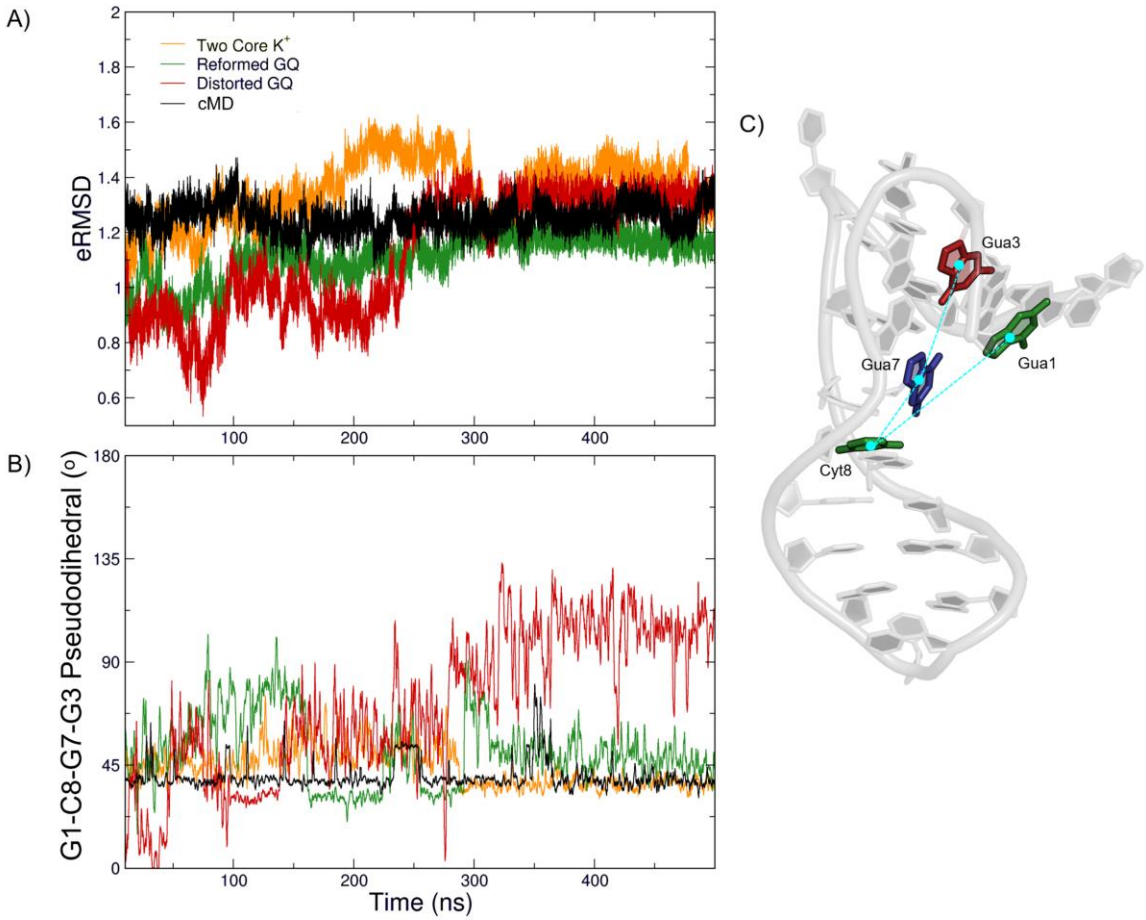

Figure 3.13 eRMSD (A) and Gua1-Cyt8-Gua7-Gua3 pseudodihedral value (B) time series for all GaMD replicate systems. (C) Bases used to calculate pseudodihedral angle are shown. Pseudodihedral values are shown as 1-ns running averages.

Ion dynamics in the GaMD simulation differed from those of the conventional MD simulations. The GaMD simulations confirmed that the *PIM-1:*Form 2 (Two K^+) GQ remained structurally stable, as noted above regarding the lack of structural perturbations in this system. Additionally, this system maintained both K^+ ions present in the core, and it is likely that this result corresponds to a more favorable conformation with the core cations providing additional stability. In both systems that began with one K^+ ion within the core, the *PIM*-1:Form 2 (cMD) and *PIM-1*:Form 2 (Reformed) GQ conformations, we observed similar core-sampling behaviors of the one cation as we saw in the conventional simulations. Additionally, this singular core K^+ ion appeared to prefer the position between tetrad 2 and the mixed tetrad in the GQ core, corresponding to our observations described in the QM analysis above. We observed that both replicates manifested an "opening" of the GQ core that resulted in the coordination of a second core K^+ ion from bulk solution at approximately 351 ns and 309 ns for the cMD and the Reformed GQ, respectively (Figure 3.14A,B).

In the *PIM-1:*Form 2 (Distorted) GQ system, we observed behaviors that largely resembled replicate 3 of the conventional MD simulations. As this simulation was initiated without K^+ ions present in the GQ core, we observed that the distorted nature of the core persisted as the swinging of the terminal Gua1:Cyt26 base pair resulted in a highly dynamic

system (Supporting Information, Figure S3.5). The GQ core remained devoid of K^+ ions and dynamic until approximately 408 ns, when a K^+ ion partitioned into the core from the bulk solution and sampled the GQ core in its entirety prior to diffusing back into the bulk solution at 417 ns. This GQ then returned to the "open" conformation as the unoccupied core allowed for increased structural dynamics until two K^+ ions from bulk solution were coordinated at 480 ns and 499 ns (Figure 3.14C). Two K^+ ions partitioned into the tetrad core in all GaMD replicates of the *PIM-1*:Form 2 GQ, indicating that such partitioning may be strongly favorable but may depend on the adoption of transient states that are more easily accessible in the GaMD simulations. These simulations further verified the findings from the QM ion interaction energy calculations, as all of the replicates had to undergo some degree of conformational rearrangement to allow ion partitioning through the mixed tetrad from the bulk solution, as indicated by the adoption of an "open" conformation, given that none of these replicates were initiated with a slipped tetrad configuration.

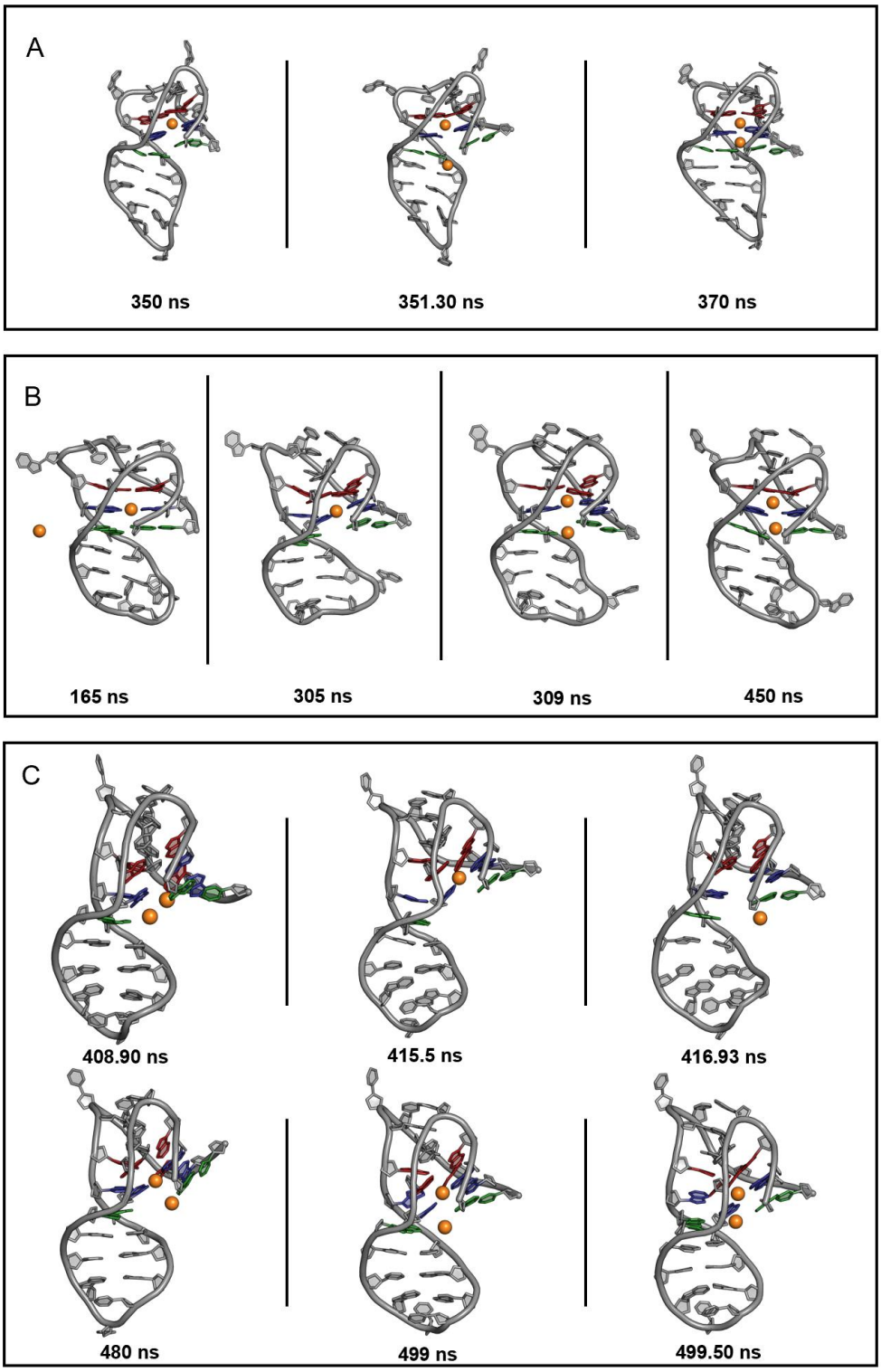

Figure 3.14 Ion dynamics for the *PIM-1*:Form 2 GQ GaMD replicates that partitioned K^+ ions into the tetrad core. Trajectory snapshots and time stamps are provided for (A) cMD, (B) Reformed, and (C) Distorted systems.

Defining the Conformational Free Energy Surface of the PIM-1:Form 2 GQ

To determine how the observations we made throughout the GaMD replicate simulations correspond to the thermodynamics associated with the various conformational states, we conducted a two-dimensional reweighting process to capture the free energy landscape associated with the various states. Reweighting is a process in which collective variables (CVs) that describe characteristic properties of a system are used to calculate the original free energy surface via unbiased probabilities. The CVs that we utilized for the reweighting process are the pseudo-dihedral and eRMSD values described in the previous section. These CVs were selected as they provide for the clearest differentiation between conformational states, while other possible CVs that we considered have more ambiguity across the conformational landscapes and are more complex for analysis purposes. For example, as we discussed in detail following the analysis of our conventional MD simulations, the angle of the bases within the mixed tetrad can shift as a result of two distinct conformations, thus leading to degeneracy in the analysis, and would likely not serve as an effective CV. Therefore, to maintain a high level of confidence in our reweighting assessment, we determined to use the stated CVs. Specifically, the pseudo-dihedral and eRMSD values directly correspond to the degree of "distortion" in the GQ associated with the "open" conformation we discussed above.

Across all GaMD replicates, we observed that the GQ coordinated two core stabilizing K^+ ions, confirming the hypothesis that the acquisition of ions from the aqueous solvent likely reflects a favorable behavior of the system and thus a low energy state or an intermediate that leads to a low free energy state. The enhanced sampling of these systems resulted in the coordination and preference for two K^+ ions within the GQ core, which differed from the observed preference for one core cation that was observed across the conventional MD systems. These observations are also aligned with our findings from the ion interaction energy calculations. When the GQ can access rare states, the energy barriers shift and the partitioning of ions into the GQ core can occur along alternate routes beyond just the tetrad axis.

Following our observations and descriptions of the GaMD replicates, we visualized the free energy landscape of replicates that behaved similarly, to obtain values from different sampled regions of the surface. Thus, we concatenated and visualized the *PIM-1*:Form 2 cMD, Two K^+, and Reformed GaMD replicates together on one surface, while analyzing the more dynamic, Distorted replicate independently. By doing so, we obtained several local minima for the *PIM*-1:Form 2 GQ as a whole. We observed enriched conformational sampling in the *PIM-1:*Form 2 (Distorted) replicate, as expected due to the large observable structural dynamics of this system. In terms of favorable states, we can confirm that in this replicate with the largest conformational ensemble, the minimum exists at an eRMSD of 0.50 and a pseudodihedral angle of 40°, which corresponds to an "open" and distorted GQ conformation described in detail above (Figure 3.15B). Additionally, when analyzing the less dynamic replicates together, the system displayed a global minimum at an eRMSD of approximately 0.75 and a pseudodihedral angle of 90°. These geometric properties correspond to a representation from the Reformed replicate in which a single core K^+ ion interacts with the mixed tetrad, resulting in the adoption of the slipped tetrad conformation (Figure 3.15A). Additionally, a local minimum was sampled in the Two K^+ system, indicating that this is also a highly favorable conformation for this *PIM-1:*Form 2 GQ.

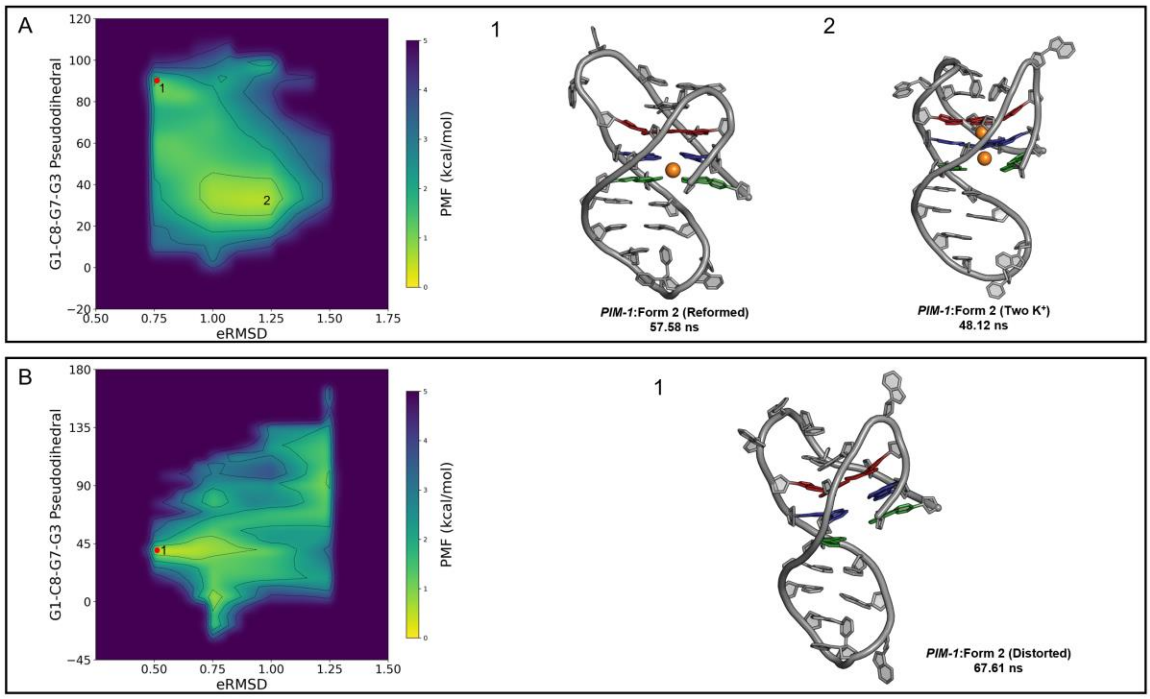

Figure 3.15 Two-dimensional reweighted free energy surface for the *PIM-1*:Form 2 GQ GaMD runs (A) concatenated replicates (cMD, Two K^+, and Reformed) and (B) *PIM-1:*Form 2 (Distorted) GQ. Global and local minima are marked and corresponding trajectory snapshots are provided with time stamps.

As shown in the free energy surfaces from the GaMD simulations, only low free energy barriers (~2 kcal mol^{-1}) exist between distinct states of the *PIM-1*:Form 2 GQ. These findings, in addition to our observations from the conventional MD simulations, provide insights into the plasticity of this GQ. We observed that this GQ sampled many distinct states and we can now attribute this level of structural plasticity to the low free energy barriers between these states. Additionally, this finding may further rationalize the preference for Form 1 over Form 2 of the *PIM-1* GQ, as there is likely some degree of unfavorability associated with the highly dynamic Form 2 GQ.

3.3.3 Mixed Solvent Simulations of *PIM*-1 GQs with SILCS

As discussed above, previous efforts to target GQs have proven to be largely unsuccessful due to limited selectivity by the ligands. GQs are commonly targeted at solvent-accessible tetrad faces, and if no other aspects of the structure are targeted to confer specificity, the compounds will likely bind to many GQs. Here, we performed SILCS simulations of the two *PIM-1* GQ conformations to determine if compounds might selectively bind to one or the other. As discussed above, the *PIM-1* gene encodes a serine-threonine kinase that is commonly implicated in cancers and can result in pro-tumorigenic downstream effects. It is hypothesized that targeting these GQs with stabilizing ligands can serve as an alternative chemotherapeutic approach by inhibiting gene expression at the transcriptional level. It is important to note that Forms 1 and 2 of the *PIM-1* GQ coexist at equilibrium, and well-informed drug design efforts are necessary to determine how to preferentially target the desired GQ given which form is present. Therefore, this study

serves as a preliminary investigation to gain insights into advantageous function group patterns that can better inform future drug-design efforts by identifying structural aspects of the GQs that may be exploited in the drug design process. Here, we applied both additive and polarizable SILCS workflows to the *PIM-1* GQs to determine the role of electronic polarization in the sampling patterns obtained and to identify potential ligand binding regions on the surfaces of these GQs.

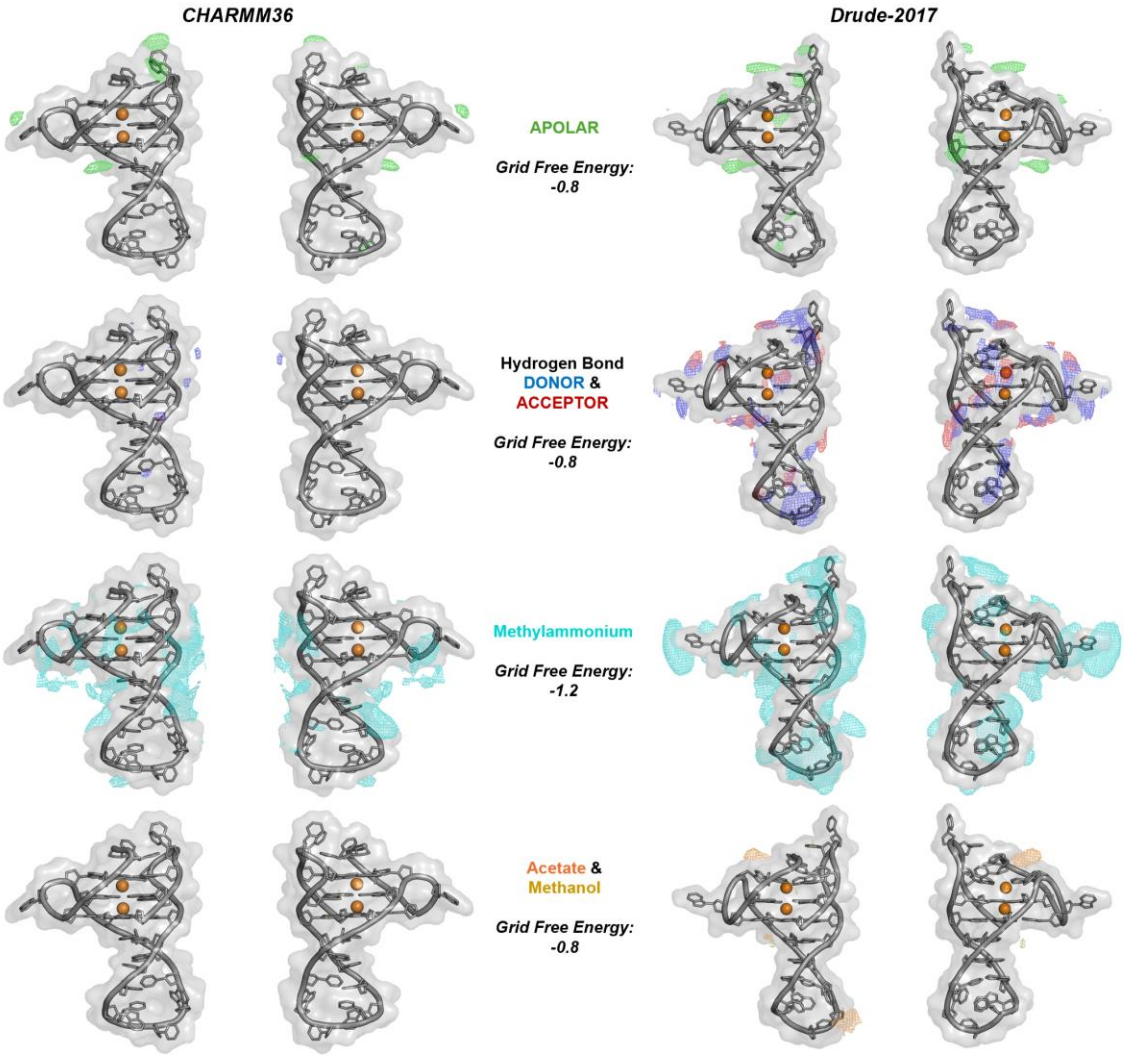

Figure 3.16 Comparison of additive and polarizable SILCS FragMaps for *PIM-1:*Form 1 GQ. The maps were generated from specific atoms of the SILCS solutes. Apolar maps represent benzene (C), propane (C), and imidazole (HC); hydrogen bond donor maps include formamide (N) and imidazole (NH); hydrogen bond acceptor maps include formamide (O), dimethylether (O), and imidazole (N). The GFE values used as a threshold to render the maps are provided.

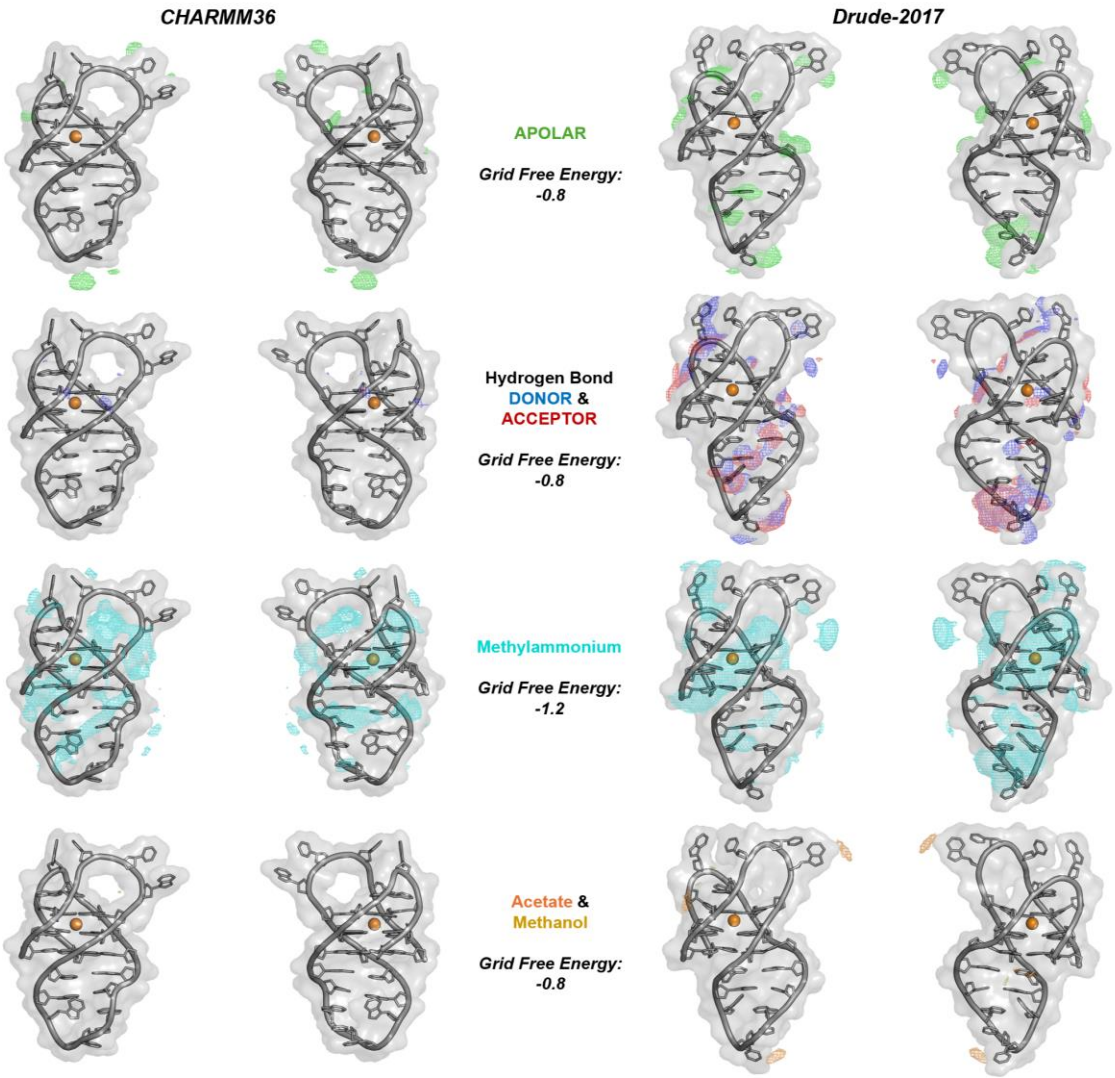

Figure 3.17 Comparison of additive and polarizable SILCS FragMaps for the *PIM-1:*Form 2 GQ. The maps were generated from specific atoms of the SILCS solutes. Apolar maps represent benzene (C), propane (C), and imidazole (HC); hydrogen bond donor maps include formamide (N) and imidazole (NH); hydrogen bond acceptor maps include formamide (O), dimethylether (O), and imidazole (N). The GFE values used as a threshold to render the maps are provided.

The SILCS simulations are typically analyzed in terms of FragMaps, which represent the probability that a given probe molecule will occupy a discretized region of space. As such, FragMaps serve as a representation for general features of the SILCS solutes, which are also common features of drug-like compounds. Therefore, the apolar map is generated from the coordinates of benzene (C), propane (C), and imidazole (HC). The hydrogen bond donor map includes contributions from: formamide (N) and imidazole (NH). The hydrogen bond acceptor map includes: formamide (O), dimethylether (O), and imidazole (N). These solute occupancies are then transformed into calculate a Grid Free Energy (GFE) score, which is a quantitative measure of affinity of the functional groups for that location. In comparing the additive and polarizable SILCS FragMaps, it is evident

that the Drude polarizable FF led to increased solute sampling around the GQ structures at equivalent GFE cutoff values. This outcome is particularly evident in the case of hydrogen bond donors and acceptors (Figures 3.16 and 3.17). As such, it appears that the inclusion of polarization by use of the Drude SILCS workflow allows for greater and more diverse probe sampling around the GQ surfaces for both *PIM-1* systems, similar to the *VEGF* systems.

While the polarizable SILCS FragMaps show considerably more solute accumulation around the GQ, the overall patterns of fragment affinity are similar across all systems. For example, we observed apolar and hydrogen bond donor and acceptor functional groups exhibit favorable sampling throughout most of the GQ structure, particularly in regions that are more solvent-accessible, including the duplex loop and the top loop bases above tetrad 1. Additionally, as hypothesized, the polyanionic backbone of the GQ attracted high concentrations of the positively charged methylammonium solute at strongly favorable GFE thresholds. This binding was nonspecific and led to the entire phosphodiester backbone being coated within the probe molecules. Additionally, we observed that the polarizable SILCS FragMaps exhibited sampling of the negatively charged acetate solute in the duplex and top loop regions of both the *PIM*-1 GQs. These anionic species are known to interact strongly with some nucleotides, such as guanine,[88] and Gua11 can be found in the outer-most location in the loop region. The most notable differences between the additive and polarizable FragMaps are focused surrounding the duplex region, which we hypothesize can serve as a potential region of specificity for ligand binding. The overall increase in solute sampling in this region in the polarizable systems is a vital consideration for drug development as designing molecules that bind in this region may result in improved pharmacophore models and therefore, drug candidates.

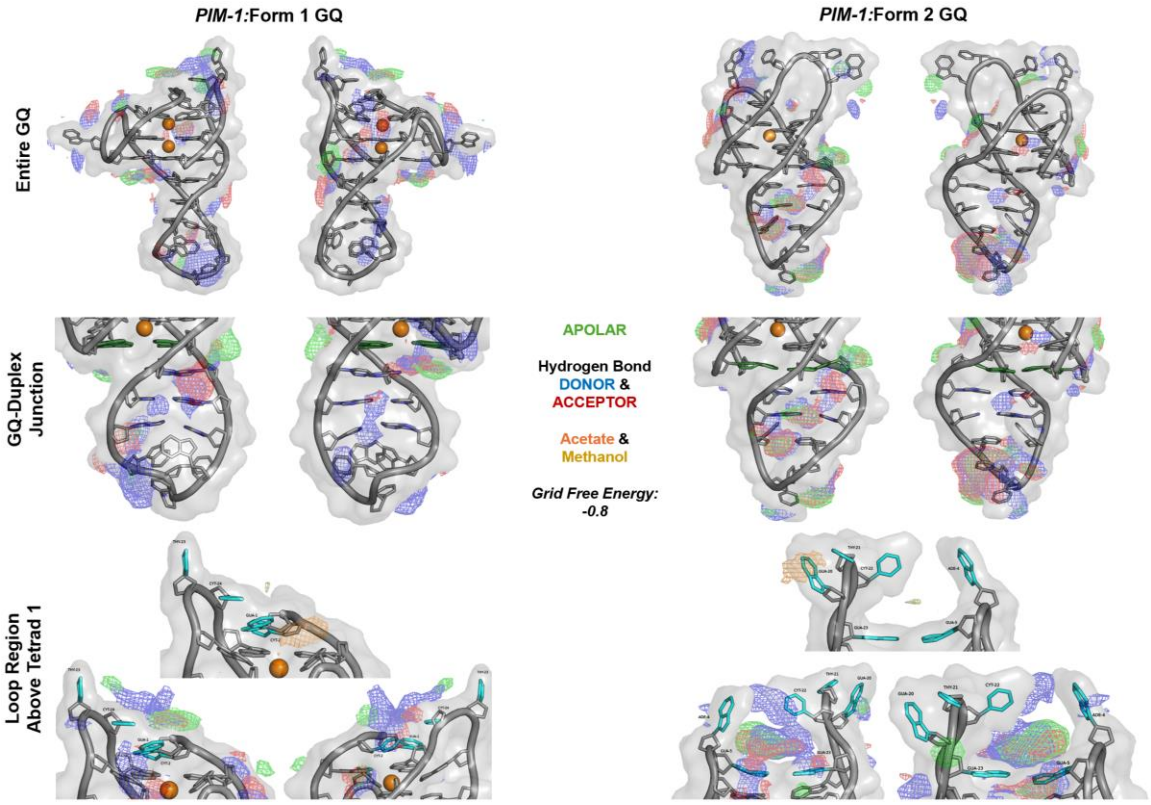

Figure 3.18 Comparison of common ligand-targeting regions from polarizable SILCS FragMaps of form 1 and form 2 of the PIM-1 promoter GQ. The maps were generated from specific atoms of the SILCS solutes. Apolar maps represent benzene (C), propane (C), and imidazole (HC); hydrogen bond donor maps include formamide (N) and imidazole (NH); hydrogen bond acceptor maps include formamide (O), dimethylether (O), and imidazole (N). The GFE values used as a threshold to render the maps are provided.

Next, we sought to explore the differences between the SILCS FragMaps of the two *PIM-1* GQs at accessible locations that may serve as regions of enhanced specificity for ligand-binding. As explained in earlier analysis of the SILCS FragMaps, due to the enhanced sampling observed in the polarizable SILCS systems, the analysis in these regions will rely on these FragMaps (Figure 3.18). We hypothesized that the presence of the mixed tetrad in the *PIM-1:*Form 2 GQ would yield different properties of putative ligands at the GQ-duplex junction relative to the *PIM-1:*Form 1 GQ, which has an all-guanine tetrad present at this location. We observed that additional apolar and hydrogen bond donor and acceptor occupancy was present at the GQ-duplex junction in the *PIM-1:*Form 1 GQ, in contrast to limited solute sampling in this region in the *PIM-1*:Form 2 GQ. Therefore, small molecules that have these chemical properties may be the basis for designing ligands to preferentially target the *PIM*-1 GQ at this junction.

When we analyzed the region above tetrad 1, which serves as a common ligand binding site in GQs, we observed similar solute sampling in both GQ systems. Interestingly, favorable sampling of the negatively charged acetate can be observed in this region in both GQ systems. Specifically, it is likely that the anionic sampling in this location is a result of the presence of Gua1 nucleotide within the loop. As previously described, it is known that acetate interacts with guanine nucleobases with a relatively high

affinity. Therefore, while this region may be useful to design ligands that bind selectively to both *PIM-1* GQs, it is unlikely the targeting in this region will provide the ability to design preferentially binding ligands that distinguish between the two forms.

3.4 Conclusions

In this study, we analyzed the topologically diverse *PIM-1* promoter GQs. These GQs exist in equilibrium with one another and gaining insight into how to preferentially target these structures serves as the first step in the drug design process. Here, we explored the electrostatics and structural dynamics of these GQs, primarily focusing on any changes to their properties that occurred as a result of the presence of the mixed cytosine-guanine tetrad in the *PIM-1*:Form 2 GQ. Overall, we observed that this mixed tetrad can adopt distinct conformations that modulate cation coordination in the GQ core, a feature that has proven to have implications on electrostatics and overall structural stability in the system. Across our conventional and enhanced sampling simulations, we identified a wide variety of conformations accessible to the *PIM-1*:Form 2 GQ, and confirmed this finding through the reconstruction of its free energy landscape. Low free energy barriers exist between these states, and therefore, these dynamic changes were accessible in both conventional MD and GaMD simulations. While our simulations of the *PIM-1*:Form 1 GQ confirmed that this structure behaves as a "prototypical" GQ system, it is likely that, when considering these findings collectively, these results may help rationalize the strong prevalence for the Form 1 GQ *in vitro* relative to Form 2, which is more polymorphic and therefore less likely to be adopted frequently.

As the primary goal of this initiative was to inform future drug design efforts, we also performed SILCS simulations to provide insights into the different structural motifs of the *PIM-1* GQs that may be exploited for preferential ligand design in the future. Our findings suggest that the GQ-duplex junction of the *PIM-1:*Form 1 GQ, the biologically dominant form, may be a suitable structural motif for the design of new small molecules. Additionally, this study served as an initial comparison between the polarizable and non-polarizable SILCS workflows in GQ drug design, and we found enhanced solute sampling in the polarizable systems, especially when capturing sampling of the negatively charged acetate solute. Thus, future efforts should be dedicated to confirming whether this improved sampling in the Drude SILCS simulations result in more detailed pharmacophore models, which in turn may improve the selectivity and efficacy of future drug candidates.

3.5 Acknowledgements

Computing time and resources were provided by Virginia Tech Advanced Research Computing and Expanse Cluster at the San Diego Supercomputer Center (SDSC) through allocation BIO230117 from the Advanced Cyberinfrastructure Coordination Ecosystem: Services & Support (ACCESS) program. which is supported by National Science Foundation grants #2138259, #2138286, #2138307, #2137603, and #2138296. This work was supported by the NIH (grant R35GM133754 to JAL) and USDA-NIFA (project VA-160092) and the Thomas F. and Kate Miller Jeffress Memorial Trust.

3.6 Supporting Information

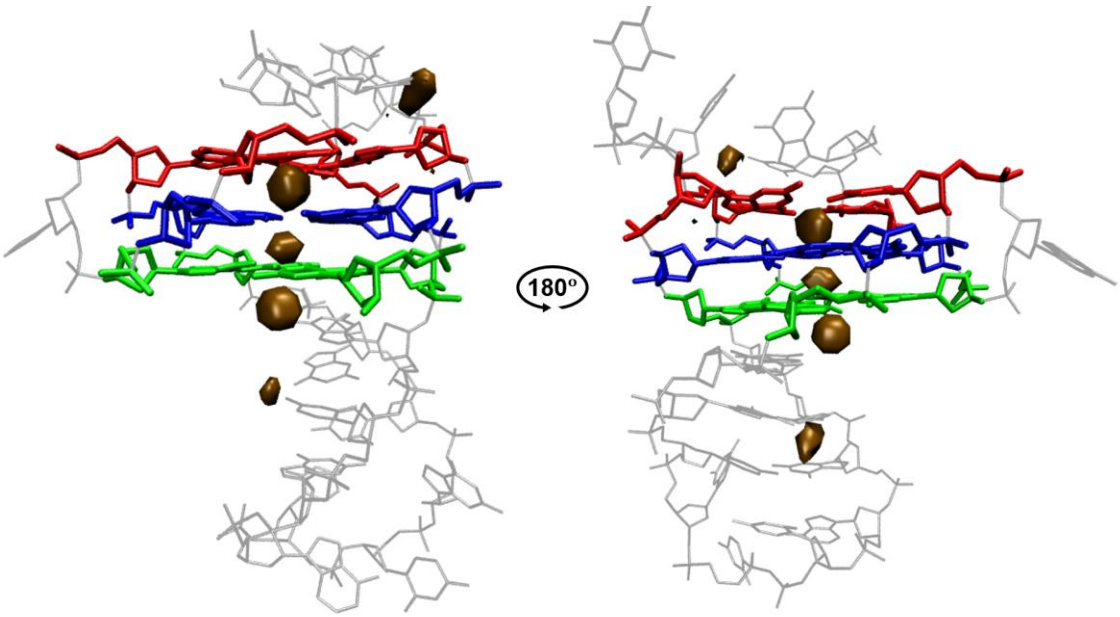

Figure S3.1 Ion Occupancy Maps for PIM-1:Form 1 GQ. Brown regions indicate positions occupied by K^+ ions for at least 1% of total simulation time across all four replicates per system (4 μs total). Tetrads are colored (1 – red, 2 – blue, 3 – green).

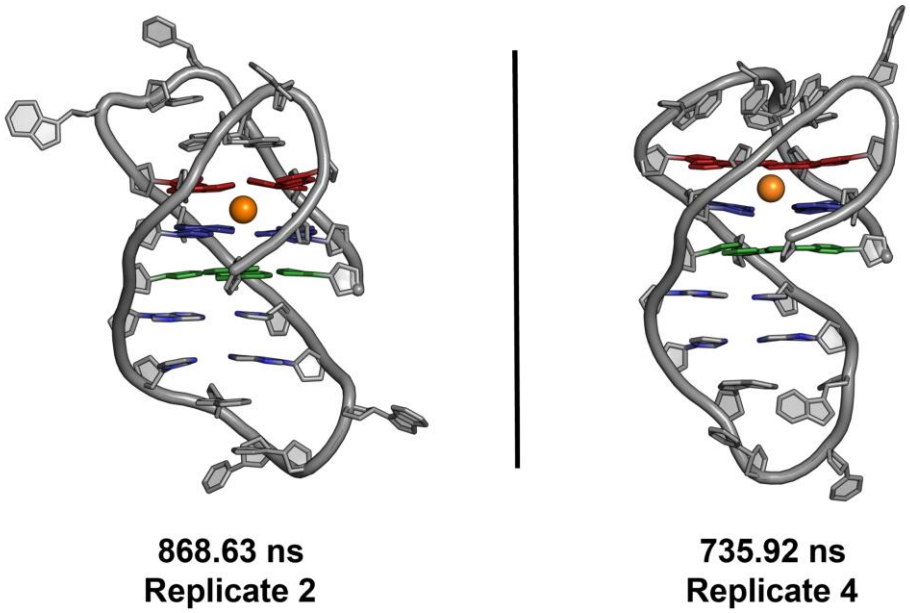

Figure S3.2. Trajectory snapshots for the *PIM-1*:Form 2 GQ replicates without further ion partitioning into the GQ core, these time stamps are representative of the comformation maintained throughout the entire simulation duration of 1 μs.

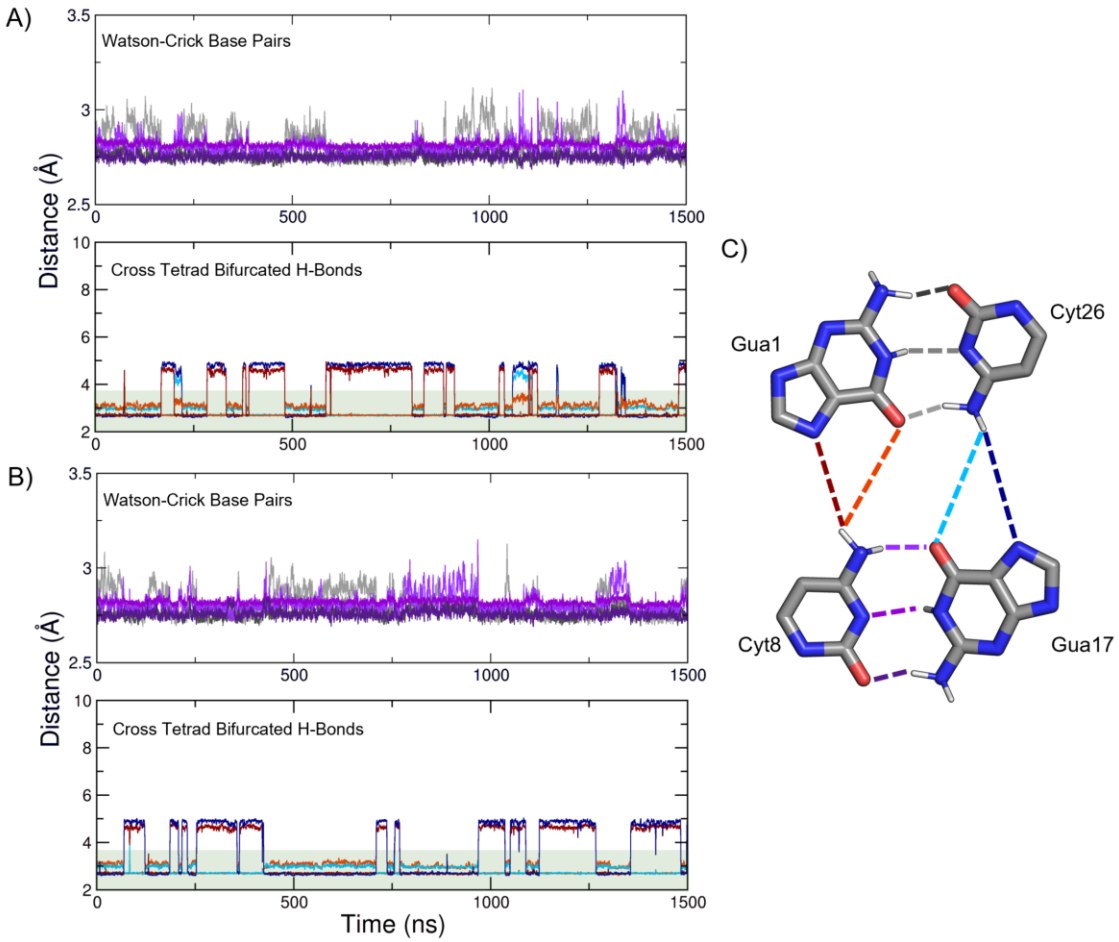

Figure S3.3 Mixed tetrad hydrogen bonding distances for (A) replicate 2 and (B) replicate 4. An approximate hydrogen bonding distance threshold of 3.5 Å is shown by the green shaded region. Watson-Crick base pair and bifurcated hydrogen bonds within the mixed tetrad are shown. (C) Watson-Crick Gua1:Cyt26 and Gua17:Cyt8 hydrogen bonds are shown in various shades of grey and purple, respectively. Bifurcated hydrogen bonds between Cyt8 and Cyt26 N4 and Gua O6 or N7 atoms are shown in shades of red and blue, respectively. The distance data series are shown as 100-point (1-ns) running averages.

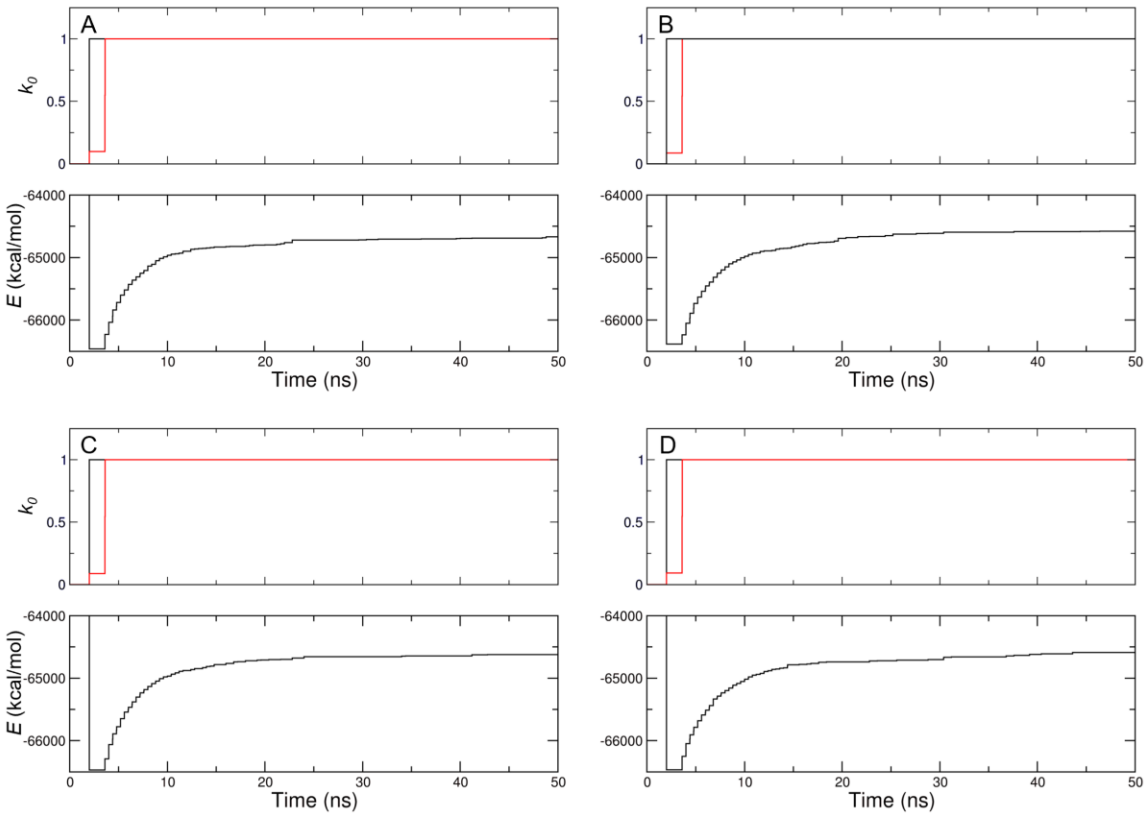

Figure S3.4 Assessing for equilibration statistics to confirm convergence in GaMD *PIM-1*:Form 2 GQ replicates (A) cMD, (B) Two K^+, (C) Reformed, and (D) Distorted. Top panels show k_0 for total (red) and dihedral (black) potential energies. Bottom panels show threshold energy, E, for the systems.

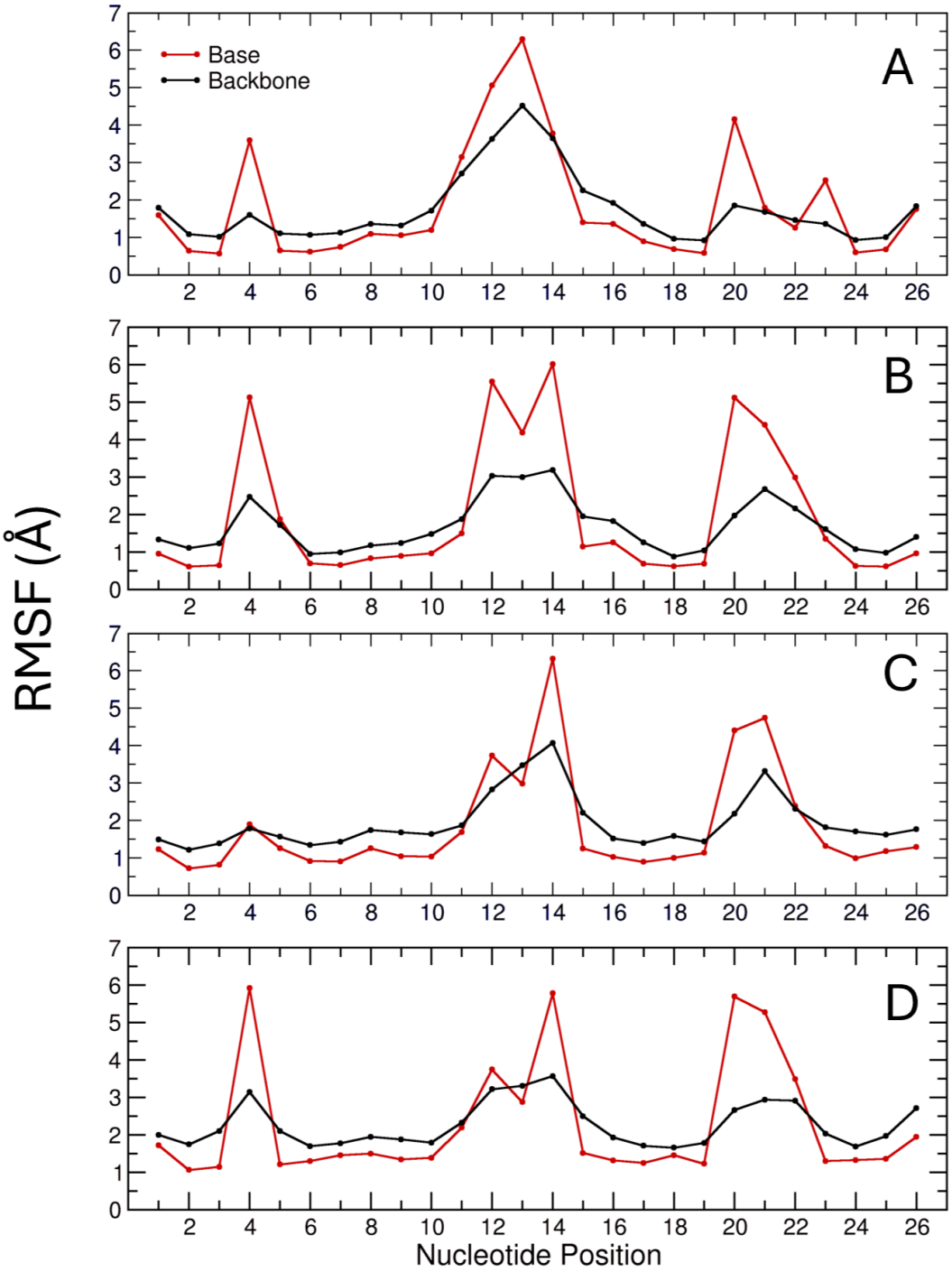

Figure S3.5 Base and backbone RMSF values for GaMD replicates of *PIM-1*:Form 1 (A) cMD, (B) Two K^+, (C) Reformed, and (D) Distorted GQ systems.

Chapter 4. Future Directions

This study provides a systematic investigation into the structural and electrostatic behaviors of two promoter GQ structures that regulate the transcription of oncogenes. It has been shown that both the *VEGF* and *PIM-1* GQs, when folded, result in the stalling of transcription of the oncogenes under their control. Therefore, interest in using these structures as targets to regulate gene expression on the genomic level has emerged as a method of chemotherapeutic treatment. This study provides foundational data that are necessary to inform future drug design efforts for these GQs.

Specifically, in our investigation of the *VEGF* GQ, we explored both the wild-type and the experimentally resolved mutant structures. The experimental structure serves as a model system for many GQ-targeting studies, therefore identifying any potential limitations that arise due to the thymine mutations is important context for further drug design efforts. Our investigation verified that, despite largely similar dynamics and behaviors across the two GQ systems, the *VEGF*-WT and *VEGF*-T12T13 GQs have some distinct properties that may impact future investigations. For example, we found that due to differences in base sizes, the accessibility of the commonly targeted region above tetrad 1 varies depending on the sequence. When the nucleotides at positions 12 and 13 are mutated to thymine, an additional binding site between Thy13 and Gua2 emerges. Additionally, we confirmed that this binding pocket may attract different functional groups through our SILCS analysis. Therefore, it is likely that the use of the mutant GQ in drug design investigations likely would result in compounds that are incompatible with the biologically relevant wild-type GQ. Thus, the next steps in this project would revolve around testing this hypothesis; by using our results from the SILCS analysis, we could conduct pharmacophore modeling and subsequent docking and derivatization of small molecules that complement the probe molecule features that we identified via SILCS. Additionally, once candidate small molecules demonstrate promise in our computational investigations, we would further validate our results through experimental techniques, such as circular dichroism (CD) and Förster resonance energy transfer melting (FRET-melting) experiments, to identify compounds with high affinity for the GQs.

Our analysis of the *PIM-1* GQs provided details of the effects of a mixed tetrad in the GQ structure. While many other GQs have recently been found to contain mixed tetrad configurations, this study serves as the first to provide insights into the implications of mixed tetrads on GQ structural dynamics and electrostatics. Therefore, the future direction of this project would primarily aim to further characterize this mixed tetrad configuration. Our simulations uncovered various conformation changes and smaller rearrangements that led to the coordination of K^+ ions from bulk solution. Therefore, further clustering analysis may prove beneficial to more rigorously characterize distinct conformations within this system and identify the most populated conformations of the *PIM-1* GQs. Additionally, as distinct ion coordination was observed across replicates, further simulations with various cationic species could provide insights on possible mixed tetrad contributions to ion preference of the GQs. We observed an enhanced structural stability that results from the presence of the mixed tetrad in both conventional and biased MD simulations. The stability of the overall GQ fold was retained despite the lack of coordinating cations in the GQ core. Therefore, further investigations into the stabilizing forces would prove to be beneficial to understanding the biological role of the mixed tetrad configuration and how it may

influence the overall *PIM-1* GQ architecture. Similarly, while we captured the free energy landscapes of the *PIM-1:*Form 2 GQ from the enhanced GaMD simulations, further comparison of those generated from various other enhanced sampling techniques could be useful, and are now feasible given the identification of appropriate CVs in this initial investigation. In addition to the structural and electrostatic properties that we analyzed in this structure, we also conducted the initial fragment screening for future drug design studies. Therefore, as in the case of the *VEGF* GQ, these insights rely on the further assessment of pharmacophore models and screens, as well as the experimental validation of our computational drug design findings.

(1) Burge, S.; Parkinson, G. N.; Hazel, P.; Todd, A. K.; Neidle, S. Quadruplex DNA : Sequence , Topology and Structure. **2006**, *34* (19), 5402–5415. https://doi.org/10.1093/nar/gkl655.
(2) Huppert, J. L.; Balasubramanian, S. G-Quadruplexes in Promoters throughout the Human Genome. **2007**, *35* (2), 406–413. https://doi.org/10.1093/nar/gkl1057.
(3) Lim, K. W.; Alberti, P.; Guédin, A.; Lacroix, L.; Riou, J.-F.; Royle, N. J.; Mergny, J.-L.; Phan, A. T. Sequence Variant (CTAGGG)n in the Human Telomere Favors a G-Quadruplex Structure Containing a G·C·G·C Tetrad. *Nucleic Acids Res* **2009**, *37* (18), 6239–6248. https://doi.org/10.1093/nar/gkp630.
(4) Yoshida, W.; Saikyo, H.; Nakabayashi, K.; Yoshioka, H.; Bay, D. H.; Iida, K.; Kawai, T.; Hata, K.; Ikebukuro, K.; Nagasawa, K.; Karube, I. Identification of G-Quadruplex Clusters by High-Throughput Sequencing of Whole-Genome Amplified Products with a G-Quadruplex Ligand. *Sci Rep* **2018**, *8* (1), 3116. https://doi.org/10.1038/s41598-018-21514-7.
(5) Rawal, P.; Kummarasetti, V. B. R.; Ravindran, J.; Kumar, N.; Halder, K.; Sharma, R.; Mukerji, M.; Das, S. K.; Chowdhury, S. Genome-Wide Prediction of G4 DNA as Regulatory Motifs: Role in Escherichia Coli Global Regulation. *Genome Res* **2006**, *16* (5), 644–655. https://doi.org/10.1101/gr.4508806.
(6) Verma, A.; Halder, K.; Halder, R.; Yadav, V. K.; Rawal, P.; Thakur, R. K.; Mohd, F.; Sharma, A.; Chowdhury, S. Genome-Wide Computational and Expression Analyses Reveal G-Quadruplex DNA Motifs as Conserved Cis-Regulatory Elements in Human and Related Species. *J Med Chem* **2008**, *51* (18), 5641–5649. https://doi.org/10.1021/jm800448a.
(7) Bochman, M. L.; Paeschke, K.; Zakian, V. A. DNA Secondary Structures : Stability and Function of G - Quadruplex Structures. *Nature Reviews Genetics* **2012**, *13* (November). https://doi.org/10.1038/nrg3296.
(8) Kettani, A.; Kumar, R. A.; Patel, D. J. Solution Structure of a DNA Quadruplex Containing the Fragile X Syndrome Triplet Repeat. *J Mol Biol* **1995**, *254* (4), 638–656. https://doi.org/10.1006/jmbi.1995.0644.
(9) Watson, J. D.; Crick, F. H. C. Molecular Structure of Nucleic Acids: A Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171* (4356), 737–738. https://doi.org/10.1038/171737a0.
(10) Gellert, M.; Lipsett, M. N.; Davies, D. R. HELIX FORMATION BY GUANYLIC ACID. *Proc Natl Acad Sci U S A* **1962**, *48* (12), 2013–2018.
(11) Pinnavaia, T. J.; Miles, H. T.; Becker, E. D. Self-Assembled 5'-Guanosine Monophosphate, Nuclear Magnetic Resonance Evidence for a Regular, Ordered Structure and Slow Chemical Exchange. *J. Am. Chem. Soc.* **1975**, *97* (24), 7198–7200. https://doi.org/10.1021/ja00857a059.
(12) Wang, Y.; Patel, D. J. Solution Structure of a Parallel-Stranded G-Quadruplex DNA. *Journal of Molecular Biology* **1993**, *234* (4), 1171–1183. https://doi.org/10.1006/jmbi.1993.1668.
(13) Špačková, N.; Berger, I.; Šponer, J. Nanosecond Molecular Dynamics Simulations of Parallel and Antiparallel Guanine Quadruplex DNA Molecules. *J. Am. Chem. Soc.* **1999**, *121* (23), 5519–5534. https://doi.org/10.1021/ja984449s.

(14) Varshney, D.; Spiegel, J.; Zyner, K.; Tannahill, D.; Balasubramanian, S. The Regulation and Functions of DNA and RNA G-Quadruplexes. *Nat Rev Mol Cell Biol* **2020**, *21* (8), 459–474. https://doi.org/10.1038/s41580-020-0236-x.
(15) Mukundan, V. T.; Phan, A. T. Bulges in G-Quadruplexes: Broadening the Definition of G-Quadruplex-Forming Sequences. *J. Am. Chem. Soc.* **2013**, *135* (13), 5017–5028. https://doi.org/10.1021/ja310251r.
(16) Lim, K. W.; Amrane, S.; Bouaziz, S.; Xu, W.; Mu, Y.; Patel, D. J.; Luu, K. N.; Phan, A. T. Structure of the Human Telomere in K+ Solution: A Stable Basket-Type G-Quadruplex with Only Two G-Tetrad Layers. *J. Am. Chem. Soc.* **2009**, *131* (12), 4301–4309. https://doi.org/10.1021/ja807503g.
(17) Liu, H.; Wang, R.; Yu, X.; Shen, F.; Lan, W.; Haruehanroengra, P.; Yao, Q.; Zhang, J.; Chen, Y.; Li, S.; Wu, B.; Zheng, L.; Ma, J.; Lin, J.; Cao, C.; Li, J.; Sheng, J.; Gan, J. High-Resolution DNA Quadruplex Structure Containing All the A-, G-, C-, T-Tetrads. *Nucleic Acids Research* **2018**, *46* (21), 11627–11638. https://doi.org/10.1093/nar/gky902.
(18) Escaja, N.; Mir, B.; Garavís, M.; González, C. Non-G Base Tetrads. *Molecules* **2022**, *27* (16), 5287. https://doi.org/10.3390/molecules27165287.
(19) Bhattacharyya, D.; Mirihana Arachchilage, G.; Basu, S. Metal Cations in G-Quadruplex Folding and Stability. *Front. Chem.* **2016**, *4*. https://doi.org/10.3389/fchem.2016.00038.
(20) Gu, J.; Leszczynski, J.; Bansal, M. A New Insight into the Structure and Stability of Hoogsteen Hydrogen-Bonded G-Tetrad: An Ab Initio SCF Study. *Chemical Physics Letters* **1999**, *311* (3), 209–214. https://doi.org/10.1016/S0009-2614(99)00821-0.
(21) Salsbury, A. M.; Lemkul, J. A. Cation Competition and Recruitment around the C-Kit1 G-Quadruplex Using Polarizable Simulations. *Biophysical Journal* **2021**, *120* (11), 2249. https://doi.org/10.1016/j.bpj.2021.03.022.
(22) Salsbury, A. M.; Dean, T. J.; Lemkul, J. A. Polarizable Molecular Dynamics Simulations of Two C-Kit Oncogene Promoter G-Quadruplexes: Effect of Primary and Secondary Structure on Loop and Ion Sampling. *J. Chem. Theory Comput.* **2020**, *16* (5), 3430–3444. https://doi.org/10.1021/acs.jctc.0c00191.
(23) Mourik, T. van; Dingley, A. J. Chemistry – A European Journal - Chemistry Europe - Wiley Online Library. *Chemistry Europe*.
(24) Chen, Y.; Yang, D. Sequence, Stability, Structure of G-Quadruplexes and Their Drug Interactions. *Current protocols in nucleic acid chemistry / edited by Serge L. Beaucage ... [et al.]* **2012**, *CHAPTER*, Unit17.5. https://doi.org/10.1002/0471142700.nc1705s50.
(25) Lodish, H.; Berk, A.; Zipursky, S. L.; Matsudaira, P.; Baltimore, D.; Darnell, J. *Molecular Cell Biology*; W.H. Freeman, 2000.
(26) Kim, N. The Interplay between G-Quadruplex and Transcription. *Curr Med Chem* **2019**, *26* (16), 2898–2917. https://doi.org/10.2174/0929867325666171229132619.
(27) Belotserkovskii, B. P.; Liu, R.; Tornaletti, S.; Krasilnikova, M. M.; Mirkin, S. M.; Hanawalt, P. C. Mechanisms and Implications of Transcription Blockage by Guanine-Rich DNA Sequences. *Proc Natl Acad Sci U S A* **2010**, *107* (29), 12816–12821. https://doi.org/10.1073/pnas.1007580107.

(28) Raiber, E.-A.; Kranaster, R.; Lam, E.; Nikan, M.; Balasubramanian, S. A Non-Canonical DNA Structure Is a Binding Motif for the Transcription Factor SP1 in Vitro. *Nucleic Acids Res* **2012**, *40* (4), 1499–1508. https://doi.org/10.1093/nar/gkr882.
(29) Armas, P.; David, A.; Calcaterra, N. B. Transcriptional Control by G-Quadruplexes: In Vivo Roles and Perspectives for Specific Intervention. *Transcription* **2017**, *8* (1), 21–25. https://doi.org/10.1080/21541264.2016.1243505.
(30) Bielskutė, S.; Plavec, J.; Podbevšek, P. Oxidative Lesions Modulate G-Quadruplex Stability and Structure in the Human BCL2 Promoter. *Nucleic Acids Research* **2021**, *49* (4), 2346–2356. https://doi.org/10.1093/nar/gkab057.
(31) Agrawal, P.; Hatzakis, E.; Guo, K.; Carver, M.; Yang, D. Solution Structure of the Major G-Quadruplex Formed in the Human VEGF Promoter in K+: Insights into Loop Interactions of the Parallel G-Quadruplexes. *Nucleic Acids Research* **2013**, *41* (22), 10584–10592. https://doi.org/10.1093/nar/gkt784.
(32) Tan, D. J. Y.; Winnerdy, F. R.; Lim, K. W.; Phan, A. T. Coexistence of Two Quadruplex–Duplex Hybrids in the PIM1 Gene. *Nucleic Acids Research* **2020**, *48* (19), 11162–11171. https://doi.org/10.1093/nar/gkaa752.
(33) Ambrus, A.; Chen, D.; Dai, J.; Jones, R. A.; Yang, D. Solution Structure of the Biologically Relevant G-Quadruplex Element in the Human c-MYC Promoter. Implications for G-Quadruplex Stabilization. *Biochemistry* **2005**, *44* (6), 2048–2058. https://doi.org/10.1021/bi048242p.
(34) Marquevielle, J.; Robert, C.; Lagrabette, O.; Wahid, M.; Bourdoncle, A.; Xodo, L. E.; Mergny, J.-L.; Salgado, G. F. Structure of Two G-Quadruplexes in Equilibrium in the KRAS Promoter. *Nucleic Acids Research* **2020**, *48* (16), 9336–9345. https://doi.org/10.1093/nar/gkaa387.
(35) Phan, A. T.; Kuryavyi, V.; Burge, S.; Neidle, S.; Patel, D. J. Structure of an Unprecedented G-Quadruplex Scaffold in the Human c-Kit Promoter. *J. Am. Chem. Soc.* **2007**, *129* (14), 4386–4392. https://doi.org/10.1021/ja068739h.
(36) Wang, Z.-F.; Li, M.-H.; Chu, I.-T.; Winnerdy, F. R.; Phan, A. T.; Chang, T.-C. Cytosine Epigenetic Modification Modulates the Formation of an Unprecedented G4 Structure in the WNT1 Promoter. *Nucleic Acids Research* **2020**, *48* (3), 1120–1130. https://doi.org/10.1093/nar/gkz1207.
(37) Drygin, D.; Siddiqui-Jain, A.; O'Brien, S.; Schwaebe, M.; Lin, A.; Bliesath, J.; Ho, C. B.; Proffitt, C.; Trent, K.; Whitten, J. P.; Lim, J. K. C.; Von Hoff, D.; Anderes, K.; Rice, W. G. Anticancer Activity of CX-3543: A Direct Inhibitor of rRNA Biogenesis. *Cancer Research* **2009**, *69* (19), 7653–7661. https://doi.org/10.1158/0008-5472.CAN-09-1304.
(38) Local, A.; Zhang, H.; Benbatoul, K. D.; Folger, P.; Sheng, X.; Tsai, C.-Y.; Howell, S. B.; Rice, W. G. APTO-253 Stabilizes G-Quadruplex DNA, Inhibits MYC Expression, and Induces DNA Damage in Acute Myeloid Leukemia Cells. *Molecular Cancer Therapeutics* **2018**, *17* (6), 1177–1186. https://doi.org/10.1158/1535-7163.MCT-17-1209.
(39) Bank, R. P. D. *RCSB PDB - 2N6C: Solution structure for quercetin complexed with c-myc G-quadruplex DNA*. https://www1.rcsb.org/structure/2N6C (accessed 2024-11-19).

(40) Islam, B.; Stadlbauer, P.; Gil-Ley, A.; Pérez-Hernández, G.; Haider, S.; Neidle, S.; Bussi, G.; Banas, P.; Otyepka, M.; Sponer, J. Exploring the Dynamics of Propeller Loops in Human Telomeric DNA Quadruplexes Using Atomistic Simulations. *J Chem Theory Comput* **2017**, *13* (6), 2458–2480. https://doi.org/10.1021/acs.jctc.7b00226.
(41) Gkionis, K.; Kruse, H.; Platts, J. A.; Mládek, A.; Koča, J.; Šponer, J. Ion Binding to Quadruplex DNA Stems. Comparison of MM and QM Descriptions Reveals Sizable Polarization Effects Not Included in Contemporary Simulations. *J. Chem. Theory Comput.* **2014**, *10* (3), 1326–1340. https://doi.org/10.1021/ct4009969.
(42) Šponer, J.; Bussi, G.; Stadlbauer, P.; Kührová, P.; Banáš, P.; Islam, B.; Haider, S.; Neidle, S.; Otyepka, M. Folding of Guanine Quadruplex Molecules–Funnel-like Mechanism or Kinetic Partitioning? An Overview from MD Simulation Studies. *Biochimica et Biophysica Acta (BBA) - General Subjects* **2017**, *1861* (5, Part B), 1246–1263. https://doi.org/10.1016/j.bbagen.2016.12.008.
(43) Song, J.; Ji, C.; Zhang, J. Z. H. The Critical Effect of Polarization on the Dynamical Structure of Guanine Quadruplex DNA. *Phys. Chem. Chem. Phys.* **2013**, *15* (11), 3846–3854. https://doi.org/10.1039/C2CP44100D.
(44) Lemkul, J. A.; Mackerell, A. D. Polarizable Force Field for DNA Based on the Classical Drude Oscillator: I. Re Fi Nement Using Quantum Mechanical Base Stacking and Conformational Energetics. **2017**, No. 1. https://doi.org/10.1021/acs.jctc.7b00067.
(45) Lemkul, J. A.; MacKerell, A. D. Jr. Polarizable Force Field for DNA Based on the Classical Drude Oscillator: II. Microsecond Molecular Dynamics Simulations of Duplex DNA. *J. Chem. Theory Comput.* **2017**, *13* (5), 2072–2085. https://doi.org/10.1021/acs.jctc.7b00068.
(46) Lemkul, J. A.; MacKerell, A. D. Polarizable Force Field for RNA Based on the Classical Drude Oscillator. *J. Comput. Chem.* **2018**, *39*, 2624–2646.
(47) Zhang, C.; Lu, C.; Jing, Z.; Wu, C.; Piquemal, J.-P.; Ponder, J. W.; Ren, P. AMOEBA Polarizable Atomic Multipole Force Field for Nucleic Acids. *J. Chem. Theory Comput.* **2018**, *14* (4), 2084–2108. https://doi.org/10.1021/acs.jctc.7b01169.
(48) Salsbury, A. M.; Lemkul, J. A. Molecular Dynamics Simulations of the C-Kit1 Promoter G-Quadruplex: Importance of Electronic Polarization on Stability and Cooperative Ion Binding. *J. Phys. Chem. B* **2019**, *123* (1), 148–159. https://doi.org/10.1021/acs.jpcb.8b11026.
(49) Lemkul, J. A.; Huang, J.; Roux, B.; MacKerell, A. D. An Empirical Polarizable Force Field Based on the Classical Drude Oscillator Model: Development History and Recent Applications. *Chem. Rev.* **2016**, *116* (9), 4983–5013. https://doi.org/10.1021/acs.chemrev.5b00505.
(50) Thole, B. T. Molecular Polarizabilities Calculated with a Modified Dipole Interaction. *Chemical Physics* **1981**, *59*, 341–350. https://doi.org/10.1016/0301-0104(81)85176-2.
(51) Harder, E.; Anisimov, V. M.; Vorobyov, I. V.; Lopes, P. E. M.; Noskov, S. Y.; MacKerell, A. D.; Roux, B. Atomic Level Anisotropy in the Electrostatic Modeling of Lone Pairs for a Polarizable Force Field Based on the Classical Drude Oscillator. *J Chem Theory Comput* **2006**, *2* (6), 1587–1597. https://doi.org/10.1021/ct600180x.

(52) Lemkul, J. A. Same Fold, Different Properties: Polarizable Molecular Dynamics Simulations of Telomeric and TERRA G-Quadruplexes. *Nucleic Acids Research* **2020**, *48* (2), 561–575. https://doi.org/10.1093/nar/gkz1154.
(53) Michel, H. M.; Lemkul, J. A. Base Pair Dynamics, Electrostatics, and Thermodynamics at the LTR-III Quadruplex:Duplex Junction. *Biophysical Journal* **2024**, *123* (9), 1129–1138. https://doi.org/10.1016/j.bpj.2024.03.042.
(54) Sertkaya, A.; Beleche, T.; Jessup, A.; Sommers, B. D. Costs of Drug Development and Research and Development Intensity in the US, 2000-2018. *JAMA Network Open* **2024**, *7* (6), e2415445. https://doi.org/10.1001/jamanetworkopen.2024.15445.
(55) Sabe, V. T.; Ntombela, T.; Jhamba, L. A.; Maguire, G. E. M.; Govender, T.; Naicker, T.; Kruger, H. G. Current Trends in Computer Aided Drug Design and a Highlight of Drugs Discovered via Computational Techniques: A Review. *European Journal of Medicinal Chemistry* **2021**, *224*, 113705. https://doi.org/10.1016/j.ejmech.2021.113705.
(56) Macalino, S. J. Y.; Gosu, V.; Hong, S.; Choi, S. Role of Computer-Aided Drug Design in Modern Drug Discovery. *Arch. Pharm. Res.* **2015**, *38* (9), 1686–1701. https://doi.org/10.1007/s12272-015-0640-5.
(57) Sadybekov, A. V.; Katritch, V. Computational Approaches Streamlining Drug Discovery. *Nature* **2023**, *616* (7958), 673–685. https://doi.org/10.1038/s41586-023-05905-z.
(58) Hassan Baig, M.; Ahmad, K.; Roy, S.; Mohammad Ashraf, J.; Adil, M.; Haris Siddiqui, M.; Khan, S.; Amjad Kamal, M.; Provazník, I.; Choi, I. Computer Aided Drug Design: Success and Limitations. *CPD* **2016**, *22* (5), 572–581. https://doi.org/10.2174/1381612822666151125000550.
(59) Childs-Disney, J. L.; Yang, X.; Gibaut, Q. M. R.; Tong, Y.; Batey, R. T.; Disney, M. D. Targeting RNA Structures with Small Molecules. *Nat Rev Drug Discov* **2022**, *21* (10), 736–762. https://doi.org/10.1038/s41573-022-00521-4.
(60) Velagapudi, S. P.; Gallo, S. M.; Disney, M. D. Sequence-Based Design of Bioactive Small Molecules That Target Precursor microRNAs. *Nat Chem Biol* **2014**, *10* (4), 291–297. https://doi.org/10.1038/nchembio.1452.
(61) Eberhardt, J.; Santos-Martins, D.; Tillack, A. F.; Forli, S. AutoDock Vina 1.2.0: New Docking Methods, Expanded Force Field, and Python Bindings. *J. Chem. Inf. Model.* **2021**, *61* (8), 3891–3898. https://doi.org/10.1021/acs.jcim.1c00203.
(62) Guvench, O.; Jr, A. D. M. Computational Fragment-Based Binding Site Identification by Ligand Competitive Saturation. *PLOS Computational Biology* **2009**, *5* (7), e1000435. https://doi.org/10.1371/journal.pcbi.1000435.
(63) Faller, C. E.; Raman, E. P.; MacKerell, A. D.; Guvench, O. Site Identification by Ligand Competitive Saturation (SILCS) Simulations for Fragment-Based Drug Design. *Methods Mol Biol* **2015**, *1289*, 75–87. https://doi.org/10.1007/978-1-4939-2486-8_7.
(64) Lind, C.; Pandey, P.; Pastor, R. W.; Alexander D MacKerell, J. Functional Group Distributions, Partition Coefficients, and Resistance Factors in Lipid Bilayers Using Site-Identification by Ligand Competitive Saturation (SILCS). *Journal of chemical theory and computation* **2021**, *17* (5), 3188. https://doi.org/10.1021/acs.jctc.1c00089.

(65) Kognole, A. A.; Hazel, A.; MacKerell, A. D. Jr. SILCS-RNA: Toward a Structure-Based Drug Design Approach for Targeting RNAs with Small Molecules. *J. Chem. Theory Comput.* **2022**, *18* (9), 5672–5691. https://doi.org/10.1021/acs.jctc.2c00381.
(66) Goel, H.; Yu, W.; Ustach, V. D.; Aytenfisu, A. H.; Sun, D.; Alexander D MacKerell, J. Impact of Electronic Polarizability on Protein-Functional Group Interactions. *Physical chemistry chemical physics : PCCP* **2020**, *22* (13), 6848. https://doi.org/10.1039/d0cp00088d.
(67) Lakkaraju, S. K.; Raman, E. P.; Yu, W.; MacKerell, A. D. Jr. Sampling of Organic Solutes in Aqueous and Heterogeneous Environments Using Oscillating Excess Chemical Potentials in Grand Canonical-like Monte Carlo-Molecular Dynamics Simulations. *J. Chem. Theory Comput.* **2014**, *10* (6), 2281–2290. https://doi.org/10.1021/ct500201y.
(68) Alexander D MacKerell, J.; Jo, S.; Lakkaraju, S. K.; Lind, C.; Yu, W. Identification and Characterization of Fragment Binding Sites for Allosteric Ligand Design Using the Site Identification by Ligand Competitive Saturation Hotspots Approach (SILCS-Hotspots). *Biochimica et biophysica acta. General subjects* **2020**, *1864* (4), 129519. https://doi.org/10.1016/j.bbagen.2020.129519.
(69) Yu, W.; Lakkaraju, S. K.; Raman, E. P.; MacKerell, A. D. Site-Identification by Ligand Competitive Saturation (SILCS) Assisted Pharmacophore Modeling. *J Comput Aided Mol Des* **2014**, *28* (5), 491–507. https://doi.org/10.1007/s10822-014-9728-0.
(70) Ho, Q. T.; Kuo, C. J. Vascular Endothelial Growth Factor: Biology and Therapeutic Applications. *Int J Biochem Cell Biol.* **2007**, *39*, 1349–1357. https://doi.org/10.1016/j.biocel.2007.04.010.Vascular.
(71) Ferrara, N.; Gerber, H.-P.; LeCouter, J. The Biology of VEGF and Its Receptors. *Nat Med* **2003**, *9* (6), 669–676. https://doi.org/10.1038/nm0603-669.
(72) Carmeliet, P. VEGF as a Key Mediator of Angiogenesis in Cancer. *Oncology* **2005**, *69*, 4–10.
(73) Tang, R. F.; Itakura, J.; Aikawa, T.; Matsuda, K.; Fujii, H.; Korc, M.; Matsumoto, Y. Overexpression of Lymphangiogenic Growth Factor VEGF-C in Human Pancreatic Cancer. *Pancreas* **2001**, *22* (3), 285.
(74) Wang, X.; Chen, X.; Fang, J.; Yang, C. Overexpression of Both VEGF-A and VEGF-C in Gastric Cancer Correlates with Prognosis, and Silencing of Both Is Effective to Inhibit Cancer Growth. *International Journal of Clinical and Experimental Pathology* **2013**, *6* (4), 586.
(75) Xia, H.; Shen, J.; Chen, S.; Huang, H.; Xu, Y.; Ma, H. Overexpression of VEGF-C Correlates with a Poor Prognosis in Esophageal Cancer Patients. *Cancer Biomarkers* **2016**, *17* (2), 165–170. https://doi.org/10.3233/CBM-160627.
(76) Baeriswyl, V.; Christofori, G. The Angiogenic Switch in Carcinogenesis. *Semin Cancer Biol* **2009**, *19* (5), 329–337. https://doi.org/10.1016/j.semcancer.2009.05.003.
(77) Bergers, G.; Benjamin, L. E. Tumorigenesis and the Angiogenic Switch. *Nat Rev Cancer* **2003**, *3* (6), 401–410. https://doi.org/10.1038/nrc1093.
(78) Guo, K.; Gokhale, V.; Hurley, L. H.; Sun, D. Intramolecularly Folded G-Quadruplex and i-Motif Structures in the Proximal Promoter of the Vascular

Endothelial Growth Factor Gene. *Nucleic Acids Res* **2008**, *36* (14), 4598–4608. https://doi.org/10.1093/nar/gkn380.
(79) Brooks, B. R.; Brooks, C.; MacKerell, A. D.; Nilsson, L. ; Petrella, R. J.; Roux, B.; Won, Y.; Archontis, G.; Bartels, C. ; B.; S.; Caflisch, A.; Caves, L.; Cui, Q.; Dinner, A. R.; Feig, M.; Fischer, S. ; Gao, J.; Hodoscek, M.; Im, W.; Kuczera, K.; Lazaridis, T.; Ma, J. ; Ovchinnikov, V.; Paci, E.; Pastor, R. W.; Post, C. B.; Pu, J. Z. ; Schaefer, M.; Tidor, B.; Venable, R. M.; Woodcock, H. L.; Wu, X. ;; Yang, W.; D. M, Y.; Karplus, M. CHARMM: Molecular Dynamics Simulation Package. *J. Comput. Chem.* **2009**, *30*, 1545–1614.
(80) Hart, K.; Foloppe, N.; Baker, C. M.; Denning, E. J.; Nilsson, L.; MacKerell, A. D. Optimization of the CHARMM Additive Force Field for DNA: Improved Treatment of the BI/BII Conformational Equilibrium. *Journal of Chemical Theory and Computation* **2012**, *8* (1), 348–362. https://doi.org/10.1021/ct200723y.
(81) Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J. ; T.; E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K. Scalable Molecular Dynamics with NAMD. *J. Comput. Chem.* **2005**, *26*, 1781–1802.
(82) Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R. Constant Pressure Molecular Dynamics Simulation: The Langevin Piston Method. *The Journal of Chemical Physics* **1995**, *103* (11), 4613–4621. https://doi.org/10.1063/1.470648.
(83) Darden, T.; York, D.; Pedersen, L. Particle Mesh Ewald: An N·log(N) Method for Ewald Sums in Large Systems. *The Journal of Chemical Physics* **1993**, *98* (12), 10089–10092. https://doi.org/10.1063/1.464397.
(84) Lamoureux, G.; Harder, E.; Vorobyov, I. V.; Roux, B. ;; MacKerell, A. D. A Polarizable Model of Water for Molecular Dynamics Simulations of Biomolecules. *Chem. Phys. Lett.* **2006**, *418*, 254–249.
(85) Eastman, P.; Swails, J.; Chodera, J. D.; McGibbon, R. T. ; Z.; Y.; Beauchamp, K. A.; Wang, L.-P.; Simmonett, A. C.; Harrigan, M.; P.; Stern, C. D. ; et al. OpenMM 7: Rapid Development of High Proformance Algorithms for Molecular Dynamics. *PLoS Comput. Biol.* **2017**, *13*, e1005659.
(86) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J Comput Chem* **2005**, *26* (16), 1701–1718. https://doi.org/10.1002/jcc.20291.
(87) Balasubramanian, S.; Neidle, S. G-Quadruplex Nucleic Acids as Therapeutic Targets. *Curr Opin Chem Biol* **2009**, *13* (3), 345–353. https://doi.org/10.1016/j.cbpa.2009.04.637.
(88) Lancelot, G.; Helene, C. Selective Recognition of Nucleic Acids by Proteins: The Specificity of Guanine Interaction with Carboxylate Ions. *Proceedings of the National Academy of Sciences of the United States of America* **1977**, *74* (11), 4872–4875.
(89) Cao, L.; Wang, F.; Li, S.; Wang, X.; Huang, D.; Jiang, R. PIM1 Kinase Promotes Cell Proliferation, Metastasis and Tumor Growth of Lung Adenocarcinoma by Potentiating the c-MET Signaling Pathway. *Cancer Lett* **2019**, *444*, 116–126. https://doi.org/10.1016/j.canlet.2018.12.015.
(90) Toth, R. K.; Warfel, N. A. Targeting PIM Kinases to Overcome Therapeutic Resistance in Cancer. *Molecular Cancer Therapeutics* **2021**, *20* (1), 3–10. https://doi.org/10.1158/1535-7163.MCT-20-0535.

(91) Yang, H.; He, K.; Dong, W.; Fang, J.; Zhong, S.; Tang, L.; Long, L. PIM-1 May Function as an Oncogene in Cervical Cancer via Activating the EGFR Signaling. *Int J Biol Markers* **2020**, *35* (3), 67–73. https://doi.org/10.1177/1724600820936295.
(92) Luszczak, S.; Kumar, C.; Sathyadevan, V. K.; Simpson, B. S.; Gately, K. A.; Whitaker, H. C.; Heavey, S. PIM Kinase Inhibition: Co-Targeted Therapeutic Approaches in Prostate Cancer. *Signal Transduct Target Ther* **2020**, *5*, 7. https://doi.org/10.1038/s41392-020-0109-y.
(93) Siu, A.; Virtanen, C.; Jongstra, J. PIM Kinase Isoform Specific Regulation of MIG6 Expression and EGFR Signaling in Prostate Cancer Cells. *Oncotarget* **2011**, *2* (12), 1134–1144.
(94) Brasó-Maristany, F.; Filosto, S.; Catchpole, S.; Marlow, R.; Quist, J.; Francesch-Domenech, E.; Plumb, D. A.; Zakka, L.; Gazinska, P.; Liccardi, G.; Meier, P.; Gris-Oliver, A.; Cheang, M. C. U.; Perdrix-Rosell, A.; Shafat, M.; Noël, E.; Patel, N.; McEachern, K.; Scaltriti, M.; Castel, P.; Noor, F.; Buus, R.; Mathew, S.; Watkins, J.; Serra, V.; Marra, P.; Grigoriadis, A.; Tutt, A. N. PIM1 Kinase Regulates Cell Death, Tumor Growth and Chemotherapy Response in Triple-Negative Breast Cancer. *Nat Med* **2016**, *22* (11), 1303–1313. https://doi.org/10.1038/nm.4198.
(95) Zhao, W.; Qiu, R.; Li, P.; Yang, J. PIM1: A Promising Target in Patients with Triple-Negative Breast Cancer. *Med Oncol* **2017**, *34* (8), 142. https://doi.org/10.1007/s12032-017-0998-y.
(96) Hubalek, M.; Czech, T.; Müller, H. Biological Subtypes of Triple-Negative Breast Cancer. *Breast Care* **2017**, *12* (1), 8–14. https://doi.org/10.1159/000455820.
(97) Magnuson, N. S.; Wang, Z.; Ding, G.; Reeves, R. Why Target PIM1 for Cancer Diagnosis and Treatment? *Future Oncol* **2010**, *6* (9), 1461–1478. https://doi.org/10.2217/fon.10.106.
(98) Zhang, X.; Song, M.; Kundu, J. K.; Lee, M.-H.; Liu, Z.-Z. PIM Kinase as an Executional Target in Cancer. *Journal of Cancer Prevention* **2018**, *23* (3), 109–116. https://doi.org/10.15430/JCP.2018.23.3.109.
(99) TURSYNBAY, Y.; ZHANG, J.; LI, Z.; TOKAY, T.; ZHUMADILOV, Z.; WU, D.; XIE, Y. Pim-1 Kinase as Cancer Drug Target: An Update. *Biomed Rep* **2016**, *4* (2), 140–146. https://doi.org/10.3892/br.2015.561.
(100) Aujla, M. Pim Kinase Inhibition and Chemoresistance. *Nat Rev Clin Oncol* **2010**, *7* (1), 3–3. https://doi.org/10.1038/nrclinonc.2009.198.
(101) Bellon, M.; Nicot, C. Targeting Pim Kinases in Hematological Cancers: Molecular and Clinical Review. *Molecular Cancer* **2023**, *22* (1), 18. https://doi.org/10.1186/s12943-023-01721-1.
(102) Cao, Y.; Yang, L.; Ding, P.; Li, W.; Pei, R. Ligand Selectivity by Inserting GCGC-Tetrads into G-Quadruplex Structures. *Chemistry – A European Journal* **2020**, *26* (64), 14730–14737. https://doi.org/10.1002/chem.202003004.
(103) Escaja, N.; Viladoms, J.; Garavís, M.; Villasante, A.; Pedroso, E.; González, C. A Minimal I-Motif Stabilized by Minor Groove G:T:G:T Tetrads. *Nucleic Acids Research* **2012**, *40* (22), 11737. https://doi.org/10.1093/nar/gks911.
(104) Ding, Y.; Xie, L.; Zhang, C.; Xu, W. Real-Space Evidence of the Formation of the GCGC Tetrad and Its Competition with the G-Quartet on the Au(111) Surface. *Chemical Communications* **2017**, *53* (71), 9846–9849. https://doi.org/10.1039/C7CC05548J.

(105) Pavc, D.; Wang, B.; Spindler, L.; Drevenšek-Olenik, I.; Plavec, J.; Šket, P. GC Ends Control Topology of DNA G-Quadruplexes and Their Cation-Dependent Assembly. *Nucleic Acids Research* **2020**, *48* (5), 2749–2761. https://doi.org/10.1093/nar/gkaa058.
(106) Szalewicz, K. Citation for: Symmetry-adapted Perturbation Theory of Intermolecular Forces. *Wiley Interdisciplinary Reviews*.
(107) Jeziorski, B.; Moszynski, R.; Szalewicz, K. Perturbation Theory Approach to Intermolecular Potential Energy Surfaces of van Der Waals Complexes. *Chem. Rev.* **1994**, *94* (7), 1887–1930. https://doi.org/10.1021/cr00031a008.
(108) Smith, D. G. A.; Burns, L. A.; Simmonett, A. C.; Parrish, R. M.; Schieber, M. C.; Galvelis, R.; Kraus, P.; Kruse, H.; Di Remigio, R.; Alenaizan, A.; James, A. M.; Lehtola, S.; Misiewicz, J. P.; Scheurer, M.; Shaw, R. A.; Schriber, J. B.; Xie, Y.; Glick, Z. L.; Sirianni, D. A.; O'Brien, J. S.; Waldrop, J. M.; Kumar, A.; Hohenstein, E. G.; Pritchard, B. P.; Brooks, B. R.; Schaefer, H. F., III; Sokolov, A. Yu.; Patkowski, K.; DePrince, A. E., III; Bozkaya, U.; King, R. A.; Evangelista, F. A.; Turney, J. M.; Crawford, T. D.; Sherrill, C. D. PSI4 1.4: Open-Source Software for High-Throughput Quantum Chemistry. *The Journal of Chemical Physics* **2020**, *152* (18), 184108. https://doi.org/10.1063/5.0006002.
(109) Pang, Y. T.; Miao, Y.; Wang, Y.; McCammon, J. A. Gaussian Accelerated Molecular Dynamics in NAMD. *J. Chem. Theory Comput.* **2017**, *13* (1), 9–19. https://doi.org/10.1021/acs.jctc.6b00931.
(110) Ratnasinghe, B. D.; Salsbury, A. M.; Lemkul, J. A. Ion Binding Properties and Dynamics of the Bcl-2 G-Quadruplex Using a Polarizable Force Field. *J. Chem. Inf. Model.* **2020**, *60* (12), 6476–6488. https://doi.org/10.1021/acs.jcim.0c01064.
(111) Reshetnikov, R. V.; Sponer, J.; Rassokhina, O. I.; Kopylov, A. M.; Tsvetkov, P. O.; Makarov, A. A.; Golovin, A. V. Cation Binding to 15-TBA Quadruplex DNA Is a Multiple-Pathway Cation-Dependent Process. *Nucleic Acids Res* **2011**, *39* (22), 9789–9802. https://doi.org/10.1093/nar/gkr639.
(112) Bottaro, S.; Di Palma, F.; Bussi, G. The Role of Nucleobase Interactions in RNA Structure and Dynamics. *Nucleic Acids Res* **2014**, *42* (21), 13306–13314. https://doi.org/10.1093/nar/gku972.