Studies Towards Thermodynamically Stable G-Quadruplexes Embedded in Canonical DNA Duplexes

A thesis presented in partial fulfilment of the requirements for the degree of

Doctor of Philosophy

in

Chemistry

At Massey University, Manawatū, New Zealand

Bruce Chilton

2023

Abstract:

DNA is the polymer responsible for the storage of genetic information and, ultimately, all processes that occur within the cell. Our understanding of DNA structure and function has developed considerably, but some areas are still unclear. In particular, a range of non-canonical DNA secondary structures such as G-quadruplexes (G4s), i-motifs and triplexes, have also been shown to form in genomic DNA sequences and these structures also appear to have a role in genome function.

Better understanding of the interactions of non-canonical secondary structures is hindered by their transient nature in the context of larger DNA structures, making it difficult to accurately study them using *in vitro* analytical techniques (e.g., NMR spectroscopy, X-ray crystallography, etc.). They are typically less thermodynamically stable than canonical DNA duplexes and are formed within the genome only in equilibrium with many secondary structures, typically favouring the canonical duplex. They can often only be formed reliably under specific conditions *in vitro* (e.g., single-stranded, low pH etc.). This thesis presents several strategies designed to stabilise non-canonical G4 secondary structures, which are of interest because they are often found in the promoter regions of oncogenes.

The most commonly used existing G4 stabilisation technique utilises small-molecule ligands which specifically bind to and stabilise G4 structures. This thesis includes an investigation into both a widely used G4-binding ligand and several newly developed ligands, but their potential to block binding sites and disrupt G4 topology makes them less suitable for our intended applications.

Hydrophobic modifications can encourage aggregation of DNA strands and therefore increase stability of secondary structures. Hydrophobic phosphate modifications in G4s proved effective at disrupting duplexes and stabilising G4s but was limited by coupling efficiency of the modification and resulting difficulties with purification. Intentional mismatches in the G4-forming sequence were introduced by inverting sequence direction or incorporating α-anomers of nucleotides. This strategy was able to completely disrupt duplex formation while preserving G4 structures, but modification sites have to be carefully considered

to avoid significant changes in G4 topology. Internal cross-links were incorporated into DNA using modified nucleotides designed for copper(I)-catalysed azide-alkyne cycloaddition. These cross-links prevent G4 structures from unfolding, but the location of these cross-links must also be carefully considered to prevent disruption of the native G4 topology and blocking protein binding sites. All three of these methods present potential routes for stabilising G4s within larger DNA structures. Furthermore, all three modifications could potentially be expanded to stabilise other non-canonical structures, such as i-motifs or triplexes.

Abbreviations:

A:	Adenine
ACN:	Acetonitrile
BLI:	Biolayer interferometry
C:	Cytosine
CD:	Circular dichroism
CD:	Chromodomain (of HP1α)
CE:	2-Cyanoethyl (protecting group)
CPG:	Controlled pore glass
CSD:	Chromoshadow domain (of HP1α)
CuAAC:	Copper(I)-catalysed azide-alkyne cycloaddition
DCM:	Dichloromethane
dG:	2′-Deoxyguanosine
DNA:	2′-Deoxyribonucleic acid
DNmt3A:	DNA methyltransferase 3A
DMSO:	Dimethyl sulfoxide
DMT:	4, 4′-Dimethoxytrityl (protecting group)
DTT:	Dithiothreitol
ESI-MS:	Electrospray ionisation mass spectrometry
EtOAc:	Ethyl acetate
EtOH:	Ethanol
G:	Guanine
G4:	G-quadruplex
HP1α:	Heterochromatin protein 1α

HPLC: High-performance liquid chromatography

- RP: Reverse-phase
- IE: Ion-exchange
- SE: Size-exclusion

MeOH: Methanol

NMR: Nuclear magnetic resonance (spectroscopy)

- COSY: Correlation spectroscopy
- HSQC: Heteronuclear single-quantum correlation spectroscopy
- HMBC: Heteronuclear multiple-bond correlation spectroscopy
- NOESY: Nuclear Overhauser effect spectroscopy

PAGE: Polyacrylamide gel electrophoresis

- dPAGE: Denaturing polyacrylamide gel electrophoresis

PDS: Pyridostatin

RNA: Ribonucleic Acid

TERRA: Telomeric repeat-containing RNA

T: Thymine

TBA: Thrombin binding aptamer

TBTA: Tris(benzyltriazolylmethyl)amine

TEA: Triethylamine

TEAA: Triethylammonium acetate

THF: Tetrahydrofuran

THPTA: Tris((1-hydroxy-propyl-1*H*-1,2,3-triazol-4-yl)methyl)amine

TINA: Twisted intercalating nucleic acid

TLC: Thin layer chromatography

TMED: *N,N,N′,N′*-tetramethylethylenediamine

TMS: Tetramethyl silane

TSP: Trimethylsilyl propanoic acid

Contributions:

All of the work presented in this thesis was completed by Bruce Chilton, except:

- His-tagged HP1α was expressed and purified by Ruby Roach, Maia Smart or Aseel Mohammed. The His-tagged HP1α hinge region was also expressed and purified by Ruby Roach.
- Some Biolayer Interferometry experiments were performed by Ruby Roach.
- Initial N^6-[2-azidoethane]-N^6-acetyl-2′-deoxyadenosine materials were provided by Dr. Hari Kurup.
- Several ^{1}H NMR experiments were performed by Dr. Pat Edwards.
- All high-resolution ESI-MS data were collected by David Lun.

Acknowledgments:

Several people have been invaluable in completing this project. In particular, my supervisors Associate Professor Vyacheslav Filichev and Dr. Tracy Hale have provided considerable assistance in developing not only the techniques and experiments I carried out, but also providing feedback while writing this thesis. Additionally, our close collaborator Professor Geoffrey Jameson has assisted with the HP1α project and provided considerable additional feedback on my work.

Other members of the HP1α group, Ruby Roach, Maia Smart, and Aseel Mohammad, advised on a number of experiments. Within the Filichev research group, Dr. Hari Kurup and Dr. Yongdong Su provided advice in the lab. Dr. Pat Edwards provided assistance with NMR experiments and David Lun with mass spectrometry experiments. I would also like to specifically thank Suraj Patel who spent a considerable amount of time reading this thesis and providing feedback.

A final thanks to my partner Teresa who has been extremely patient during the stressful final stages of completing this thesis.

Table of Contents:

List of Figures

List of Tables

List of Schemes

1. Introduction

Deoxyribonucleic acid (DNA) polymers are used by all living organisms to store their genetic information. DNA is contained within the nucleus of eukaryotic cells and interacts with a range of proteins to transcribe and replicate the information during the various stages of the cell cycle. These interactions with proteins are central to cellular function as they preserve the integrity of the genome and define the pattern of gene expression. Mutations can arise during the normal processes within a cell, such as during DNA replication, or due to mutagens, giving rise to health conditions such as cancer. Ultimately, understanding the structure and function of DNA and related proteins is vital for effective treatment.

Interactions of various DNA secondary structures with proteins are an area of significant interest, as our understanding of them is still limited. This limitation arises from the variety of secondary structures DNA can fold into within the cell. To further develop our understanding, it is necessary to build accurate models of *in vivo* secondary structures outside of cells. This enables us to study specific secondary structures and their interactions with proteins more effectively in the context of larger DNA assemblies which include canonical duplexes. Developing *in vitro* models of various secondary structures in large DNA complexes is further complicated by the long-term stability of some secondary structures, which we aim to address by introducing short regions of modified oligonucleotides to stabilise selected secondary structures.

1.1. DNA Secondary Structure

DNA's presence in cells has been recognised for over 150 years but the composition and structure of the molecule have been the subject of considerable investigation during that time. The accepted model of DNA primary structure took prolonged investigation to determine. In 1909 RNA was deconstructed and shown to contain three components: a negatively charged phosphate group, a ribose sugar and a range of nucleobases (adenosine, cytosine, thymine and guanine) as shown in Figure 1.1.[1] The existence of the 2′-deoxyribose sugar, and therefore DNA, was not recognised until several years later.[2] Each nucleoside is composed of a nucleobase and a ribose sugar while a nucleotide also includes a phosphate group. This DNA primary structure allows for a variety of possible

DNA secondary structures to be formed, and in subsequent decades a range of possibilities were considered.

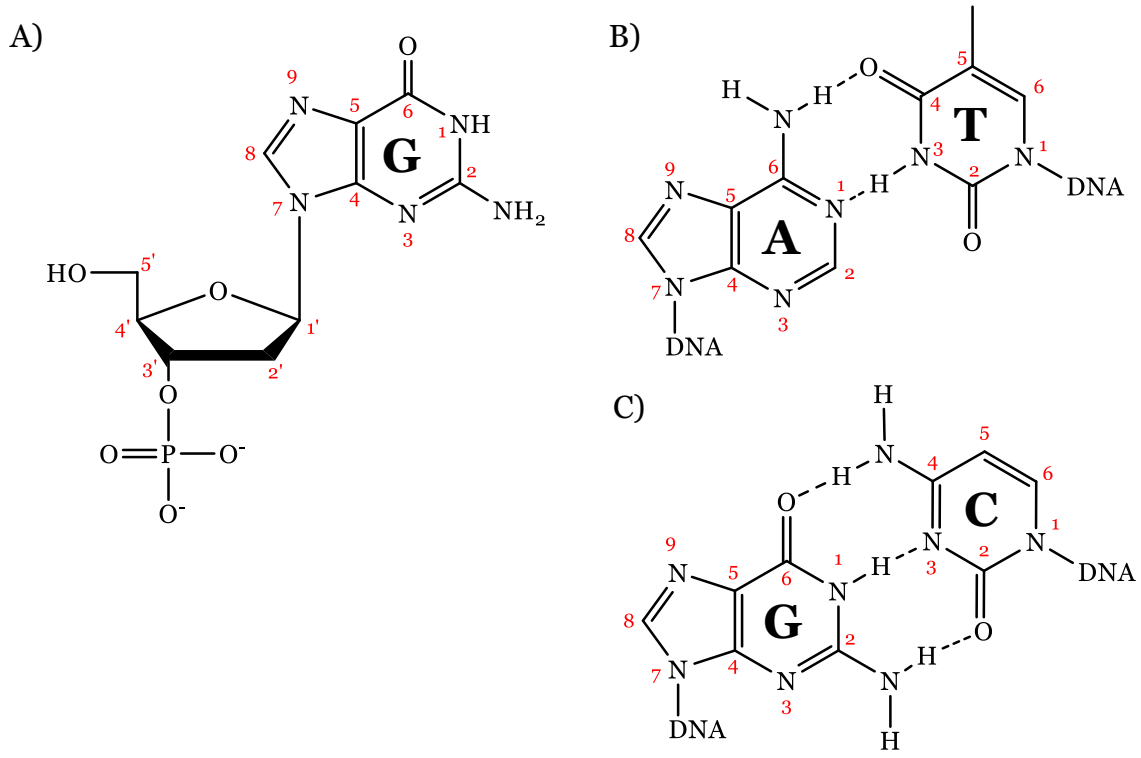


Figure 1.1. A) A standard DNA nucleotide (dG) containing a phosphate group, ribose sugar and a nucleobase. B) Canonical base pair between adenine and thymine. C) Canonical base pair between guanine and cytosine.

Griffith's experiment in 1928 demonstrated that DNA contained the genetic information of the cell, as opposed to other cell components such as proteins, by showing that genetic information could be transferred from heat-treated virulent bacteria to living nonvirulent bacteria.[3] However, this experiment did not conclude that DNA was responsible for the transfer of genetic information. This was confirmed by the Avery-Macleod-McCarty experiment in 1943. The molecule responsible for the transfer of information was successfully precipitated and the composition analysed, proving that it was DNA.[4] In 1950 the exact ratio of nucleobases was determined by hydrolysis of DNA and this was used to establish the canonical base pairs (Figure 1.1B and C).[5] X-ray crystallography was eventually used to provide the first complete description of the canonical, double-helical structure of DNA in 1953.[6, 7]

1.1.1. Canonical DNA Structures

The structure first described by Watson and Crick, referred to as the canonical duplex, featured long polymers of the previously described nucleotides, connected from the 5′-OH of one sugar to the 3′-OH of the next by phosphate

groups. Individual strands are directional, with a 5′-end and a 3′-end referring to the position of the free OH group at each end of that sequence. Two strands intertwine in an antiparallel orientation, meaning they each run from the 5′-end to 3′-end in the opposite direction to each other. The spaces between each sugar-phosphate backbone are referred to as the major and minor grooves, which serve as important binding sites for many ligands and proteins. The presence of this structural feature is not limited to canonical duplexes and is an important factor in the function of all secondary structures.

The foundation of this secondary structure is the two base pairing interactions, formed through hydrogen bonds, between adenine and thymine or cytosine and guanine (Figure 1.1B and C). However, other forces also play a significant role in the canonical duplex secondary structure. To reduce electrostatic repulsion between negatively charged phosphate groups each nucleotide is offset by approximately 36 degrees. This causes the DNA to wind into the right-handed double helical structure, containing 10 – 10.5 base pairs in each turn. The hydrophobic nucleobases are arranged towards the centre of this structure, while the charged phosphate and ribose sugar backbone face outwards, towards the aqueous solvent. Some contribution to duplex stability also comes from π-π stacking interactions between base pairs. Entropically, forming any secondary structure from large flexible oligonucleotides is unfavourable. The combination of interactions driving duplex formation (hydrogen-bonding between nucleobases, electrostatic repulsion of phosphates, hydrophobicity of nucleobases, etc.) is necessary to make formation of canonical duplex structures enthalpically favourable. However, many other DNA secondary structures are possible based on these types of interactions, and the relative enthalpies of the interactions is a determining factor in their formation.

The most common DNA duplex found in the cell nucleus is referred to as B-DNA[8] (Figure 1.2A[9], PDB: 1K8J), but other secondary structures featuring canonical base pairing are also present. A- and Z-DNA (Figure 1.2B[10], PDB: 440D and Figure 1.2C[11], PDB: 7ATG, respectively) are the most commonly observed alternative canonical secondary structures. Moreover, all three forms appear to be present in cells. A-DNA features a more compact right-handed double helix that is otherwise similar to B-DNA. It typically forms in dehydrated DNA samples

in vitro but has also been shown to form in dehydrating conditions *in vivo*.[12] Formation of A-DNA is potentially a method of protecting DNA from damage under extreme conditions.

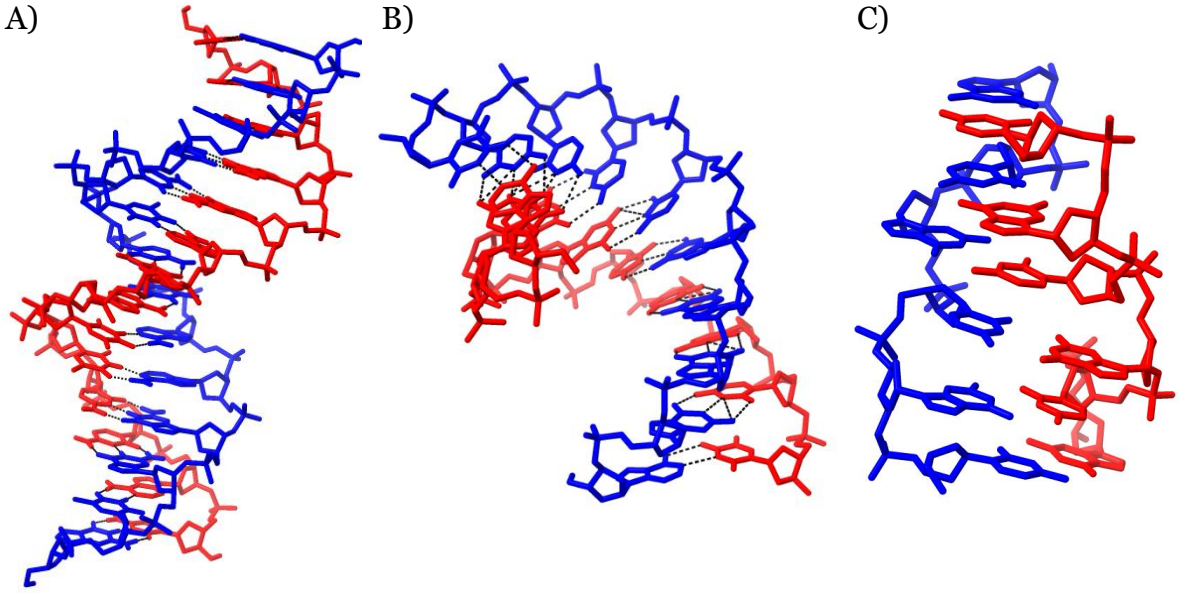


Figure 1.2. X-ray crystal structures of the most common DNA duplex conformations, showing individual strands in red and blue. A) B-DNA, the most commonly observed structure, featuring a right-handed double helix. B) A-DNA, a more compact right-handed double helix often observed in dehydrated DNA samples. C) Z-DNA, a left-handed helix typically observed in CG-rich DNA and at high salt concentrations due to the proximity of charged phosphate groups.

Z-DNA instead features a left-handed helix with a zigzag pattern, primarily found in sequences of alternating purine and pyrimidines (particularly CG). This structure is considerably different from both A- and B-DNA, with the negatively charged phosphate groups packed closer together (8 Å as opposed to 11.7 Å) than in the B-DNA.[13] This means that the presence of cations, and therefore salt concentration, is an important driving factor in Z-DNA formation. The exact role of Z-DNA in genomic function is not fully understood but it appears to form transiently in canonical duplexes during specific biological processes, such as transcription, and potentially plays a regulatory role in gene expression.[14, 15]

The driving forces of A- and Z-DNA formation are primarily the same as B-DNA, i.e., hydrogen-bonding, electrostatic repulsion and the hydrophobic effect. However, the presence of specific nucleobases as well as salt and hydration conditions appear to also play a role in their formation. The regulation of these conditions within the cell appears to be crucial to controlling secondary structure formation.

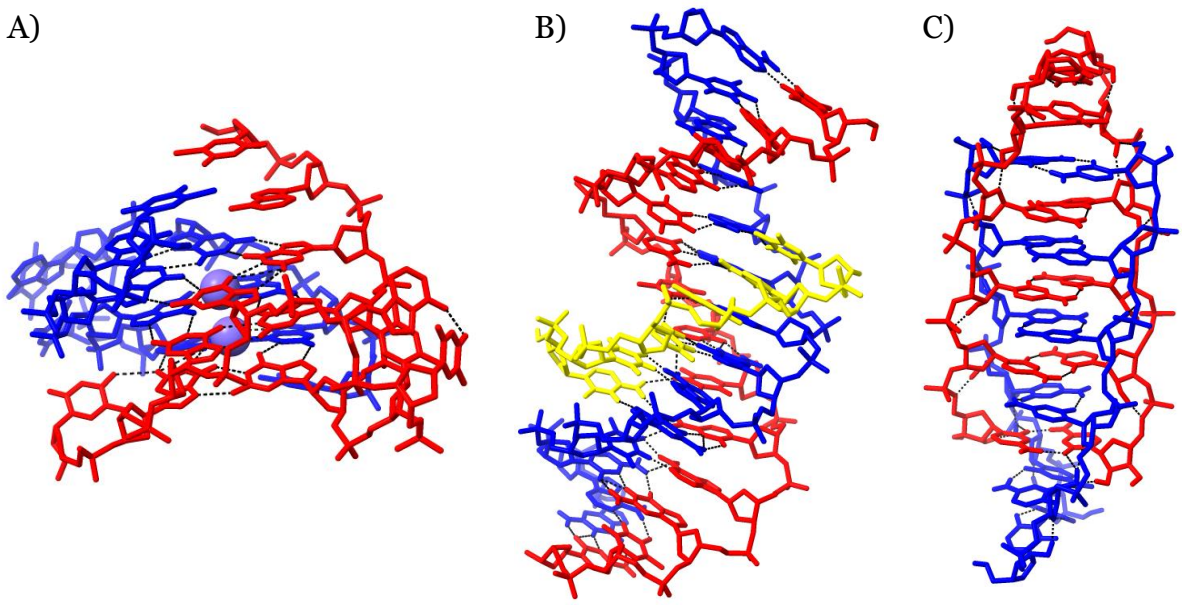


Figure 1.3. Representations of X-ray crystal structures of some common non-canonical DNA structures, showing individual strands in red, blue and yellow. While these structures are often represented as uniform shapes, X-ray structures clearly indicate they have more flexible arrangements. A) G-quadruplex, central cations (K^+) shown in purple at centre. B) Triplex, third invading strand is shown in yellow, sitting in the groove of a DNA duplex. C) i-Motif structure formed with C^+:C base pairs.

1.1.2. Non-Canonical DNA Structures

Since 1953 the number of described DNA secondary structures has grown considerably.[16, 17] Some are observed primarily outside of cells, but over time, many have been shown to exist inside cells. Canonical base pairing is not the only possibility, nor does it lead to the only structures present in genomic DNA. Several other unique secondary structures, such as G-quadruplexes (G4s), i-motifs and triplexes have been shown to form as a result of alternate, but typically less favourable, hydrogen-bonding interactions. Previously, the structure of B-DNA and its resulting properties were considered solely responsible for the functions of the genome, but it is now evident, as described below, that non-canonical DNA and RNA secondary structures also play important roles in genome function.

G-quadruplexes (G4s, Figure 1.3A[18], PDB: 1K8P) are secondary structures formed through non-canonical hydrogen-bonding arrangements in G-rich sequences. A G-tetrad is the most basic unit of a G4, formed by four guanosine nucleobases as shown in Figure 1.4A. Multiple G-tetrads are stacked to form a G-quadruplex (Figure 1.4A – C and Figure 1.3A). This arrangement results in several electronegative carbonyl groups on guanosine residues positioned close together, which would normally destabilise the structure due to electrostatic repulsion.

This is overcome through the presence of cations, such as Na^+ or K^+, which are positioned within the tetrad and coordinate with the carbonyl groups in the centre of the structure. The specific cation influences G4 stability, with thermal

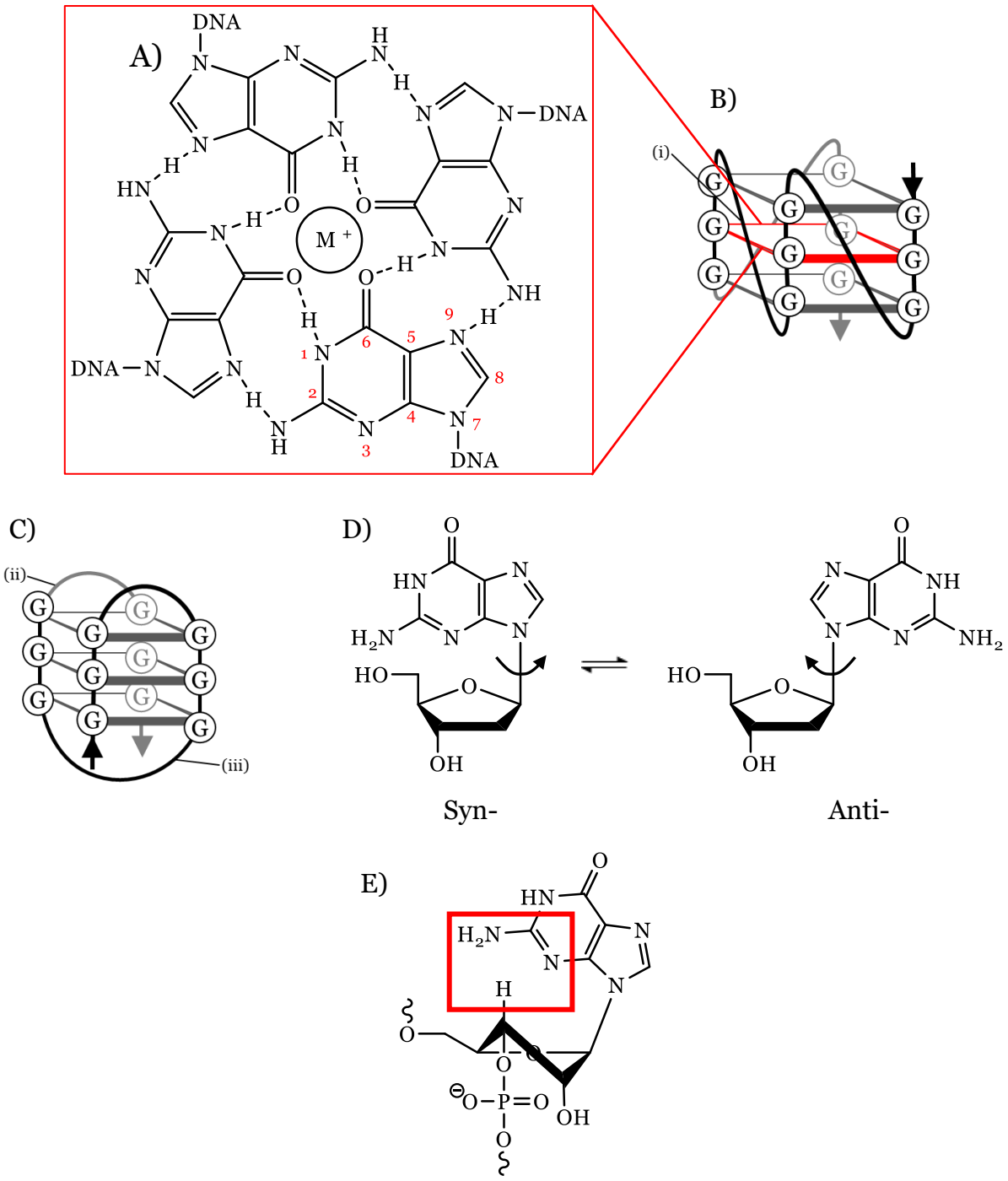


Figure 1.4. Basic interactions of G-quadruplex secondary structures. A) A G-tetrad, the basic component of the G4 structure, composed of four guanosine nucleobases interacting through non-canonical hydrogen bonds, arranged around a central cation (M^+, typically Na^+ or K^+). B) A parallel and C) an antiparallel topology of G4. Loop orientations are related to these topologies: (i) propeller in parallel G4s and (ii) lateral and (iii) diagonal in antiparallel G4s. D) Left: Syn configuration of guanosine, found only in antiparallel G4s, Right: anti-configuration of guanosine found in both parallel and antiparallel G4s. E) The syn-configuration of guanosine results in an unfavourable interaction between H-3′ and N-3/N² in RNA due to the adoption of north sugar-puckering, meaning RNA strongly prefers to form parallel G4s which contain only anti-guanosine.

stability and rate of formation increasing in the order of $Li^+ < Na^+ < K^+$.[19] Free energy of Na^+ binding is more favourable than K^+ but is counteracted by the increased cost of Na^+ dehydration. The exact difference between ions is sequence dependent, but the preference for K^+ binding is significant, despite relatively small differences in thermal stability between the different forms of the G4.

While the topology of B-, A- and Z-DNA is reasonably consistent (i.e., antiparallel with differences primarily related to left vs right-handed helices and the density of packing) G4-DNA topology can vary considerably. Unlike duplexes, strand directionality and corresponding loop arrangements can differ between parallel, antiparallel and hybrid structures, as shown in Figure 1.4B and C. While canonical structures always contained two strands orientated in opposite 5′ to 3′ directions, the four strands composing G4s could all be orientated in the same direction (parallel) or could have one or two strands in opposite directions (hybrid and antiparallel, respectively). Within these topologies various loop arrangements are possible, such as propeller, lateral and diagonal loops. Specifically, propeller loops are seen in parallel G4s (Figure 1.4B), while lateral and diagonal loops are seen in antiparallel structures (Figure 1.4C). Hybrid structures can contain a mixture of loop arrangements. These loop variations affect physical properties such as the stability of the G4 structure but can also influence their interaction with proteins.

While the formation of different topologies can be sequence dependent, factors such as salt concentration are also important. Guanosine nucleobases in parallel and antiparallel G4s also have two conformations of the glycosidic bond, the *syn* and *anti*-configurations shown in Figure 1.4D. Parallel topologies contain only *anti*-guanosine, while antiparallel topologies contain a mixture of both. This allows topologies to be readily distinguished using circular dichroism (CD) spectrometry as the interaction between tetrads containing different combinations of dG conformation results in different signals when circularly polarised light is passed through a sample. Each topology, therefore, has distinct characteristic spectra in CD profiles, which are described in more detail in Chapter 7.2.6.

RNA specifically favours the *anti*-configuration due to the presence of a 2′-OH. Changes in the shape of the ribose sugar contributes to the stability of these

secondary structures, but the 2′-OH in RNA results in a preference for the shape shown in Figure 1.4E, referred to as north sugar-puckering. This subsequently results in an unfavourable interaction between the 3′-H of the ribose sugar and the N-3 and N^2-amine of the base when dG is in the *syn*-configuration, meaning RNA shows a strong preference for the *anti*-configuration and therefore parallel topologies.[20] Unlike canonical duplex structures, a single G-rich oligonucleotide could be capable of folding into multiple G4 topologies, depending on a range of conditions such as pH and salt concentration, presence of monovalent and divalent cations and crowding conditions.

Computational analysis of the human genome initially identified over 250,000 potential G4-forming sequences,[21-23] but when genome sequencing data was also incorporated a further 450,000 were identified.[24] It is unclear if G4s are formed in all of these regions, but many of these G-rich sequences have been shown to form G4s. In most cases G4-forming regions appear to only form G4s transiently, in competition with the canonical duplex. One of the most common G4-forming sequences is the $(G_3TTA)_n$ repeat found in telomeric ends of eukaryotic chromosomes. This sequence is commonly studied as a G4-forming sequence, although it can also form a canonical duplex. Telomeres shorten during cell division, and become shorter with age. While telomere shortening can act as a tumor suppresor, telomere length has also been linked to genome instability and increased risk of cancer development.[25] Over or underexpression of the Pif1 helicase, which is known to unwind G4s, resulted in telomere shortening and lengthening, respectively. This suggests that G4 formation could be a significant factor in telomere and therefore genomic stability.[26, 27] G4-forming sequences are also enriched in the promoter regions of many genes, including oncogenes such as BCL-2,[28] c-KIT,[29] c-MYC[30] and KRAS.[31-34] An oncogene is a coding sequence in DNA whose increased expression is often linked to the development of cancer. The exact role of G4s in these genes is not entirely understood, but the secondary structure specific interactions with a range of proteins, including DNA methyltransferases,[35, 36] Pif1 helicase[37] and a chromatin architectural protein, HP1α,[38] indicate potentially critical roles in regulation and genome maintenance. Increased levels of G4s have also been observed in cancer tissue.[39] While the existence of G4s in the cell is clear, the roles of these structures is not obvious in

every instance. Investigations into these roles are a major focus of the ongoing research on G4 structures.

DNA triplexes (Figure 1.3B[40], PDB: 1BWG) are another non-canonical DNA secondary structure and the most similar to canonical duplexes. They are usually formed by the non-canonical hydrogen-bonding of a third DNA strand to a canonical duplex within the major groove. Several Hoogsteen bonding arrangements are possible, such as C:G*C$^+$, C:G*G, T:A*T and T:A*A (Figure 1.5A – D, respectively), which allow for the formation of triplexes with additional strands orientated in both the same direction as the purine strand in the

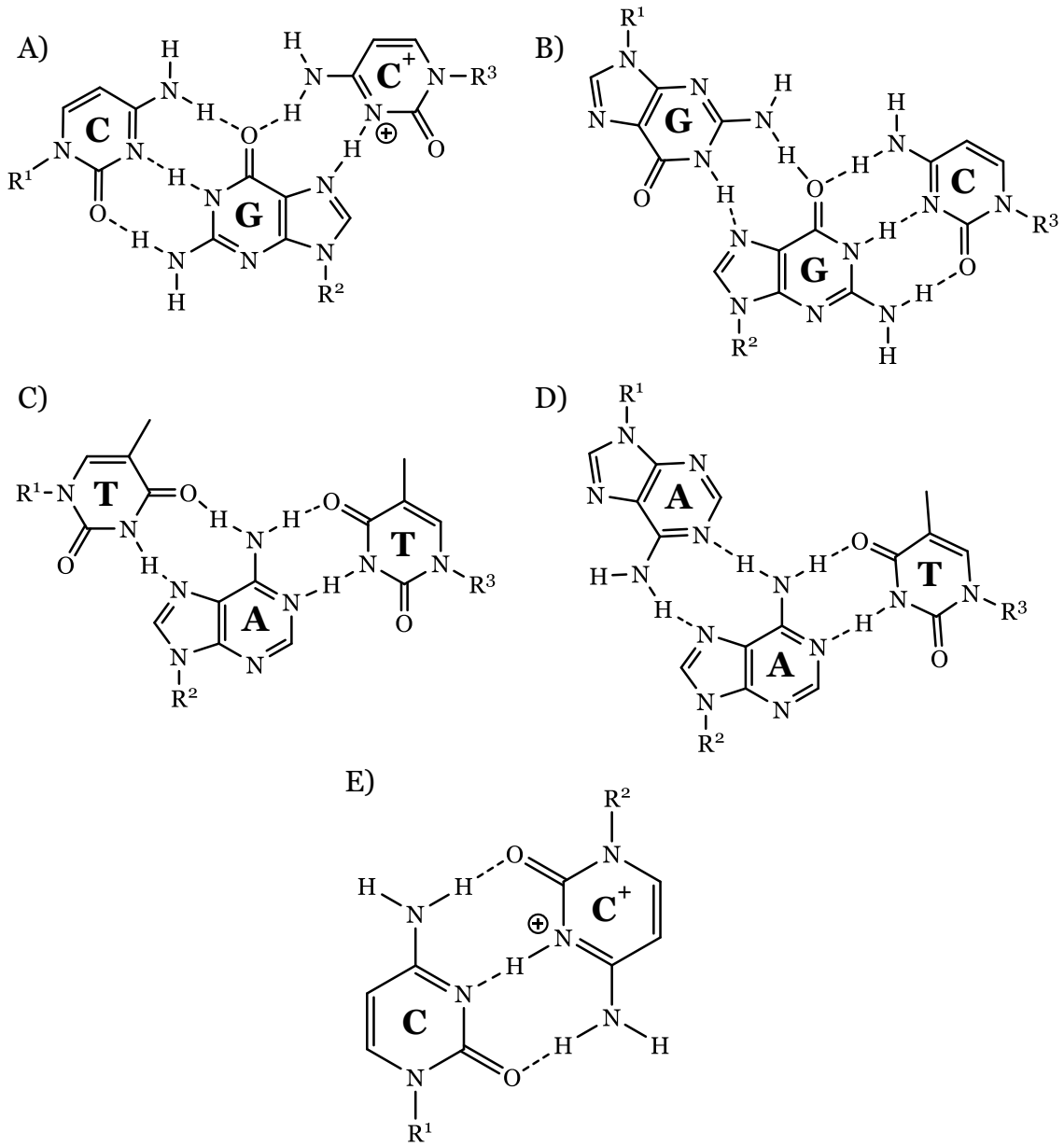


*Figure 1.5. Basic base-pairing interactions of other non-canonical secondary structures. A – D Hoogsteen hydrogen-bonding arrangements seen in triplex formation: A) T:A*T base pair. B) G:C*G base pair. C) T:A*A. D) C:G*C$^+$. E) C:C$^+$ Base-pairing interaction involved in i-motif formation.*

canonical duplex (parallel) and the opposite direction (antiparallel). However, the need to protonate cytosine for the formation of C:G*C$^+$ means that parallel triplexes usually require acidic conditions. The biological implications of triplex formation are not entirely clear, but it appears that it could have a role in multiple genomic functions, including genome stability, replication, and transcription.[41, 42]

i-Motifs (Figure 1.3C[43], PDB: 1I9K) are structures formed in C-rich DNA, although typically only at sufficiently low pH (e.g. pH < 5.5) for cytosine to be hemi-protonated, allowing the formation of C:C$^+$ base pairs (Figure 1.5E).[44] However, there has been some indication that it may be possible for i-motifs to form in higher pH ranges including physiological pH around pH 7.4.[45] As this forms in C-rich DNA, they can often form in the complementary strand to sequences forming G4s. Therefore i-motif formation must be considered whenever we form G4s with both strands present.

Within the nucleus, non-canonical secondary structures exist in equilibrium with the canonical duplex and to obtain an accurate picture of genomic activity it is insufficient to consider only duplex DNA. Going forward, our primary interest is in G4 DNA, its comparison to the canonical duplex and its interactions with proteins. We also consider, in some circumstances, i-motifs because of their potential formation in strands complementary to G4s.

1.1.3. Competition Between DNA Secondary Structures

Similar forces drive the formation of both canonical and non-canonical secondary structures. Hydrogen-bonding between nucleobases has an important role in the formation of all secondary structures. Structures such as the triplex are partially built upon canonical AT and GC base pairs, but non-canonical Hoogsteen hydrogen-bonding provides an additional basis for triplex, G4 and i-motif formation. Each structure also minimises the interactions of hydrophobic nucleobases with aqueous solvents by positioning them towards the centre. The arrangement of nucleobases in all secondary structures results in π-π stacking interactions contributing to their formation. Electrostatic repulsion between the negatively charged phosphates also results in each structure having a twisted shape, similar to the double helix in canonical duplexes (see Figure 1.2 and Figure 1.3). Overall, this means the relative thermal stabilities of the different structures

are derived from the strengths of the forces which drive their formation. Other forces, such as the repulsion of the carbonyl groups positioned in the centre of G4 structures also play a role in determining the stability of specific secondary structures, but in general they are formed using the same interactions.

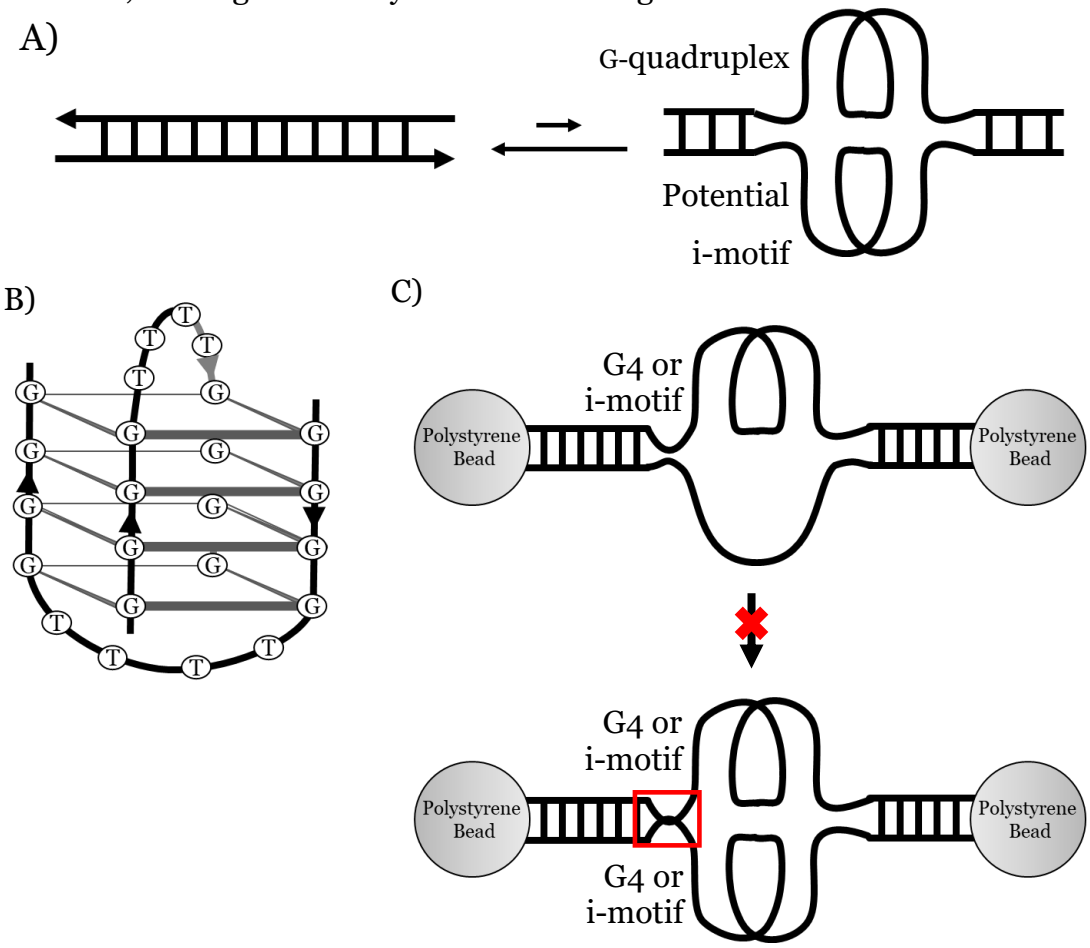


Figure 1.6. A) Formation of secondary structures could occur in many genomic DNA sequences, but as shown here, the canonical duplex is thermodynamically favoured over the G4 structure in most scenarios. However, different structures exist in equilibrium and as discussed, this equilibrium may shift throughout the cell cycle. B) The bimolecular structure of a $G_4T_4G_4$ G-quadruplex. C) A simplified representation of the model proposed by Cui et al. highlighting their conclusion that steric hindrance resulting from overlap of the two strands adjacent to non-canonical structures, inhibits the formation of multiple non-canonical secondary structures simultaneously.

Within the cell the formation of these different secondary structures is often an equilibrium related to the favourability of these interactions. G4 and duplex competition specifically is defined primarily by the differences in their hydrogen-bonding arrangements. ΔG° values obtained by *in vitro* studies typically indicate that duplex formation between complementary strands is favoured. At near physiological pH, temperature and salt concentrations, telomeric DNA sequences, such as $AG_3(TAG_3)_3$ (22AG), typically form predominantly duplex structures. UV-melting analysis of the 22AG G4 showed that ΔG° at 37°C in 0.1 M KCl was -3.8 kcal mol^{-1}, while ΔG° for duplex formation of 22AG with the

complementary 22CT sequence under identical conditions was less than -5 kcal mol^{-1}.[19] Additionally, the G4-forming sequence $G_4T_4G_4$ forms a duplex in the presence of the complementary $C_4A_4C_4$ strand in 150 mM Na^+ buffer but forms a bimolecular G4 (Figure 1.6B) in K^+ buffer. These results suggest that in buffers used to mimic physiological conditions duplexes are often slightly favoured over G4s, but adjusting the conditions can encourage the formation of a different secondary structure. Factors such as increased molecular crowding or variations in salt concentration could result in G4 formation consistent with *in vitro* results.

Outside of physiological conditions G4s begin to compete with duplex formation. As described in Chapter 7.2.3, imino proton peaks in ^1H NMR can be used to identify DNA secondary structure. ^1H NMR analysis of 22AG and its complementary 22CT sequence showed only imino peaks corresponding to a duplex structure at pH 7.0, as expected. As pH was decreased, more peaks corresponding to G4 and i-motif formation were present and by pH 4.5 a mixture of i-motif, G4 and duplex peaks were observable.[46] Similar results were obtained using CD spectroscopy to identify the various secondary structures.[47] Lower pH increases the stability of the i-motif and destabilises the GC-base pair due to the increased protonation of the cytosine nucleobase. These effects result in a corresponding increase in G4 stability. Overall, this suggests that variations in cellular conditions could result in increased formation of G4s and other secondary structures. The role of this equilibrium is often not reflected in *in vitro* studies using buffers intended to mimic physiological conditions because complementary strands are not present.

While cellular pH is relatively constant, other conditions, such as salt concentration, may vary throughout the cell cycle, potentially increasing G4 formation. In Shirude *et al.*[29], single molecular FRET experiments indicated that with no K^+ cations present solutions containing the G4-forming c-KIT sequences and the complementary strand at pH 7.4 contained almost entirely canonical duplexes. However, with 100 mM KCl added several subpopulations corresponding to multiple G4 topologies were observed. While the canonical duplex was still the primary structure, the non-canonical G4 was also present to some degree, suggesting an equilibrium between both structures. Molecular crowding could also have a role in secondary structure formation in cells but is

often not considered in *in vitro* experiments. Molecular crowding occurs in environments with high concentrations of macromolecules, such as proteins and DNA, and has been shown to alter the properties of molecules by reducing the available solvent volume, effectively increasing their concentration, and driving complex formation. Some studies have simulated a crowded environment by including high concentrations of polymers like polyethylene glycol. These results have indicated that G4s, and other structures formed by Hoogsteen H-bonds, may be more favourable under these conditions.[48] Overall, this suggests that differences in cellular conditions could encourage the formation of a range of secondary structures.

G4/i-motif interdependence studies also indicate how cellular conditions could affect secondary structure formation. Cui *et al.*[49] produced DNA sequences containing both i-motif and G4-forming regions. These sequences were incorporated into longer DNA sequences with double-stranded DNA handles allowing them to be attached to polystyrene (Figure 1.6C). Optical tweezers were then used to stretch the molecules and observe the unfolding of the secondary structure under strain. In buffer conditions where only one structure was expected to form, single-molecule force ramp assays showed extensions expected to correspond to G4 and i-motif unfolding. However, similar extensions were also observed in buffer conditions when both G4 and i-motif structures were expected to be formed. This implies that the two structures are mutually exclusive even in sequences where formation of both structures is possible (such as in cells). Cui *et al.*[49] suggested that this observation was caused by changes in the shape of adjacent single-stranded DNA connecting the non-canonical secondary structure to the canonical duplex. When one non-canonical secondary structure was formed, the connecting region could form a bulging shape, disrupting formation of the other secondary structure (Figure 1.6C). These regions on complementary strands could potentially interact and cause steric strain if both structures are formed simultaneously meaning only one structure is formed at a time. This effect was not observed when the two structures were completely offset. This is unintuitive as the formation of one structure prevents nucleobases on the

opposite strand from hydrogen-bonding, which would be circumvented if both structures were folded simultaneously.

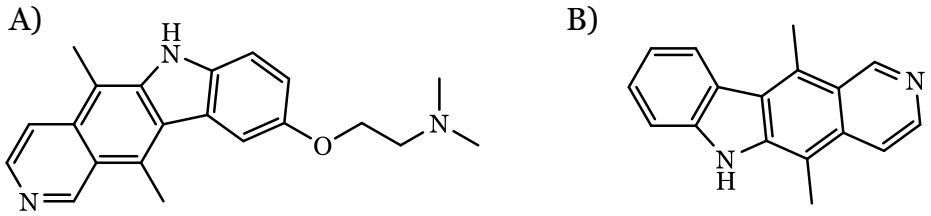


Figure 1.7. Ligands used to stabilise A) G4s (GQC-05) and B) i-motifs (71795).

These results are supplemented by observations made by King *et al.*[50] Immunofluorescence microscopy using antibodies directed against G4 and i-motifs showed both secondary structures were present in proliferating MCF-7 cells, although their distribution varied through the cell cycle. i-Motifs were mainly present during G1, the growth phase, where cells prepare for DNA replication (or S-phase), while G4s were primarily present during S-phase.

Further investigation used two ligands, GQC-05 and 71795 (Figure 1.7), which stabilise G4s[51] and i-motifs[50], respectively. These ligands were independently added to cells arrested during G1 and S-phase. When G4s were stabilised during the G1 phase almost no i-motifs were visible. A similar result was seen for G4s when i-motifs were stabilised in the S-phase. This result indicates that the mutually exclusive nature of these secondary structures, as observed by Cui *et al.*[49] is consistent with the behaviour of G4s and i-motifs in cells. These findings support the idea that while G4 and i-motif formation occurs in cells, despite the preference for canonical duplex formation, these structures are rarely formed simultaneously. Furthermore, formation or G4 and i-motif structure may be related to the cellular functions occurring before and during DNA replication.

Current research suggests that varying conditions such as salt concentration, pH or molecular crowding could significantly encourage the formation of different secondary structures. The differences between cellular conditions and *in vitro* experiments mean that obtaining completely accurate depictions of G4-protein interactions is challenging. Most *in vitro* oligonucleotide experiments use individual sequences that fold into the desired secondary structure. However, as the previously discussed studies have demonstrated, this is not how DNA exists within the cell. To develop a better understanding of cellular behaviour it is

necessary to study these interactions in the context of larger DNA structures, but under physiological conditions canonical duplexes are the dominant species present. Simply recreating conditions containing mixtures of secondary structures is not sufficient to guarantee that results accurately depict interactions with proteins when alternative structures could be present. Instead, we require large DNA duplex structures which reliably form the desired secondary structure under all conditions, regardless of the presence of complementary DNA.

1.2. Heterochromatin Protein 1α (HP1α)

Chromatin is a nucleoprotein complex within the cell nucleus, formed when genomic DNA undergoes hierarchical levels of folding.[8, 52] The DNA is wrapped tightly around large, highly basic proteins called core histones to form nucleosomes. Arrays of nucleosomes undergo further levels of folding, until architectural proteins such as a Heterochromatin protein 1 (HP1) help partition the genome into higher order domains of more densely packed, transcriptionally silent heterochromatin, or less condensed and transcriptionally active euchromatin. This packing and partitioning of chromatin is important for several reasons, including space efficiency, gene expression, and genome stability.[52] Depending on cell plasticity, between 25% and 90% of eukaryotic genomes may be stored as heterochromatin.[53, 54]

The HP1 family are responsible for establishing and maintaining heterochromatin. There are three paralogs in mammalian cells, HP1α, HP1β and HP1γ. These paralogs consist of a structurally conserved chromo-domain (CD) that interacts with di- and tri-methylated lysine 9 of histone H3 (H3K9me2/3) and a structurally related chromo-shadow domain (CSD) required for homodimerisation and protein interaction. These domains are connected by an intrinsically disordered hinge domain shown to interact with DNA and RNA. [55-58]

To date, HP1α is the only paralog shown to interact with nucleic acids with structural specificity.[38] HP1α is key to ensuring the transcriptional silencing of vast regions of repetitive DNA including the centromere and telomere. It is also shown to interact with G4 in telomeric repeat-containing RNA (TERRA) transcribed from mammalian telomeres, suggesting it may be involved in

targeting HP1a to telomeres to maintain heterochromatin. Further evidence has suggested that the presence of RNA also plays a role in HP1 binding.[59, 60]

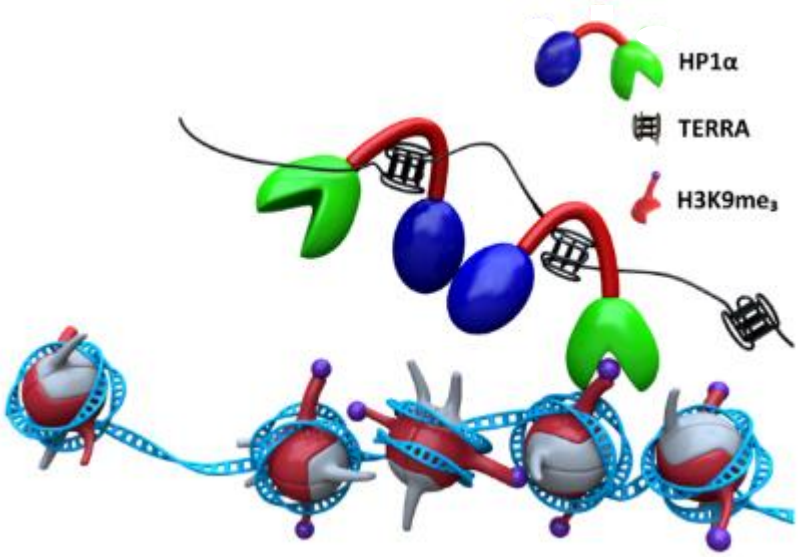


Figure 1.8. Proposed mechanism of HP1α. Two HP1α proteins dimerise though the chromoshadow domain (blue), the chromodomain (green) interacts with histones through the H3K9me3 mark of heterochromatin and the unstructured hinge (red) interacts with DNA or RNA.

Recent data indicated that binding between HP1α and DNA is driven by the specific affinity of the HP1α hinge region for parallel RNA and DNA G4s.[38] The role of other domains in this specificity is unknown, although currently unpublished research within our group has suggested that at least some of HP1α's specificity may be attributed to the chromoshadow domain.

Biolayer Interferometry (BLItz) experiments showing the affinity of HP1α for G4s (see Chapter 7.2.8 for details on the BLItz protocol) gave three key values: rate constants of association and dissociation (k_{on} and k_{off}, respectively) and the overall dissociation constant (K_D), which is the ratio of k_{off}/k_{on}. This means that a large K_D represents significantly faster dissociation than association, while a small K_D represents slower dissociation than association. The results of these experiments are verified using electrophoretic mobility shift assays (EMSA). HP1α affinity for a variety of secondary structures was tested, including duplexes, i-motifs and both parallel and antiparallel G4s. These experiments indicated that HP1α had little or no affinity for short DNA duplexes (e.g., dsTel45) and i-motifs (e.g., DAP i-motif) (Table 1.1). Antiparallel G4s sometimes showed no binding (e.g., Oligo G), but when BLItz data indicated binding did occur, they gave slow k_{on}, but also slow k_{off} resulting in reasonably high K_D (e.g., src2,), which indicated

low association but tight binding (Table 1.1). On the other hand, parallel G4s consistently gave relatively high k_{on} values and low k_{off} resulting in low K_D values (e.g., TERRA45, Table 1.1). This indicated good affinity for HP1α. Duplex and i-motif structure consistently showed no binding, while antiparallel G4s showed weak binding, if any. This specificity for G4s, particularly parallel G4s, suggests that the presence of these structures in cells could be an important factor in the formation of heterochromatin.

Table 1.1. Examples of DNA and RNA sequences tested for HP1α affinity with a variety of secondary structures.

Name	Sequence	Secondary Structure	K_D (nM)
dstel45	$(G_3T_2A)_7G_3 + (C_3TA_2)_7C_3$	Duplex	NA
DAP i-motif	$(C_5G)_4C_5$	i-motif	NA
Oligo G	$T(G_3T_2A)_3G_3TG_3(T_2AG_3)_3T$	Antiparallel G4	NA
src2	$G_4CAGCTG_3TCGCTCG_4A_2CG_4$	Antiparallel G4	410
TERRA45	$(rG_3rU_2rA)_7rG_3$	Parallel G4	74

1.3. Project Aims and Methodology

Genomic DNA folds into multiple secondary structures. Our understanding of the behaviour of DNA and proteins in cells increasingly depends upon interactions of proteins with both non-canonical secondary structures and canonical duplexes. Within the cell nucleus these structures exist simultaneously and often transiently, but *in vitro* studies often use single-stranded DNA forming only a single secondary structure. To better understand these interactions, it is necessary to build accurate *in vitro* models of DNA secondary structures as they exist in the cell, but many non-canonical structures are less thermodynamically favourable compared to canonical duplexes. Formation of these structures can be induced by changing conditions such as salt concentration or pH, but factors such as pH can also influence the stability and function of proteins. A more desirable strategy is to create thermodynamically stable structures within a larger DNA duplex array. The primary aim of this thesis is to establish several potential methods of producing these types of models. These methods will focus on stabilising G4 structures intended for incorporation into large DNA duplexes intended to study interactions of proteins with chromatin, but we anticipate that these models could be expanded to other DNA and RNA secondary structures. Ultimately, these models could be used for a range of purposes, including:

1. Understanding the role of G4s in the organisation of the genome.

2. Understanding the interactions G4s have in the context of longer duplexes with other cellular components (HP1α, transcription factors, helicases etc.).
3. A general method of producing modified DNA to stabilise other secondary structures and provide similar insights into the roles and interactions of those structures.
4. Development of drugs designed to specifically target G4-DNA interfaces and proteins, with potential to also target other secondary structures.

In this context, it is also critical to distinguish between kinetically trapped and thermodynamically stable G4s. Some existing strategies described throughout this thesis are capable of leading to stable G4s, which do not unfold in the presence of the complementary strand for days or weeks. However, in these cases, when samples are heated the G4 unfolds, and duplexes form when the sample is cooled. We refer to these structures as kinetically trapped G4s because the kinetic product of mixing the two strands, but not the thermodynamic product, is a G4. For some purposes these structures are acceptable, but in larger DNA complexes determining which secondary structure or structures are present after storage or transport of a sample is challenging. For this purpose, G4 structures are required that will only fold into a G4 in the presence of complementary DNA, even when the sample is heated and cooled.

We investigated several strategies for producing modified oligonucleotides which form thermodynamically stable G4 structures and destabilise canonical duplexes. The resulting modifications can then be incorporated into any sequence to produce the desired model. Each modification strategy is assessed based on their applicability to the goals stated above. These methods have not been expanded to larger DNA structures, which would require additional experimentation with DNA ligation beyond the scope of my research. I also do not present any modifications to non-G4 structures, but methods will also be evaluated on the ease of their translation into other secondary structures. The four primary hypotheses investigated in this thesis are as follows:

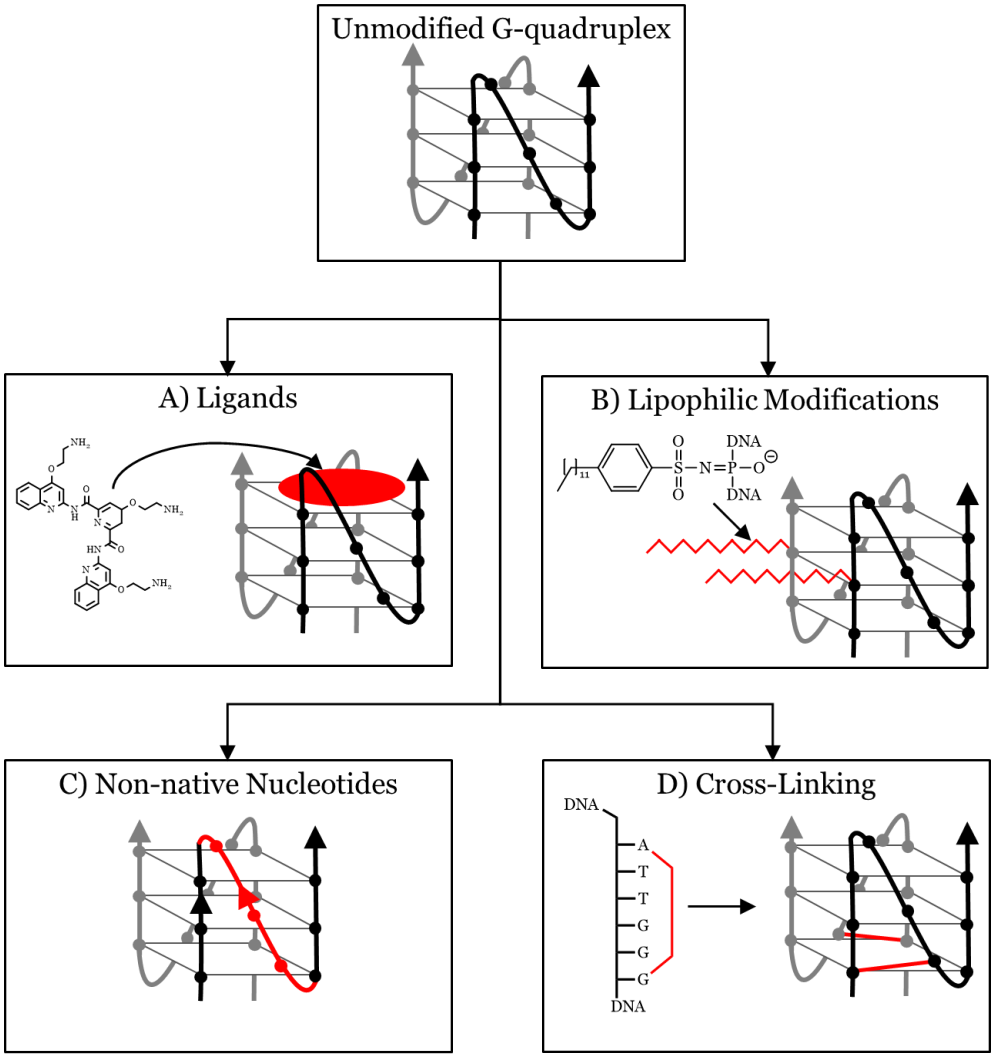


Figure 1.9. Various potential strategies for G4 stabilisation. A) Small-molecule ligands, which bind to G4s and have been shown to stabilise G4s. B) Lipophilic modifications which aggregate in aqueous solution, encouraging the formation of particular secondary structures. C) Introduction of mismatches to DNA duplexes using inverted or α-nucleotides which encourage alternative strand directionalities. D) Introduction of chemical cross-links between two sites in the sequence to enforce folding of particular G4 topologies.

1. Coordination of small-molecule ligands and G4s (Chapter 2, Figure 1.9A) improves G4 stability, but also alters G4 topology.
2. Hydrophobic moieties introduced to phosphates via Staudinger ligation with 4-dodecylbenzenesulfonyl azide will encourage aggregation of hydrophobic DNA strands, thereby stabilising G4s (Chapter 3, Figure 1.9B).
3. Incorporating inverted or α-nucleotides into DNA sequences will have minimal impact on G4 stability and topology but will create mismatches in canonical duplexes (Chapter 4, Figure 1.9C).

4. Internal cross-links in DNA sequences will encourage formation of G4 structures and prevent unfolding into other DNA secondary structures, such as canonical duplexes. (Chapter 5, Figure 1.9D).

1.3.1. Existing Strategies to Stabilise DNA Secondary Structure

Several strategies already exist for stabilising secondary structures. As mentioned above, adjusting salt concentration or pH can increase the formation of G4s and i-motifs. We have identified several potential problems with existing strategies for secondary structure stabilisation with respect to the goals of this project. These include:

(i) Inducing a change in G4 topology.
(ii) Disrupting the interactions between proteins and G4s.
(iii) The modification is not stable in the presence of the protein.
(iv) The modification only stabilises the G4 for a limited period of time (i.e., kinetically trapped vs thermodynamically stable structures).

A variety of ligands has been synthesised to stabilise different secondary structures. These ligands are useful for studying a specific secondary structure, but they introduce several issues when studying more complex systems. Small-molecule ligands, such as GQC-05 (Figure 1.7A) or pyridostatin (PDS) (Figure 1.10A), have been shown to stabilise G4s.[61] In Chapter 2 we investigate the use of pyridostatin for G4 stabilisation and demonstrate the issues of this strategy. These issues could be partially mitigated with careful ligand selection but make it hard to compare *in vitro* results to the physiological behaviour.

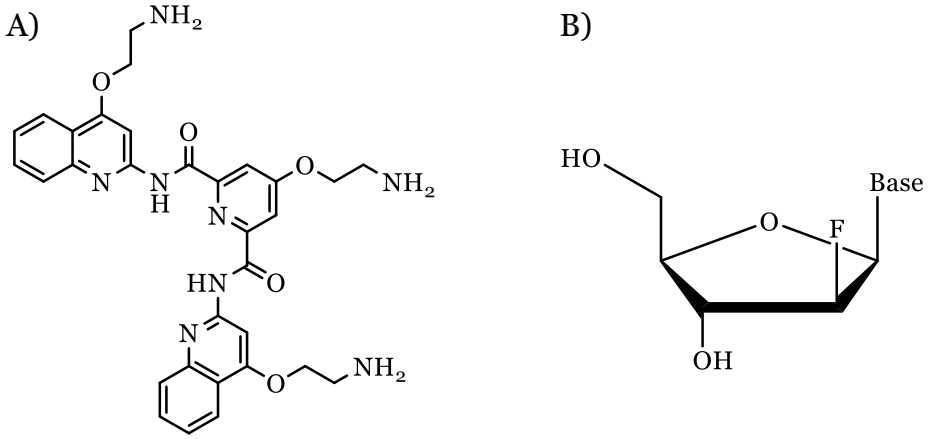


Figure 1.10. Several existing methods of stabilising G4s: A) Pyridostatin, a small-molecule ligand, shown to improve G4 stability. B) 2′-Fluoroarabinonucleic acid, a non-native nucleoside used to create kinetically trapped, but not thermodynamically stable G4s.

In some cases, chemical cross-links have already been used to stabilise other secondary structures. For example, DNA hairpins were stabilised by triazole linkers created through copper(I)-catalysed azide-alkyne cycloaddition[62] and DNA duplexes were stabilised by disulphide bridges.[63] These modifications could theoretically allow for the specific reinforcement of desirable G4 topologies. We had previously experimented with disulphide bridges and in addition to challenges in forming cross-links after synthesis, the disulphide bridges could be reduced by dithiothreitol or other reagents commonly used in the preparation and storage of proteins. The use of triazole linkers for G4 stabilisation is explored in Chapter 3.

Kinetically trapped G4s had been obtained from sequences in which unique nucleotide modifications were used to prevent G4 unfolding. One particularly successful modification featured 2′-fluoroarabinonucleic acids (Figure 1.10B).[64, 65] These modifications increased the favourability for parallel G4s by enforcing the *anti*-conformation of dG at modification sites. This is seemingly due to the interaction between the 2′-F and the sugar of adjacent guanine residues, indicated by 2D NMR, particularly NOESY, and by X-ray crystallography (PDB: 1K8P).[18] Telomeric G4 sequences containing the modification were examined using ^{1}H NMR in the presence of the complementary strand. ^{1}H NMR can be used to distinguish secondary structures because of the distinct imino-proton peaks for G4s (11 ppm), duplexes (13-14 ppm) and i-motifs (15 ppm) (see Chapter 7.2.3.). ^{1}H NMR spectra were taken periodically over multiple days for mixtures of both G-rich and C-rich modified and unmodified sequences. Native sequences reformed into duplexes within 1 hour, indicated by the appearance of signals at approximately 13-14 ppm, but sequences containing 2′-fluoroarabinonucleic acids typically maintained a G-quadruplex structure for up to a month. However, while these structures were kinetically trapped, they would slowly revert to duplexes over time, and this process can be accelerated when samples were heated to obtain the thermodynamic product.[66]

In general, the existing methods fall into two categories. The first are strategies which have shown promise in stabilising other secondary structures but need to be applied to G-quadruplexes. These include chemical cross-linking and lipophilic modifications. These methods require translation into G4-forming

sequences and analysis of the resulting properties. As will be demonstrated throughout this thesis, these properties are not always predictable and specific modifications made to G4s do not behave identically to similarly modified canonical duplexes. The second are strategies which may not produce thermodynamically trapped G4s. Both small-molecule ligands and modified nucleotides have been applied to G4s in the past, but only result in kinetically trapped structures or result in significant changes in G4 topology. In these cases, we look for alternative strategies which could overcome existing limitations.

1.3.2 Sequence Selection

The sequences chosen for our experiments were determined based on several factors. Primarily, sequences were chosen because they resemble G-rich sequences found in the human genome. However, this includes a significant number of potential sequences. To further narrow this selection, sequences were chosen based on length as it is easier to produce modified sequences of shorter DNA. Ultimately, longer DNA sequences will need to be synthesised for models, but shorter sequences are easier to produce quickly and with high yield. For this purpose, the tel sequence, $dTAG_3TTAG_3T$ (12 nucleotides), is the most suitable sequence used for non-native nucleotide (Chapter 4) and cross-linking (Chapter 5) modifications. This sequence closely resembles the G_3TTA repeat found in the telomeres of human chromosomes, and can fold into a bimolecular, parallel G4. Previous experiments had shown that non-native nucleotides stabilised the tel sequence to form a bimolecular, parallel G4.[64] Some of our experiments use slightly elongated forms of this sequence to add flexibility. In these cases, we introduce nucleotides to maintain the G_3T_2A repeat (i.e., adding TA to the 3′-end of tel rather than any other combination of nucleotides to complete TTA for this section of the sequence).

In some cases, we used other sequences to further development methodologies tested using the tel sequence. In these cases, c-MYC, $G_3TG_3TAG_3TG_3$ (16 nucleotides), and c-KIT, $G_3CG_3CGC_2AG_3AG_4$ (20 nucleotides), were used. Both sequences form parallel, unimolecular G4s. As mentioned previously, these sequences are found in the promoter regions of oncogenes,[29, 30] giving them obvious relevance to cancer research, and the role of G4s in the expression of these genes is therefore an important area of

investigation. These sequences are slightly longer than tel but are short enough to be readily synthesised and modified.

In the case of lipophilic modifications, we used two sequences, TG_4T, known to form a tetramolecular parallel G-quadruplex, and $G_4T_4G_4$, known to form a bimolecular antiparallel G-quadruplex. These sequences did not resemble native G4s as closely as tel, c-MYC or c-KIT. However, they were chosen because their properties are well studied[67-71] and they are easily synthesised. As discussed in Chapter 3, the challenges in synthesising oligonucleotides containing multiple lipophilic modifications made tetra- or bi-molecular G4-forming sequences a more suitable alternative to the unimolecular sequences used in other experiments.

We also used the Pu39 sequence, $AG_4CG_3CGCG_3AG_2A_2G_5CG_3AGCG_4CTG$ (39 nucleotides), found in the promoter region of the BCL-2 gene, due to the interest in this sequence within the HP1α research group. Pu39 contains 6 G-tracts and has therefore been shown to form several different G-quadruplexes composed of different combinations of these G-tracts.[72] These structures are primarily parallel although CD indicates some mixed topologies are present. While HP1α showed affinity for Pu39,[38] there was some interest in the potential for differences in binding affinity between the possible G-quadruplexes. While this was being investigated by other members of the HP1α research group at Massey University, Pu39 also remained an interesting sequence for future modification. However, for testing synthetic modifications, the primary focus of this thesis, Pu39 was too long. It was assessed, alongside other sequences, for its interactions with small-molecule ligands specifically because no modification was necessary. In future, Pu39 could be an important target for modifications which are successful in preliminary experiments.

Three of the native sequences (c-MYC, c-KIT and Pu39) were all known to bind HP1α, based on previous work within this group. Those experiments had also shown an affinity of HP1α for telomeric RNA (TERRA22, 45 and 96), which are similar to the tel sequence (although no bimolecular G4s had been tested previously). While this thesis is primarily concerned with demonstrating different possible modification methods and their advantages, this was also a consideration in their selection. However, DNA sequences were ultimately

chosen for their ability to demonstrate that modifications were possible and gave positive outcomes rather than specific interactions with proteins. Eventually it would be preferable to produce sequences which can be used for experiments with proteins, and in some cases preliminary results have been obtained and are presented here.

2. Stabilisation of G-Quadruplexes using G4-Targeting Ligands

2.1. Introduction

Since the presence of G4-forming sequences in telomeres and the promoter region of oncogenes[28-34] was discovered there has been considerable interest in these sequences as targets for anticancer drugs. Many proteins are unable to process G4s, and by stabilising them it is possible to prevent the functioning of those proteins. The location of G4-forming sequences in oncogene promoters means that this could be an effective method of targeting tumour cells. One of the most popular strategies for accomplishing this is utilising small molecule ligands which have been shown to stabilise G4 structures.[61] Successful G4-targeted ligands have typically shown preference for G4 binding over canonical duplexes and result in increased G4 thermal stability. Stabilising G4s could disrupt cellular processes because most enzymes interact primarily with the canonical duplex. Several ligands are shown in Figure 2.1 and discussed below, but a wide variety of ligands has been produced primarily with a focus on therapeutic applications. Despite extensive development, no G4 ligands have advanced past phase II clinical trials and the development of new ligands is ongoing.

Binding models of ligands are generally similar, although some ligands feature additional functionality to improve binding. Aromatic moieties interact with G-tetrads through π-π interactions, lone pairs and protons on nitrogen atoms allow for hydrogen-bonding within the G4 groove, and the potential protonation of side chains can allow for electrostatic interactions with negatively charged phosphates. The prevalence of these functional groups can be seen in all the example ligands in Figure 2.1. The acridine-based C_8 ligand (Figure 2.1A) contains an 8-carbon long flexible chain between aromatic rings and polar functional groups. This added flexibility was intended to allow additional interaction with the tops of G-tetrads, in addition to the groove. NMR and docking studies showed evidence of both loop and π-π stacking interactions, suggesting that it was able to introduce these interactions. The ligand was specifically designed to bind to KRAS22-RT, a 22-nucleotide G4-forming sequence found in the promoter of the KRAS gene and showed remarkably strong binding to that sequence.[73] The KRAS gene is responsible for making K-Ras, a

protein which transfers information from outside the cell to the nucleus, which has been identified as an oncogene. C_8 resulted in an approximately 40 °C increase in KRAS22-RT's T_m according to CD melting experiments and gave an apparent dissociation constant in fluorescence titration experiments of 4.4×10^{-7} M.[73] The binding modes of a wide range of ligands have been described with similar detail. Binding modes are similar across ligands, primarily based on the interactions of pyridine rings with nucleobases and charged phosphates, with specific ligands introducing additional features to improve G4 binding and selectivity, such as the flexible alkyl chain in C_8.

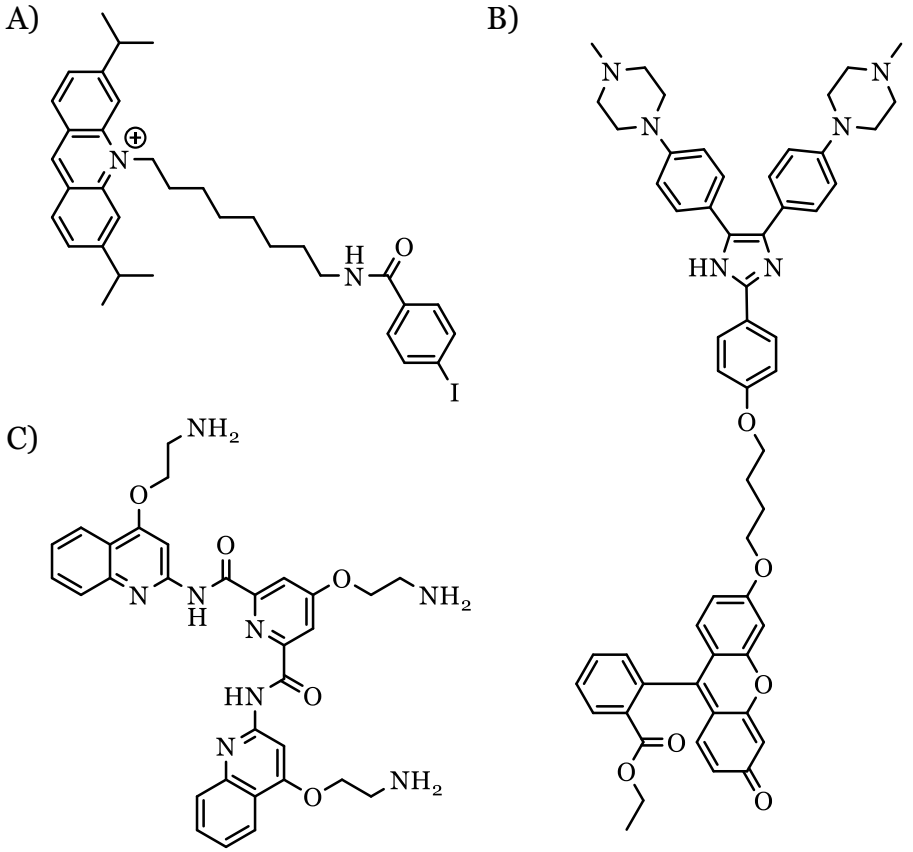


Figure 2.1. Examples of existing G4-binding ligands: A) Acridine-based C_8 ligand, shown to include π-π stacking and loop interactions. B) IZFL-2, a fluorescent ligand which shows specificity for c-MYC over other G4s. C) Pyridostatin, ligand used to inhibit polymerase activity.

Initially most ligands were designed to target cancer cells, but more recently ligands have also been designed as targeted G4 labels. Fluorescent tags are attached to ligands, meaning they both bind specifically to G4s and contain detectable fluorophores, indicating G4 localisation in cells. The IZFL-2 ligand (Figure 2.1B) is an interesting example because in addition to binding to and stabilising G4s, it showed specificity for the c-MYC G4 over other G4 sequences.[74] Its binding is based on both π-π stacking and hydrogen-bonding interactions.

This was first studied using UV-Vis absorption titrations, wherein the change in absorption was measured as ligand was added. However, experimental results were inconclusive, and the binding mode was further verified computationally using long-time multiscale molecular dynamics simulations. The specificity of IZFL-2 allows it to identify not only the location of G4s, but specifically the formation of G4s in the c-MYC promoter region, which is a known oncogene. Many ligands have been produced containing other fluorescent groups and these highly specific G4 binding ligands are useful tools in identifying G4 formation in cells. The use of one such G4 ligand, GQC-05 (Figure 1.7A), is illustrated in the experiments by King *et al.* discussed in Chapter 1.

The pyridostatin (PDS) ligand (Figure 2.1C), which we will use in our subsequent experiments, is an example of a ligand known to induce polymerase stalling.[61, 75] By stabilising G4 structures it is possible to stop the processing of DNA by enzymes that specifically interact with canonical duplex structures. While this ligand has been shown to improve G4 stability, it is unclear whether coordination of the ligand to the G4 would affect G4 topology and, as a consequence, block or change the shape of protein-binding sites. Our initial hypothesis is that ligands such as PDS will induce changes in G4 structure, which would therefore interfere with studies of protein binding intended to mimic conditions in cells.

To test this hypothesis, we performed CD and ^{1}H NMR experiments on several G4-forming sequences before and after addition of PDS. These experiments will clarify the effect that PDS has on G4 stability in the buffer we use for ^{1}H NMR and will then allow for comparison between the widely used existing method of stabilising G4s with small-molecule ligands and the new methods of G4 stabilisation described in this thesis. The effect of PDS binding to several G4 sequences will be assessed using the same techniques used for our modifications, both in terms of thermal stability (see Chapter 7.2.6 for details of CD analysis technique) and disruption of duplex formation (see Chapter 7.2.3 for details of ^{1}H NMR analysis).

Additionally, we present our investigation of several new, G4-targeted ligands developed by our collaborator, Dr. Masayuki Fujii of Kindai University, Osaka (Japan). These ligands are based on short peptide chains intended to

match the shape and distribution of charges in the backbone of G4 structures. However, we encountered problems with their use, primarily a significant decrease in the solubility of resulting G4-ligand complexes, which is described in Chapter 2.3.

2.2. Evaluation of the Effect of Pyridostatin on G4 Stability and Topology

The investigation of PDS binding focused on the four natively occurring sequences discussed in Chapter 1.3.2, tel, c-KIT, c-MYC and Pu39 (Table 2.1). In the case of tel we instead use the slightly elongated Ttel sequence to allow these results to be compared with ^{1}H NMR experiments which were run simultaneously and is discussed in Chapter 4. Each sequence was tested in Na^{+} and K^{+}-containing buffers, before 2 mole equivalents of PDS were added to the K^{+}-containing stock solution and sequences were tested again. Literature indicated that 2 equivalents were sufficient to see significant changes in G4 behaviour.[76] Circular dichroism (CD) spectroscopy was used to determine the apparent topology and stability of each G4, and ^{1}H NMR spectroscopy was used to observe changes resulting from PDS complexation with G4s and to challenge PDS-G4 complexes with complementary sequences (see Chapter 7.2.3. and 6.2.6 for protocols of CD and ^{1}H NMR experiments in G4 studies).

2.2.1. Topology and Stability of G4-PDS Complexes

In Na^{+}-containing buffer these sequences formed G4s with a range of topologies. Ttel and c-KIT appeared to be only partially folded in Na^{+} buffer (Figure 2.2A and C), while c-MYC and Pu39 formed G4s with mixed topologies (Figure 2.2B and D). In K^{+}-containing buffer c-MYC formed parallel G4s (Figure 2.2B), while Ttel, c-KIT and Pu39 formed mostly parallel G4s, but each contained a shoulder at approximately 290 nm (Figure 2.2A, C and D). These topologies were mostly preserved upon addition of PDS, except for Ttel, which instead formed an antiparallel G4 in the presence of PDS (Figure 2.2A). However, in all instances we observed a slight decrease in the intensity of the peaks corresponding to G4 formation. It is unclear if this is due to decreased G4 formation or a small amount of complex aggregation and precipitation, although minimal precipitate was visible in solution. In most cases this meant that the

addition of PDS did not have a significant impact on the overall topology, according to CD spectroscopy.

Table 2.1. DNA Sequences and G4 thermal stability in presence and absence of Pyridostatin

Name	Sequence	Na^+ (T_m, °C)	K^+ (T_m, °C)	K^+ + Pyridostatin (T_m, °C)
c-KIT	$G_3CG_3CGC_2AG_3AG_4$	Not folded	44 (p)	>90 (p)
c-MYC	$G_3TG_3TAG_3TG_3$	41 (p), 45 (ap)	73 (p)	>90 (p)
Ttel	$T_2AG_3T_2AG_3T$	30 (p) (i)	46 (p)	>90 (p), 55 (ap)
Pu39	$AG_4CG_3CGCG_3AG_2A_2G_5CG_3AGCG_4CTG$	51 (ap)	58 (p), 53 (ap)	>90 (p), 60 (ap)
i-c-KIT	$C_4TC_3TG_2CGC_3GC_3$	---	---	---
i-c-MYC	$C_3AC_3TAC_3AC_3$	---	---	---
i-Ttel	$AC_3TA_2C_3TA_2$	---	---	---
i-Pu39	$CAGC_4GCTC_3GC_5T_2C_2TC_3GCGC_3GC_4T$	---	---	---

10 μM strand concentration, 20 μM pyridostatin, 20 mM sodium phosphate, 10 mM KCl (in K+ samples), pH 7.0. (p) parallel topology, (ap) antiparallel topology. (i) Structure is only partially folded at 15 °C, so Tm is estimated.

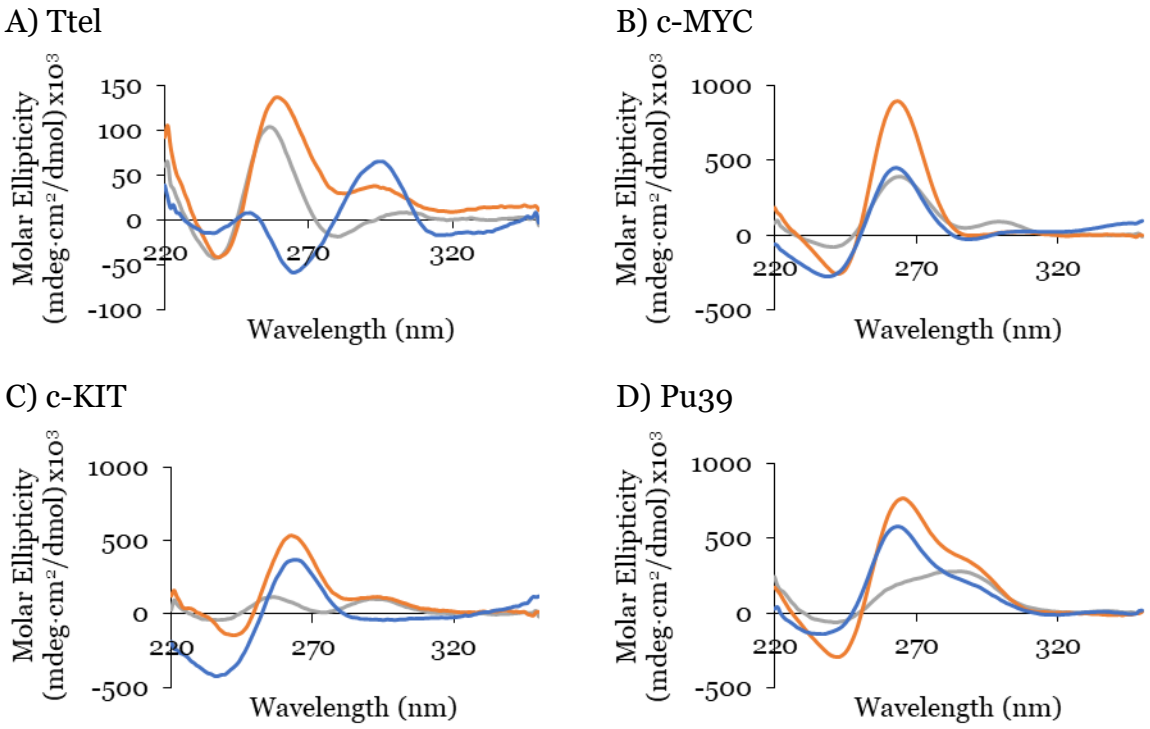


Figure 2.2. CD profiles of G4-forming sequences in Na^+ buffer (grey), K^+ buffer (orange) and K^+ buffer in the presence of PDS (blue). Conditions: 10 μM strand concentration, 20 μM ligand concentration, 20 mM sodium phosphate, 10 mM KCl (in K^+-containing samples), pH 7.0.

However, the initial results of CD spectroscopy were partially contradicted by ^{1}H NMR spectra. The changes in G4 topology upon PDS addition were also observed by ^{1}H NMR spectroscopy (Figure 2.3), which could be compared to CD spectra. In all cases, significant changes in the imino protein peaks between 10 – 12 ppm were observed when PDS was added. Ttel and c-MYC sequences, which

had the most clearly defined peaks in this region before addition of PDS, moved to a lower chemical shift and no longer had such clearly defined peaks (Figure 2.3A and B, respectively). The less clearly defined peaks suggest that the addition of PDS induced the formation of a wider range of G4 topologies and the change in chemical shift may indicate aggregation of G4s into larger structures. This may explain why we observed a decrease in signal intensity after PDS addition in CD experiments. The PDS control (Figure SI-1) showed no peaks in this region, indicating that any changes are not a result of overlapping peaks, and therefore the observed changes must reflect changes in G4 structure resulting from the formation of a G4/ligand complex. These results were also observed for c-KIT and Pu39, although these sequences had less clearly defined imino proton peaks in K^+ buffer before ligand additions (Figure 2.3C and D, respectively). While CD spectroscopy indicates that G4 topology is mostly conserved, ^{1}H NMR spectra suggest that at least minor topological changes have occurred. Alternatively, broad imino proton peaks in ^{1}H NMR can indicate increased aggregation of G4s,

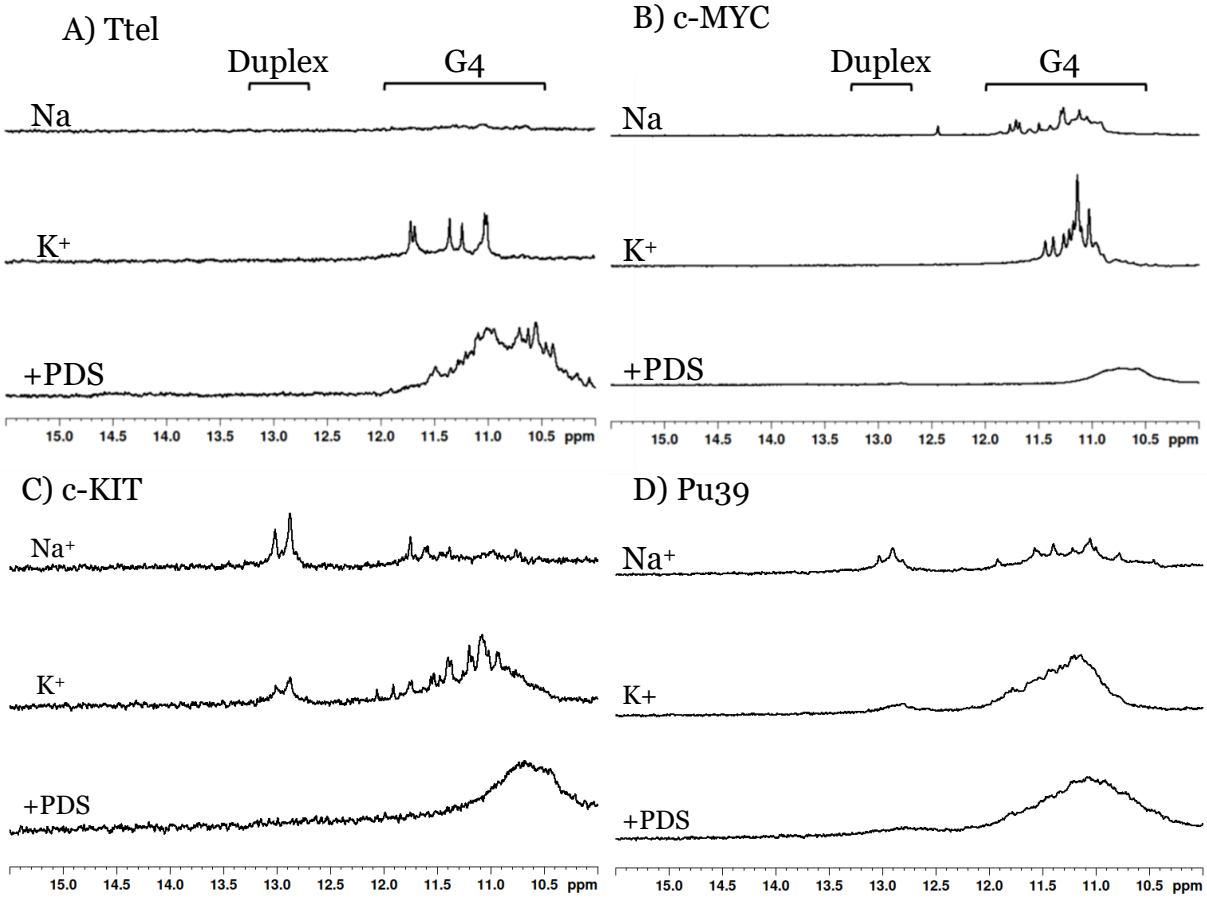


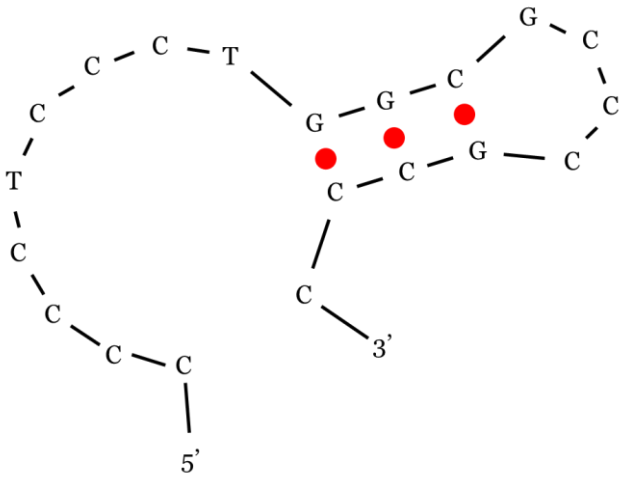
Figure 2.3. Change in topology, as indicated by ^{1}H NMR spectroscopy, of G4-forming sequences in Na^+ and K^+-containing buffer, followed by addition of PDS to K^+ buffer samples. A) Ttel. B) c-MYC. C) c-KIT. D) Pu39. Conditions: 200 µM strand concentration, 400 µM PDS, 20 mM sodium phosphate, 10 mM KCl (in K^+-containing samples), 10% D_2O, 1% TSP, 25 °C, pH 7.0.

caused by the PDS ligand, but the change in chemical shift suggests that the topology may also have changed. This shift may also be induced by a shielding effect of PDS in NMR, but the change still implies a significant difference between G4s before and after PDS addition. Broadly, this suggests that if a parallel topology is more thermodynamically favoured it may be preserved, but the presence of a ligand such as PDS potentially leads to multiple topologies and aggregates.

PDS is reported to increase G4 stability,[75] but after observing the significant change in topology suggested by NMR measurements, it was necessary to assess thermal stability of the G4s in the NMR buffer. The T_m values indicated results consistent with literature: addition of PDS caused a significant increase in thermal stability (Table 2.1, Figure SI-2 (Na^+), Figure SI-3 (K^+) and Figure SI-4 (K^+ + PDS)). Control samples, without PDS, showed increased stability in K^+-containing buffer compared to Na^+, which is the case for most G4-forming sequences. c-MYC was considerably more stable than the other sequences. After PDS addition c-KIT, c-MYC and Pu39 were all stable up to 90 °C when PDS was present. The Pu39 sample showed some change in CD spectra at close to 90 °C but was not completely melted. c-KIT and c-MYC showed no significant change. Interestingly, the Ttel spectra appeared to change from the characteristic spectra of an antiparallel G4 to the spectra of a parallel G4 as it was heated. By approximately 65 °C, a large positive peak at 265 nm is visible, and the intensity of the positive peak at 290 nm has decreased considerably. These spectra then remained unchanged at 90 °C. This suggests either a complete change in G4 topology was induced by heating, or PDS stabilises two G4 structures with different thermal stabilities. This confirms that the increase in thermal stability caused by forming G4-PDS complexes is significant, which agrees with reported results that addition of PDS causes a noticeable increase in G4 stability. Even c-KIT, the G4 with the lowest stability in K^+ buffer with $T_{m\,of}$ 44 °C, was stable at 90 °C in the presence of the ligand.

2.2.2. Duplex Formation of PDS-G4 Complexes and Complementary DNA

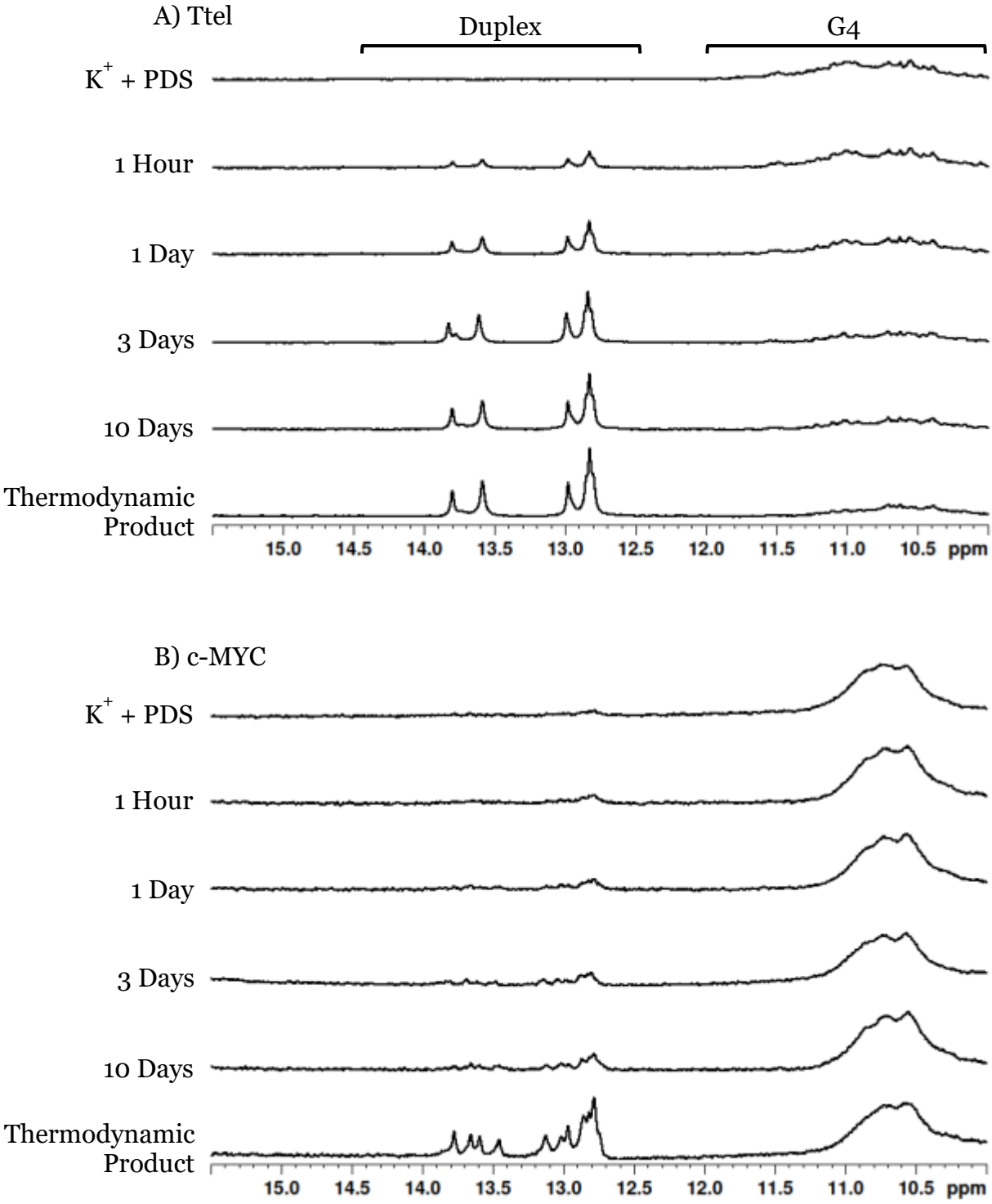
While thermal stability changes significantly in the presence of PDS, the effect on G4 competition with the corresponding canonical duplex is unclear. To test this, we challenged G4-PDS complexes with their complementary sequence (i.e., i-c-KIT was used with c-KIT, i-c-MYC was used with c-MYC, etc.). ^{1}H NMR spectra were analysed over several days and then the thermodynamic product was produced by heating the sample at 90 °C for 5 min and slowly cooling to room temperature. Each sample contains an internal standard of 1% trimethylsilylpropanoic acid (TSP), which has a ^{1}H NMR peak at 0 ppm, used to evaluate the ratio of each structure present in solution. In the absence of PDS, all peaks in the 10 – 12 ppm range (corresponding to G4 formation) were no longer visible after the thermodynamic product was formed (Figure SI-5), and new peaks appeared around 13 – 14 ppm (corresponding to duplex formation. In the case of Ttel and c-KIT this conversion was almost complete after 1 hour, but c-MYC was not completely converted until the thermodynamic product was obtained. In the presence of PDS, all four sequences also immediately began to form some duplex structure (Figure 2.5A – D). However, the changes typically occurred more slowly over several days and after the thermodynamic product was obtained, some peaks corresponding to the G4 structure were still present in all samples. It should also be noted that i-c-KIT forms a DNA hairpin (Figure 2.4[77]) and peaks around 13.2 ppm correspond to the base pairs formed by this sequence (Figure 2.5C). Additionally, all samples had similar peaks between 13 and 14 ppm



ΔG = -2.67 kcalmol^{-1}

Figure 2.4. Stable hairpin structure predicted using IDT sequence analyser for i-c-KIT sequence.

due to the duplexes observed without PDS present. This suggests that in the presence of PDS, these sequences reach an equilibrium containing a mixture of duplex and G4 structures, while native sequences form primarily canonical duplexes in the absence of PDS.



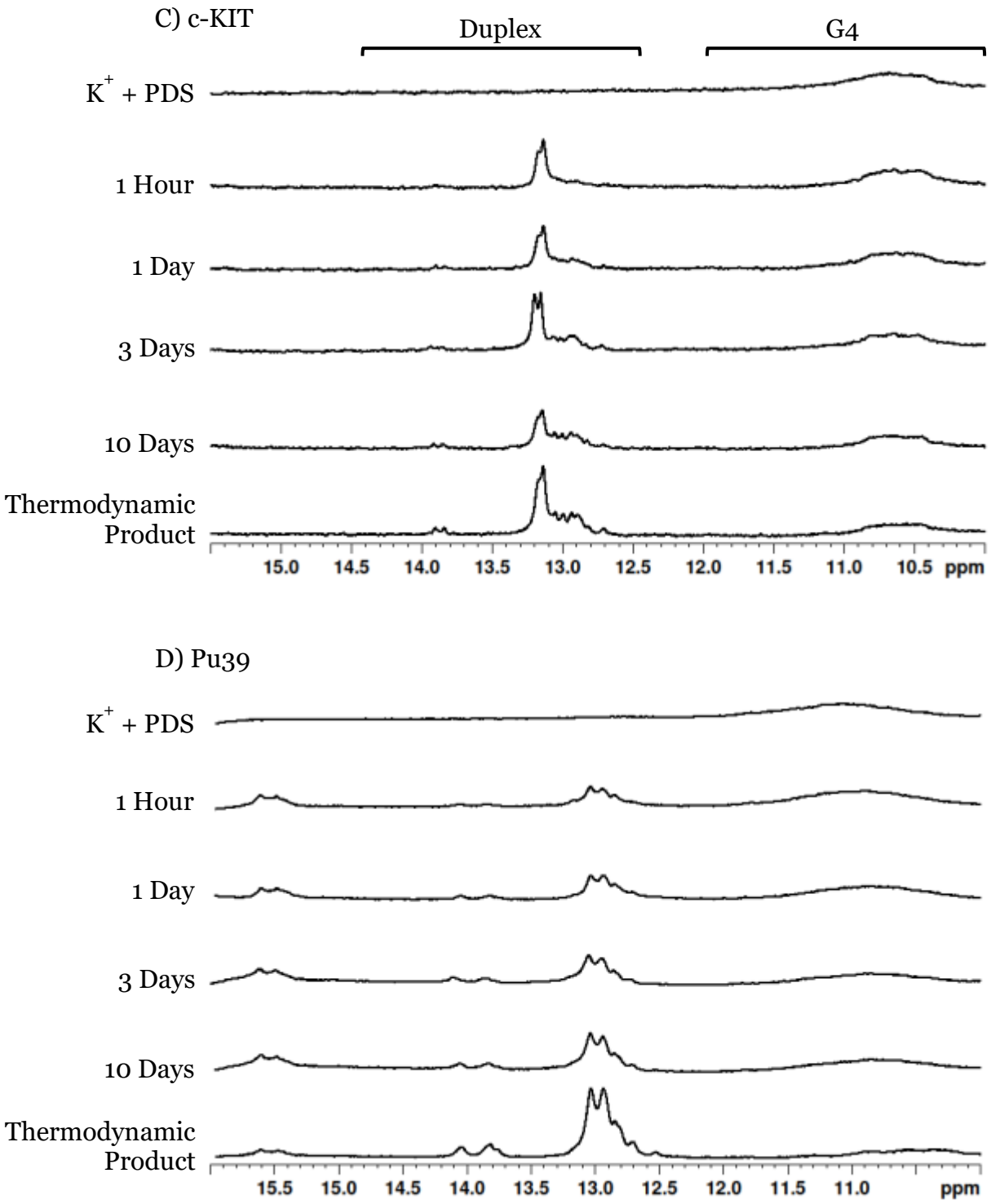


Figure 2.5. Evaluation of duplex formation for G4-forming sequences in the presence of PDS and the complementary strand using ^{1}H NMR. A) Ttel + i-Ttel. B) c-MYC + i-c-MYC. C) c-KIT + i-c-KIT. D) Pu39 + i-Pu39. Conditions 200 μM strand concentration, 400 μM ligand concentration, 20 mM sodium phosphate, 10 mM KCl, 10 % D_2O, 1% TSP, pH 7.0.

In samples without complementary DNA present the addition of PDS increased thermal stability significantly. However, ^{1}H NMR indicates that PDS addition also induces significant topological changes to the G4 and the improvements in G4 stability when challenged with the complementary sequence

are limited and sequence dependent. G4-stabilising ligands are an area of considerable interest in the G4 field and are potentially a useful tool for G4 stabilisation in other contexts. However, with respect to the specific aims of this project, addition of ligands to G4 samples did not sufficiently inhibit duplex formation, induced changes in G4 topology and molecular state, and could potentially compete with the binding of G4 samples with proteins intended for study. This suggests that more effective alternatives are necessary for developing large nucleosomal DNA models intended to study the interactions of proteins with G4s in cells.

2.3. Evaluation of Peptide-based Ligands as G4 Binders

Ligands derived from existing proteins are another interesting possibility due to their diverse structures and functions, as well as their proven ability to bind to DNA and RNA. In a previous study our collaborator, Dr. Masayuki Fujii of Kindai University, Osaka (Japan), developed several peptide-based ligands and showed their ability to bind to duplex DNA.[78-81] Specifically, Peptideβ7 (sequence: $(RL)_7$) and peptideα6 (sequence: $(LRAL)_6$) respectively form structures composed of an antiparallel β-sheet or α-helices in the presence of duplex DNA or RNA. These peptides are able to penetrate the cell membrane and form an siRNA-peptide complex. Formation of this conjugate resulted in silencing of the chimeric BCR/ABL gene in the Human Chronic Myelogenous Leukaemia Cell Line K562.[82] This suggests a potential strategy for targeting diseases such as leukaemia. Based on this result, our collaborators suggested that similar peptides could be developed for targeting G4s and therefore target regions known to form G4s in oncogenes, including the sequences used to test PDS.

Our collaborator provided us with several polypeptides designed specifically as G4 ligands. We performed several experiments to assess their viability as G4 ligands. This allows us to draw comparisons between the new ligands, derived from proteins, and a non-biologically derived ligand, such as PDS. There is also some interest in their therapeutic application, regardless of the limited applicability to our immediate aims.

Each ligand is composed of a central four-glycine repeat enclosed by regions containing four alternating repeats of two amino acid residues, either arginine/leucine (RL), arginine/serine (RS) or arginine/glycine (RG) (Table 2.2).

These ligands are referred to by the non-uniform section of their peptide sequences, i.e., RL4, RS4, and RG4. The cationic arginine residues (pK_a = 13.8) are intended to interact with the anionic phosphate backbone of DNA and form a stable DNA/ligand complex. The peptides all featured a four-glycine region in the centre, intended to align with the G-tetrad. RL4 uses a similar model to peptideβ7 which both contain predominantly leucine and arginine, while both RG4 and RS4 were designed with glycine and serine respectively with the expectation of increasing hydrophilicity and therefore improving solubility.

Table 2.2. Sequences and T_m's of G4-forming sequences in the presence of peptide-based ligands

Name	Sequence	K^+	+RG4	+RL4	+RS4
c-KIT	$G_3CG_3CGC_2AG_3AG_4$	$44°C^{(p)}$	$58°C^{(p)}$	$41°C^{(p)}$	$48°C^{(p)}$
c-MYC	$G_3TG_3TAG_3TG_3$	$73°C^{(p)}$	$71°C^{(p)}$	$77°C^{(p)}$	$74°C^{(p)}$
Ttel	$T_2AG_3T_2AG_3T$	$46°C^{(p)}$	$43°C^{(p)}$	$44°C^{(p)}$	$44°C^{(p)}$, $28°C^{(ap)}$
Pu39	$AG_4CG_3CGCG_3AG_2A_2G_5CG_3AGCG_4CTG$	$58°C^{(p)}$, $53°C^{(ap)}$	$64°C^{(p)}$	$63°C^{(p)}$	$67°C^{(p)}$
RG4	H-RGRGRGRGGGGRGRGRGR-OH				
RL4	H-RLRLRLRGGGGRLRLRLR-OH				
RS4	H-RSRSRSRGGGGRSRSRSR-OH				

Strand concentration: 10 μM; Ligand concentration: 20 μM; buffer: 20 mM sodium phosphate, 10 mM KCl, pH 7.0. (p) parallel topology, (ap) antiparallel topology.

2.3.1. Initial Assessment of Ligand Affinity using CD Spectroscopy and Thermal Stability

Initially, the topology and thermal stability of the G4/ligand complexes were tested using CD spectroscopy. The same sequences were used to test peptide-based ligands that were used for PDS. CD spectra are normally corrected for small variations due to differences in concentration using the Chirascan software provided by Applied Photophysics, the supplier of the CD spectrophotometer, but in this case, we observed significant changes in both CD and UV-Vis signal when ligands were added to some samples. In particular, the Ttel sequence in the presence of RS4 (Figure 2.6A) showed a significant increase in intensity, while Pu39 in the presence of RS4 (Figure 2.6D) showed a considerable decrease. c-MYC samples were more varied, with increased signals when RG4 or RS4 were added, and decreased signal when RL4 was added (Figure 2.6B). c-KIT signal was generally consistent, except in the presence of RG4, when the signal decreased (Figure 2.6C).

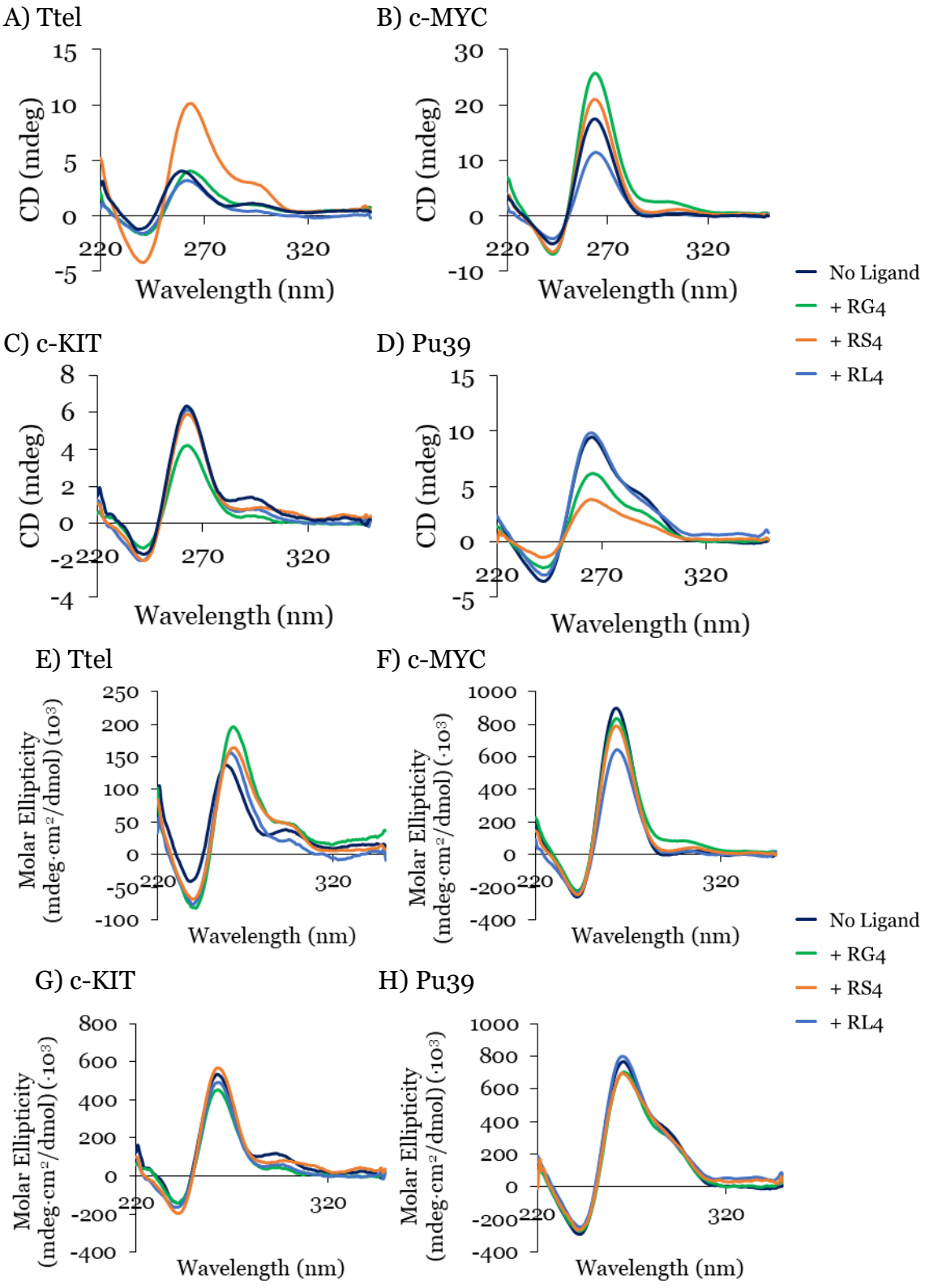


Figure 2.6. A – D) CD profiles of G4-forming sequences in the presence of peptide-based ligands before adjusting for differences in concentration: A) Ttel, B) c-MYC, C) c-KIT, D) PU39. E – H) CD profiles of G4-forming sequences in the presence of peptide-based ligands after adjusting for differences in concentration: A) Ttel, B) c-MYC, C) c-KIT, D) PU39. Conditions: 20 μM strand concentration, 40 μM ligand concentration, 20 mM sodium phosphate, 10 mM KCl, pH 7.0.

The spectra shown in Figure 2.6E – F have been adjusted based on the concentration calculated from UV-Vis signal at 260 nm when the sample was

recorded (see Chapter 7.2.5.). In these spectra, the difference in intensity between G4 samples with or without any of the ligands is minor, and they generally have the same G4 topology as the sequence without ligand present (Figure 2.6E - H). Ttel is a minor exception, upon addition of the ligands, particularly RG4, the CD signal around 260 nm shifts towards 265 nm (Figure 2.6E). Overall, this is not a large change, but it does suggest that the ligand offers some support to the G4 structure. In PDS experiments we observed a considerable change in the CD spectra, and a more pronounced decrease in signal intensity for other samples. This could suggest that these ligands are better designed to support the existing G4 topology, but significantly broader screening experiments would be required to determine if this trend is consistent for all peptide ligands vs conventional ligands.

In terms of thermal stability, most samples indicated only minor changes in T_m of 4 – 5 °C (Table 2.2, Figure SI-6, Figure SI-7 and Figure SI-8). Compared to ligands such as PDS this was not a particularly significant increase. For some samples, such as c-KIT + RG4 and all Pu39 samples, we observed a larger increase in thermal stability of around 10 – 15 °C.

These preliminary results suggest that the ligands provided some support for the G4 structure. However, many of the samples contained a considerable amount of white precipitate, which appeared only after addition of the ligand. It was therefore proposed that a G4/ligand complex may be formed but was subsequently precipitating from solution due to insufficient solubility. This would explain some of the observations made in the initial CD spectra, with precipitation causing a decrease in CD signal. The minor changes seen for melting experiments might therefore be inaccurate if the only G4 remaining in solution is not interacting with the ligand. In particular, the c-KIT + RG4 samples contained the most visual precipitate and CD data indicated a decrease in concentration (Figure 2.6C). Recording CD data at a range of RG4 concentrations shows a clear decrease in CD signal as ligand is added (Figure 2.7). Of the c-KIT samples, this sample also showed the largest increase in thermal stability. It is unclear if these results support the binding of G4s to these ligands and if that binding is structure specific. We present the subsequent experiment testing precipitation of various secondary structures in the presence of the RG4 ligand below, but we are

currently discussing the future directions of these ligands with our collaborator. Most likely, further investigation will require similar ligands with a greater focus on improving their solubility.

2.3.2. Titration of Various DNA Secondary Structures with Peptide-Based Ligands

To evaluate further, we performed titrations of increasing quantities of RG4 against a c-KIT G4 in intervals of 0.2 equivalences, recording a CD spectrum after each addition. A final volume was added to achieve a ratio of 2:1 ligand to DNA. A linear decrease in CD signal was observed as peptide was added, as shown in Figure 2.7A, supporting the theory that G4/peptide complexes precipitated from the solution. To simplify these results, the CD signal at the λ_{max} was extracted and used to plot the decrease in CD signal as ligand was added, shown in Figure 2.7B. This indicated that some interaction between the G4-forming sequence and the peptides was occurring but did not determine if those interactions were structure specific. Since the ligands could theoretically interact with any sequences through electrostatic interactions with the phosphate backbone this titration was repeated with other DNA secondary structures. This included a parallel RNA G4, antiparallel G4, i-motif, single-stranded DNA and duplex DNA. The c-MYC sequence was also tested, to determine if this interaction was specific to c-KIT or possible with any G4. As shown in Figure 2.8A, a linear disappearance of CD signal at 265 nm, as observed for c-KIT, was not seen with other G4s. However, c-MYC signal did decrease, although less consistently and the peak did not disappear entirely. Additionally, c-MYC signal did initially seem to increase at 265 nm (shown in Figure 2.8A), possibly indicating some stabilisation of the G4

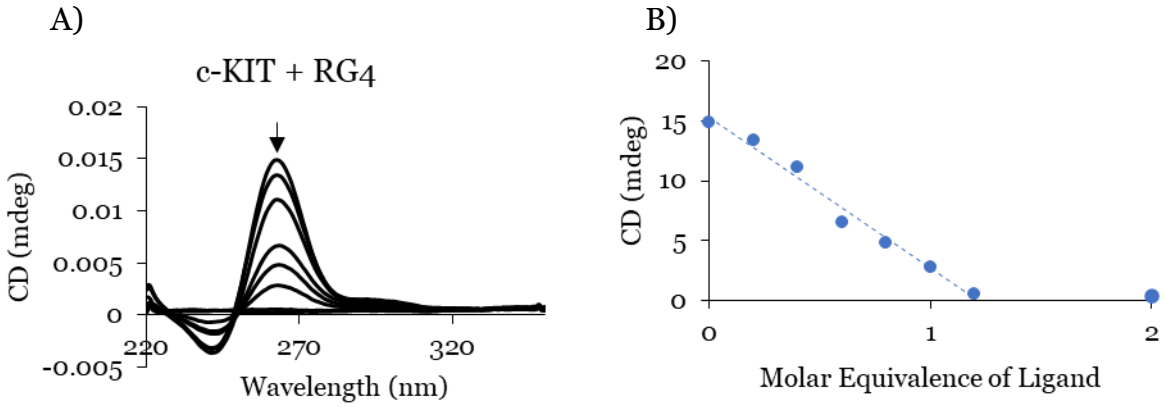


Figure 2.7. A) Titration of c-KIT sequence with RG4 ligand. CD spectra recorded with additional 0.2 mole eq. of RG4 added. Linear data is extrapolated from the λ_{max} (262 nm). B) CD signal at 265 nm vs ligand mole ratio. Conditions: 10 μM strand concentration, 2 – 20 μM ligand concentration, 20 mM sodium phosphate, 10 mM KCl, pH 7.0.

before precipitation occurred. A steep decrease in TERRA45 signal was observed at around two equivalences, with minimal change prior to that addition. The unimolecular, antiparallel G4, thrombin binding aptamer (TBA, Table 2.3), did not exhibit significant changes. Instead, the signal decreased slowly as ligand was added, beginning around 0.8 eq. It appears to cause some precipitation of each sequence, but c-KIT in particular appears to precipitate at lower molar ratio of G4 to ligand. The length of sequences such as TERRA45 could mean they require a significant amount of ligand to increase hydrophobicity, but these results could also indicate a specific preference for c-KIT DNA.

In addition to G4s, several other DNA structures were tested (Table 2.3). Several duplex sequences, with a range of lengths were tested, as shown in Figure 2.8B. Some decrease in signal was observed for the DNA 1 duplex, which was the shortest structure tested. The c-KIT duplex also exhibited some decrease in signal at 267 nm, but this wavelength also corresponds to a parallel G4, which could indicate precipitation of a G4 present in equilibrium. This would then induce a shift in equilibrium towards the G4 and more precipitation. In Figure 2.8C, further equivalents were added to the c-KIT duplex and CD signal disappeared entirely, suggesting that complete precipitation occurred, but at higher DNA to ligand ratio than with G4 structures. Herring sperm DNA is a large naturally occurring duplex. It was tested on the basis that the short DNA 1 duplex may have decreased in signal only because it was a short sequence, and greater quantities of ligand might be necessary to precipitate such large sequences. As expected, no significant changes were observed in CD spectra with these quantities of ligand added.

Non-duplex or non-G4 structures were also tested, shown in Figure 2.8D. For single-stranded DNA, no significant change in signal was observed when RG4 was added. For the i-motif sample of ctel45 at pH 5.5, however, the signal decreased considerably above one equivalent of the ligand. This decrease was not observed for ctel45 at pH 7.0, when i-motifs are not formed according to CD (Figure SI-9C). This could suggest that the ligand interacted with i-motifs as well, but less efficiently than with c-KIT G4. While functionally very different, G4 and i-motif structures are slightly similar in shape, so the ligands designed to target the G4 may have some interaction with i-motifs as well in the absence of any

Table 2.3. DNA sequences tested with peptide-based ligands and their secondary structures.

Structure	Sequence Name(s)	Sequence	λ_{max}
Antiparallel G4	TBA	$G_2T_2G_2TGTG_2T_2G_2$	295 nm
Parallel G4	c-KIT	$G_3CG_3CGC_2AG_3AG_4$	267 nm
	c-MYC	$G_3TG_3TAG_3TG_3$	265 nm
Parallel RNA G4	TERRA45	$(G_3U_2A)_7G_3$	265 nm
i-motif	ctel45 (pH 5.5)	$(C_3TA_2)_7C_3$	285 nm
Single-stranded	DNA 1	CA_7G	265 nm
	ctel45 (pH 7.0)	$(C_3TA_2)_7C_3$	283 nm
	tRNA	---	268 nm
Duplex	DNA 1	CA_7G	280 nm
	i-DNA 1	CT_7G	
	c-KIT	$G_3CG_3CGC_2AG_3AG_4$	267 nm (G4), 285 nm
	i-c-KIT	$C_4TC_3TG_2CGC_3GC_3$	(duplex)
	Herring Sperm DNA	---	275 nm

See Figure SI-9 for CD spectra showing secondary structures.

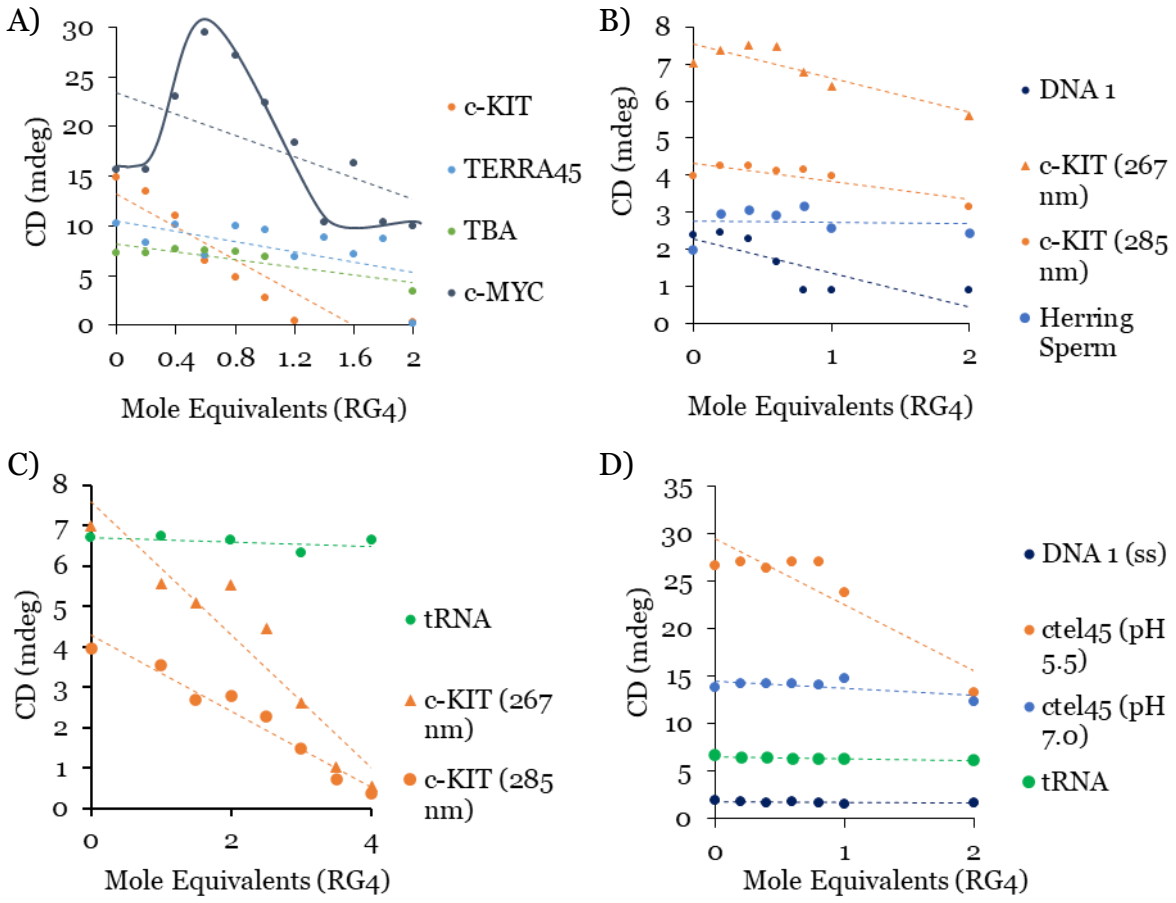


Figure 2.8. CD titration of various DNA secondary structures with the RG4 ligand showing CD signal (mdeg) at the λ_{max} for each secondary structure (Table 2.3). A) G4 structures (c-MYC, c-KIT and TERRA45, TBA. B) Duplex structures. C) Extended titrations of tRNA and the c-KIT duplex. D) Single-stranded DNA and i-motif sequences. Conditions: 10 μM strand concentration, 2 – 20 μM ligand concentration, 20 mM sodium phosphate, 10 mM KCl, pH 7.0.

functionality giving greater structural specificity. The tRNA sequence was another longer sequence, intended to evaluate the relationship between the number of phosphate groups and the precipitation of DNA. Once again, no significant change in CD signal was observed upon addition of RG4 and this was also true when additional ligand was added, as shown in Figure 2.8C.

The reliability of these titration experiments is unclear. Upon repetition, similar overall decreases in CD signal were observed, but the exact point at which precipitation begins varies between titrations. These results provide clear evidence that an interaction between RG4 and some DNA secondary structures is occurring, but it is insufficient to conclude whether this interaction is structure specific. More effective methods of determining this would require higher quantities of the ligand than we were provided to perform experiments such as ^{1}H NMR, as demonstrated with PDS. Alternatively, labelled or tagged peptides could allow for experiments such as electrophoretic mobility shift assays (EMSA) or biolayer interferometry (BLItz).

Based on CD experiments, the ligands induced minimal change in G4 topology and potentially provide some increase in thermal stability. However, assessing the extent of these changes is challenging due to the precipitation of any resulting complex. Titration experiments indicated that significant precipitation occurred in samples containing G4 sequences, particularly c-KIT but also in other parallel and antiparallel G4 sequences to lesser extents. However, some precipitation was also observed for some duplex and i-motif sequences. Ultimately, if we accept that precipitation indicates the formation of an insoluble or low solubility DNA/ligand complex, these results could suggest a preference, but not specificity, of the ligands for G4 secondary structures.

Combined with our collaborator's previous experiments with duplex structures, these peptide ligands are a unique possibility for future targeted G4 ligands. We are currently in correspondence with our collaborators with regards to potential future experiments with this type of ligand. There remains an interest in further modification of these peptides to include additional G4 specific features, potentially increasing specificity for G4 structures. Following the results obtained in our experiments, there is also some interest in the addition of features intended to improve hydrophilicity. Further exploration of these ligands and their

properties, or the properties of similar ligands is necessary before their applicability to G4 stabilisation or therapeutics can be accurately assessed. However, they do not appear to be more applicable to the overall aims of this thesis than PDS or other ligands and, similar to existing ligands, potential therapeutic application will likely be the focus of future investigations.

2.4. Conclusions

A considerable number of G4-binding ligands were synthesised that were shown to improve G4 stability with the possibility of high sequence specificity. In most cases these ligands are intended for therapeutic applications. In these applications, ligand-induced G4 stabilisation results in inhibition of enzymes such as polymerases or transcriptases because they are unable to process non-canonical structures. This strategy is effective for these purposes, but conflicts with the aim of this project. We require G4 stabilisation strategies which preserve the G4 topology and the ability of G4 specific proteins to bind. In this regard we find several issues with small-molecule G4 ligands as a stabilisation method.

As mentioned in Chapter 2.1., ligands with a range of binding modes have been developed, but they tend to involve interactions between the ligand and the G-tetrad or phosphates in the groove. This means that many ligand-binding sites compete directly with the binding of G4-specific proteins. For therapeutic purposes, this is generally not a concern, as the ligands are intended to act as inhibitors, but it makes them less optimal when investigating protein binding. One proposed mechanism for HP1α binding to G4s involves interactions with the G-tetrad from above. This is primarily suggested by the reduced affinity of HP1α for G4s with lateral and diagonal loops (i.e., antiparallel vs parallel G4s), as the loops could block the interaction of HP1α with G-tetrads from that direction. Testing HP1α binding using ligands with a similar binding mode could completely disrupt binding, preventing analysis using techniques such as X-ray crystallography.

Furthermore, the results observed with the PDS ligand highlight another issue with ligands for our intended application. While CD spectra indicate that G4 topology is similar to the naturally occurring G4, ^{1}H NMR spectra show that the structure of the G4 is changed considerably. While this change improves thermal stability and inhibits formation of the canonical duplex, as expected, it

also means that results obtained in the presence of these ligands cannot be used as analogues for G4/protein interactions occurring in the cell. Specifically, their potential for blocking protein-binding sites and influencing G4 topology means that other methods of modification are preferable when creating stabilised G4 complexes for protein binding.

Despite this, the results obtained using PDS and peptide-based ligands presented here provide a useful comparison for the strategies described in the following chapters. Small-molecule ligands are the most widely used method of G4 stabilisation, and in developing new strategies it is useful to compare with the preferred existing method.

3. Effect of Lipophilic Phosphate Modifications on G4 Stability

3.1. Introduction

The small-molecule ligands discussed in the previous chapter are currently a popular method of stabilising G4s, with new ligands being constantly designed and synthesised, but alternative methods have been developed simultaneously. Phosphate modification is a method which has commonly been used to produce phosphorothioates, which are used to protect synthesised DNA from enzymatic degradation (Figure 3.1A).[83, 84] Other modifications are possible using a range of synthetic techniques, but the Filichev group, in collaboration with Dr Dmitry Stetsenko of Novosibirsk State University, has focused on developing modified oligonucleotides using the Staudinger reaction. These modifications are intended

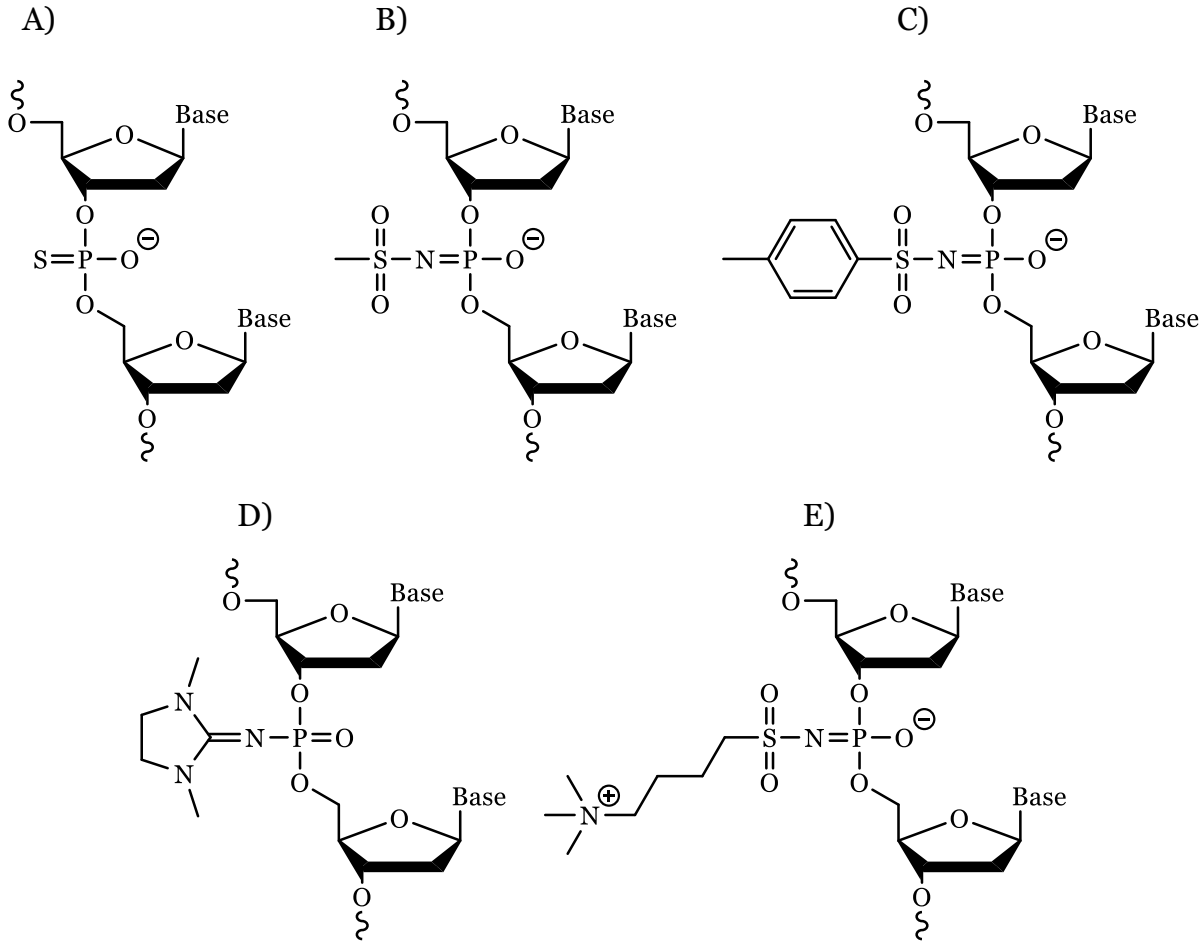
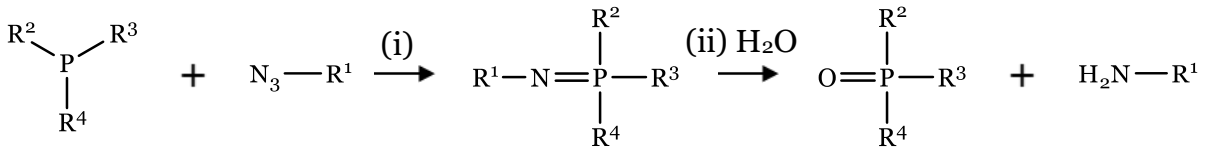


Figure 3.1. Examples of existing phosphate modifications. A) phosphorothioates, used to protect DNA from enzymatic degradation. B) μ-modification. C) Tosyl-modification. Both μ- and Tosyl-modification were used as controls for Staudinger modifications due to their minimal impact on DNA secondary structures. D) PG-modification, results in a charge-neutral phosphate allowing secondary structure formation independent of salt concentration. E) N+ modification, results in a zwitterionic DNA sequence, giving similar properties to D).

to be used to encourage duplex, triplex and G4 formation and eventual drug development following cell uptake studies.



Scheme 3.1. Standard Staudinger reaction

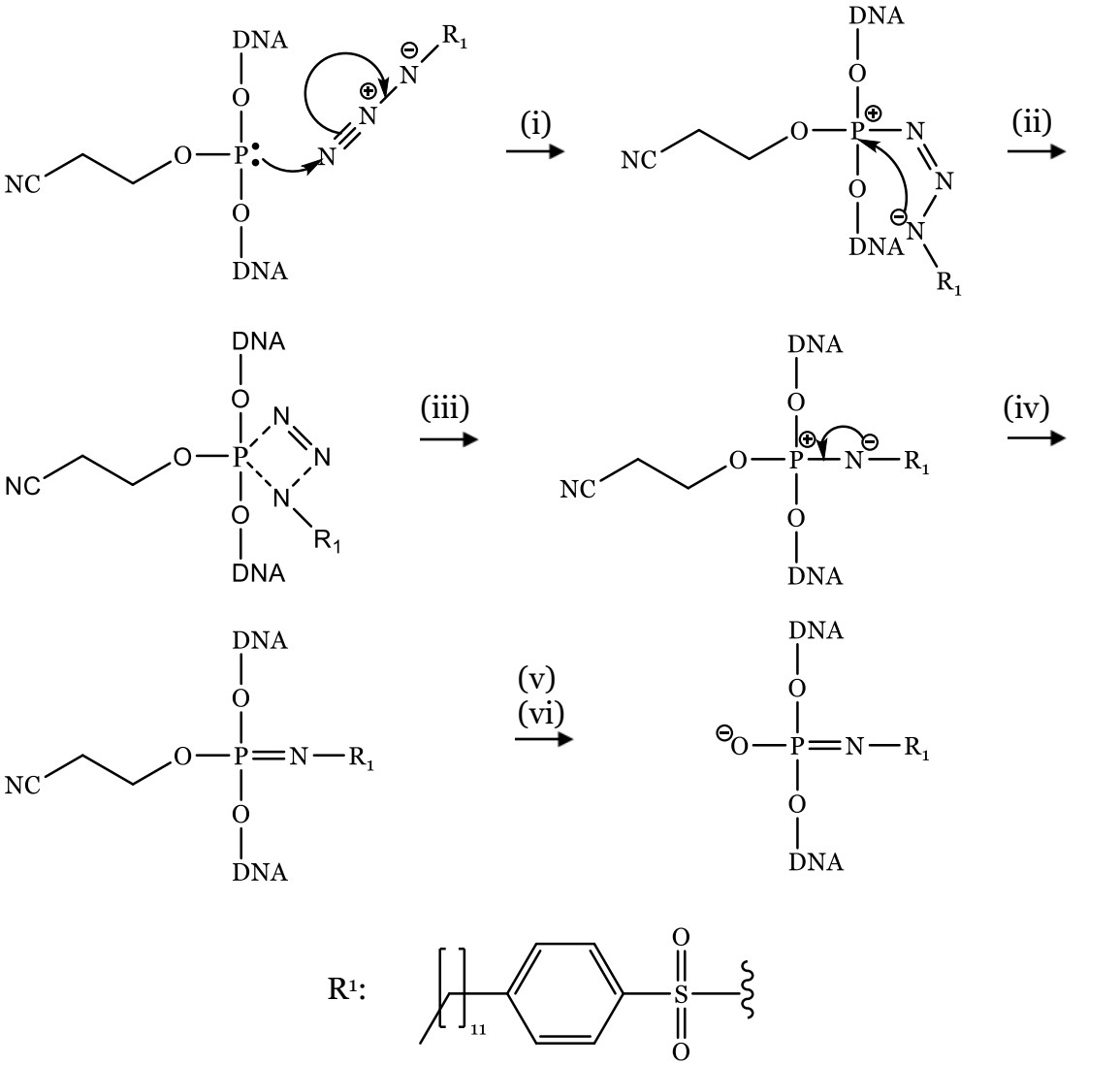


Figure 3.2. Mechanism of Staudinger reaction with the 4-dodecylbenzenesulfonyl R-group used in this synthesis. (i) Phosphorus lone pair attacks terminal nitrogen of azide. (ii – iv) Rearrangement to expel N_2 and form iminophosphorane. (v) DNA synthesis (see Chapter 7.2.1.). (vi) deprotection: 28% aq. ammonia, 55 °C, 12 hours. Conditions for Staudinger reaction: 0.1 M 4-dodecylbenzenesulfonyl azide solution in dry ACN, 37 °C, 15 mins.

Typically, Staudinger ligation is a reaction used to reduce an azide to an amine while introducing a range of functional groups. The organic azide reacts with a phosphane or phosphite, commonly triphenylphosphane, and produces an iminophosphorane (Scheme 3.1 (i)). The lone pair of the trivalent phosphorus

attacks the terminal nitrogen of the azide (Figure 3.2(i)). A subsequent rearrangement results in the expulsion of N_2 and the formation of the iminophosphorane (Figure 3.2 (ii - iv)). This would then be hydrolysed in the presence of water to produce an amine, as shown in Scheme 3.1 (ii). However, the iminophosphorane can be stabilised through resonance by including a sulfonyl group in the modification (i.e., R^1 in Scheme 3.1 and Figure 3.2). The trivalent phosphorus formed during chemical DNA synthesis (see Chapter 7.2.1. for a detailed description) is a suitable reactant for this process prior to the oxidation step. With the addition of the sulfonyl group, the entire modification is stable during DNA synthesis and deprotection (Figure 3.2 (v – vi)).[85]

This procedure has been used to produce oligonucleotides containing several modifications shown in Figure 3.1 The μ- and Ts-modification are intended to have minor impact on G4 formation and allow comparison with other modifications. Neutral oligonucleotides were produced either by changing the phosphate structure to not include a negative charge after rearrangement of the iminophosphorane (Figure 3.1D)[86] or by incorporating positively charged functional groups onto phosphates, resulting in a net charge of zero (Figure 3.1E).[85] These modifications are primarily of interest because they form secondary structures independently of salt concentration due to the considerable reduction in electrostatic repulsion between strands. This makes them ideal for encouraging formation of secondary structures such as triplexes because they negate the repulsion between the invading strand and canonical duplexes when forming Hoogsteen hydrogen-bonding arrangements. The primary method of preventing hydrolysis of the iminophosphorane in these cases was the inclusion of a sulfonyl group, as mentioned above. However, the imine shown in Figure 3.1D is also stable as it can form an iminophosphorane through resonance with the imidazolidine ring. The R-group attached to sulphur can theoretically be functionalised with a range of possible modifications allowing for a variety of chemical modalities to be introduced to DNA via this method.

This strategy was further expanded by our collaborator to stabilise canonical duplexes by incorporating a 4-dodecylbenzenesulfonyl group in the sequenced DNA. In this case, the R-group attached to the sulfonyl azide is a lipophilic carbon chain, and aggregation due to the hydrophobic effect encouraged duplex

formation, resulting in duplexes with high thermal stability.[87] This modification presented a promising new method of stabilising G-quadruplexes. The duplexes had been modified with lipophilic groups at the terminal ends, so that they would aggregate in aqueous solution and stabilise the secondary structure. We proposed it would be possible to stabilise other secondary structures rather than duplexes by incorporating lipophilic modifications at positions where only formation of a specific secondary structure would allow for aggregation. Using this strategy, we could discourage duplex formation and create thermodynamically trapped G4s.

3.2. Methodology

The investigation of G4s was based on two DNA sequences, dTGGGGT (dTG_4T) and dGGGGTTTTGGGG ($dG_4T_4G_4$) because both sequences have well reported properties.[67-71] In the presence of both Na^+ and K^+ cations TG_4T forms a parallel tetramolecular G4 (this topology is shown in Figure 3.3A), while $dG_4T_4G_4$ forms an antiparallel bimolecular G4 (this topology is shown in Figure 3.3B). dTG_4T is characterised by guanines arranged in *anti*-glycosidic configuration, while $dG_4T_4G_4$ contains mixed *anti* and *syn* glycosidic configurations. Both sequences are characterised by slow association kinetics, although $dG_4T_4G_4$ typically associates more quickly than dTG_4T due to being bimolecular.

Table 3.1. dTG_4T and $dG_4T_4G_4$ sequences synthesised and thermal stability of G4s

Name	Sequence	T_m (±1 °C, Na^+ Buffer)
TG_4T	TG_4T	70 °C
TG_4XT	$TG_{4X}T$	> 90 °C
TXG_4T	T_XG_4T	> 90 °C
TG_2XG_2T	$TG_{2X}G_2T$	52 °C$^{(ap)}$
$G_4T_4G_4$	$G_4T_4G_4$	64 °C$^{(ap)}$
$G_4T_4G_3XG$	$G_4T_4G_{3X}G$	54 °C $^{(ap)}$, 74 °C $^{(p)(i)}$
$GXG_3T_4G_4$	$G_XG_3T_4G_4$	52 °C $^{(p)}$
$GXG_3T_4G_3XG$	$G_XG_3T_4G_{3X}G$	> 90 °C $^{(ap, p)}$

All samples were stable in K^+ buffer (100 mM KCl added). Conditions: 20 μM strand concentration, 10 mM lithium cacodylate, 100 mM NaCl, pH 7.2. (p) parallel topology. (ap) antiparallel topology. $X/_X$ indicates position of phosphate modified with 4-dodecylbenzenesulfonyl imine. (i) This T_m is approximate as the sample is not completely melted at 90 °C.

We intended to synthesise several modified sequences containing the 4-dodecylbezenesulfonylimino modification. Three dTG_4T sequences were synthesised, with lipophilic moieties on phosphates between T and G at either end of the sequence and between the two central G residues in the sequence

(Table 3.1, Figure 3.3A). As described below, the results of experiments with these sequences meant we did not synthesis $dG_4T_4G_4$ with a central modification. Instead, we focused on $dG_4T_4G_4$ sequences containing lipophilic moieties at the ends of the sequence, as well as a sequence containing the modification at both ends. (Table 3.1, Figure 3.3B and C). The modifications of each sequence had slightly different goals. With TG_4T sequences we expected that aggregation of the hydrophobic chains would reinforce the naturally occurring structure because the parallel topology would result in lipophilic modifications being arranged together (Figure 3.3A). In the presence of the complementary dAC_4A sequence, the ability for lipophilic moieties to aggregate within the structures would be reduced, destabilising the duplex. $dG_4T_4G_4$ sequence would contain modifications at opposite ends of the naturally occurring G4 structure (Figure 3.3B). We expect

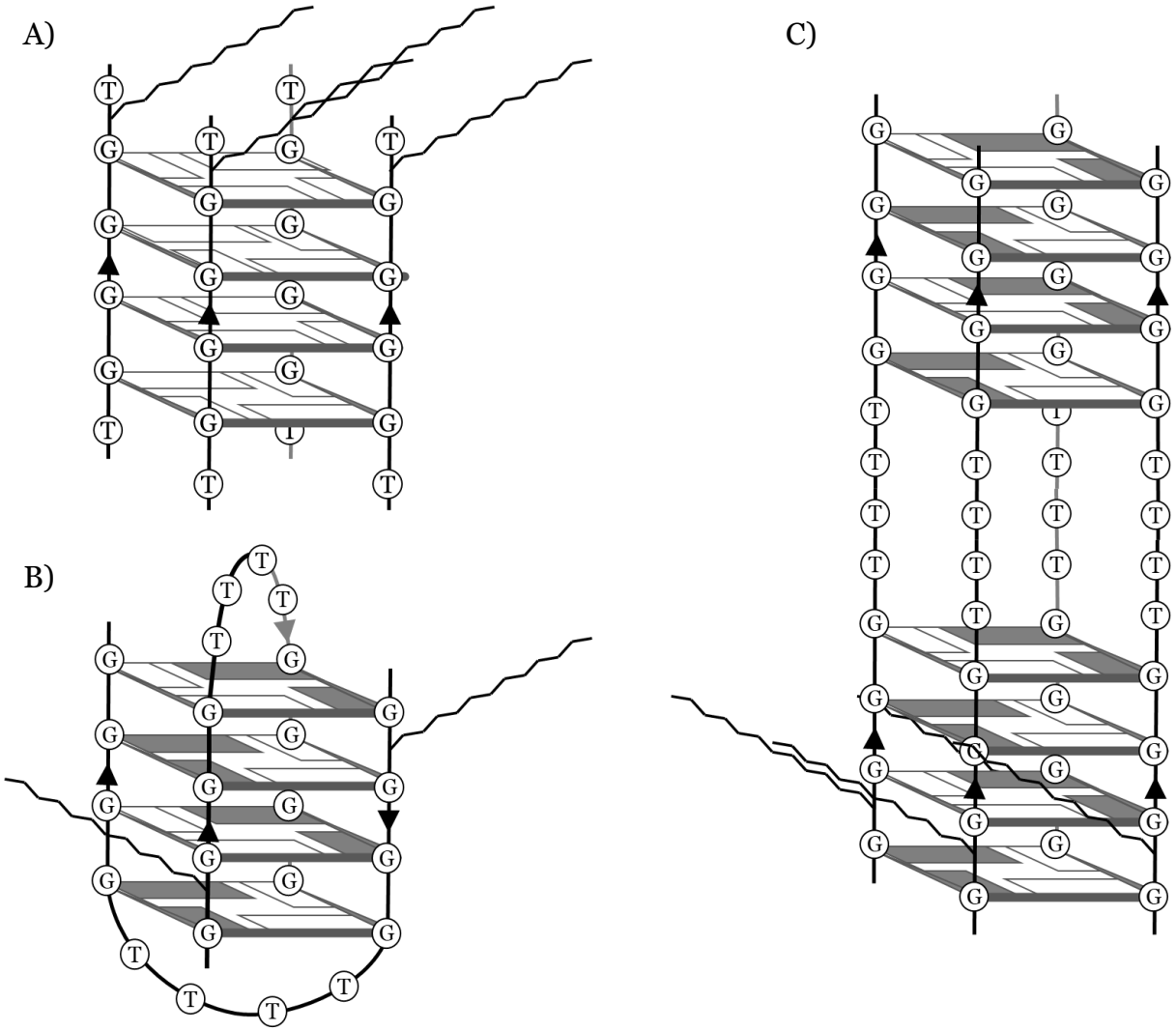


Figure 3.3. Depiction of modified G4 structures: A) Modified dTG_4T allows for aggregation of lipophilic moieties in the naturally occurring parallel G4. B) Modified $dG_4T_4G_4$ does not allow for aggregation in the naturally occurring antiparallel G4. C) A suggested alternative $dG_4T_4G_4$ topology allowing for aggregation of lipophilic moieties.

this to encourage the formation of a different topology, such as a parallel G4, which results in lipophilic moieties arranged closer together (Figure 3.3C). The expected topology is unknown since a range of new topologies might allow for more aggregation, but the naturally occurring antiparallel G4 topology is expected to be less favoured.

Staudinger ligation was carried out using a 0.1 M solution of 4-dodecylbenzenesulfonyl azide in dry acetonitrile (ACN) (Scheme 3.1, Figure 3.2). This solution was added to the column in three 700 µL injections instead of a standard iodine/pyridine oxidation step during DNA synthesis (see Chapter 7.2.1. for details of general DNA synthesis and 7.4.1. for details of DNA synthesis using 4-dodecylbenzenesulfonyl azide). This solution was passed though the column with evenly separated vacuum pulses over 15 minutes.

The modified G-rich sequences were obtained as dimethoxytrityl (DMT)-off sequences following cleavage with 28% aqueous ammonia (Figure 3.2 (ii)) and purified using reverse-phase HPLC (see Chapter 7.2.2.), which allowed the separation of the target products from unmodified and truncated sequences due to high affinity of dodecyl containing sequences to the C18 phase. However, high hydrophobicity meant that modified sequences had retention times near the end of the HPLC programme (r_t = approx. 24 mins).

When multiple modifications were attempted (i.e., $GXG_3T_4G_3XG$), this caused difficulty in separating fully modified sequences from partially modified sequences. This could be partially mitigated by maximising the yield of the Staudinger reaction by increasing the concentration of sulfonyl azide solution or performing reactions over longer times or at higher temperatures. However, as more modifications are added, the risk of partially modified sequences appearing increases. Ion-exchange HPLC does not offer a more effective purification method, as partially and fully modified species have identical net charges. At this stage it is possible to anticipate difficulties with modifying and purifying unimolecular G4s containing three or more modifications. For these applications size-exclusion HPLC or preparative dPAGE could be considered as alternative methods of purification. Alternatively, longer reverse-phase HPLC programmes or different buffer conditions could be developed to try and improve separation of highly hydrophobic modified DNA sequences.

The composition of all sequences was confirmed by ESI-MS (see Chapter 7.2.4.). To assess G4-topology we used nuclear magnetic resonance and circular dichroism spectrometry (see Chapter 7.2.3 and 7.2.5., respectively). Thermal stability studies were performed by monitoring CD profiles against increased temperature. Molecularity of G4-topologies were assessed by native ESI-MS.

3.3. Analysis of Tetra- and Bi-molecular G4 Sequences Containing Lipophilic Moieties

3.3.1. Assessment of G4 Formation

After DNA synthesis and purification, modified G-rich sequences showed the presence of a variety of alkyl chain lengths (C_{11} – C_{16}) in ESI-MS (Figure SI-33 – Figure SI-38), which was a consequence of using commercially available 4-dodecylbenzenesulfonyl azide of technical grade (90% purity, AK Scientific Inc.). These products were not separable on reverse phase and ion-exchange HPLC, so sequences were used without further purification. The effect of hydrophobic modifications on G4 stability would be similar regardless of the presence of a small quantity of shorter or longer alkyl chains.

Purity of oligonucleotides was further determined using 20% denaturing PAGE (see Chapter 7.2.7.). All dodecyl-containing TG_4T sequences showed multiple bands on the gel (Figure 3.4A). which were not present for native sequences or modified $G_4T_4G_4$ potentially indicating that these sequences were impure. We also noted that when TXG_4T was folded into G4s using a standard protocol (heating of the sample in Na^+-containing buffer, pH 7 for 5 min at 90 °C, followed by slow cooling and leaving at 4 °C overnight) an unusual CD profile for G4 structures was observed with positive ellipticities at 230, 260 and 300 nm (Figure 3.4B). By using preparative 20% dPAGE (1 mm thickness) individual bands of TXG_4T were cut out and extracted. The isolated fractions were then run again on an analytical 20% dPAGE gel and showed a single identical band, as well as the presence of multiple less concentrated bands (Figure 3.4A, right). This suggested that the isolated fractions present on dPAGE are more likely a result of aggregation of dodecyl containing TG_4T sequences than the result of impurities from synthesis. This would be particularly pronounced for dTG_4T because the hydrophobic interactions between lipophilic chains have a larger effect on the shorter oligonucleotides compared to $dG_4T_4G_4$. To disrupt these hydrophobic

aggregates, samples were heated at 90 °C for at least 1 hour in the folding buffer. Formation of the expected G4 structures was then detected by CD (Figure 3.4B) after slow cooling (100 µM strand concentration, 1 °C/min from 90 to 4 °C). Initial CD samples were retested after several months at 4 °C and formation of the G4 was also observed. This suggests that the hydrophobic aggregates are kinetic products, but the G4 is the thermodynamic product.

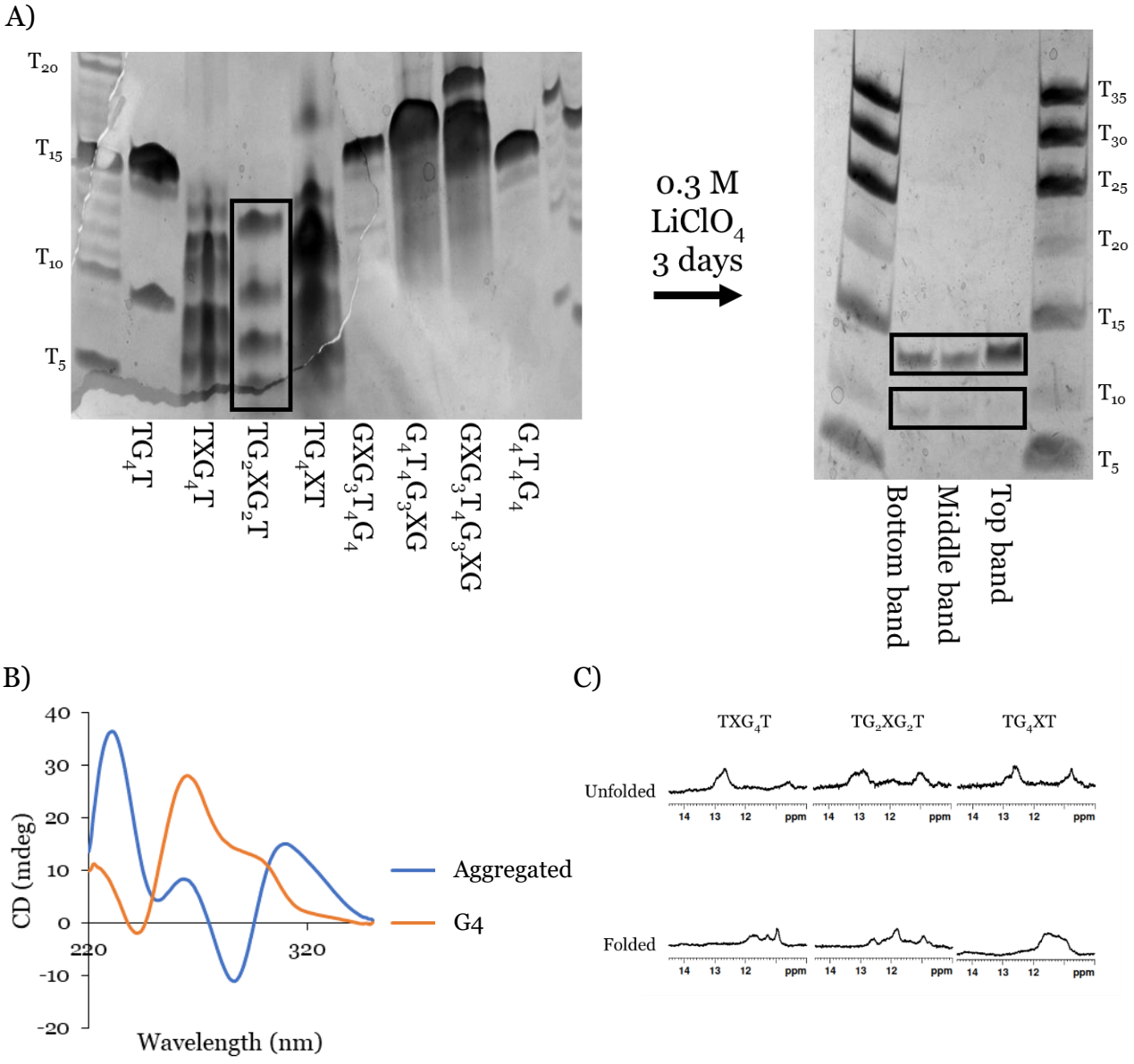


Figure 3.4. Unique structures of modified dTG_4T. A) Initial dPAGE shows multiple bands. When dTG_2XG_2T was extracted, each band gives the same result on a subsequent dPAGE, suggesting that this is the result of an equilibrium of multiple structures rather than impurities in the solution. B) CD profile of dTG_4XT. Initial formation shows primarily non-G4 secondary structure. After heating for approximately 1 hour, formation of G4 structure is observed. C) ^{1}H NMR spectra of structures shown in CD spectroscopy. Conditions: 10 µM strand concentration, 10 mM lithium cacodylate, 100 mM NaCl, pH 7.2. dPAGE: 20% acrylamide, 7 M urea, 1x TBE running buffer. NMR: 200 µM strand concentration, 20 mM sodium phosphate, 10 mM KCl, 10% D_2O, 1% TSP, pH 7.0. X = 4-dodecylbenzenesulfonyl imino phosphorane.

The presence of hydrophobic aggregates of modified dTG_4T sequences was also investigated using ^1H NMR spectrometry. G-quadruplexes typically show imino proton peaks in the 10 – 12 ppm range, while Watson-Crick base pairing gives peaks in the 12 – 14 ppm range. The initial aggregated structures primarily showed peaks around 13 ppm, indicating non-G4 hydrogen-bonding (Figure 3.4C). Broad peaks were also observed at approximately 11 ppm, suggesting some G4 formation had occurred (Figure 3.4C). These spectra suggest that some hydrogen-bonding arrangements do exist, most likely driven by the aggregation of the dodecyl modification, but not complete formation of the G4s. After heating to 90 °C and slow cooling, as described above, the CD spectra changed considerably to more closely reflect the expected spectra of a G-quadruplex (Figure 3.4B). ^1H NMR spectra of these samples agreed with the change observed in the CD spectra with a significant increase in the intensity of the broad peaks in the 11 – 12 ppm range and a corresponding disappearance of the peaks observed previously around 13 ppm. This corresponds to the pattern in ^1H NMR spectra usually observed for G-quadruplex structures and agrees with the analysis of CD spectra and dPAGE. Further analysis of the non-G4 structures is challenging because they did not fly in ESI-MS under native conditions (15% aq. MeOH, 15 mM NH_4OAc, pH 7.4).

3.3.2. Assessment of G4 Topology

With the changes to folding protocol described above, we obtained G4 assemblies containing the hydrophobic linkers. Initial observations of these structures showed significant topological changes in modified G4s. Native TG_4T gives a purely parallel topology in the presence of both Na^+ and K^+ cations. Both TG_4XT and TXG_4T showed a significant positive peak at 265 nm (Figure 3.5A and B), indicating the presence of *anti*-guanosines and a similar parallel topology. However, they also both had significant positive shoulders at 295 nm, indicating the presence of some *syn*-guanosines. The relative orientation of all DNA strands in these complexes could not be determined but both parallel and antiparallel G4 structures appeared to be present. As mentioned above, ^1H NMR gave broad imino proton peaks, suggesting that these structures are not highly symmetrical and likely contain a variety of secondary structures which results in multiple overlapping peaks. We proposed that the increased hydrophobic effect caused by the introduction of our modification was counteracted by a significant increase in

steric hindrance at the ends of sequences due to the bulky carbon chains being forced into close proximity by the shape of the G4 structure. This may result in some strands adopting the opposite orientation and explains the overall mixed G4 topology.

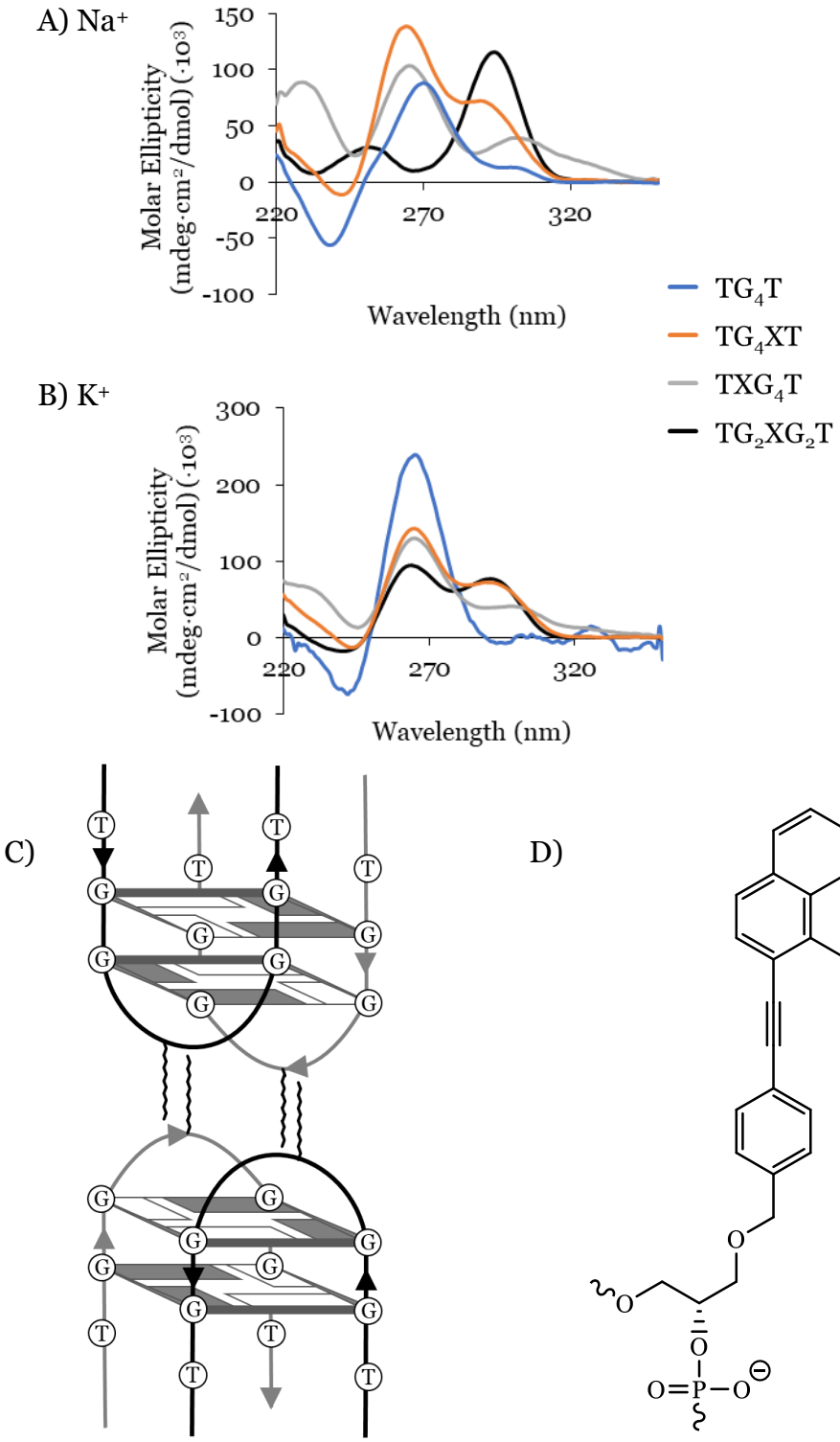


Figure 3.5. dTG_4T and modified sequences in A) Na^+ buffer and B) K^+ buffer. (C) Proposed structure of dTG_2XG_2T G4. (D) Structure of twisted intercalating nucleic acid (TINA) moiety Conditions: 10 µM strand concentration, 10 mM lithium cacodylate, 100 mM NaCl or KCl, pH 7.2.

TG_2XG_2T, with a dodecyl moiety in the middle of the G-tract, showed an almost complete change in topology. In the presence of K^+, it showed a similar mixed topology to the other modified TG_4T sequences, with positive peaks at 265 and 295 nm, suggesting the presence of both *syn*- and *anti*-guanosines. However, in the presence of Na^+ an almost complete switch to a positive peak at 295 nm was observed (Figure 3.5A). This indicates the presence of *syn*-guanosines and resembles the CD profile of an antiparallel G4. We proposed that this indicated the formation of a bimolecular antiparallel dimer (Figure 3.5C). Similar CD results were obtained previously for comparable sequences containing twisted intercalating nucleic acid (TINA) modification (Figure 3.5D),[88] which is even more hydrophobic. These sequences, when analysed using ^{1}H NMR spectroscopy, were determined to fold into G4 structures as shown in Figure 3.5C. The central hydrophobic groups meant that this topology was more favourable than the expected native tetramolecular parallel structure, as it better allows for the hydrophobic groups to aggregate. CD spectra obtained of modified TG_4T sequences did not show the expected increased favourability for the parallel G4, but they do demonstrate that this type of hydrophobic modification is capable of inducing a significant change in G4 topology.

The outcome observed for $G_4T_4G_4$ modifications was closer to expectations, with some sequences shifting towards a parallel topology. Native $G_4T_4G_4$ forms a characteristic antiparallel G4 as indicated by CD with positive peaks at 295 and 240 nm and a negative peak at 260 nm. In Na^+ buffer the CD spectrum of $GXG_3T_4G_4$ closely resembles this native topology (Figure 3.6A), but when K^+ was present, a new positive peak at 265 nm appeared (Figure 3.6B), indicating

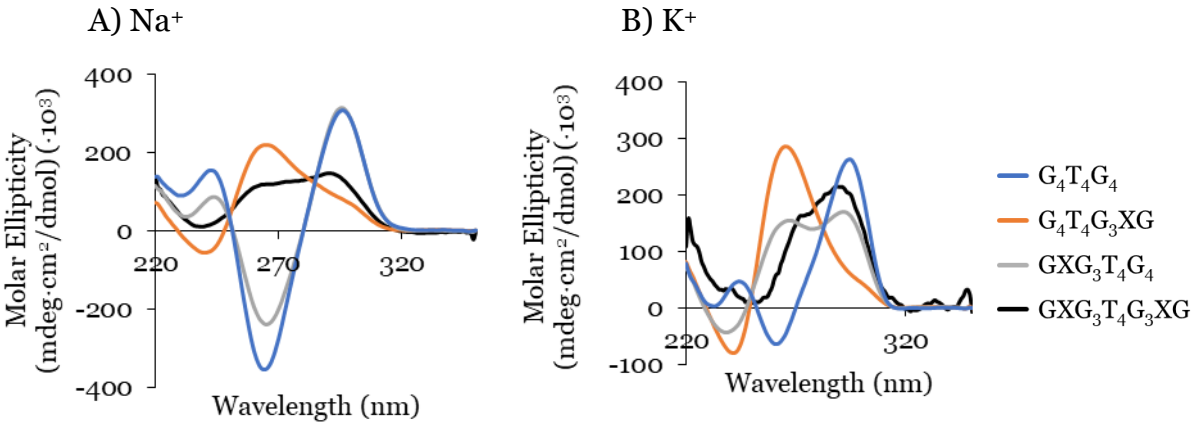


Figure 3.6. $dG_4T_4G_4$ and modified sequences in A) Na^+ buffer and B) K^+ buffer. Conditions: 10 μM strand concentration, 10 mM lithium cacodylate, 100 mM NaCl or KCl, pH 7.2.

formation of G4s of parallel topology. Similarly mixed topologies were also observed for $G_4T_4G_3XG$ and $GXG_3T_4G_3XG$ according to CD spectra, even in the presence of Na^+. $GXG_3T_4G_3XG$ also showed mixed topology in K^+ buffer, but $G_4T_4G_3XG$ had a CD spectrum resembling a G4 of parallel topology. Surprisingly, the 3′-modification showed the greatest change in the topology of those tested for both Na^+ and especially K^+, rather than the sequence containing two modifications. Once again, we propose that this was due to interactions between the lipophilic chains at the ends of sequences, similar to the observations of TXG_4T and TG_4XT.

Overall, this suggests that careful selection of hydrophobic modification sites is more important than the overall number of modifications for stabilising parallel G4 topologies. Both TG_4T and $G_4T_4G_4$ modifications resulted in distortions of the expected G4 when multiple lipophilic chains were added in the structures, which could suggest that steric effects are a more important consideration than expected. Native ESI-MS (see Figure SI-10) was used to test the molecularity of both modified G4s. Obtaining peaks corresponding to DNA complexes in mass spectrometry is challenging because complexes often dissociate during ionisation. Better results can sometimes be obtained by reducing the temperature and voltage used during ionisation, but this can also result in samples not appearing in mass spectrometry at all. The results presented in Figure SI-10 show the clearest indication of the G4 molecularity obtained for modified TG_4T and $G_4T_4G_4$ sequences. TG_4XT showed masses corresponding to a tetramolecular complex, while $GXG_3T_4G_4$ showed masses corresponding to bimolecular G4s. Other sequences gave significantly less indication of G4 complexes being present, even under the native conditions described. It is likely that if tetramolecular arrangements of lipophilic $G_4T_4G_4$ were formed, they dissociate under mass-spectrometry conditions and therefore are not detectable. Typically, more reliable determination of secondary structure molecularity can be obtained using native PAGE, which is discussed in Chapter 3.3.4. However, G4 structures tend to behave differently on gels depending on topology, meaning native PAGE is similarly unreliable for obtaining precise molecularities.

An initial assessment of the G4s formed from G-rich sequences containing lipophilic modifications showed that they did not support formation of the

expected topologies. Specifically, it appears that unfavourable interactions of the lipophilic chains may induce some change in guanosine conformation and potentially G4 topology. While this did not support our initial expectations of these modifications, it does suggest that hydrophobic moieties could be an effective strategy for controlling G4 topology in some sequences. In particular, modification of $dG_4T_4G_4$ seemed to encourage formation of the parallel topology, with $G_4T_4G_3XG$ in particular forming a mostly parallel G4 in K^+ buffer.

3.3.3. Assessment of G4 Thermal Stability

The thermal stability of DNA secondary structures is a useful metric for determining their overall stability. Increased thermal stability of G4 structures could indicate a corresponding reduction in formation of the canonical duplex when complementary DNA is added. Our expectation was that hydrophobic modifications would increase G4 thermodynamic stability by providing additional support for the structure through the hydrophobic effect.

No change in CD signal was observed for any of the modified G4-forming sequences in K^+ buffer, even at 90 °C. The thermal stability of TXG_4T and TG_4XT increased significantly in Na^+-containing buffer, both giving T_m values in excess of 90 °C (Table 3.1, Figure SI-11B and C). The uniquely structured TG_2XG_2T was an exception, with a significant decrease in T_m. This is likely due to a change from four G-tetrads per G4 to just two. While the structure was changed by these modifications, CD melting profiles showed the peaks decreasing simultaneously (Figure SI-11D). This suggested that we were observing a single mixed topology, rather than distinct structures with purely parallel and antiparallel topologies. This agrees with our previous conclusion that the dodecyl monomer in the middle of a dTG_4T sequence converts a predominantly parallel tetramolecular G-quadruplex into an assembly composed of two identical antiparallel G-quadruplex subunits stacked via dodecyl-dodecyl interfaces (Figure 3.5C). Overall, these results suggested that addition of large hydrophobic moieties has the expected effect on TXG_4T and TG_4XT G-quadruplex stability.

The T_m value for $G_4T_4G_3XG$ was determined using the peaks in CD spectra corresponding to both antiparallel and parallel topologies, because each topology had different properties. Both $GXG_3T_4G_4$ and $G_4T_4G_3XG$ showed decreased thermal stability in Na^+ buffer for the peaks corresponding to an antiparallel G4

at 295 nm, from 64 °C to 50 and 54 °C respectively (Table 3.1, Figure SI-12B and C). On the other hand, the peak of $G_4T_4G_3XG$ at 265 nm, corresponding to a parallel G4 topology, gave slightly higher thermal stability of 74 °C compared to the native sequence (Table 3.1, Figure SI-12C). The 3′ + 5′-dodecyl sequence had a mixed topology, but both structures had T_m values exceeding 90 °C in Na^+ buffer, indicating a significant increase in thermal stability (Table 3.1, Figure SI-12D). While antiparallel G4s were formed, the thermal stability of the native topologies was decreased relative to the unmodified sequence. Conversely, when parallel G4s were observed, the thermal stability increased compared to the unmodified sequence. Despite the decrease in thermal stability for the naturally occurring topology, introduction of lipophilic moieties was only sufficient to stabilise the parallel topology when a 3′-modification was present. This suggests that we may be observing competition between the more favourable hydrogen-bonding arrangement of the antiparallel G4, and the more favourable hydrophobic interactions allowed by a parallel topology of $G_4T_4G_4$ due to the lipophilic chains. At higher temperatures, the hydrogen bonds begin to break, and hydrophobic interactions become the dominant force driving secondary structure formation, resulting in the parallel topology having a higher T_m. We do not see such a dramatic change for modified TG_4T sequences because the hydrophobic interactions were intended to reinforce the hydrogen-bonding arrangements rather than challenge them. This could mean that this is not an ideal method for driving the formation of unfavourable topologies but is more useful as a method for reinforcing existing structures. This is seen, for example, when these modifications are used to disrupt duplex formation, as described below.

3.3.4. Effect of Lipophilic Modification on Duplex Formation

While lipophilic chains had a significant effect on G4 topology, I also hypothesised that increasing the thermal stability of G4s would reduce the formation of WC duplexes in the presence of the complementary strand. Native PAGE was used to compare the mobility of native $G_4T_4G_4$ and the modified G4s in the presence of the unmodified $C_4A_4C_4$ strand (Figure 3.7). Native $G_4T_4G_4$ (lane 2) migrates as a single band below the T_{10} component of the ladder in both Na^+ and K^+ buffers. Upon addition of the complementary stand $C_4A_4C_4$ a new retarded band was observed at around the T15 component of the ladder suggesting some formation of the duplex (lane 3 and 4). The single-stranded complementary sequence, $C_4A_4C_4$, can be seen in lane 8, with a similar mobility to the $G_4T_4G_4$ G4. When the sample was heated, an equilibrium was established between G4 and duplex structures, as seen in lane 3. The modified G4s in the absence of complementary strand can be seen in lanes 5, 11 and 14, but these sequences are far less mobile on the gel, with considerable aggregation in the well. However, some duplex formation can be observed in Na^+ buffer, with new bands appearing around the T_{20} ladder component in lanes 6, 7, 9, 10, 12 and 13 (Figure 3.7A). However, in K^+ buffer this effect is significantly reduced. While the native sequence forms a similar equilibrium after heating in the presence of the

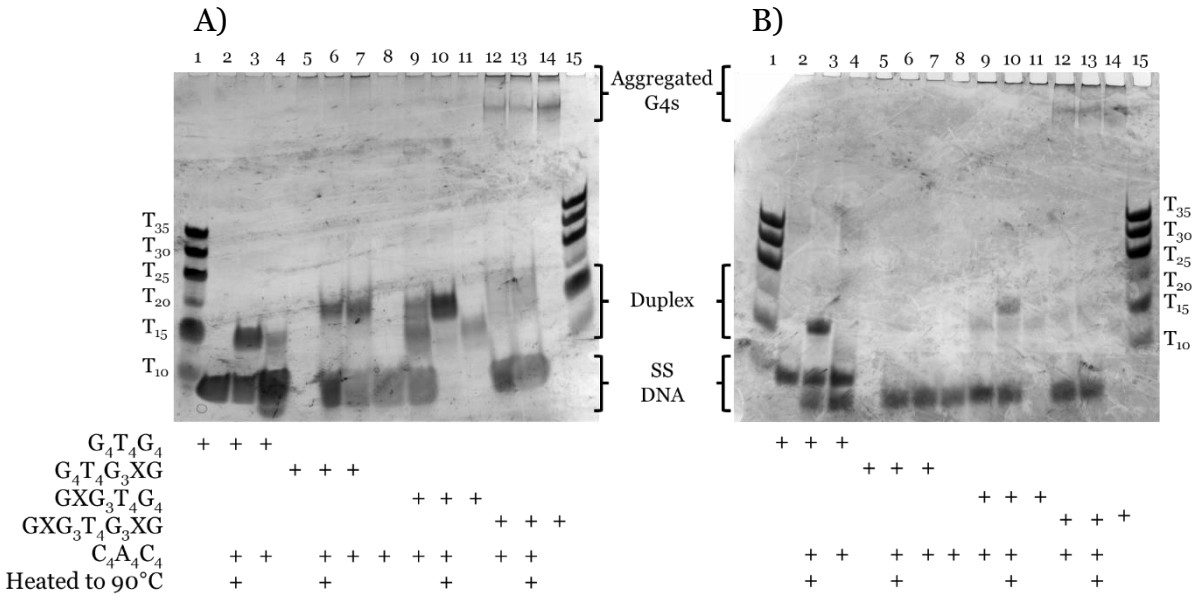


Figure 3.7. Native PAGE of $dG_4T_4G_4$ sequences challenged with complementary DNA. A) Na^+ buffer. B) K^+ buffer. Formation of duplex is observed for unmodified sequence, but only minimal duplex formation is observed for modified sequences. However, modified sequences, particularly $GXG_3T_4G_3XG$, show considerable aggregation in wells due to their hydrophobicity. Conditions: 100 μM strand concentration, 10 mM lithium cacodylate, 100 mM NaCl or KCl, pH 7.2, 20% acrylamide gel.

complementary strand, very little change is observed for modified sequences (Figure 3.7B). A faint band can be seen around T_{20} in lanes 9 and 10, suggesting some duplex formation with $GXG_3T_4G_4$, but not in any of the other samples. Overall, this suggests that, as expected, dodecyl modifications had the effect of stabilising G4 structures and, consequently, reducing duplex formation.

3.4. Conclusions and Perspectives

The Staudinger reaction was successfully used to modify G4 DNA sequences with 4-dodecylbenzenesulfonyl azide. The introduction of lipophilic moieties to the native phosphates altered the topology of G-quadruplexes depending on the site of modification. The modified TG_4T sequences TXG_4T and TG_4XT both showed significantly increased thermal stability, but also preference for *syn*-guanosines. Additionally, TG_2XG_2T completely changed topology and based on previous experiments with hydrophobic TINA modifications, we proposed a dimeric structure containing bimolecular antiparallel G4s. This new structure showed a drastic decrease in thermal stability compared to the native sequence.

The modified $G_4T_4G_4$ sequences all showed significant topological changes, giving mixtures of the antiparallel topology seen in native sequences and a parallel topology. However, the thermal stability of the antiparallel topology was lower than the native sequence, while the thermal stability of the parallel topology, when it could be measured independently, was higher than that of the native sequence. Sequences containing the phosphate modification at the 3′-end, particularly $G_4T_4G_3XG$, tended to show greater favourability for the parallel topology,

When these sequences were challenged with a complementary sequence, they showed some duplex formation in Na^+ buffer, but minimal duplex formation in K^+ buffer. This suggested that while these modifications did not encourage the formation of particular G4 topologies as effectively as we had expected, they did disrupt the formation of canonical duplexes. The modifications appeared, in many cases, to encourage a mixed G4 topology (e.g., a 3 + 1 topology containing three parallel strands and one antiparallel strand).

Analysis of these results indicates that 4-dodecylbenzenesulfonyl imine or similar hydrophobic phosphate modifications could be ideal methods of

stabilising selected DNA secondary structures or topologies. However, we encountered two significant problems with these modifications for achieving our aims, and these problems prevented further experimentation.

Firstly, we encountered issues purifying highly hydrophobic modified sequences using HPLC. While TG_4T is a useful model for studying these interactions, because it is easy to synthesise and data on its properties are readily available, many of the proteins that these strategies are intended to study target unimolecular G4s. Synthesising unimolecular G4-forming sequences, such as the Pu39, c-MYC or c-KIT, would require at least two modifications in each sequence. For sequences such as Pu39 (39 nucleotides) or TERRA45 (45 nucleotides) even more modifications could be necessary. Further optimisation of Staudinger ligation conditions could overcome this issue, but it also presents challenges for purification. Purification of DNA is, in part, necessary to remove short sequences which result from failed coupling reactions. Sequences containing fewer nucleotides have different mobility in both reverse-phase and ion-exchange HPLC due to their difference in length and therefore hydrophobicity and charge. However, inclusion of hydrophobic modifications meant that these sequences were not separable using standard HPLC purification conditions. Both TG_4T and $G_4T_4G_4$ are short enough that the number of incomplete sequences is low, but this becomes a greater concern with sequences such as Pu39 or TERRA45. Ultimately, these issues can be overcome with careful consideration of Staudinger ligation reaction protocol and purification conditions but expanding modifications with 4-dodecylbenzenesulfonyl azide to more complex sequences will require extensive testing.

Secondly, lipophilic moieties did not afford the level of topological control that we had anticipated. We did observe a shift towards parallel topologies for $G_4T_4G_4$ modifications, but antiparallel topologies were still present. TG_4T modifications were intended to reinforce the parallel G4, but CD spectrometry indicated the presence of *syn*-guanosine, possibly the result of partial antiparallel G4 formation. We suggested that these issues were caused by steric hindrance, as the formation of some G4 topologies necessitated dense packing of the hydrophobic groups. These issues could be addressed for unimolecular G4s by adjusting the number of hydrophobic groups added, but the results of $G_4T_4G_4$

modifications suggested that this might not be enough to completely control the G4 topology.

Additionally, a significant issue we highlighted with G4 ligands is their tendency to bind to similar sites to many G4-specific proteins. A similar issue is possible with lipophilic modifications, as their size potentially leads to them blocking protein binding regardless. As mentioned previously, the specificity of proteins such as HP1α for parallel G4s is potentially due to the lateral and diagonal loops of antiparallel G4s blocking the top and bottom of G4 structures, and therefore minimising protein interactions. Large moieties, such as 4-dodecylbenzenesulfonylimino phosphate modifications could present the same issue for blocking protein binding.

Further experimentation with these modifications, both in terms of G4 structures and other DNA secondary structures, could be warranted. Shorter carbon chains could be incorporated to minimise unfavourable interactions between chains. Less hydrophobic modifications may also allow for more effective separation of oligonucleotides using HPLC. Alternatively, further investigation could focus on improving the synthetic and purification conditions in the context of multiple phosphate modifications within unimolecular sequences. This may require longer HPLC programs than the standard purification method used here, or the use of different buffer conditions to improve separation. There is also some potential to test interactions of the modified G4s with other proteins. This thesis focuses primarily on HP1α, which did not show affinity for multimeric G4s structures in other experiments, but modified TG_4T, $G_4T_4G_4$ or similar sequences could be investigated as inhibitors of other proteins. A more detailed discussion of the potential development of this strategy is included in Chapter 6.

Overall, hydrophobic modifications are a useful method of controlling the formation of DNA secondary structures, but the challenges with purification and limited and unpredictable topological control afforded made them unappealing to investigate further in this thesis. Compared to the ligands discussed in Chapter 2 we observed a similar ability to partially disrupt duplex formation, but also similar disruption of the naturally occurring G4 topology, and several synthetic challenges which could limit their application to other G4-forming sequences.

4. Incorporation of Inverted- and α-Nucleotides to Introduce Canonical Base-Pairing Mismatches

4.1. Introduction

The use of ligands and hydrophobic modifications involves changes exterior to the standard G4 structure. One disadvantage of this approach is the potential to block protein-binding sites. A more effective possible approach is to change the nucleotides themselves to create G4s which most closely mimic the original G4 structure. In this and the next chapters we consider these strategies from two angles. Non-native nucleotides, in this context, refer to the incorporation of any nucleotides which do not appear in DNA within the genome. One example, which will be explored later, are α-nucleotides (Figure 4.1A) which feature the opposite stereochemistry at the 1′-position to standard nucleotides. Sequences containing these nucleotides prefer to form parallel duplexes rather than antiparallel.[89-94] In previously published research we had shown that the inclusion of α-D-2′-deoxyguanosine (dG) resulted in formation of parallel G4s over antiparallel G4s.[95] Other nucleotides are produced synthetically, such as 2′-fluoroarabinonucleic acid (Figure 4.1B) or 5-bromouracil (Figure 4.1C). 2′-Fluoroarabinonucleic acid has been used to produce kinetically trapped G4s.[64, 65] 5-Bromouracil is used to obtain some X-ray structures as it has a minimal effect on DNA structure and acts as an anomalous scatterer to improve contrast.[18] In Chapter 5 we also explore the use of modified nucleotides intended for further modification after DNA synthesis.

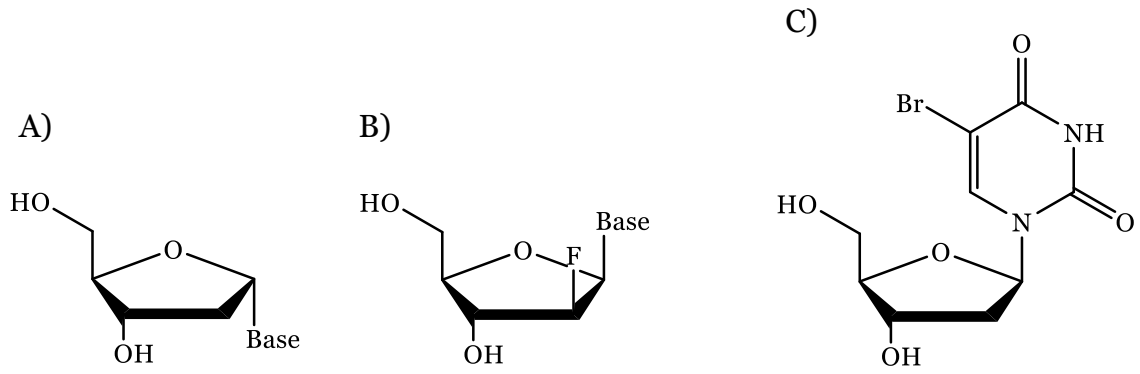


Figure 4.1. Non-native nucleosides with useful properties for DNA modifications. A) α-Anomers of standard nucleosides, featuring inverted stereochemistry at the 1′-position. B) 2′-Fluoroarabinonucleic acid. This modification has previously been shown to result in kinetically trapped G4s, but not thermodynamically trapped. C) 5′-Bromouracil. This modification has been used to obtain X-ray crystal structures of DNA secondary structures because it behaves as an anomalous X-ray scatterer.

It was proposed to replace native nucleotides with non-native nucleotides in G4-forming sequences, in positions that stabilise G4s but introduce mismatches in DNA duplexes. As mentioned above, G4-forming sequences containing α-nucleotides had previously shown the ability to form parallel G4s and duplex-forming sequences composed of α-nucleotides preferred to form parallel duplexes with sequences containing native nucleotides. I therefore hypothesised that synthesising G4-forming sequences containing some α- or inverted-nucleotides would destabilise duplex formation significantly more than G4 formation. This could then produce thermodynamically trapped G4s, as shown in Figure 4.2A and B. Additionally, because the base, sugar and phosphate are effectively unchanged compared to unmodified G4s, protein interactions should

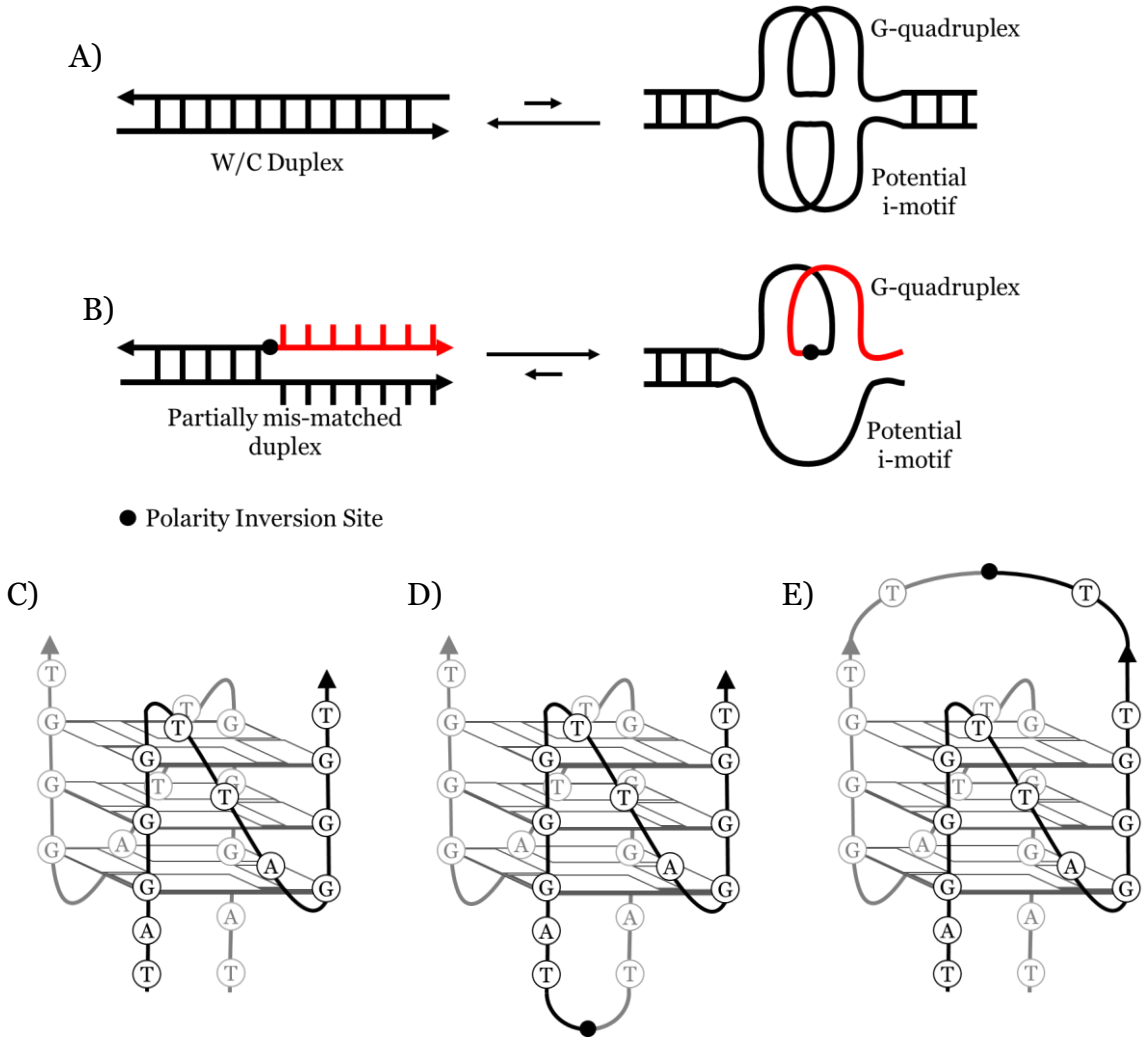


Figure 4.2. A) Unmodified DNA sequences fold into an equilibrium favouring canonical duplex structures, as discussed in Chapter 1. B) Introduction of modified sites creates partially mismatched DNA, resulting in destabilised canonical duplexes and more favourable G4 structures. C) Unmodified tel sequence, TAG_3TTAG_3T. D) Example of proposed tel modification containing a 5′-inversion, 5′inv-tel. E) example of proposed tel modification containing a 3′-inversion, 3′inv-telT.

also be unchanged, avoiding many of the problems highlighted for ligands and hydrophobic modifications in Chapters 2 and 3.

Native nucleotides are incorporated into DNA using standard phosphoramidites containing a 5′-DMT protecting group and a 3′-phosphoramidite (Figure 4.3A, see Chapter 7.2.1 for full details of DNA synthesis.). Multiple non-native nucleotides are also commercially available as phosphoramidites, but in this chapter we focus on two: inverted 3′-DMT-5′-phosphoramidites (Figure 4.3B) and anomeric 5′-DMT-3′-α-phosphoramidites (Figure 4.3C). To test the ability of these modifications to disrupt duplex formation, I incorporated inverted nucleotides into a telomeric G4 sequence. In canonical duplexes, the two strands are arranged in opposite directions with the 5′-end of one strand opposite the 3′-end of the other, with each strand only running in one direction. DNA synthesis typically starts at the 3′-end and proceeds to the 5′-end using 5′-DMT-3′-phosphoramidites (Figure 4.3A). Between each nucleotide addition the 5′-DMT group is removed, allowing the new nucleotide to be added to the deprotected 5′-OH of the previous nucleotide (i.e., at the 5′-end of the sequence). However, 3′-DMT-5′-phosphoramidites (Figure 4.3B) are commercially available and make it possible to instead react a 5′-phosphoramidite with a free 5′-OH (or a free 3′-OH with a 5′-DMT-3′-phosphoramidite) reversing the direction of the sequence. Oligonucleotides containing inverted thymidine nucleotide at the 3′-end are commercially available and widely used to prevent 3′-exonuclease activity by effectively concealing the 3′-end of a sequence.[96, 97] However, it can also be used to change polarity for other purposes, and this has already been applied to G4s in some

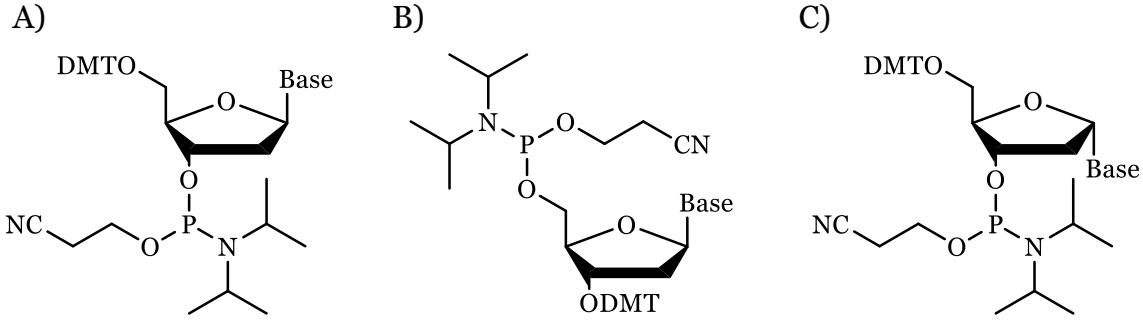


Figure 4.3. Three commercially available types of phosphoramidite which can be used in DNA synthesis: A) standard 5′-DMT-3′-phosphoramidites. B) inverted 3′-DMT-5′-phosphoramidites. C) 5′-DMT-3′-α-phosphoramidites.

contexts. We aim to explore broader applications of these modifications in creating G4 models.

The effect of polarity inversions on G4 stability and topology has been investigated in the context of tetramolecular G4s formed by TG_nT (n = 3-6) and the monomeric thrombin binding aptamer (TBA, $TG_2T_2G_2TGTG_2T_2G_2T$).[98] In the case of TBA, which forms a unimolecular, antiparallel G4, a single G-repeat was modified with inverted nucleotides. NMR data indicated the topology changed from antiparallel to a 3+1 hybrid arrangement, but CD data indicated no change. This suggests that while the effective orientation of the guanosine nucleotides was changed, the overall G4 topology and glycosidic bond conformation was conserved. This is of particular importance to our project because it suggests that, even if modified guanosine residues result in different hydrogen-bonding arrangements, a consistent G4 topology is still obtained. Modified TG_nT sequences showed similar properties,[98-101] but these sequences are less interesting in a biological context, where short tetramolecular G4s are uncommon. The studies of TG_nT did focus on changes in glycosidic bond conformation of adjacent guanosine residues. While some changes in glycosidic bond angles were observed, modified sequences ultimately formed parallel tetramolecular G4s, retaining the topology of the unmodified strand. It is unclear if minimal shifts in glycosidic bond angle will have a significant impact on the G4 topology of other sequences or their interactions with proteins. Regardless, the existence of multiple examples of topological conservation in G4 sequences modified with inverted nucleotides is promising for the ability of inversions to maintain G4 topology while disrupting duplex formation.

The initial experiments focused on the tel sequence (Table 4.1, Figure 4.2C) introduced in Chapter 1 and studied in relation to ligand interactions in Chapter 2. The symmetrical nature of this assembly makes it an ideal G4 to study as it provides simplified NMR spectra. The two sequences run in opposite directions from the 5′-end and visual analysis of the characterised structure indicates that the terminal 5′-dT nucleotides are close to each other in space. We proposed to effectively link the two strands with a native phosphate by incorporating a polarity inversion. This would potentially create a new diagonal loop between the ends of the two strands (Figure 4.2D, sequence 5′inv-tel, Table 4.1). We expected

that the original G4 structure would be maintained while reinforcing the parallel topology and reducing formation of the canonical duplex in the presence of the complementary strand (Figure 4.2).

Table 4.1. Native and inverted G-Quadruplex forming sequences based on human telomeric repeat

Abbreviation	Sequence	Complementary Sequence	Abbreviation of Complementary Sequence
tel	5′-$(TAG_3T)_2$-3′	5′-$(AC_3TA)_2$-3′	i-tel
Ttel	5′-T$(TAG_3T)_2$-3′	5′- $(AC_3TA)_2$A-3′	i-Ttel
telT	5′-$(TAG_3T)_2$T-3′	5′-A$(AC_3TA)_2$-3′	i-telT
telTA	5′-$(TAG_3T)_2$TA-3′	5′-TA$(AC_3TA)_2$-3′	i-telTA
5′inv-tel	3′-$(TG_3AT)_2$-5′—5′-$(TAG_3T)_2$ -3′	5′- $(AC_3TA)_2(ATC_3A)_2$-3′	i-5′inv-tel
5′inv-Ttel	3′-$(TG_3AT)_2$T-5′—5′-T$(TAG_3T)_2$ -3′	5′- $(AC_3TA)_2$AA$(ATC_3A)_2$-3′	i-5′inv-Ttel
3′inv-telT	5′-$(TAG_3T)_2$T-3′—3′-T$(TG_3AT)_2$-5′	5′- $(ATC_3A)_2$AA$(AC_3TA)_2$-3′	i-3′inv-telT
3′inv-telTA	5′-$(TAG_3T)_2$TA-3′—3′-AT$(TG_3AT)_2$-5′	5′- $(ATC_3A)_2$ATTA$(AC_3TA)_2$-3′	i-3′inv-telTA
Ttel-2rev	5′-$(TG_3AT)_2$TT$(TAG_3T)_2$-3′	5′- $(AC_3TA)_2$AA$(ATC_3A)_2$-3′	i-5′inv-Ttel

To account for any increase in steric hindrance caused by the introduction of this additional loop, we also designed an elongated sequence to allow added flexibility (sequence 5′inv-Ttel, Table 4.1). We also tested sequences that contained an inversion at the 3′-end instead (Figure 4.2E), with both one and two-nucleotide extensions since the 3′-ends of the native sequence were further apart than at the 5′-end (sequences 3′inv-telT and 3′inv-telTA, Table 4.1). The unmodified sequences, forming bimolecular G4s, were used as controls in our study. We refer to the non-inverted telomeric sequence, as shown in Table 4.1, as tel and the inverted sequences according to the location of their inversions (i.e., 5′inv-tel or 3′inv-telT). Elongated sequences are referenced by the location and nucleotide(s) used for elongation (e.g., Ttel is the tel sequence with an additional thymidine at the 5′-end, whereas telT refers to the tel sequence having extra T at the 3′-end). We also anticipated that the increased strand length and formation of uni- rather than bimolecular G4s would increase thermal stability of resulting complexes. To ensure that the observed effects were a result of polarity inversion, we used a control sequence, called Ttel-2rev in which two Ttel sequences are running in opposite directions (sequence Ttel-2rev, Table 4.1) with no polarity inversions but mimicking the nucleotide composition of 5′inv-Ttel. DNA

sequences complementary to G4s are provided in Table 4.1. Their names include prefix “i-” to indicate C-rich sequences with the potential to form i-motifs.

These experiments focused on the modification of a bimolecular G4 to create a new loop. I hypothesised that DNA duplexes containing a folded thermodynamically stable G4 in the presence of a complementary C-rich strand could be obtained by introducing a polarity inversion in the G-rich sequence, thereby creating mismatches in a canonical duplex without affecting G4 formation (Figure 4.2). Later modifications (see Chapter 4.3) would include modifications of unimolecular sequences with more specifically targeted modifications. In canonical duplexes all bases are involved in hydrogen-bonding, whereas G4s generally only contain hydrogen bonds within the G-tetrad (with some exceptions), so we proposed that modifications made within loop regions would have less effect on G4 stability while significantly destabilising duplex DNA. Initial tests focus on bimolecular tel G4s to determine if inverted nucleotides have the intended effect on G4 structure and topology as well as duplex formation, while later tests focus on developing additional basic sequences to be developed into models for experiments with proteins.

4.2. Modification of a Bimolecular Telomeric G4 with Inverted Nucleotides

4.2.1. Synthesis of Modified tel Sequences and Investigation of their G4 Structure

Initially, sequences containing internal polarity inversions were obtained from Integrated DNA Technologies Inc., to perform initial tests on modified sequences. Preliminary experiments, primarily CD spectroscopy and native PAGE (see Chapter 7.2.6. and 6.2.7., respectively), were promising, but larger quantities were needed for repetition of PAGE experiments and further experiments using ^{1}H NMR (see Chapter 7.2.3.). At this stage, 3′-DMT-5′-phosphoramidites were obtained from Chemgenes Corporation for DNA synthesis with inverted phosphoramidites. Ultimately, addition of inverted phosphoramidites was achieved using standard DNA protocols, but coupling times for each inverted base was increased from two minutes to five minutes, as described in Chapter 7.4.2. Oligonucleotides were obtained with reasonable yields. Mass spectrometry (protocol described in Chapter 7.2.4., Figure SI-39)

and denaturing PAGE (dPAGE, 7 M urea, Figure SI-13) were used to confirm the composition and purity of modified G-rich sequences. On dPAGE all sequences had mobility corresponding to their size with mobility close to T_{12-14} oligothymidylates for native sequences (Lanes 2, 5 and 9, Figure SI-13) and T_{24-28} for inverted sequences (Lanes 3, 6 and 10, Figure SI-13). Mobility of both short and long complementary sequences was also evaluated on dPAGE (short: Lanes 4, 7 and 11; long: Lanes 8 and 12, Figure SI-13).

The initial investigation focused on the influence of the inverted nucleotides on the G4 topology, molecularity and thermal stability using circular dichroism (CD) spectroscopy and non-denaturing PAGE. The unmodified G4s showed the characteristic CD peaks of antiparallel G4s in Na^+ buffer (Figure 4.4A), but they formed a mixed topology in K^+ buffer with a positive peak at 265 nm, negative peak at 240 nm and a positive shoulder or a peak at 290 nm (Figure 4.4B). A control sequence Ttel-2rev had the signature spectra of an antiparallel G4 in both buffers, a significant difference to the native bimolecular G4s. In Na^+ buffer, the

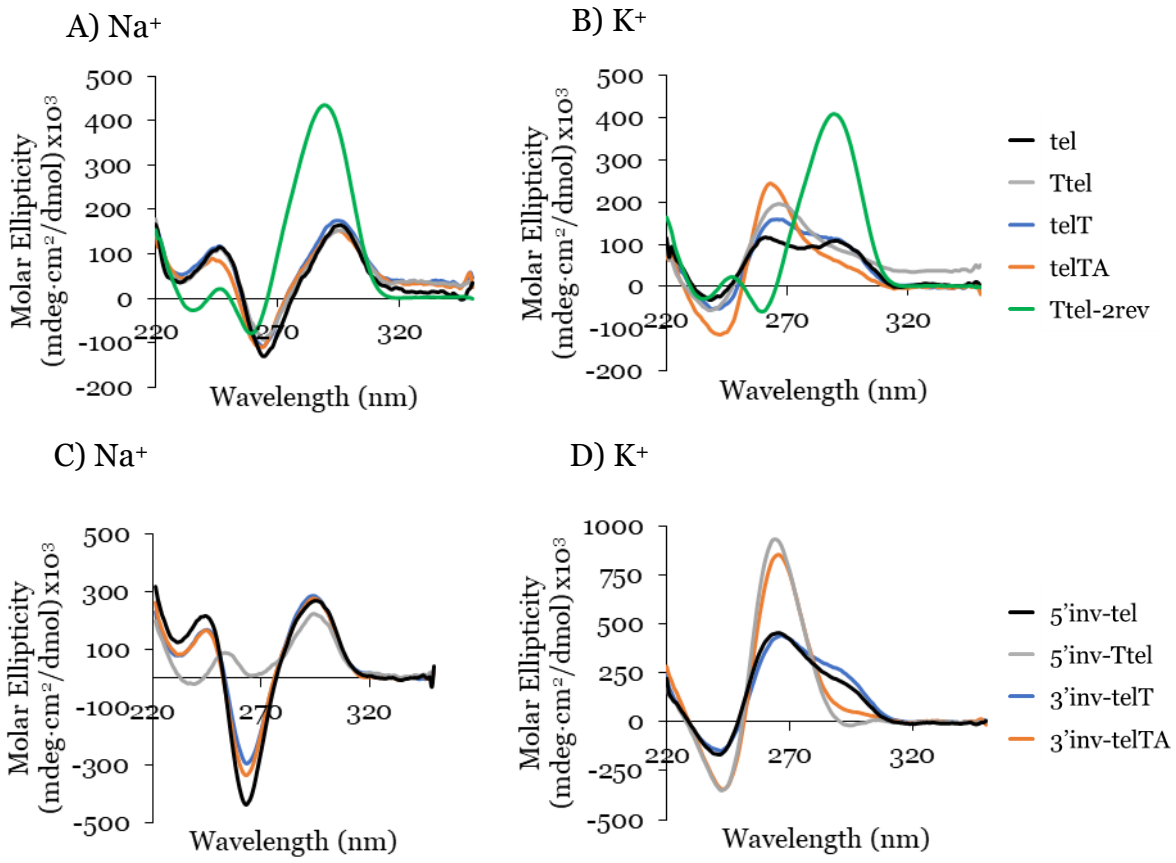


Figure 4.4. CD profiles of G4-forming sequences based on the telomeric repeat. A) Unmodified sequences, Na^+ buffer. B) Unmodified sequences, K^+ buffer. C) Modified sequences, Na^+ buffer. D) Modified sequences K^+ buffer. Conditions: 10 mM lithium cacodylate, 100 mM NaCl or KCl, pH 7.2.

sequences containing inversions mostly gave antiparallel CD profiles similar to the parent unmodified sequences, except for 5′inv-Ttel, which gave an additional positive peak at 255 nm (Figure 4.4C). This peak most likely indicates the presence of two competing G4 topologies but could also indicate the presence of a non-G4 DNA secondary structure. A large negative peak at 265 nm resulting from the antiparallel topology and a large positive peak at 265 nm from the parallel topology might cancel out, resulting in the strange CD profile of 5′inv-Ttel in the 240-270 nm region.

However, as with the unmodified sequences, the addition of K^+ ions resulted in a significant shift towards the parallel topology. The 5′inv-tel and 3′inv-telT sequences gave topologies similar to the unmodified parent sequences, but 5′inv-Ttel and 3′inv-telTA shift completely to a parallel topology, lacking the shoulder at 290 nm observed in the non-inverted strands (Figure 4.4D). This result supports the conclusion that the unusual profile seen for 5′inv-Ttel in Na^+ is a result of a mixture of topologies formed due to the significantly increased stability of the parallel G4. The parent tel sequence only adopted this topology in K^+ buffer, but the introduction of polarity inversions resulted in a more stable parallel topology. This shift was not observed for Ttel-2rev, which was antiparallel in both buffers, indicating that the inversions in the sequence led to formation of G4s of the parallel topology. This result also supports our original assumption that a longer loop would be required to offset increased strain on the G4 resulting from the creation of new loops, since the significant changes in topologies between Na^+ and K^+ buffers were only observed for inverted sequences with additional nucleotides in the middle loop.

On non-denaturing PAGE (Figure 4.5 and Figure SI-18), the lanes containing only G-rich sequences (Lane 2 for native Ttel and Lane 13 for 5′inv-Ttel in Figure 4.5) showed very similar mobilities (approx. T_{10}) regardless of strand length, which is commonly seen when compact G-quadruplexes are formed.

We determined the thermal stability (Table 4.2, Figure SI-14 – Figure SI-17) using the protocol described in Chapter 7.2.6. T_m values for unmodified sequences in Na^+ buffer were all between 35 – 40 °C, significantly lower than T_m of Ttel-2rev (51 °C, ΔT_m = +17 °C), which was the expected outcome resulting

Table 4.2. T_m *values for unmodified and modified telomeric G4-forming sequences.*

Sequence	T_m (Na^+, °C)	T_m (K^+, °C)
tel	37 (ap)	37 (ap), 48 (p)
Ttel	34 (ap)	48 (p)
telT	38 (ap)	40 (ap), 45 (p)
telTA	38 (ap)	50 (p)
5′inv-tel	76 (ap)	83 (p)
5′inv-Ttel	34 (ap), 75 (p)	84 (p)
3′inv-telT	68 (ap)	74 (p)
3′inv-telTA	62 (ap)	73 (p)
Ttel-2rev	51 (ap)	57 (ap)

Conditions: 10 μM strand concentration, 10 mM lithium cacodylate, 100 mM NaCl or KCl, pH 7.2. (p) Parallel topology. (ap) Antiparallel topology.

from the change in molecularity (bi- to mono-molecular) and strand length (12 nucleotides to 24 nucleotides). T_m was also higher in K^+ than in Na^+ buffer. It is noticeable that T_m values of the modified sequences were all higher than Ttel-2rev. Ttel-2rev showed an increase of only 9 °C relative to the Ttel sequence, while 5′inv-tel and 5inv-Ttel had T_m values of 83 °C (ΔT_m = +35 °C) and 84 °C (ΔT_m = +36 °C), respectively in K^+ buffer compared to tel and Ttel's T_m values of 48°C (for the peak at 265 nm). These results were consistent across all modified sequences, but 3′inv-telT and 3′inv-telTA show smaller increases relative to their unmodified counterparts, with T_m at 74 °C (ΔT_m = +29 °C) and 73 °C (ΔT_m = +23 °C). Additionally, in Na^+ buffer the 5′inv-Ttel sequence gave two melting points, as there was a significant difference in stability for antiparallel and parallel topologies. This also supports our prediction that this sequence had formed a parallel G4 in Na^+-containing buffer, which none of the other tested sequences did. We expected that conversion of a bi- to uni-molecular G4 should result in higher thermal stability. G4s containing inverted nucleotides had uniformly higher thermal stability than the Ttel-2rev control, which lacked inverted nucleotides. This is likely a result of the antiparallel topology adopted by Ttel-2rev being less thermodynamically favoured than the parallel G4 topology formed by the modified sequences. Initial CD results had already shown that the presence of inversions encouraged formation of stable G4s of parallel topology, and we now considered the effect this would have on duplex formation in the presence of a complementary strand.

4.2.2. Competition of Inverted tel G4s with Canonical Duplexes

We analysed G4 stability in the presence of the complementary strand using non-denaturing PAGE (Figure 4.5A), focusing on three sequences 5′inv-tel, 5′inv-Ttel and 3′inv-telTA. The first sequence closely resembles the original bimolecular G4 whereas the latter two had shown the greatest shift towards the parallel topology according to CD spectra. The complementary strands were selected to complement both the shorter native strands (12-14 nucleotides) and longer modified strands (24-28 nucleotides), as shown in Figure 4.5B - E. The native G4s, modified G4s and complementary strands were run on non-denaturing PAGE as controls alongside the mixtures of G4-forming sequences and their complements. To evaluate if G4s form duplexes at room temperature with their complements, G4 was formed first and then a complementary strand was added. To form a thermodynamic product, this mixture was heated at 90 °C for 5 min and then slowly cooled down to 4 °C.

The results shown in Figure 4.5A and Figure SI-18 suggest that all inverted sequences disrupted duplex formation. The unmodified G4 controls, shown in Lane 2 on each gel, had mobility around that of the T_{10} marker, with new bands appearing with lower mobility in lanes 3-6 (3-4 for 5′-tel) indicating formation of the duplex products shown in Figure 4.5B and C in the presence of

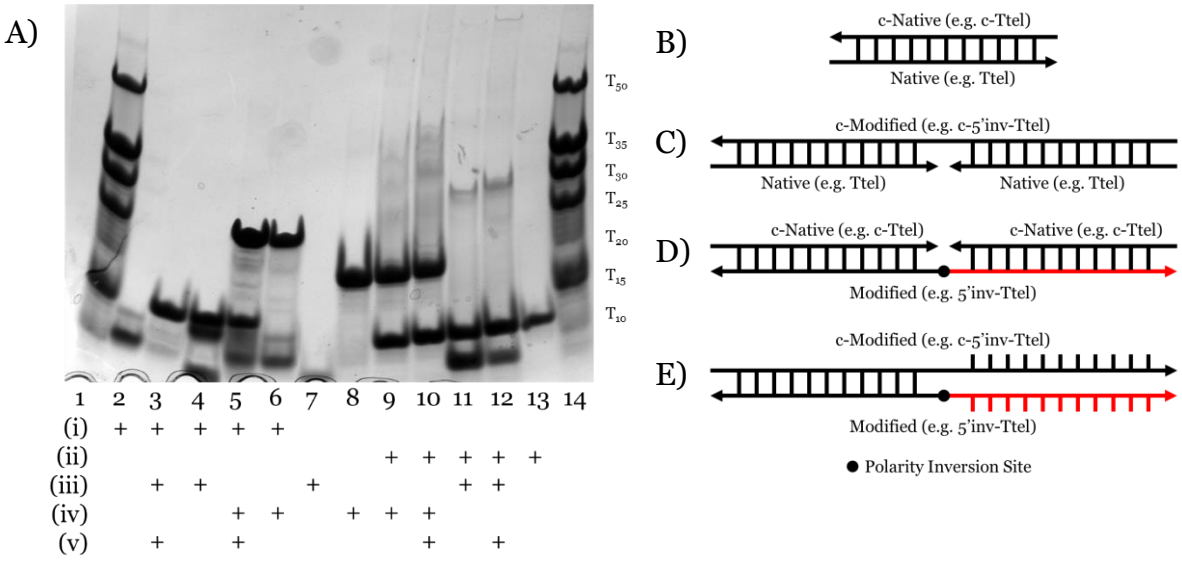


Figure 4.5. A) Native gel of telomeric G4s Ttel and 5′inv-Ttel with their complementary sequences (Table 4.1). (i) Ttel, (ii) 5′inv-Ttel, (iii) c-Ttel, (iv) c-5′inv-Ttel, (v) Thermodynamic product formed (heated at 90 °C for 5 mins and cooled at 4 °C overnight). Unmodified sequences appear to form two duplexes: B) Structure proposed to form in Lanes 3 and 4. C) Structure proposed to form in lanes 5 and 6. Modified sequences could form two duplexes, but formation appears to be minimal: D) Possible structure formed in lanes 11 and 12. E) Possible structure formed in lanes 9 and 10. Conditions: 100 μM strand concentration, 20% polyacrylamide gel, 10 mM lithium cacodylate, 100 mM KCl, pH 7.2.

complementary DNA. These bands are obviously distinct from the native G4 and complementary strand, indicating formation of a new species. This is most obvious in the Ttel sample with new bands appearing around the T_{15} and T_{25} components of the ladder for short and long complementary sequences, respectively. In contrast, upon addition of a complementary strand to 5′inv-Ttel, multiple bands corresponding to individual components were observed, i.e., lanes 9-12 versus lane 13 for G4 and lanes 7 and 8 for its complements on Figure 4.5A. Even for thermodynamic product (lanes 10 and 12), only faint low retarding bands were observed suggesting that putative duplex products shown in Figure 4.5D and E were not formed. In contrast, for 3′inv-telTA and 5′inv-tel (Figure SI-18A and B, respectively), formation of duplex products C or D along with individual components were already observed at room temperature (lanes 9 and 11 for 3′inv-telTA and lane 6 for 5′inv-tel). Duplex products were more dominant after samples were heated and cooled down in the lanes for thermodynamic products. This indicates that 5′inv-Ttel hinders duplex formation for most of the series studied. Although these results support the idea that the use of inverted nucleotides in G4s can inhibit duplex formation, supporting G4 stability, bands on native gels could be unclear leading to inconclusive results. We decided, then, to use ^{1}H NMR spectrometry as a more effective method of accurately ascertaining the secondary structure composition of our DNA mixtures.

For ^{1}H NMR analysis we focused on the region from 10.5 to 14.5 ppm, which corresponds to the imino protons of guanosine. The chemical shift of these peaks differs considerably depending on hydrogen-bonding arrangements allowing for secondary structures to be readily distinguished, making this an ideal technique for comparing various secondary structures.[102] G4s typically have imino proton chemical shifts around 11-12 ppm, whereas for the canonical Watson-Crick duplex the imino protons appear around 13-14 ppm, and for i-motifs around 15-16 ppm. We also narrowed our initial focus to the 5′inv-Ttel sequence which had shown a significant preference for parallel G4 topology, the greatest increase in T_m compared to the unmodified parent sequence, and almost no duplex formation on non-denaturing PAGE in the presence of its complements. The ^{1}H NMR spectra shown in Figure 4.6 for the Ttel sequences in Na^+ and K^+ buffers are all consistent with those expected for G-quadruplexes. In Na^+ buffer a mixture of G4

topologies is present according to the ^{1}H NMR spectra as evident from the appearance of multiple peaks of various intensities at 10.5 – 12 ppm (Figure 4.6A). Upon addition of K^+ (as KCl) a change in topology was observed for native Ttel, resulting in the symmetrical G4 topology reported for bimolecular tel sequences previously (Figure 4.6D).[64] A similar NMR profile is also observed for 5′inv-Ttel, suggesting that the polarity inversions did not significantly affect the G-tetrad hydrogen-bonding arrangement (Figure 4.6F). On the other hand, the native control Ttel-2rev does not exhibit a single conformation, instead forming a mixture of topologies in both Na^+ and K^+ buffers (Figure 4.6B and E).

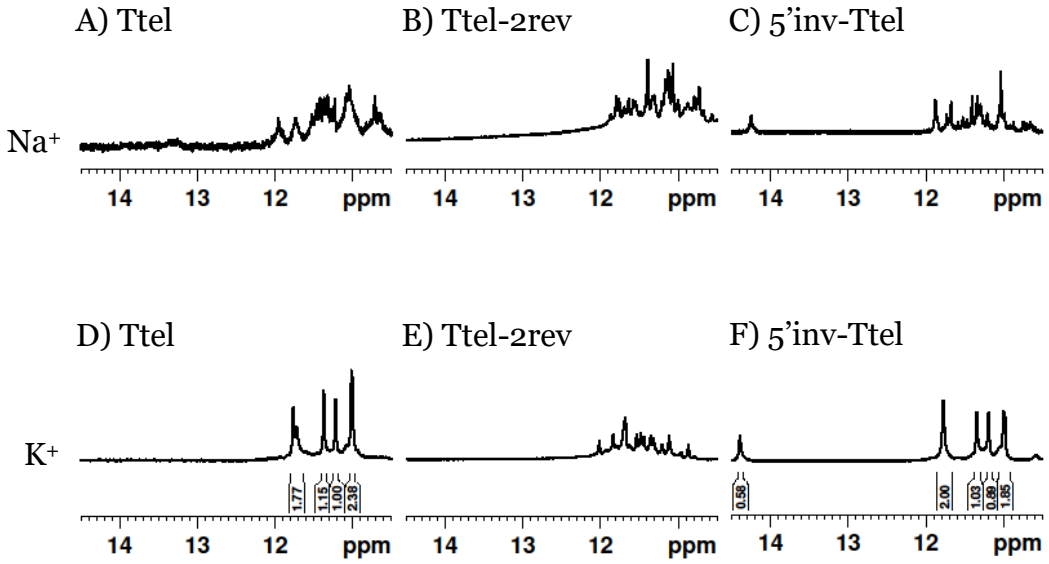


Figure 4.6. 1H NMR profiles of modified telomeric G4s. A) Ttel, Na+ buffer. B) Ttel-2rev, Na+ buffe.; C) 5′inv-Ttel, Na+ buffer. D) Ttel, K+ buffer. E) Ttel-2rev, K+ buffer. F) 5′inv-Ttel, K+ buffer. Conditions: approx. 200 μM strand concentration, 20 mM sodium phosphate, 10 mM KCl (in K+ samples only), 10% D2O, 1% TSP, pH 7.0.

Formation of symmetrical G4s is evident from the appearance of distinct imino signals having 2:1:1:2 ratio in integrated signals at 10 – 12 ppm which is halved for what can be expected for a G4 formed by 12 guanosines (Figure 4.6D and F). Slight changes in the distribution of these peaks potentially indicates minor adjustments to H-bonds to accommodate the additional loop, but otherwise Ttel and 5′inv-Ttel appear to form G4s with very similar topology (see Figure 4.2C – E for examples of these topologies). Interestingly, 5′inv-Ttel also contains an additional peak at 14.5 ppm, which indicates the presence of Watson-Crick A-T base pairing occurring in the new elongated central loop. This peak is not observed for Ttel-2rev, suggesting that this base pairing is only possible with

the inversions present. Integration of this signal gives a value of 0.58 which means that there is approximately one A-T base pair in a G4 with twelve guanines.

Next, the duplex formation experiments analysed by non-denaturing PAGE were replicated using ^{1}H NMR spectroscopy. When Ttel and i-Ttel were mixed, as shown in Figure 4.7A, peaks began to appear around 13-14 ppm immediately, and after 3 days almost no peaks were visible in the 11-12 ppm region. Thus, G4 structures began to refold into duplexes almost instantly. Heating of samples followed by slow annealing resulted in a complete switch to the duplex showing that both the kinetic and thermodynamic product of mixing the two sequences is the WC-duplex. However, when 5′inv-Ttel and i-5′inv-Ttel were mixed, as shown Figure 4.7B, almost no change was observed, and this was also the case after sequences were heated and then cooled down. Peaks of low intensity were observable from 13-14 ppm, suggesting that an equilibrium is established, but this equilibrium strongly favours the G4 and not the duplex. One notable shift is the disappearance of the peak at 14.5 ppm, suggesting that introduction of the complementary sequence disrupted formation of H-bonding in the loop. Overall, these results, on adding (partially) complementary strands establish that the introduction of polarity inversions in G4s significantly diminishes WC-duplex formation, thereby increasing the stability of alternative secondary structures such as G4s.

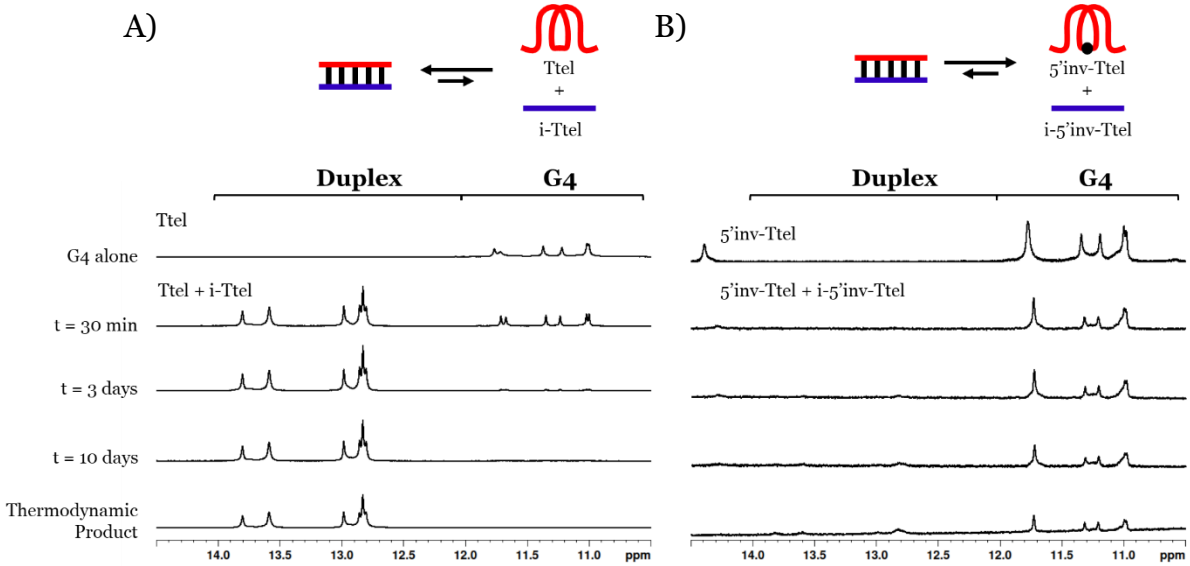


Figure 4.7. ^{1}H NMR of A) Ttel and B) 5′inv-Ttel challenged with complementary DNA. Spectra are obtained over several days to observed formation of a kinetic product, before being heated to 90 °C for 5 mins, then cooled to 4 °C overnight. Subsequent experiments showed no further change was observed. Conditions: approx. 200 μM strand concentration, 20 mM sodium phosphate, 10 mM KCl, 10% D_2O, 1% TSP, pH 7.0.

4.2.3. Extension of Modified Telomeric G4s with Duplex-Forming Tails

The intended application of these thermodynamically stable G4s required that they are incorporated into longer DNA sequences, as models for G-rich regions in genomic DNA. To accomplish this, we synthesised new sequences containing 5′inv-Ttel inserted into the centre of the linker region of the widom601 nucleosomal array[103] incorporating the ends of this region as tails at either end of the 5′inv-Ttel sequence (Table 4.3, Figure 4.8). Sequence names now include either a suffix of 5′-tail or 3′-tail to indicate at which end of the 5′inv-Ttel the new duplex-forming region is incorporated. Non-complementary T4 spacers were inserted between the G4 and duplex-forming regions to allow for additional flexibility.

Table 4.3. Ttel and 5′inv-Ttel sequences with tails for duplex formation.

Abbreviation	Sequence
Ttel-5′-tail	5′-GCATGTATTGT$_4$(TG$_3$AT)$_2$TT(TAG$_3$T)$_2$-3′
5′inv-Ttel-5′-tail	5′-GCATGTATTGT$_4$-3′—3′-(TG$_3$AT)$_2$T-5′—5′-T(TAG$_3$T)$_2$-3′
i-Ttel-5′-tail	5′-(AC$_3$TA)$_2$AA(ATC$_3$A)$_2$T$_4$CAATACATGC-3′
Ttel-3′-tail	5′- (TG$_3$AT)$_2$TT(TAG$_3$T)$_2$T$_4$GCGGCCGCC-3′
5′inv-Ttel-3′-tail	3′-(TG$_3$AT)$_2$T-5′—5′-T(TAG$_3$T)$_2$T$_4$GCGGCCGCC-3′
i-Ttel-3′-tail	5′-GGCGGCCGCT$_4$(AC$_3$TA)$_2$AA(ATC$_3$A)$_2$-3′
5′-Tail	5′-GCATGTATTG-3′
i-5′-Tail	5′-CAATACATGC-3′
3′-Tail	5′-GCGGCCGCC-3′
i-3′-Tail	5′-GGCGGCCGC-3′

CD spectra, shown in Figure 4.8A and B, revealed that the native sequences Ttel-5′-Tail and Ttel-3′-Tail correspond to a mixture of structures and topologies resembling those seen for Ttel-2rev. In Na^+ buffer they both most likely adapt antiparallel G4 topologies, but in K^+ buffer the CD spectra differ. The Ttel-5′-tail shifts towards a parallel topology, whereas Ttel-3′-tail remains an antiparallel G4. Inverted sequences with a 3′- or 5′-tail favour a parallel G4 topology, even in Na^+ buffer. In all cases they appear to have the shoulder observed at 290 nm although this is least pronounced for 5′inv-Ttel-5′-tail. This shoulder is possibly caused by the presence of single-stranded DNA, which can give low intensity CD signals at a range of wavelengths. The sequences added at either end were also tested independently with their complements and were shown to form stable duplexes,

as shown by CD spectroscopy in Figure 4.8C and ^{1}H NMR spectroscopy in Figure 4.9 (i).

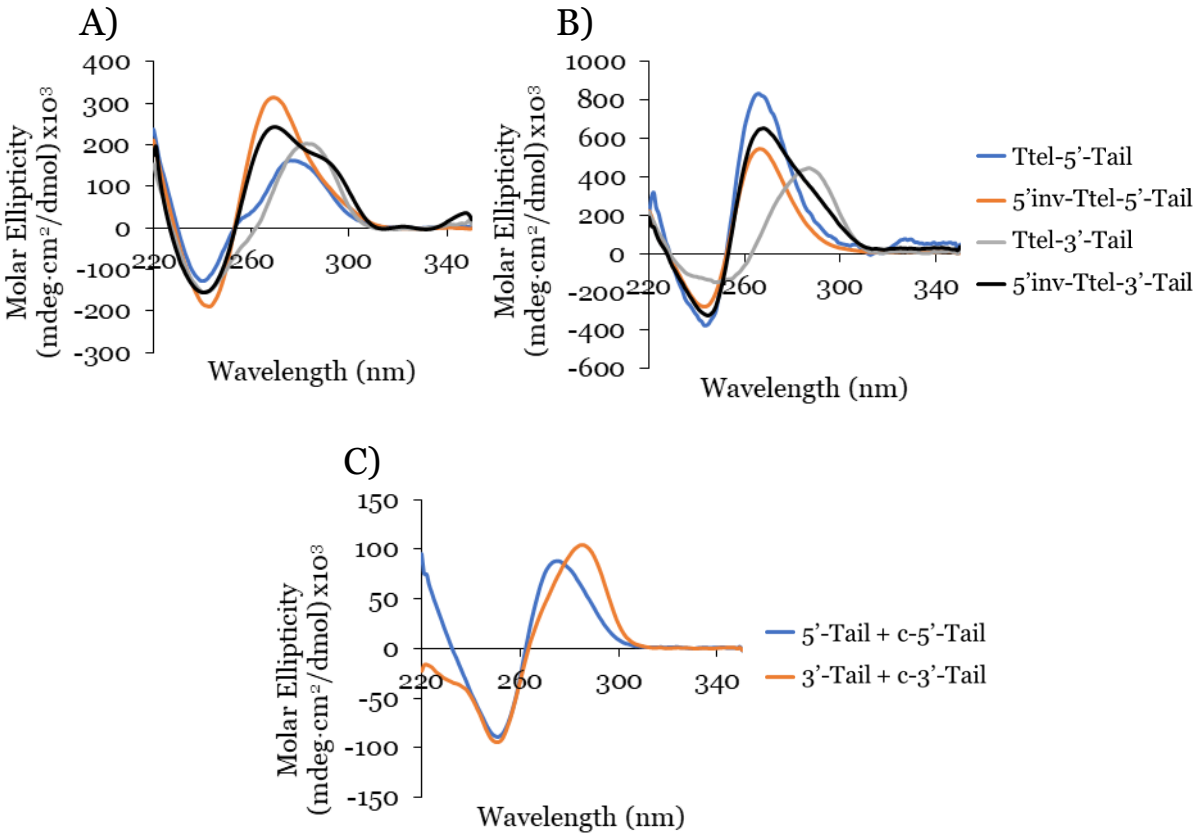


Figure 4.8. CD profiles of Ttel and 5'inv-Ttel sequences with additional tails for duplex formation. A) Na^+ buffer, B) K^+ buffer, C) Duplexes formed by tails alone, K^+ buffer. Conditions: 20 μM strand concentration, 20 mM sodium phosphate, 10 mM KCl (K^+ buffer only), pH 7.0.

In ^{1}H NMR experiments, the unmodified controls do not have the same symmetrical features observed previously for the bimolecular G4s, instead giving spectra more similar to Ttel-2rev (Figure SI-19A, C, E and G). The symmetrical structure is still partially preserved in the case of inverted G4s (Figure SI-19B, D, F and H), although individual peaks in the imino region are not as clearly defined as for the sequences without tails. The 9-mer 3′-tail has five cytosines which contribute to formation of a self-complementary duplex resulting in distinct peaks at 13 – 13.5 ppm for G4 alone (Ttel-3′-Tail, Figure 4.9A, spectrum i). This structural feature is suppressed in the inverted sequence (5′inv-Ttel-3′-Tail, Figure 4.9B, spectrum iii). Addition of the complementary strand to native sequences (Ttel-3′-tail and Ttel-5′-tail) showed similar results to the bimolecular G4s, rapidly unfolding G4 and forming antiparallel duplexes (Figure 4.9A and C). In both cases the duplex appears to be both the kinetically and thermodynamically favoured product. Addition of complementary C-rich strands

to the inverted sequences (Figure 4.9B and D) showed the appearance of peaks at 12.5 – 14 ppm, corresponding to duplex formation, but the peaks in the 11 – 12 ppm range remained mostly unchanged. This indicates that duplexes are formed, but comparing these peaks to the tail only controls suggests that duplex formation occurs primarily within the tails. This means that non-G4 regions formed antiparallel duplexes, whereas the inverted G-rich regions formed stable G4s and duplex formation in this region was disrupted. Heating and slow annealing of these samples also caused very little change in ^{1}H NMR, suggesting that mixed

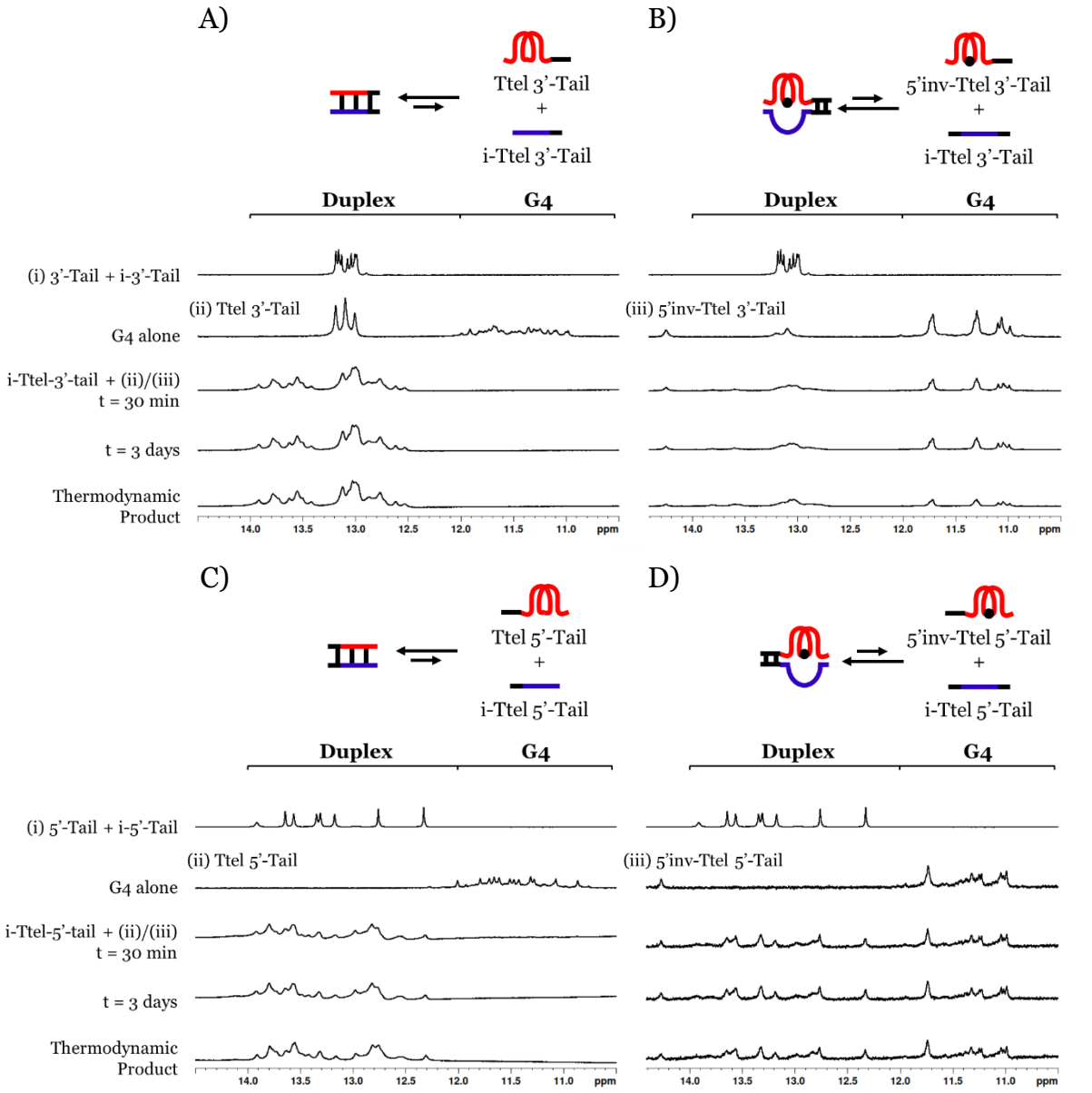


Figure 4.9. ^{1}H NMR spectra of A) Ttel-3′-tail, B) 5′inv-Ttel-3′-tail, C) Ttel-5′-tail and D) 5′inv-Ttel-5′-tail challenged with complementary DNA. Spectra are obtained over several days to observed formation of a kinetic product, before being heated to 90 °C for 5 mins, then cooled to 4 °C overnight. Subsequent experiments showed no further change. Conditions: approx. 200 μM strand concentration, 20 mM sodium phosphate, 10 mM KCl, 10% D_2O, 1% TSP, pH 7.0.

G4/duplex structures are the thermodynamically favoured products. Additional ^{1}H NMR spectra taken after several months showed no further change.

4.2.4. Interaction of Modified Telomeric G4s with HP1α

Finally, we investigated the binding of inverted G-quadruplexes to HP1α. HP1α is known to favour binding to parallel G4 topologies.[38] We investigated the HP1α-G4 complex formation with G4s containing inverted nucleotides. Table 4.4 shows the sequences for Oligo B and Oligo 2G, which were assessed earlier[38] and used as positive and negative controls, respectively. To determine the ability of HP1α to bind to G4s with inverted polarity, biolayer interferometry (BLItz, see Chapter 7.2.8.) was performed with purified his$_6$-tagged HP1α immobilised on nickel sensor tips immersed in a solution containing oligonucleotides. Figure 4.10A indicates that HP1α has similar affinity for the inverted sequences without tails added as for the antiparallel G4, Oligo 2G. Typically, HP1α has a higher affinity for parallel G4s, with HP1α showing reasonably strong affinity for Oligo B. However, as shown in Figure 4.10B, HP1α shows no affinity for the modified G4 containing a duplex-forming tail at the 3′-end.

Table 4.4. G4 controls for HP1α binding

Name	Sequence	G4 Topology
Oligo B	5′-TG$_3$T$_2$G$_3$T$_2$G$_3$T$_2$G$_3$T$_2$G$_3$T$_2$G$_3$T$_2$G$_3$T-3′	Parallel
Oligo 2G	5′-TG$_3$T$_2$AG$_3$T$_2$AG$_3$T$_2$AG$_3$TG$_3$T$_2$AG$_3$T$_2$AG$_3$T$_2$AG$_3$T-3′	Antiparallel

HP1α's binding to DNA and RNA has been shown to derive from the hinge domain that has two charged lysine patches at residues 89-91 and 104-106.[58, 60] To determine whether HP1α interacts with native and not inverted nucleotides in the DNA sequence, we decided to test how the hinge of HP1α binds to our synthetic constructs. A his$_6$-tagged hinge of HP1α was recombinantly expressed in *E. coli* and purified by Ruby Roach (PhD student in T. K. Hale Laboratory) (Figure SI-20). The hinge was immobilised on nickel sensor tips and used to perform further BLItz experiments with a K^+ solution containing a G4 with inverted sequence or a control native G4 or a duplex. In this case, we observed strong binding of the HP1α hinge domain for G4s with an inverted sequence

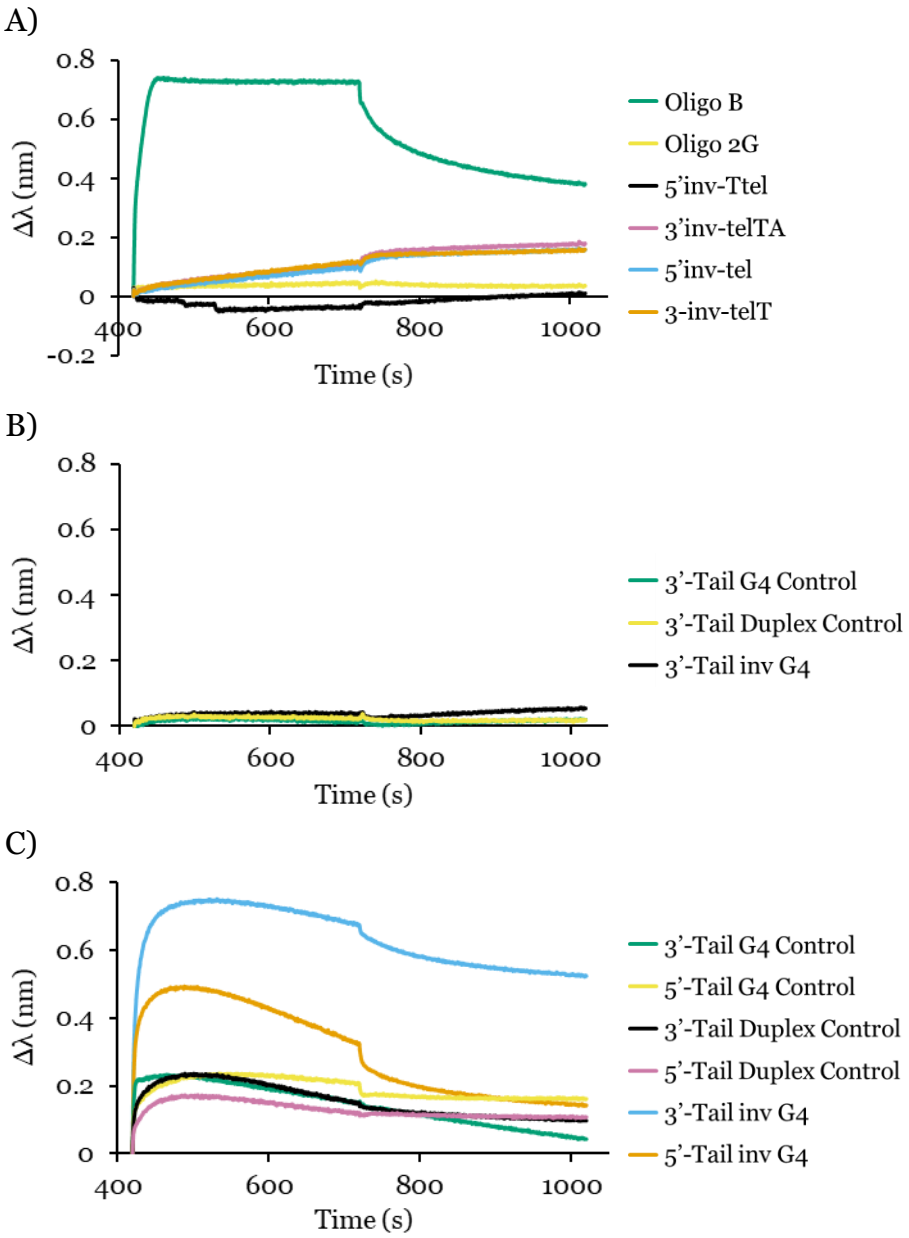
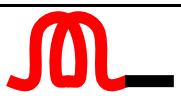
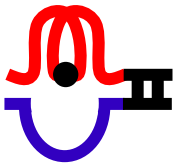
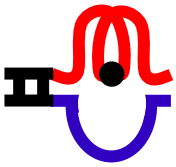


Figure 4.10. Biolayer interferometry (BLI) analysis of binding of: A) Immobilised HP1α to modified telomeric G4 sequences compared to Oligo B and Oligo 2G control. B) immobilised HP1α to modified Ttel with 3′-tails compared to the unmodified G4 and duplex. C) Immobilised HP1α hinge domain to modified Ttel with duplex-forming tails compared to unmodified Ttel G4s and duplexes. Conditions: 100 μg/mL protein concentration, 2 μM strand concentration, buffer: 20 mM sodium phosphate, 10 mM KCl, pH 7.0.

(Figure 4.10C). Table 4.5 shows the dissociation constants, K_D, which were highest for duplexes (i.e., very weak binding), followed by the native G4s, but, surprisingly, inverted G4s showed almost a 10-fold increase in binding affinity over unmodified G4s (smallest K_D). This q suggests that poor binding of inverted G4s to HP1α are the result of HP1α selectivity rather than the use of inverted nucleotides specifically. It was proposed that several antiparallel G4s did not bind to HP1α due to the presence of lateral or diagonal loops that blocked binding sites.

Table 4.5. Association and dissociation rates, and dissociation constants for oligonucleotides with the isolated hinge domain of HP1α from BLI measurements.

Name	Sequences	k_{on} / 1000 ($M^{-1}s^{-1}$)	k_{off} x 1000 (s^{-1})	K_D (nM)
3′-Tail G4 control	 Ttel 3′-tail	85.6 (19)[a]	0.92 (11)	10.8 (13)
5′-Tail G4 control	 Ttel 5′-tail	39 (3)	15.4 (7)	390 (3)
3′-Tail duplex control	 Ttel 3′-tail + i-Ttel 3′-tail	64 (8)	53 (4)	840 (13)
5′-Tail duplex control	 Ttel 5′-tail + i-Ttel 5′-tail	32 (11)	28 (3)	900 (3)
3′-Tail Inv-G4	 5′inv-Ttel 3′-Tail + i-Ttel 3′-Tail	85.2 (9)	4.85 (9)	56.9 (12)
5′-Tail Inv-G4	 5′-inv 5′-Tail + i-Ttel 5′-Tail	149 (9)	39.1 (15)	263 (19)

(a) The error in the least significant digit is provided in the parenthesis.

I tentatively conclude that the addition of a diagonal loop through introduction of inversions causes a similar effect, as our modified sequences behave similarly to antiparallel G4s, preventing full-length HP1α binding. Our telomeric G4s based on inverted sequences form a symmetrical G4 of parallel topology but to inclusion of the inversion site means they contain an additional diagonal or lateral loop that shields one of the external G-tetrads (seen in Figure 4.2D and E). This feature, which is more typical of antiparallel G4s, might prevent our inverted G4 constructs from interacting with full-length HP1α. Based on this

we would suggest that the selectivity of HP1α may be determined not only by G4 topology, but also by the loop arrangements seen primarily in antiparallel G4s.

4.3. Further Modifications of c-KIT and c-MYC G4-forming Sequences with Inverted- and α-Nucleotides

4.3.1. Synthesis of Modified c-MYC and c-KIT G4s

Internal polarity inversions had introduced several unique properties to the telomeric DNA sequences, which offered a useful baseline for new modified G4s with controlled topology and increased thermal stability. Most importantly, these inverted sequences did not form duplexes in the presence of the complementary strand, instead remaining as thermodynamically stable G4s. However, we needed to solve one problem before this strategy could be explored further. While the previously modified G4s had the desired thermodynamic stability, HP1α showed no affinity for the chosen sequence. We used the tel sequence and its derivatives because it was easily synthesised and was a useful model of telomeric G4s, but it is not an ideal sequence for testing the ability of sequences containing these modifications to interact with HP1α.

Experiments with the intrinsically disordered HP1α hinge domain suggest that the lack of binding observed for our previous modifications was due to HP1α specificity rather than a complete lack of affinity for oligonucleotides containing polarity inversions. Instead, we selected alternative G4-forming sequences with known affinity for HP1α to modify. The first was a 20-mer c-KIT sequence, $dG_3CG_3CGC_2AG_3AG_3$, which forms a parallel G4 (Figure 4.11A, PDB: 2KQH)[104] and had shown reasonable affinity to HP1α in previous experiments (K_D = 199 nM).[38] This sequence was only slightly longer than the telomeric sequence explored previously meaning it was not significantly more difficult to synthesise, but it was unimolecular meaning that if it retained its topology we would avoid adding any new loops which could disrupt complex formation. I synthesised a c-KIT sequence containing a single central polarity inversion site similar to the Ttel modification carried out previously (Table 4.6). Because the initial testing of this sequence, described below, showed it also did not bind to HP1α, sequences

containing more precise modifications were synthesised, intended to minimise any impact on G4 formation.

It was proposed that the development of sequences with only G-tract or loop nucleobases modified (Table 4.6). Modified sequences would contain either no modified guanosines or all modified guanosines, and theoretically hydrogen-bonding interactions would be unaffected (like the example shown in Figure 4.11C). Therefore, modified G4s would have G-tetrad and loop arrangements consistent with the unmodified G4 structures and retain their ability to form complexes with proteins. We also decided, at this stage, to expand the scope of this work to include another short, parallel G4 with known affinity to HP1α, c-MYC (Figure 4.11B, PDB: 1XAV).[105] This sequence is a 16-mer, $dG_3TG_3TAG_3TG_3$, which showed slightly less affinity to HP1α (K_D = 610 nM)[38] than c-KIT but was shorter and therefore easier to synthesise.

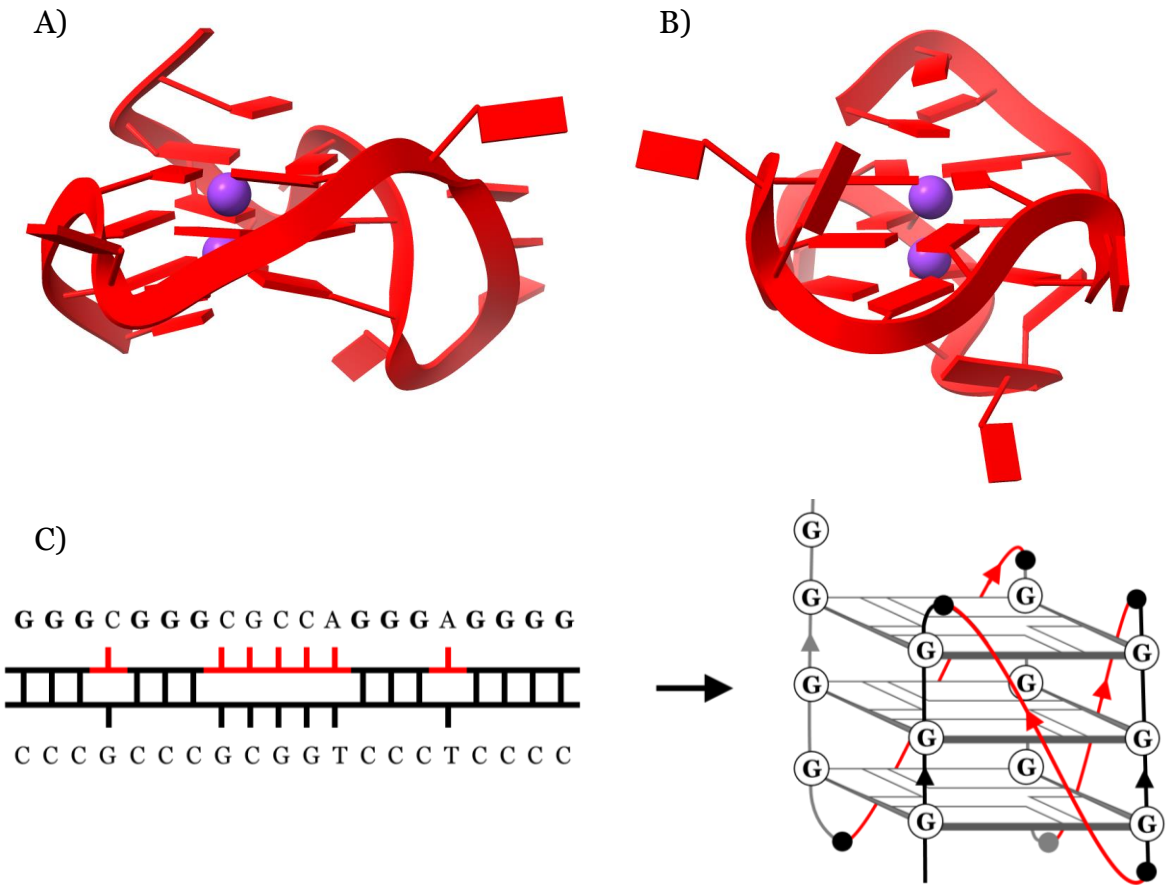


Figure 4.11. NMR solution structure of A) c-KIT and B) c-MYC monomers. C) Example of proposed modifications to c-KIT and c-MYC G4s. A similar number of nucleotide mismatches are introduced to the duplex structures relative to previously tested tel sequences but limiting modifications to loops or modifying all dG residues minimises changes to G4 topology.

The synthetic process was the same as for inverted nucleotides described briefly in Chapter 4.2.1 and in more detail in the experimental section (Chapter 7.4.2). For synthetic reasons, sequences requiring a 3′-inverted nucleotide (i.e., c-KIT-inv-dG and c-MYC-inv-dG) included an additional non-inverted thymidine at the 3′-end and used dT-CPG support. Longer coupling times for modified nucleotides were required for both types of modification, but sequences could otherwise be obtained with good yields. The sequences were cleaved from the solid support in 28% aqueous ammonia at room temperature overnight. However, under these conditions we encountered a recurring problem during mass spectrometry analysis. Masses were consistently observed with masses of +70/+140 etc. This corresponds to the *N*^{2}*-iso*butyryl protecting group used for standard dG phosphoramidites. Subsequent deprotections in 28% aq. ammonia were performed at 55 °C for 6 hours. Sequences were purified using reverse-phase HPLC and then tested using similar experiments to the telomeric sequences explored previously. In this case we moved immediately to ^{1}H NMR spectroscopy rather than non-denaturing PAGE as it had previously given similar, but more quantifiable, information.

4.3.2. Evaluation of c-KIT-inv Topology, Stability and Protein Interaction

The unmodified c-KIT sequence was folded in K^+-containing buffer and formed characteristic parallel G4s. The c-KIT-inv sequence folded into a parallel G4 in both buffers (Figure 4.12A and B), indicating that introduction of inversions already improved the stability of the naturally occurring G4 topology. The inversion of half of the sequence also had a significant impact of increasing the T_m of modified c-KIT in both Na^+- and K^+-containing buffer. The modified sequence had a T_m of approximately 79 °C in Na^+, although it was not completely melted at 90 °C (Table 4.6, Figure SI-21A). This was a significant increase in thermal stability compared to the unmodified sequence which was unfolded in Na^+ buffer. The T_m also increased from 44 °C to >90 °C in K^+ buffer (Table 4.6, Figure SI-21B).

Table 4.6. Sequences and T_m *values of c-KIT and c-MYC G4-forming sequences containing inverted and α-nucleotides.*

Name	***Sequence***	***Na^+ (T_m, °C)***	***K^+ (T_m, °C)***
c-KIT	$G_3CG_3CGC_2AG_3AG_4$	Not Folded	$44^{(p)}$
c-KIT-inv	$\underline{G_3CG_3CG}C_2AG_3AG_3$	$\sim79^{(p)(a)}$	$>90^{(p)}$
c-KIT-inv-loops	$G_3\underline{C}G_3\underline{CGC_2A}G_3\underline{A}G_4$	$49^{(ap)}$	$>90^{(p)}$, $57^{(ap)}$
c-KIT-α-loops	G3CG3CGC2AG3AG4	$54^{(ap)}$	$71^{(p)}$, $40^{(ap)}$
c-KIT-inv-dG	$\underline{G_3}C\underline{G_3}CGC_2A\underline{G_3}A\underline{G_4}T$	$43^{(ap)}$	$69^{(p)}$
c-KIT-α-dG	$\underline{G_3}C\underline{G_3}CGC_2A\underline{G_3}A\underline{G_4}$	$\sim53^{(p)}$	$74^{(p)}$
c-MYC	$G_3TG_3TAG_3TG_3$	$25^{(p)}$, $48^{(ap)}$	$71^{(p)}$
c-MYC-inv-loops	$G_3\underline{T}G_3\underline{TA}G_3\underline{T}G_3$	$50^{(p)}$, $34^{(ap)}$	$55^{(p)}$
c-MYC-α-loops	$G_3\underline{T}G_3\underline{TA}G_3\underline{T}G_3$	$52^{(p)}$	$\sim83^{(p)(a)}$
c-MYC-inv-dG	$\underline{G_3}T\underline{G_3}TA\underline{G_3}T\underline{G_3}T$	$46^{(p)}$, $34^{(ap)}$	$54^{(p)}$
c-MYC-α-dG	$\underline{G_3}T\underline{G_3}TA\underline{G_3}T\underline{G_3}$	$33^{(p)}$	$74^{(p)}$
i-c-KIT	$C_4TC_3TG_2CGC_3GC_3$	---	---
i-c-MYC	$C_3AC_3TAC_3AC_3$	---	---

Non-native nucleotides are bolded and underlined. Inv- refers to sequences containing inverted nucleotides, α- refers to sequences containing α-nucleotides. All sequences are written in the 5′ to 3′ direction if standard nucleotides were used. (p) Parallel topology and (ap) antiparallel topology refer to topology of G4 being melted. (a) these sequences were not completely melted at 90 °C, T_m *value is estimated. Conditions: 20 μM strand concentration, 20 mM sodium phosphate, 10 mM KCl (K^+ buffer only), pH 7.0.*

BLItz experiments were performed with these sequences to see if the modified c-KIT sequence could bind to HP1α (see Chapter 7.2.8.). Once again, we used TERRA45, as a positive control.[38] HP1α showed lower, but comparable, affinity for unmodified c-KIT to TERRA 45, but showed no affinity for modified c-KIT (Figure 4.12C). We proposed that this might be due to the modification sites chosen for this sequence as a similar strategy for c-KIT as for previous telomeric sequences. The tel modification was based on our observation of the short distance between the 5′-ends, which made it an ideal location for the introduction of this type of cross-link. c-KIT-inv also contained an internal polarity inversion in the middle of the sequence but this was not based on the same observations as tel regarding the positions of nucleobases. c-KIT is a unimolecular G4, meaning that rather than effectively introducing a cross-link between two strands, as had been achieved with the tel sequences, we had made a significant modification to part of the G4-forming region. The new orientations of dG nucleotides within the G-tetrads could result in a topological change to preserve the original hydrogen bonding arrangement.

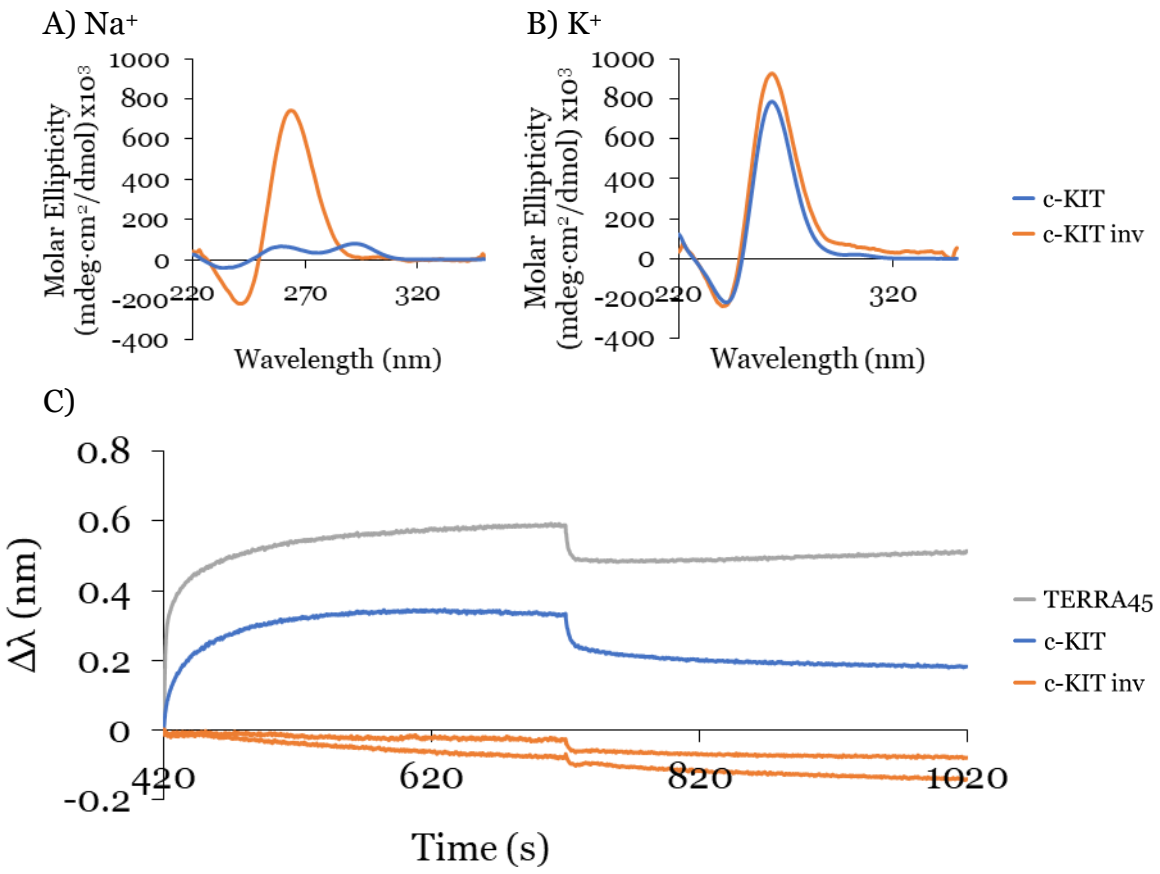


Figure 4.12. A) CD profile of c-KIT and c-KIT-inv in Na⁺ buffer. B) CD profile of c-KIT and c-KIT-inv in K⁺ buffer. C) Biolayer interferometry (BLI) analysis of binding of immobilised HP1α to c-KIT-inv compared to c-KIT and TERRA45 controls. Conditions: CD: 20 μM strand concentration, 10 mM lithium cacodylate, 100 mM NaCl or KCl, pH 7.2. BLItz: 2 μM strand concentration, 100 μg/mL protein concentration, 20 mM sodium phosphate, 10 mM KCl, pH 7.0.

CD spectroscopy of c-KIT-inv showed a similar parallel topology to the unmodified G4 (Figure 4.12A and B). While this is usually a good indicator of G4 topology, it does not allow complete characterisation of the G4 structure. This would only be possible using techniques such as 2D NMR spectroscopy or X-ray crystallography. These techniques potentially give more detailed information regarding the conformational changes induced by inverted nucleotides. Crystallisation of the G4 structure was attempted by vapour diffusion of ethanol, but ultimately this did not result in crystallisation of the modified G4. 2D NMR experiments were also attempted. Full characterisation of the G4 structure required ROESY experiments to identify cross-peaks between adjacent nucleobases, which could be used to assign peaks to each nucleotide in the sequence using other 2D NMR experiments, such as TOCSY and HMBC. ROESY data also needed to include cross-peaks for the imino protons in the G-tetrad, to identify the H-bonding arrangement within the G4, and therefore the complete

secondary structure. While ROESY data taken with these samples were sufficient to identify cross-peaks between H-8 positions of adenosine and guanosine, as well as methyl protons of thymidine, it was insufficient to identify imino proton cross-peaks. The quantity of DNA obtained by this synthesis was too low to obtain sufficiently high resolution ROESY data for a complete characterisation of the G4 secondary structure using the available instrument. In the absence of the NMR data required for complete characterisation, CD spectroscopy and BLItz are used to infer that the G-tetrad stacking must be consistent with the unmodified G4 to produce a similar CD spectrum, but the loops must be orientated differently to allow for that G-tetrad arrangement, and these loops disrupt interaction with the protein.

4.3.3. Synthesising c-KIT and c-MYC Sequences Containing G-Tract and Loop Specific Modifications to Preserve Topology

We addressed the issues observed with c-KIT-inv by carrying out more precise modifications intended to preserve the hydrogen-bonding arrangements, as described in Chapter 4.3.1. By only modifying the nucleobases in the loops we could ensure that the orientation of guanosine-guanosine H-bonds did not change between modified and unmodified sequences, as no modified guanosine would be involved in hydrogen-bonding. Similarly, we tested to see if modifying only guanosine nucleobases would cause the same conformational change across all guanosine-guanosine H-bonds and result in an effectively identical structure to the unmodified sequence. Our initial hypothesis was that modifying only the residues in the loop would be most successful, as this eliminates any changes to the parts of the sequence involved in G4 hydrogen-bonding. These experiments would also determine if multiple short, modified regions resulted in comparable disruption of duplex formation to the longer regions tested with 5′inv-Ttel sequences.

However, these sequences had an additional synthetic challenge. During DNA synthesis, trityl levels in the waste following DMT cleavage are measured using a UV-Vis detector. In previous sequences, a slight (approx. 10%) decrease in trityl level was recorded during DNA synthesis for the first coupling step immediately following a change in strand direction. The decrease was not observed for subsequent base couplings (i.e., coupling of inverted nucleotides to

other inverted nucleotides was unaffected). This suggested that the 3′-3′ and 5′-5′ coupling of the phosphoramidite and alcohol was less efficient than the standard 3′-5′ coupling. In previous sequences this had a minimal effect on the final yield because most sequences contained only a single inversion site, but the decrease in yield was noticeable for the 5′inv-Ttel-5′-tail sequence which was both longer than the initial Ttel sequences and contained two inversion sites. The new sequences would require 6 – 8 inversion sites each, potentially resulting in a significant decrease in yield throughout DNA synthesis. Previously, coupling times for modified nucleotides were increased to 5 minutes. For synthesis of the new sequences, coupling times of the modified nucleotides were further increased to 10 minutes, and coupling times of unmodified dG were also increased to 5 minutes (see chapter 7.4.2 for details). This also caused the total synthesis duration to increase significantly. To ensure that coupling efficiency remained high throughout the synthesis freshly dried molecular sieves were added to phosphoramidite bottles and P_2O_5 was added to the sample chamber periodically (approx. every 2 – 3 hours) to minimise additional moisture in the system. With these changes made, sequences were obtained with comparable yields to the 5′inv-Ttel tail sequences which was sufficient for CD and ^{1}H NMR studies.

4.3.4. Evaluation of G4 Formation in c-KIT and c-MYC Sequences Containing G-tract and Loop Specific Modifications

The synthesised sequences were folded into G4s, as described previously. CD spectroscopy experiments were carried out to ensure that modified sequences would fold into G4s with the same topology as unmodified sequences. In Na^+-containing buffer unmodified c-MYC gives primarily a mixed topology (Figure 4.13A), with a large peak at 265 nm and a smaller peak at 300 nm. As seen previously, unmodified c-KIT was unfolded in Na^+-containing buffer. Modified c-KIT sequences showed significant differences. All sequences except c-KIT-α-dG appeared to be folded into G4s of mixed topology with positive peaks at 265 nm and 295 nm (Figure 4.13C). c-KIT-α-dG instead gave the CD profile of a parallel G4 with a single positive peak at 265 nm, even in Na^+-containing buffer. This peak also had significantly higher intensity than the other sequences, which is expected when a sequence forms a single topology, rather than multiple topologies. This corresponded to the previously reported data showing that the inclusion of α-dG

nucleotides encouraged the formation of parallel G4 topologies.[95] Modified c-MYC sequences behaved similarly except the difference between c-MYC-α-dG and other c-MYC sequences is less pronounced because the unmodified sequence is already parallel in Na^+-containing buffer (Figure 4.13E). Moreover, c-MYC-α-dG does not have the small peak at 300 nm observed for the unmodified and other modified sequences and does have a slightly higher intensity for the peak at 265 nm. This indicates that, as expected, the sequence containing α-dG favours parallel G4 topologies.

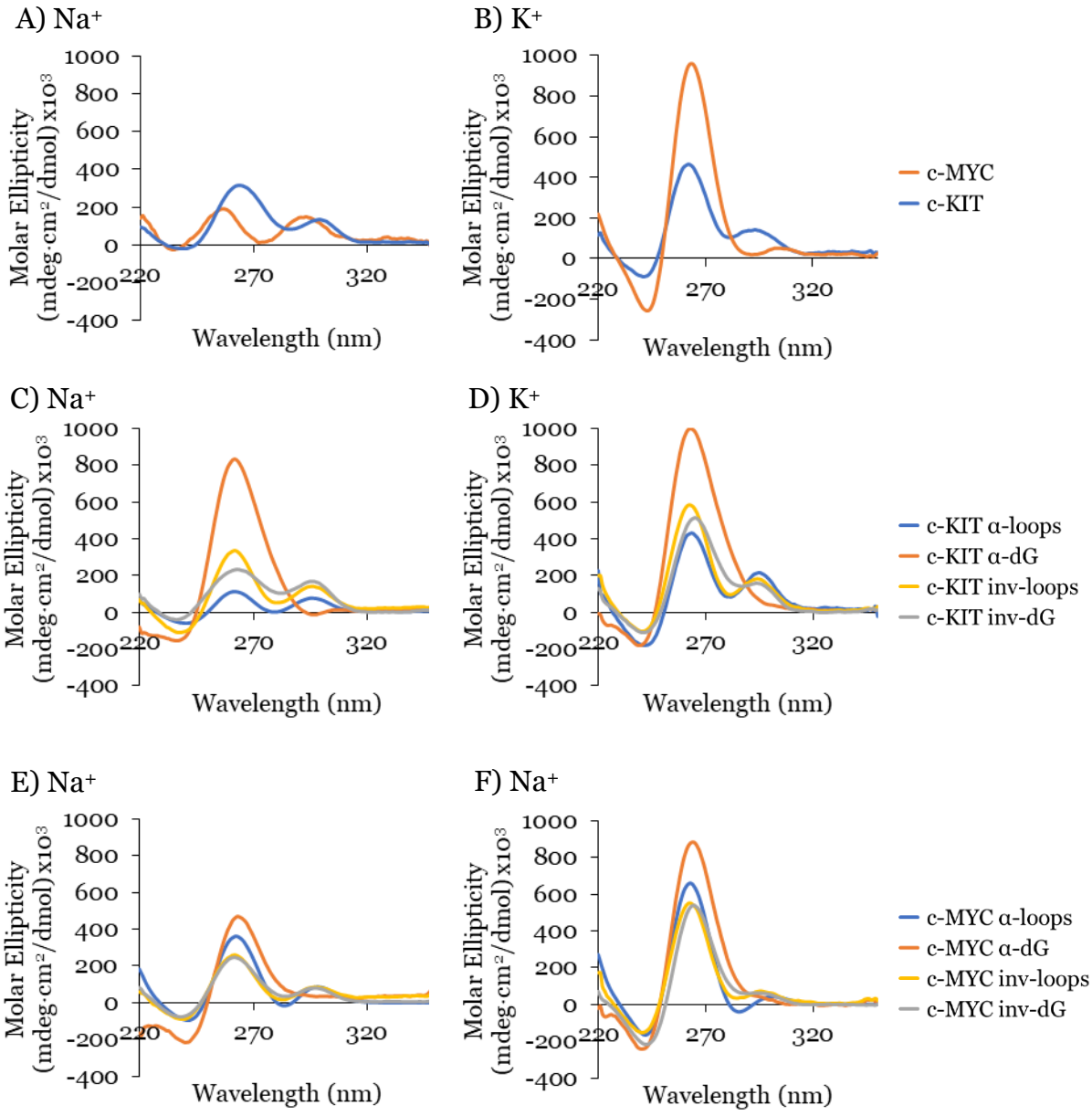


Figure 4.13. CD profiles of modified and unmodified c-MYC and c-KIT sequences in Na^+ and K^+ buffers. A) Unmodified sequences in Na^+ buffer. B) Unmodified sequences in K^+ buffer. C) Modified c-KIT sequences in Na^+ buffer. D) Modified c-KIT sequences in K^+ buffer. E) Modified c-MYC sequences in Na^+ buffer. F) Modified c-MYC sequences in K^+ buffer. Conditions: 20 μM strand concentration, 20 mM sodium phosphate, 10 mM KCl (K^+ buffer only), pH 7.0.

Unmodified c-KIT forms a G4 with mixed topology in K^+ buffer with a large positive peak at 265 nm and a smaller positive peak at 295 nm (Figure 4.13B). This is consistent for all modified sequences except c-KIT-α-dG, which has no peak at 295 nm and a larger positive peak at 265 nm (Figure 4.13D). In K^+-containing buffer, the unmodified c-MYC sequence forms a parallel G4, although it does still have a small peak at around 300 nm (Figure 4.13B). This is also consistent for all c-MYC modifications except for c-MYC-α-dG. Similar to c-KIT-α-dG, this sequence does not give a peak at 300 nm and has a higher intensity peak at 265 nm (Figure 4.13F). CD spectra confirm that these modified sequences retained the native G4 topology, but increased favourability of the parallel topology with α-dG containing sequences was also observed. HP1α prefers to bind parallel G4s so this could result in improved interactions with HP1α. However, it could also indicate a topological change which might disrupt interactions with proteins.

Thermal stability experiments were carried out to investigate any changes in stability resulting from these modifications. All modified c-KIT sequences showed increases in thermal stability (Table 4.6, Figure SI-22 –Figure SI-26). In Na^+-containing buffer c-KIT was unstructured (Figure SI-22B), while modified sequences were not (Figure SI-23), so no T_m value was obtained. However, in K^+-containing buffer most sequences showed an increase in T_m of approximately 30 °C, although c-KIT-inv-loops had the largest increase in T_m and was not completely unfolded at 90 °C (Table 4.6, Figure SI-22C and Figure SI-24). More variation was observed for c-MYC sequences. In Na^+-containing buffer increased T_m was consistently observed, with the smallest for c-MYC-α-dG (ΔT_m = +8 °C, Figure SI-25C) and the largest for c-MYC-α-loops (ΔT_m = +27 °C, Figure SI-25D) (Table 4.6). In K^+-containing buffer inv-loops and inv-dG showed decreases in T_m of -16 °C and -17 °C, respectively (Figure SI-26A and B), while α-dG was comparable to the unmodified sequence (ΔT_m = +3 °C, Figure SI-26C) and α-loops were not fully melted at 90 °C ($\Delta T_m \approx$ +12 °C, Figure SI-26D). These results were not consistent between buffers, with sequences containing inverted nucleotides showing increases in Na^+-containing buffer and decreases in K^+-containing buffer. The c-MYC-α-loops sequence showed the largest increase in thermal stability in both buffers. Overall, changes in T_m are a minor consideration

as particularly large changes could indicate significant changes in G4 topology which may disrupt protein binding. Melting data is mostly considered as context for the results obtained using NMR and BLI.

4.3.5. Evaluation of Duplex Formation in c-KIT and c-MYC Sequences Containing G-Tract and Loop Modifications

^{1}H NMR spectroscopy experiments were performed similarly to the previous experiments with 5′inv-Ttel. Unmodified c-KIT and c-MYC sequences were used as controls for duplex formation, along with the complementary i-c-KIT and i-c-MYC sequences (Figure 4.14A and B). Modified and unmodified G4s were folded and mixed with their complementary strand and ^{1}H NMR were recorded regularly over the course of several days, particularly focusing on the presence of peaks between 10.5 and 14.5 ppm, which correspond to the imino protons of G4s and duplexes. After three days, samples were heated at 90 °C for five minutes, cooled to 5 °C and stored overnight to obtain the thermodynamic product. The i-c-KIT control sequence gave an ^{1}H NMR spectra with two peaks at approximately 13.2 ppm (Figure SI-28). These peaks are consistently seen in duplex formation experiments and are therefore not considered as evidence of duplex formation with ou G-quadruplexes. These peaks are most likely the result of a DNA hairpin. This was described in Chapter 2.2.2. and shown in Figure 2.4. The presence of both guanosine and cytosine to form G-C base pairs allows i-c-KIT to form a hairpin more readily than i-c-MYC.

The unmodified c-MYC and c-KIT sequences give ^{1}H NMR spectra with peaks around 11 – 11.5 ppm, with more clear resolution of peaks for c-MYC (Figure 4.14 A and B). This corresponds to our expectation for a G-quadruplex, and the more resolved c-MYC spectra is an indicator of a single G4 topology, rather than the mixed or unfolded topology of c-KIT. Upon addition of complementary strands, these peaks for c-KIT quickly decrease in intensity while new peaks appear between 12.5 and 14 ppm, indicating formation of A-T and G-C base pairs consistent with duplex formation. Additionally, we do not see evidence of the peaks at 13.2 ppm which were present for the i-c-KIT control, suggesting that this hairpin unfolds rapidly once the G4-forming sequence is added. Immediately upon addition of complementary DNA unmodified c-MYC did not completely refold into a duplex. After heating and cooling, the G4 peaks are almost

completely gone and new peaks have appeared between 12.5 and 14 ppm. These results indicate that for unmodified c-MYC and c-KIT G4s, the duplex is the thermodynamically favoured product when in the presence of the complementary strand. The transition of c-MYC to the thermodynamic product is considerably slower than for c-KIT and required heating to reach an equilibrium between the G4 and duplex structure.

The results for modified G4 structures were consistent regardless of modification type and location. Peaks corresponding to G4 structures were constantly observed during both the kinetic and thermodynamic experiments, indicating that each modification strategy stabilised G4 structures. Some duplex formation was observed, but at all stages the major species present was the G4 (Figure 4.14C – J). These modifications appear to be slightly more effective in c-KIT samples. All modified c-MYC sequences have new low intensity peaks appearing around 13 ppm (Figure 4.14C, E, G and I), while c-KIT samples only have the doublet corresponding to hairpin formation in the complement (Figure 4.14D, F, H, and J). ^{1}H NMR clearly indicates that the G4 structure is the dominant structure in the final equilibrium for all modified sequences. The use of shorter modified regions had a comparable impact on duplex formation to the contiguous modifications tested previously. This allows for tuneable, high-precision modifications with a more controllable impact on G4 topology, which retain the ability to prevent duplex formation. Developing a larger array of models containing these types of modification would also allow for further experimentation with the required number of modifications to achieve duplex disruption. Modifications of a single c-KIT loop region (i.e., 5 nucleotides, CGC_2A) could still provide considerable duplex disruption while further minimising the effect on G4 stability. However, the complete modifications of loop and dG regions carried out here provides sufficient evidence that this type of controlled modification can still disrupt duplex formation.

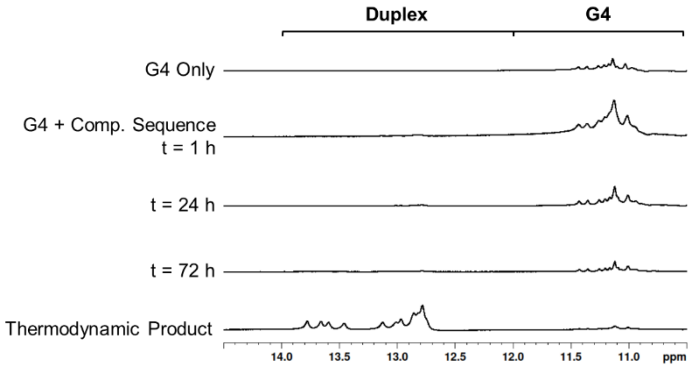
One notable feature of these spectra is that the peaks observed in the 10.5 – 12 ppm range were not always consistent with the unmodified G4. While we saw similar indicators for topology in CD spectroscopy, the overall structure of many sequences appeared different according to ^{1}H NMR spectroscopy. Of note are c-MYC-inv-dG and c-MYC-α-loops (Figure 4.14C and I, respectively) which have

far more resolved peaks than the unmodified G4. This suggests that these sequences form a single G4 topology rather than other sequences which contain a mixture of G4 topologies. These peaks seem to correspond to the dominant structure in the unmodified sample, but fewer other peaks appear, and the peaks are better resolved for modified sequences. While not as pronounced, c-KIT-inv-loops and c-KIT-α-loops (Figure 4.14F and J, respectively) resemble the native c-KIT sequence more closely than the other modified sequences. These results suggest that modifications of the loops have a smaller impact on the hydrogen-bonding within the G-tetrad, which could be expected. However, it is unclear if the differences in ^{1}H NMR spectra are correlated to a difference in G4 topology sufficient to disrupt protein interactions.

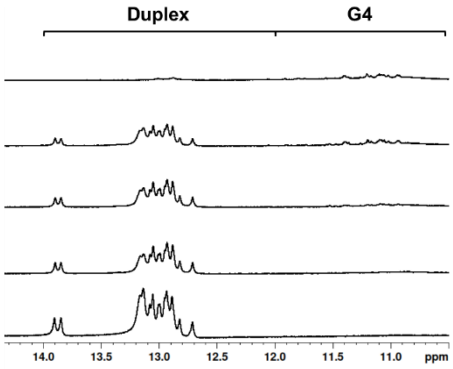
Conversely, ^{1}H NMR spectra of c-MYC-inv-loops and c-KIT-α-dG (Figure 4.14E and H, respectively) were less resolved than the native sequence, suggesting that these oligonucleotides may enable the formation of additional topologies. In c-MYC-inv-dG and c-MYC inv-loops samples (Figure 4.14C and E, respectively) specifically we also observed the appearance of peaks around 15 – 16 ppm in the presence of complementary DNA (Figure SI-27A). These peaks are consistent with i-motif formation, suggesting that, without the possibility of forming a stable DNA duplex, the complementary strand was able to fold into an i-motif instead. These peaks were consistent throughout kinetic and thermodynamic experiments, although they were not present for all samples.

Based on the combination of ^{1}H NMR and CD spectroscopy we can conclude that α-dG modifications favoured parallel topology G4s as expected, while inverted nucleotides maintained similar G4 topology to unmodified G4s. However, ^{1}H NMR spectra indicate that while topology is consistent, the arrangement of H-bonding guanosine nucleobases in the G-tetrads differs slightly. Previous CD experiments with 5′inv-Ttel showed that this G4 favoured a parallel topology more than the unmodified G4. Moreover, the imino proton peaks corresponding to the G-tetrads had a similar distribution to unmodified sequences, suggesting a similar, symmetrical structure. This was not the case for modified c-MYC and c-KIT sequences. The imino proton regions of these spectra

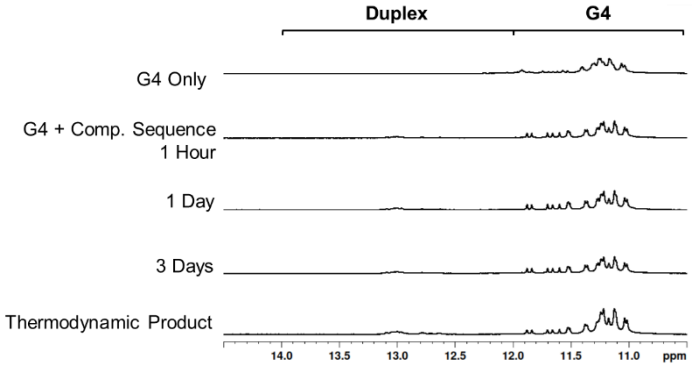
A) c-MYC



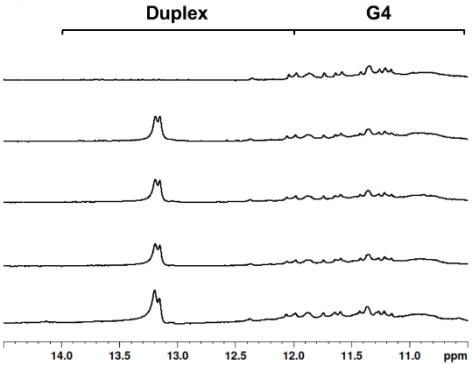
B) c-KIT



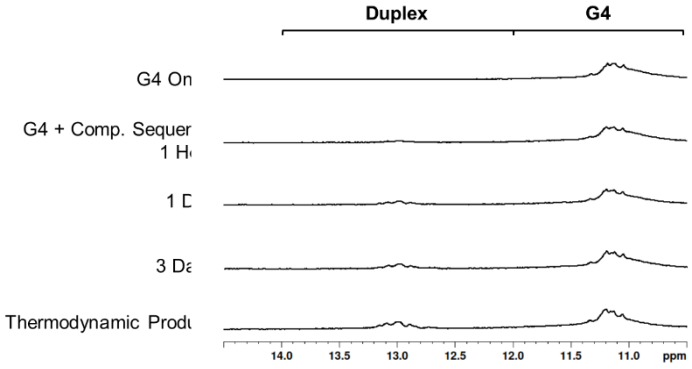
C) c-MYC-inv-dG



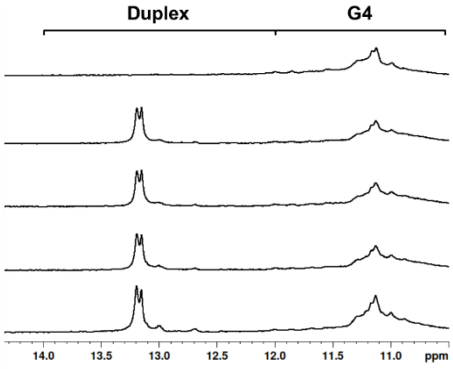
D) c-KIT-inv-dG



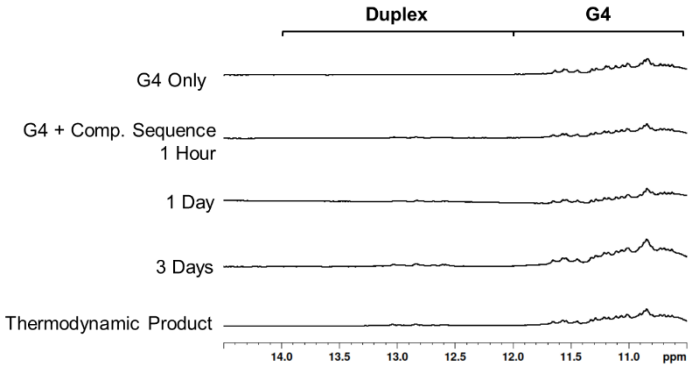
E) c-MYC-inv-loops



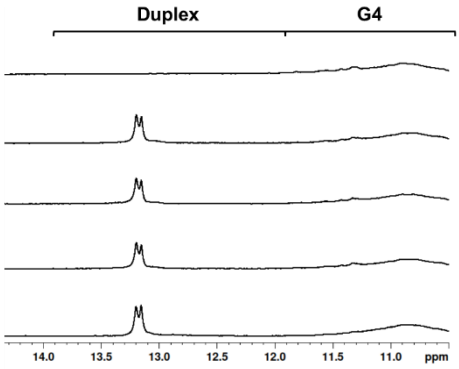
F) c-KIT-inv-loops



G) c-MYC-α-dG



H) c-KIT-α-dG



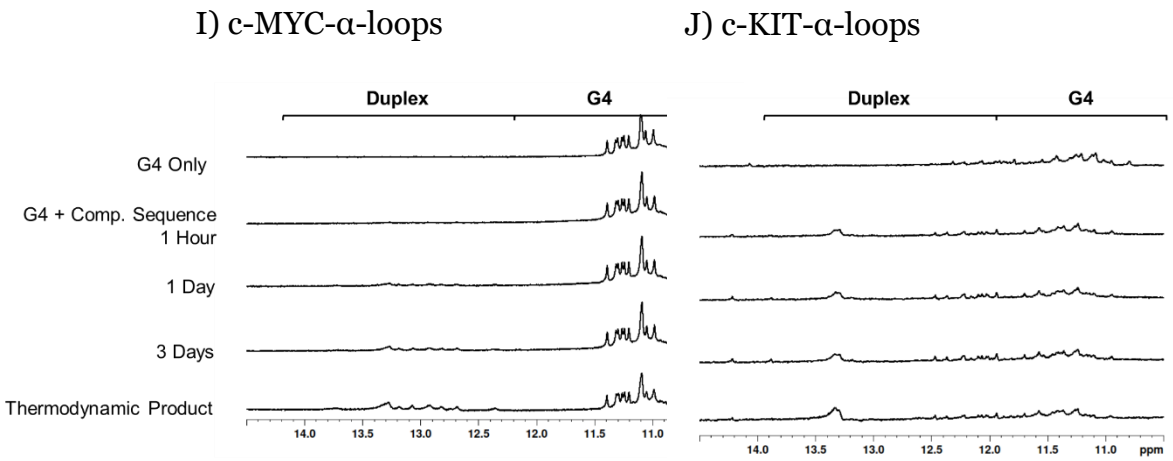


Figure 4.14. ^{1}H NMR profiles of duplex formation for c-MYC and c-KIT sequences containing inverted- and α-nucleotides. A) Unmodified c-MYC. B) Unmodified c-KIT. C) c-MYC-inv-dG. D) c-KIT-inv-dG. E) c-MYC-inv-loops. F) c-KIT-inv-loops. G) c-MYC-α-dG. H) c-KIT-α-dG. I) c-MYC-α-loops. J) c-KIT-α-loops. Conditions: 200 μM strand concentration, 20 mM sodium phosphate, 10 mM KCl, 10% D_2O, 1% TSP, pH 7.0.

changed considerably for sequences containing modified nucleotides. Despite this, these sequences display a similar ability to prevent duplex formation when complementary i-c-MYC and i-c-KIT strands were added. The changes in topology for sequences containing α-dG are explained by the preference of sequences containing α-dG to form G4s of parallel topology. For other modifications it is possible to identify similarities between the native G4 imino proton peaks and the peaks for modified sequences. For c-MYC, the inv-dG and α-loop sequences appear to have similar arrangements of imino proton peaks, with better resolution, resulting from greater preference for a single topology. The inverted loop sequence also has peaks in a similar region to the unmodified sample, but resolution of these peaks is worse. For c-KIT, this effect is similar, but less obvious. c-KIT-inv-dG gives imino proton peaks which appear to be partially matched with several lower intensity peaks in the unmodified sample. In both cases, inverted dG modifications appear to result in better resolved structures, but the modified c-KIT sequence favours a structure which is less dominant for the unmodified sequence. c-KIT-α-loops and inv-loops resemble the spectra of the unmodified species more closely, with imino proton peaks primarily around 11 ppm. However, the intensity of these peaks is considerably reduced, particularly c-KIT-inv-loops. According to ^{1}H NMR spectra inverted dG sequences and α-loop sequences, particularly c-MYC, seemed to most closely resemble the original G4 topology. However, analysis of these imino protons in

^{1}H NMR indicated that this modification caused more substantial changes to the G4 topology of c-MYC and c-KIT sequences than was observed for Ttel sequences.

4.3.6. Interactions of c-KIT and c-MYC Sequences Containing G-Tract and Loop Modifications with HP1α

Finally, we performed BLItz experiments (see Chapter 7.2.8.) with modified sequences to determine if they showed affinity to HP1α (Figure 4.15). In the initial experiments, the protein showed similar affinity to modified c-KIT sequences and the unmodified control (Figure 4.15A). c-KIT-α-loops showed slightly higher affinity to HP1α compared to unmodified c-KIT, but the difference was relatively small. Modified c-MYC sequences also showed affinity to HP1α (Figure 4.15B). However, dissociation data, particularly for c-MYC sequences were not reliable due to degradation of the protein samples used. These experiments require repetition when new protein is purified, but this was not included here due to time constraints. These samples also indicated initially strong interactions between HP1α and most c-KIT and c-MYC sequences, but this was followed by an

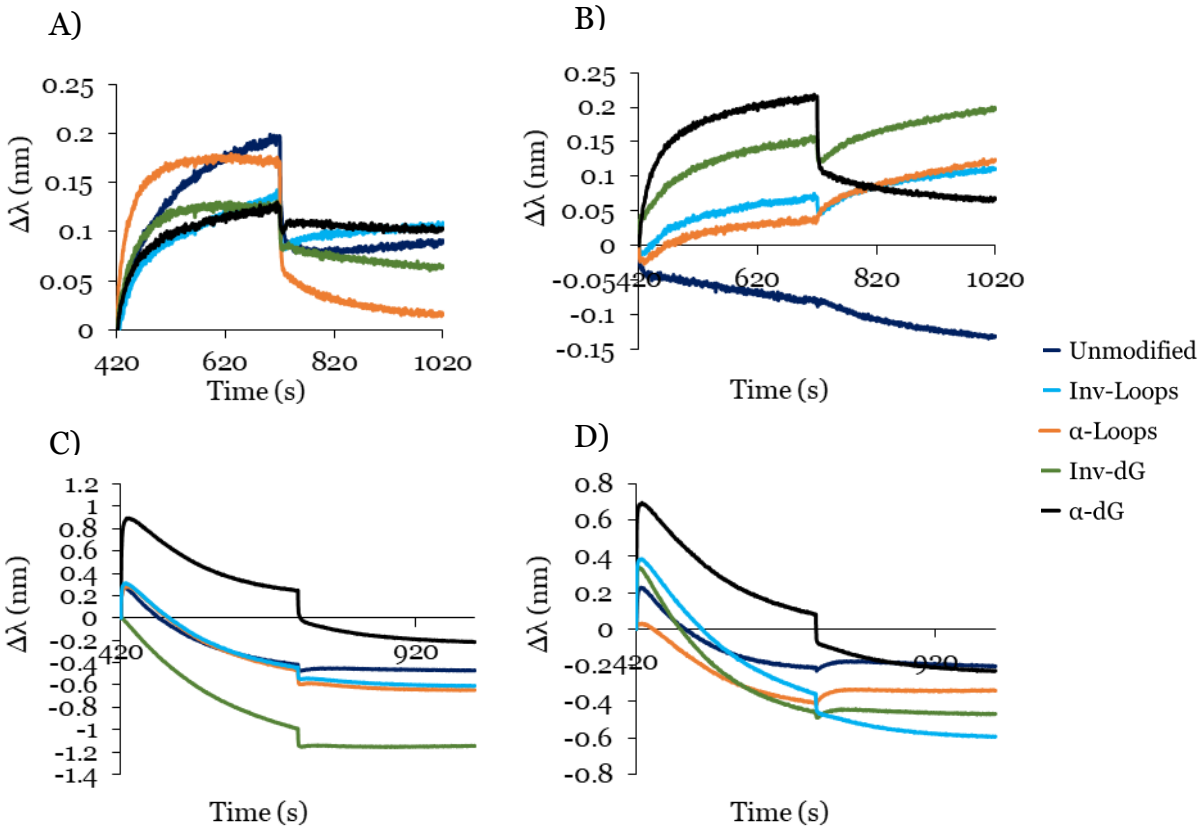


Figure 4.15. Two examples of biolayer interferometry (BLI) results of immobilised HP1α with c-KIT (A/C) and c-MYC (B/D). A) and B) were initially run using a his$_6$-tagged HP1α stock stored at -20 °C. The experiment was repeated in C) and D) with new stock, which had been stored at -80 °C. Conditions: 2 μM strand concentration, 100 μg/mL protein concentration, 20 mM sodium phosphate, 10 mM KCl, pH 7.0.

unusual decrease in binding (Figure 4.15C and D). Apart from the initial c-KIT experiment, all sequences containing α-dG nucleotides showed higher affinity to HP1α. Results for inv-dG, inv-loops and α-loops were inconsistent but generally showed some binding to HP1α.

These results are most likely due to degradation of the available protein stocks over time. The protein concentration measured using UV-Vis had decreased significantly and BLItz experiments consistently showed less protein bound to the tip during initial association compared to previous experiments and reference experiments showed decreases in protein bound to the tip over time (similar to the association steps in Figure 4.15C and D), although this was more pronounced in the presence of G4 DNA. However, the preliminary results suggest that more precise modification of the c-KIT and c-MYC sequences had limited the changes in G4 topology suggested to have occurred for c-KIT-inv, therefore allowing binding to the protein. We attempted to confirm these results using an electrophoretic mobility shift assay (EMSA, see Chapter 7.2.7.) in case the Ni-NTA tips used for BLItz were faulty. However, the modified and unmodified sequences could not compete for HP1α binding to the labelled TERRA45 probe. This indicates that HP1α for these sequences is significantly lower than for TERRA45, which agrees with previous BLItz results.[38] This EMSA could be improved by instead using labelled c-MYC and c-KIT probes. Overall, preliminary BLItz data, while not consistent enough to obtain accurate K_D values, does suggest that c-MYC and c-KIT samples with more selective nucleotide modifications resulted in sequences that not only disrupted duplex formation, but were also able to interact with HP1α.

4.4. Conclusions and Perspectives

Non-native DNA nucleotides present an effective method of building thermodynamically stable models of genomic G4s and other non-canonical secondary structures, but further investigation is necessary to determine if they could be developed into inhibitors of G4-specific proteins. Throughout these experiments we have consistently observed modified sequences which conserve the existing G4 topology, while disrupting the formation of duplex DNA. In terms of creating thermodynamically-trapped G4s in duplex structures to study their interactions with proteins, these properties are ideal, allowing us to potentially

develop large DNA duplexes containing a G-rich region which closely resembles native DNA. As expected, the inclusion of both inverted- and α-nucleotides was sufficient to disrupt duplex formation, while allowing formation of stable G4s. However, in the case of inverted telomeric sequences we did see some change in G4 behaviour potentially resulting from the introduction of a new loop to accommodate the internal polarity inversion site or distortion of the G-tetrads. This was also seen for the c-KIT sequence which was inverted in the centre of the structure.

Initial experimentation with modification sites selected to minimise impact on the G4 structure resulted in sequences that still disrupted duplex formation but preserved the topology of the original G4 more closely. The ability to disrupt duplex formation was consistent across all sequences, but, unlike Ttel, c-MYC and c-KIT sequences containing more precise modifications showed preliminary evidence of interactions with HP1α. Originally, we proposed that changing only the loops would give the best results, as this would cause no change to the hydrogen-bonding arrangements of the G4. This did not appear to be the case, as HP1α showed no preference for sequences with modified loops or G-tetrads.

Inverted G-tetrads had a greater effect on resolving a single G4 topology than inverted loops. We assumed that α-dG would provide a similar result due to α-dG G4s previously documented preference for parallel topology. While CD does seem to indicate this, NMR did not indicate only a single topology was formed. There are several reasons why this might be the case. Multiple different parallel topologies might be possible, but it could also indicate distortion of the G-tetrad or increased aggregation of G4s resulting from the modified nucleotides. Of all the modified c-MYC and c-KIT sequences, only c-MYC inv-dG and c-MYC α-loops gave clearly defined peaks. The lack of consistency between similar modifications makes it difficult to predict outcomes for other sequences. It is also unclear if these factors are important for HP1α binding, but the effect of additional loops of HP1α has provided insight into why it may favour parallel G4s and further investigation of the role of G4 loop structures in HP1α binding is ongoing.

Relative to other strategies, non-native nucleotides seem to provide better inhibition of duplex formation than PDS or hydrophobic modifications did in similar experiments. The simplicity of synthesising and purifying sequences

containing these modifications compared to developing new ligands or incorporating phosphate modifications make this an attractive strategy for developing a wide range of G4 models, though it will require further testing to determine the optimal sequences and modification sites for these models.

This strategy needs to be expanded to a wider range of sequences to further these findings. Pu39, used for testing ligands previously, is an interesting sequence for studying G4 interactions because of its location in the promoter region an oncogene, BCL-2,[28, 106-108] and inverted nucleotides are one possibility for modifying and studying this sequence. It contains six G-rich regions, meaning it folds into multiple different G4 structures incorporating different combinations of these regions. The sequence is also longer, making it more of a synthetic challenge. Inverting loop regions and some G-tracts to manipulate the G4 topology that can form is a potentially unique new application of this method. Similar strategies could be applied to other longer G4 sequences.

While this strategy could be expanded to other sequences, possibly a more interesting application is the use of this method to develop models of the G4 structure as it appears in cells. This strategy is an effective method of addressing the issues with G4s incorporated into large DNA duplexes, and it is necessary to test these properties in those structures. This has been partially addressed in the 5′inv-Ttel tail experiments, which demonstrated the ability to produce DNA structures containing canonical duplexes adjacent to thermodynamically trapped G4s. Next steps for this project involve synthesising sequences, possibly c-KIT and c-MYC, which contain loop or G-tetrad modifications, as well as adjacent duplex-forming regions. In addition, these sequences require overhangs at the 5′-end to allow for DNA ligation to allow for thermodynamically stable G4s to be incorporated into larger DNA structures. This is a process in which two DNA molecules are joined by ligation and will be required to make DNA sequences of sufficient size to create the protein/DNA complexes that we intend to study, including nucleosomal arrays. Accomplishing this goal is a significant undertaking and would likely be coupled with a more detailed study of the resulting complex.

Internal polarity inversions were able to successfully create thermodynamically stable G4s. They show more promising properties for their

study in large DNA structures than ligands and hydrophobic modifications, and with thoughtful implementation these modifications could create the intended thermodynamically stable DNA-G4s within canonical duplexes for studying protein interactions with chromatin.

5. Chemical Cross-linking of G4s using Copper(I)-Catalysed Azide-Alkyne Cycloaddition

5.1 Introduction

Chemical cross-linking is a method for secondary structure stabilisation which has been used for both DNA and proteins. This strategy introduces a chemical link between two nucleotides or amino acids, effectively enforcing a particular fold in the structure and encouraging the formation of the preferred secondary structure. While the examples mentioned below demonstrate the application of this type of modification to proteins, canonical duplex DNA and some single-stranded DNA structures, it had not previously been employed successfully within G4 structures.

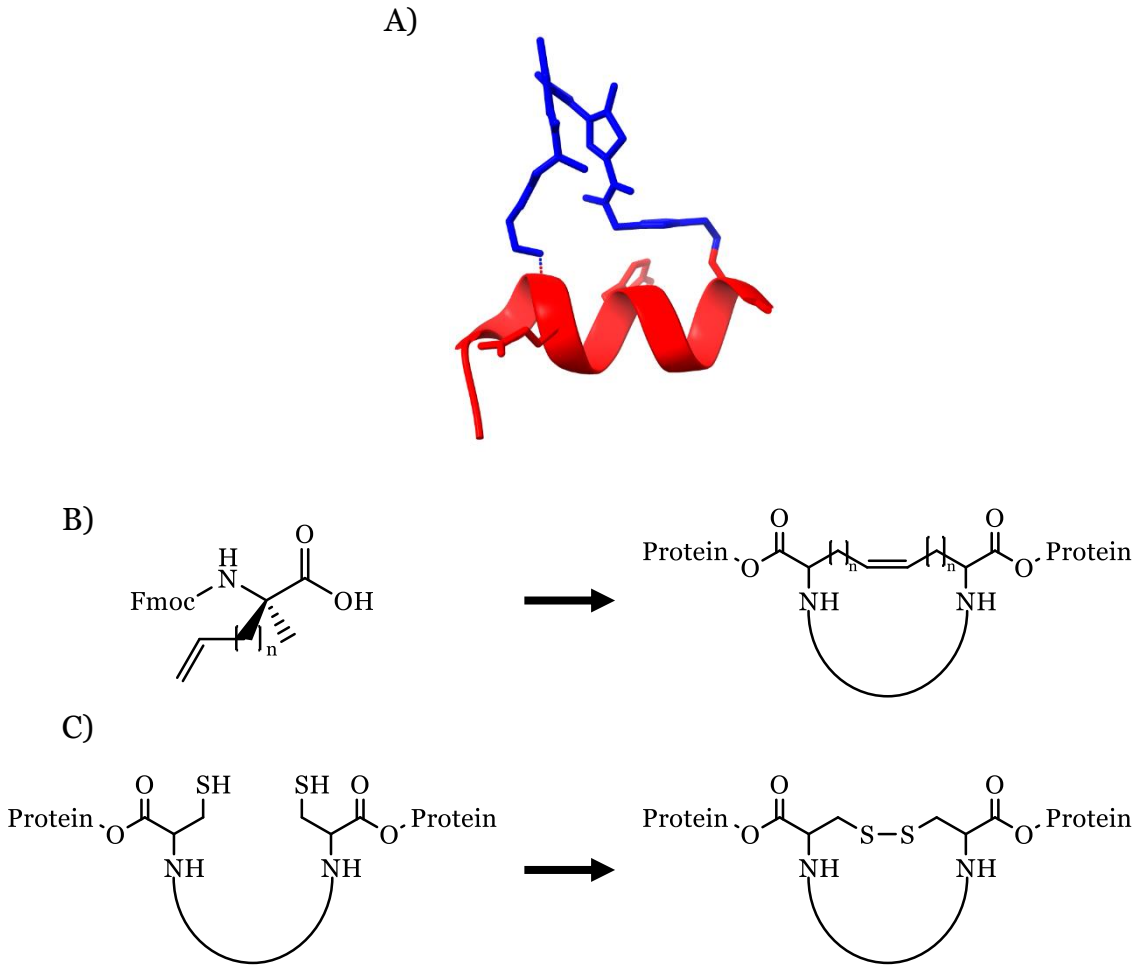


Figure 5.1. A) X-ray crystal structure of a stapled peptide using a diarylethene moiety to create a photosensitive cross-link. B) O-Allylserine's application for introducing peptide staples. Two modified amino acids are incorporated into the peptide during chemical synthesis and cross-linked post-synthetically via ring-closing metathesis. C) Another method of obtaining a cross-linked peptide by forming disulphide bonds between cysteines.

Stapled peptides are a common method of stabilising protein secondary and tertiary structures, particularly α-helices, utilising cross-linking, as shown in Figure 5.1A,[63] which is part of a peptide containing a photosensitive peptide staple. A common method of introducing these cross-links is ring-closing cross metathesis wherein two alkene functional groups at different positions in the polypeptide are reacted to form a large ring. A range of modifications, such as O-allylserine (Figure 5.1B)[109] or substituted non-native amino acids,[110, 111] can be used to introduce alkene functionality, and these are then cross-linked using ruthenium-catalysis based reactions. However, other modifications, such as disulphide bond formation between cysteines (Figure 5.1C),[112] have also been tested as useful methods of stabilising α-helical structures with comparable outcomes. These structures have been shown to improve formation of α-helical structures, increase protease resistance and offer interesting possibilities as targeted drugs for binders of α-helices.

Similar strategies have been applied to oligonucleotides. Originally, our interest in this strategy was based on the post-synthetic formation of disulphide bonds reported by Prestinari and Richert[63], which was shown to stabilise canonical duplexes (Figure 5.2A). They utilised Sonogashira coupling to modify 5-iodouridine with 3-butyn-1-ol or 5-hexyn-1-ol, followed by formation of a thioester using thiobenzoic acid. The sulphur atoms could then be used to form disulphide bridges after the modification was incorporated into an oligonucleotide. Several sequences were synthesised containing these modifications. Modified sequences formed duplexes with complementary DNA and had increased thermal stability. Additionally, sequences tested which contained base pairing mismatches showed smaller decreases in stability than unmodified sequences. Prestinari and Richert also highlight the advantages of disulphide bridges over other modification strategies. Specifically, they mention reductive amination and amide formation as alternatives considered early in developing the disulphide modification strategy. Disulphides were selected because they were formed spontaneously post-synthetically in the presence of oxygen, removing the requirement for additional synthetic steps. They also have limited side reactions, introduce minimal steric strain, are compatible with DNA synthesis and deprotection steps.

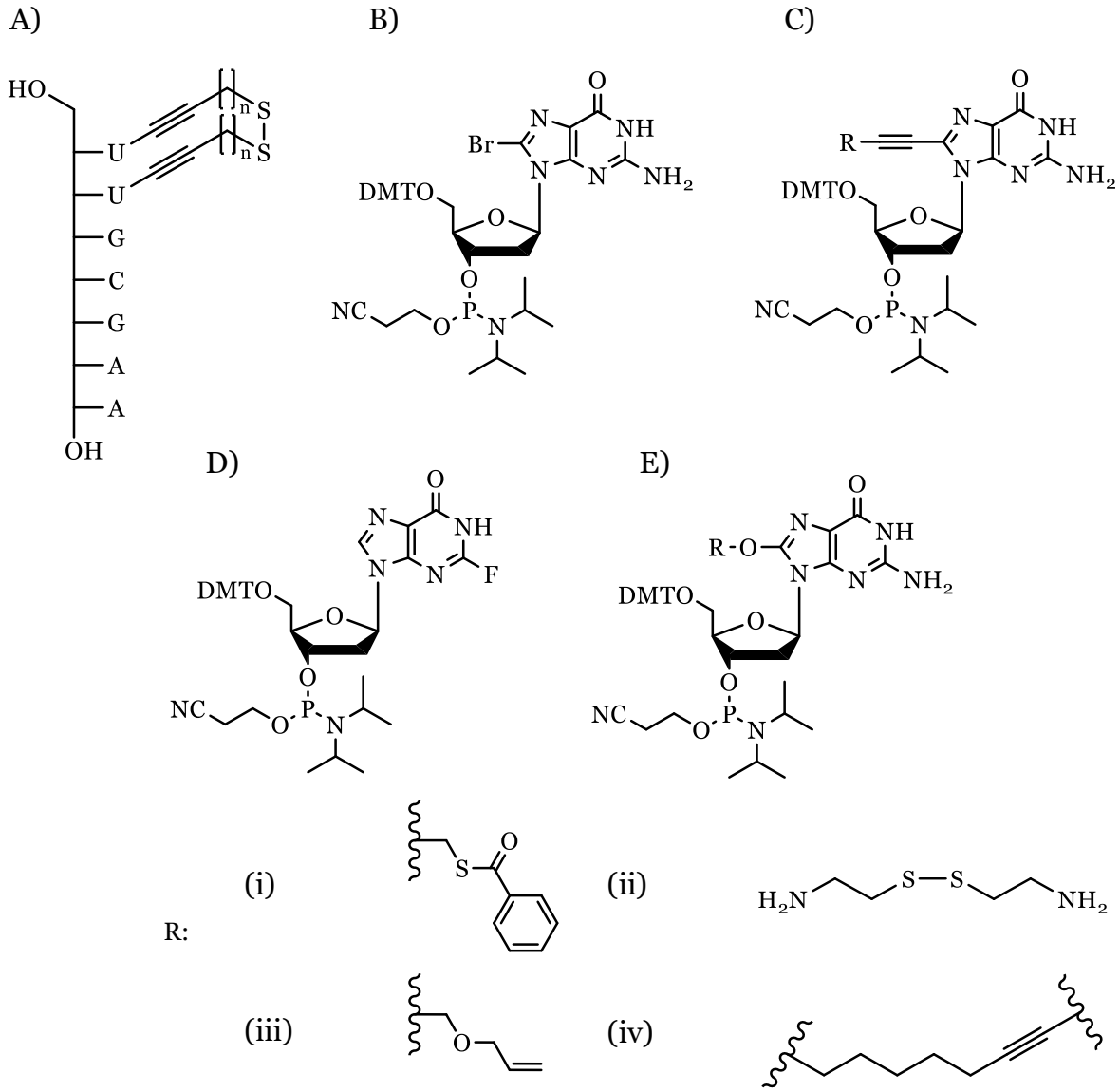


Figure 5.2. Examples of previously explored methods of introducing chemical cross-links to DNA. A) A disulphide linkage introduced to duplex DNA, shown to improve duplex stability. B) 5′-DMT-8-bromo-2′-deoxyguanosine phosphoramidite used as a precursor for DNA synthesis and Sonogashira coupling. C) A proposed guanosine modification with a functional group added at the 8-position via Sonogashira coupling. D) 5′-DMT-2-fluoroinosine phosphoramidite used to post-synthetically incorporate functionality via an amine substitution reaction. E) A proposed guanosine modification using nucleophilic substitution of 8-bromo-2′-deoxyguanosine. Potential modifications: (i) A benzothioate, used for introducing disulphide bridges; (ii) cystamine, an amine also used for introducing disulphide bridges; (iii) 3-(prop-2-yn-1-yloxy)prop-1-ene, proposed as a method of introducing an alkene for cross-metathesis; (iv) crosslink produced from 1,7-octadiyne, proposed to introduce cross-link post-synthetically with 8-bromo-2′-deoxyguanosine phosphoramidites.

Initially I had attempted to expand this strategy to guanosine and G4s.[113] 2′-Deoxyguanosine (Figure 5.2B) was modified via bromine substitution at the 8-position. 8-bromo-2′-deoxygaunosine was then modified using a scheme based on the modification strategy developed by Prestinari and Richert.[63] The final product would include a thioester at the 8-position of guanosine, shown in Figure 5.2C, (R = (i)), which would be used in G-rich sequences similarly to the modified

uridine described previously. The translation of this strategy proved unsuccessful due to the formation of a large number of side products during Sonogashira coupling. While this product was characterised by NMR and mass spectrometry, successful formation of the product was unreliable and the subsequent reaction with thiobenzoic acid was unsuccessful. A variety of reagents was tested to reduce formation of Sonogashira side products, but ultimately this approach was abandoned.

Instead, a new method of introducing sulphur containing functional groups to oligonucleotides were explored.[113] 2-Fluoroinosine phosphoramidite (Figure 5.2D) is commercially available and can be used as a convertible nucleobase. The fluoride could be substituted post-synthetically with various amines. Several sulphur-containing amines were utilised (e.g., cystamine (Figure 5.2 (ii)) to synthesise sulphur-containing G4-forming sequences. This method was more successful, and we were able to obtain G-rich sequences containing sulphur groups. However, disulphide bridges were reported to form spontaneously when exposed to air, which was not observed in this case. DTT was used to cleave any existing disulphides and oxidising conditions were used to encourage disulphide formation. These challenges led us to conclude that this modification did not allow for sulphur atoms to interact sufficiently to form disulphide bridges. It was also noted that disulphide bridges are cleaved in the presence of dithiothreitol (DTT), a strong reducing agent, which is also used during protein preparation and storage. We were therefore concerned that disulphide bridges would be unstable if they were introduced to solutions containing proteins. Ultimately, we determined that this strategy was not effective in our hands and would not result in useful models.

Several other possibilities were also explored as alternatives to disulphide bridges. We suggested that 8-bromo-2′-deoxyguanosine could be converted to a phosphoramidite (Figure 5.2E) and incorporated during DNA synthesis. This would allow Sonogashira coupling to be carried out post-synthetically. 1,7-Octadiyne (Figure 5.2C, R = (iv)) was tested as a potential method of joining two 8-bromo-2′-deoxyguanosine sites in the DNA via two Sonogashira coupling reactions, but the large number of side products with similar retention times in HPLC made it impossible to isolate and identify the desired oligonucleotides.

Alkene functionality was considered to allow for ring-closing metathesis using Grubb's Catalyst, similar to the preferred strategy for peptide stapling. 3-(Prop-2-yn-1-yloxy)prop-1-ene (Figure 5.2C, R = (iii)) was synthesised and used in Sonogashira coupling but this approach encountered similar issues to previous Sonogashira coupling reactions. Sonogashira coupling had been the primary challenge of these strategies, so alternatives such as nucleophilic aromatic substitution of 8-bromo-2′-deoxyguanosine were considered (Figure 5.2E). However, test reactions with methanol and butanol returned only the starting materials.

Eventually, we considered a strategy using copper(I)-catalysed azide-alkyne cycloaddition (CuAAC). This method was developed concurrently with the previous experiments by Dr. Hari Kurup within the Filichev lab.[62] The method creates a triazole that links two distant modified nucleotides in the DNA (Figure 5.3A), each containing either the azide or the alkyne functional group. Once synthesised and purified, CuAAC was performed to create the cross-link (see Chapter 5.2.1 below). The resulting sequence formed a cross-linked single-

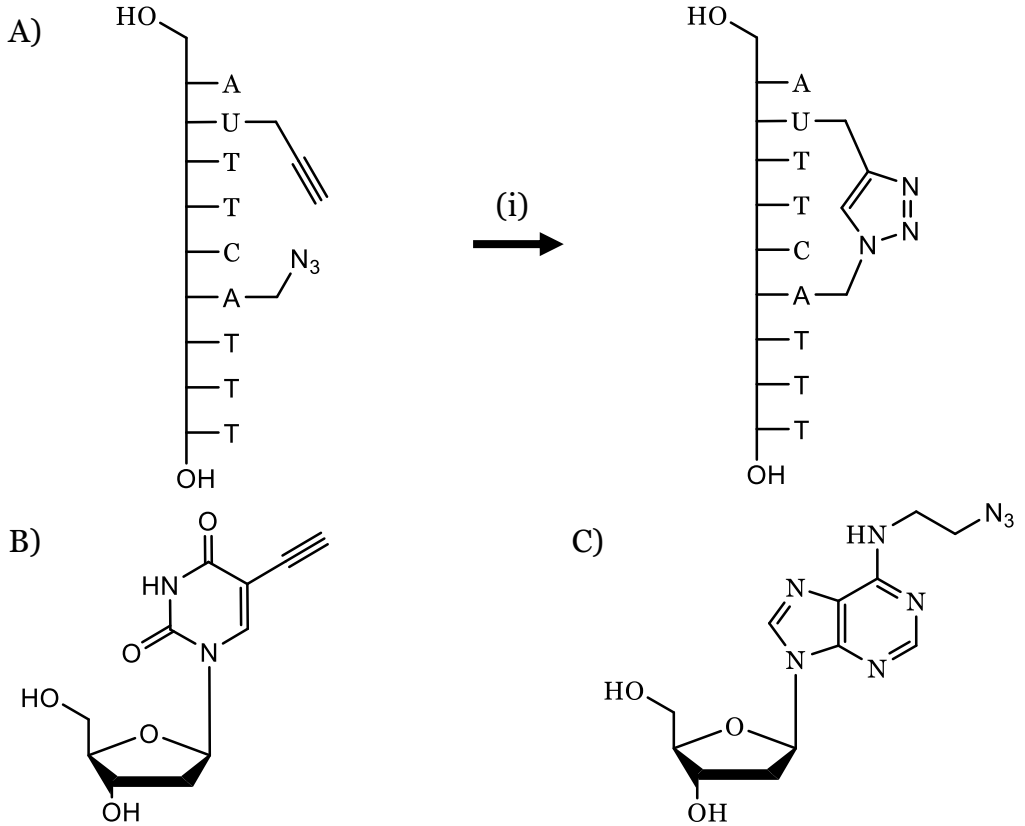


Figure 5.3. A) Formation of oligonucleotide containing a triazole linker using adenosine modified with azide and alkyne functionality for copper(I)-catalysed azide-alkyne cycloaddition. (i) THPTA, $CuSO_4$, sodium ascorbate, solvent: H_2O. Previous modifications of single-stranded DNA, forming a triazole cross-link: B) 5-Ethynyl-2′-deoxyuridine. C) N^6*-Ethylazide-2′-deoxyadenosine.*

stranded DNA, which was shown to bind to and inhibit the APOBEC3A enzyme with low nM potency. Like the strategy discussed in Chapter 4, cross-linking relies on modified nucleotides which are incorporated during DNA synthesis. In previous chapters a major concern with each strategy was the effect modifications would have on protein binding. The cross-linked single-stranded DNA produced by Dr. Kurup had resulted in better substates for the enzyme, depending on the location of modifications. We hypothesised that this strategy could also be applied to controlling the topology and thermodynamic stability of G4s and could produce thermodynamically stable G4s useful for creating larger DNA structures.

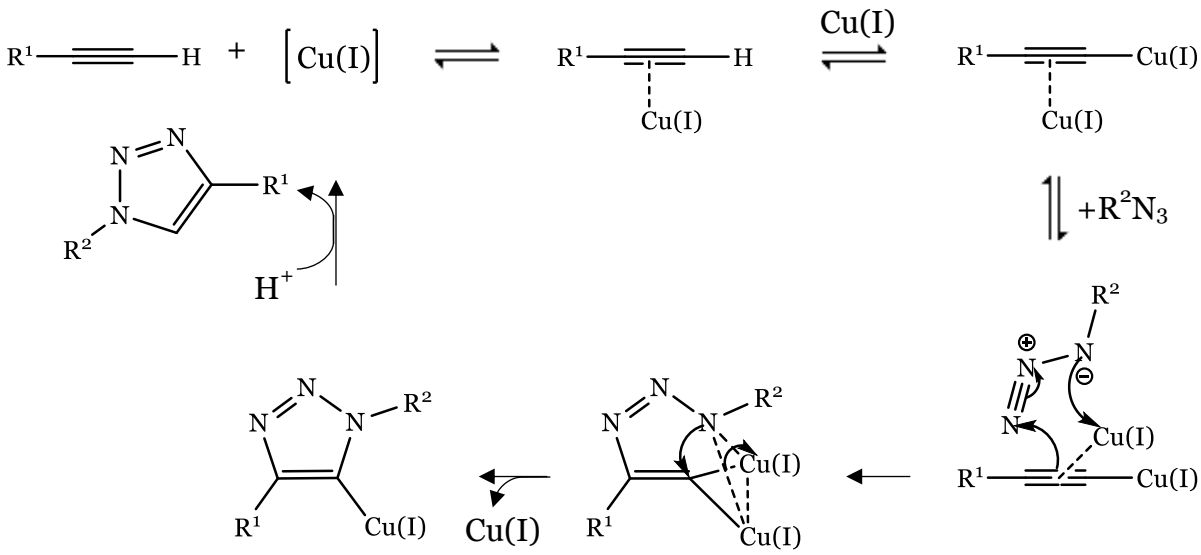
5.2. Methodology

Dr. Hari Kurup's original modifications had used two modified nucleotides (Figure 5.3B and C) containing either an alkyne or azide, which were created specifically to target the APOBEC3 enzyme. After incorporation into DNA, post-synthetic copper(I)-catalysed azide-alkyne cycloaddition produced a cross-linked DNA sequence (Figure 5.3A). The nucleotides were modified at the N^6 position of adenosine with a two-carbon chain between the functional group and the nitrogen of adenosine. The functional groups were stable during DNA synthesis and could be reliably cross-linked post-synthetically to create the triazole linkage between two nucleotides. Previously, some adenosine nucleosides modified with similar long alkyl azide linkers had also been synthesised using other approaches,[114] but were not converted to phosphoramidites and H-phosphonates. Longer linkers allow for greater flexibility in modification sites but are not necessary for cross-linked G4s to be produced.

5.2.1. Copper(I)-Catalysed Azide-Alkyne Cycloaddition

Click-reactions are a group of reactions commonly used for conjugating two molecules with compatible functional groups. These reactions are characterised by being highly thermodynamically favoured, meaning they have high yields and few or no side products. CuAAC is a useful click-reaction for joining alkyne and azide functional groups, forming 1,4-triazoles. A copper(I) catalyst is necessary to ensure the formation of the 1,4-triazole, as a 1,5-triazole variant can also form when samples are heated. Two Cu(I) ions coordinate to the alkyne, which facilitates an electron rearrangement in the presence of the azide and results in the formation of a 1,4-triazole and the regeneration of the copper catalyst

(Scheme 5.1).[115, 116] Copper(I) ions are provided by a $CuSO_4$ solution, with Cu(II) reduced by the addition of freshly prepared sodium ascorbate. Tris((1-hydroxy-propyl-1*H*-1,2,3-triazol-4-yl)methyl)amine (THPTA) is used as a water-soluble ligand for Cu(I) which stabilises Cu(I) in aerobic conditions. THPTA was used in place of the previously used tris(benzyltriazolylmethyl)amine (TBTA) to avoid the necessity of adding DMSO to the reaction mixture as TBTA is insoluble in water. 2 M Triethylammonium acetate (TEAA) is also added as a buffer.



Scheme 5.1. Mechanism of copper(I)-catalysed azide-alkyne cycloaddition (CuAAC)

This reaction is ideal for DNA modification because the functional groups it utilises are compatible with DNA synthesis. Both the azide and alkyne are stable in the conditions through all steps of DNA synthesis. Click reactions have high yields with few side products, simplifying purification of the final product. The reaction also has no effect on the structure of native nucleotides.[117-120] This hopefully contrasts with some of the strategies discussed previously, which were shown to result in differences in G4 topology, potentially as a result of changes to native nucleotides. However, azides can react with trivalent phosphorus, as discussed in Chapter 3. Phosphoramidites synthesised previously tended to cyclise if stored for extended duration, meaning they were usually prepared immediately prior to DNA synthesis. To avoid this problem, we synthesised an azide modified H-phosphonate, which contains a pentavalent phosphorus, and is therefore unreactive with the azide. Additionally, modified nucleotides were cleaved from the solid support and deprotected at 30 °C for 24 hours and freeze-

dried to prevent formation of undesirable 1,5-triazole linkages, which could form when samples were heated and concentrated under vacuum.

To implement this methodology for G4s we used 2′-O-propargylguanosine phosphoramidite (Figure 5.4A), a commercially available phosphoramidite containing a terminal alkyne at the 2′-O-position, which can be used instead of dG.[121] This alkyne could then be cross-linked with a modified adenosine in the G4 loop which contains an azide.

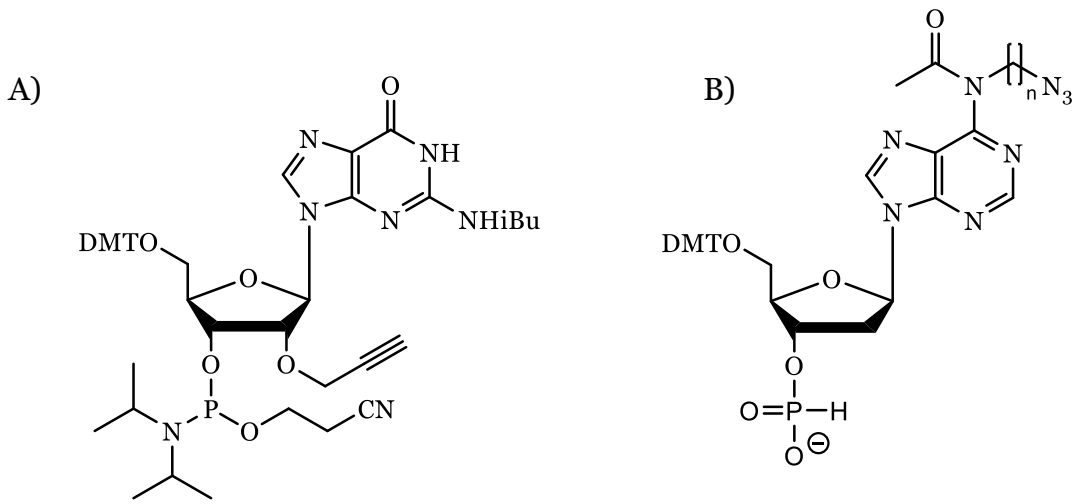


Figure 5.4.; A) Commercially available 2′-O-propargyl-5′-DMT-guanosine phosphoramidite. B) N⁶-*Modified adenosine containing variable length azide linkers.*

5.2.2. Structural Considerations for Creation of G4 Cross-Links

We proposed two modifications to the tel sequence, intended to test the ability of chemical cross-links to stabilise both a parallel and an antiparallel topology. The tel sequence contains an adenosine nucleotide in the loops, making it potentially ideal for dA-dG cross-link formation. The structure of several telomeric sequences have been determined using NMR, which we used to determine modification sites and lengths of cross-links. We used RNA-based G4s when structures were available because the inclusion of 2′-O-propargyl guanosine mimics the presence of the 2′-OH found in RNA, and we therefore obtain more accurate distances and orientations from the RNA models. The approximate length of various linkers containing triazoles was calculated using a *6-31 G** basis set and compared to the distance between potential modification sites in the structure of telomeric G4s. The two-carbon linker (Figure 5.5A, n = 2) had a length from of approximately 7.7 Å from the nitrogen of the amine (N⁶ of adenosine) to the oxygen of the ether (2′-position of guanosine). Four- and six-

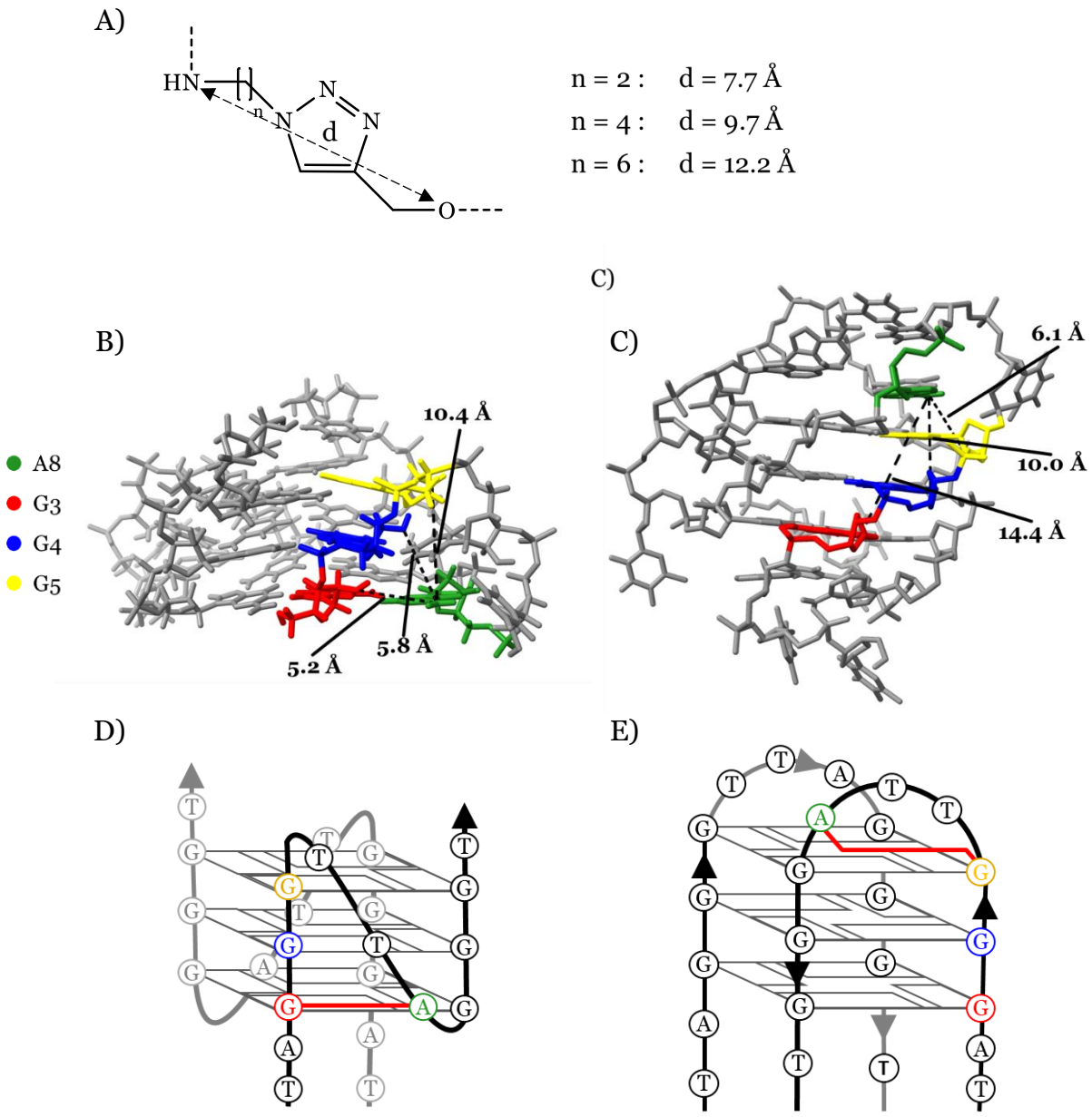


Figure 5.5. Measurement between intended cross-link modification sites. A) Structure of cross-link indicating distance measured in X-ray structures of G4s. B) Parallel telomeric G4 structure, 2M18, indicating distances between potential modification sites. C) Antiparallel telomeric G4 structure, 2MBJ, indicating distances between potential modification sites. D) Expected parallel telomeric G4 topology. E) Expected antiparallel telomeric G4 topology, showing relative positions of potential and optimal cross-links.

carbon (Figure 5.5A, n = 4 and n = 6, respectively) linkers gave lengths of 9.7 Å and 12.2 Å, respectively.

For both parallel and antiparallel G4s we found that the A8 nucleotide (TAG3T2AG3T) was a preferable location for adenosine modification compared to the A2 position (TAG3T2AG3T), due to the orientation of the N6 position and its proximity to the 2′-position of relevant guanosine residues. Based on the parallel TERRA G4 structure (PDB: 2M18)[122] (Figure 5.5B) the orientation of the

N6 position of A8 relative to the 2′-OH of guanosine was optimal for a G3—A8 cross-link (red in Figure 5.5D, Table 5.1 (sequences tel-G3A8 and tel-G3A8-X)), with a distance of approximately 5.2 Å (red in Figure 5.5B). The G4—A8 cross-link also appears possible at 5.8 Å. While the G5—A8 distance of 10.4 Å is theoretically possible with a longer linker, the relative orientation of the 2′-OH of dG and N6 of adenosine means a cross-link would have to pass through the G4 structure and disrupt G4 formation.

Table 5.1. Cross-link modified G4-forming sequences derived from the human telomeric repeat.

Name	Sequence	T_m (Na^+, °C)	T_m (K^+, °C)
tel	$TAG_3T_2AG_3T$	---	57
tel-G3A8	TA**Y**G_2T_2**Z**G_3T	---	58
tel-G3A8-X	TA**Y**G_2T_2**Z**G_3T	49*	42
tel-G5A8	TAG_2**Y**T_2**Z**G_3T	---	50
tel-G5A8-X	TAG_2**Y**T_2**Z**G_3T	---	---

**Structure is only partially folded initially, so this Tm is inaccurate. However, this structure is more folded in Na+ buffer than any other sequence tested. Y = 2′-O-propargylguanosine, Z = N^6-ethylazide-2′-deoxyadenosine. Sequence names containing -X indicates that Cu(I)-catalysed azide-alkyne cycloaddition has been performed.*

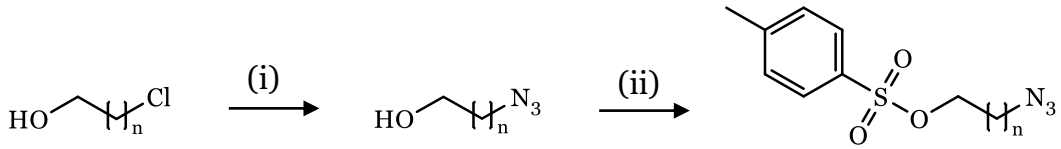
Conversely, the antiparallel G4 topology (PDB: 2MBJ)[123] (Figure 5.5C) suggests that the G5—A8 cross-link is optimal (yellow in Figure 5.5E, Table 5.1 (sequences tel-G5A8 and tel-G5A8-X)), with a distance of 6.1 Å. The G4—A8 cross-link is also possible. It shows a distance of 10.0 Å but passes partially through the G4 structure. The G3—A8 cross-link is also possible with a longer linker, with a distance of 14.4 Å, but the orientations of the groups would be problematic, as with the G5—A8 cross-link for the parallel topology. Sufficiently long and flexible linkers could allow cross-links to be formed between modification sites which are incorrectly orientated, but these could also allow formation of other topologies. We avoid using the G4—A8 modification because it potentially allows for either topology to form. Instead, the G3—A8 and G5—A8 cross-links were selected (Table 5.1) with sequences names including -X indicating that cross-linking has been performed. We proposed that the G3–A8 cross-link will reinforce the parallel topology, while the G5–A8 cross-link will result in a normally unfavourable antiparallel G4.

Both cross-link locations are theoretically close enough for the two-carbon linker to be viable. However, most G4-forming sequence contain few adenosine nucleotides meaning that careful selection of adenosine modification sites would

not be possible for many sequences. Therefore, in addition to the synthesis of Dr Kurup's original adenosine modification with 2-chloroethanol, modifications were attempted with 4-chlorobutanol and 6-chlorohexanol (i.e., the products shown in Figure 5.4B). Longer linkers would provide additional flexibility, but this is not necessarily preferable as it could counteract the topological restriction resulting from the use of shorter linkers. Availability of longer linkers would, however, simplify the inclusion of chemical cross-links in other G4-forming sequences.

5.3 Synthesis of Modified Phosphoramidites and H-Phosphonates

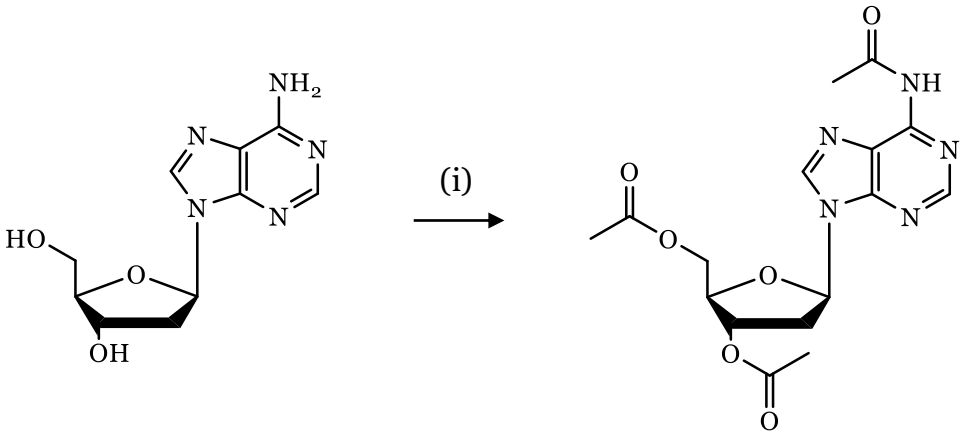
The three modified phosphoramidites were synthesised starting from 2-chloro-1-ethanol, 4-chloro-1-butanol and 6-chloro-1-hexanol. An S_N2 substitution of the chloride with an azide was carried out with NaN_3 in water (Scheme 5.2 (i)). This reaction was reasonably slow and required refluxing for 3 days. The characteristic stretch of an organic azide was observed at approximately 2100 cm^{-1} in IR spectroscopy and the appearance of this peak was used to indicate formation of the product (Figure SI-29). Completion of the reaction was determined by the disappearance of peaks belonging to the starting material in ^1H NMR. The products were characterised primarily ^1H NMR spectroscopy. This showed a change in the chemical shift of H-2/4/6 from around 3.5 ppm to around 3.25 ppm. The resulting materials were reacted with toluenesulfonyl chloride in dichloromethane (DCM) in the presence of triethylamine (TEA) (Scheme 5.2 (ii)). Reaction progress was monitored overnight by TLC (10% ethyl acetate/hexane) and more toluenesulfonyl chloride and TEA were added until the reaction was



1a	n = 1	**2a**	n = 1	**3a**	n = 1
b	n = 3	**b**	n = 3	**b**	n = 3
c	n = 5	**c**	n = 5	**c**	n = 5

Scheme 5.2. Synthesis of tosyl azide precursors for incorporation into adenosine. n = 1, 3, 5. (i) NaN_3, reflux, 3 days, solvent: H_2O. (ii) p-Toluenesulfonyl chloride, TEA, 18 hours, r.t., solvent: DCM.

complete. The products were extracted, purified by silica gel column chromatography (0 – 20% EtOAc/hexane) and characterised by ^{1}H and ^{13}C NMR spectroscopy. In addition to the peaks corresponding to the starting material, peaks corresponding to the aromatic toluene ring appear at approximately 7.8 and 7.3 ppm in ^{1}H and 150 and 120 ppm in ^{13}C indicating the reactions were successful. Additionally, a peak corresponding to the CH_3 group is visible at approximately 2.5 ppm in ^{1}H NMR and 22 ppm in ^{13}C NMR.

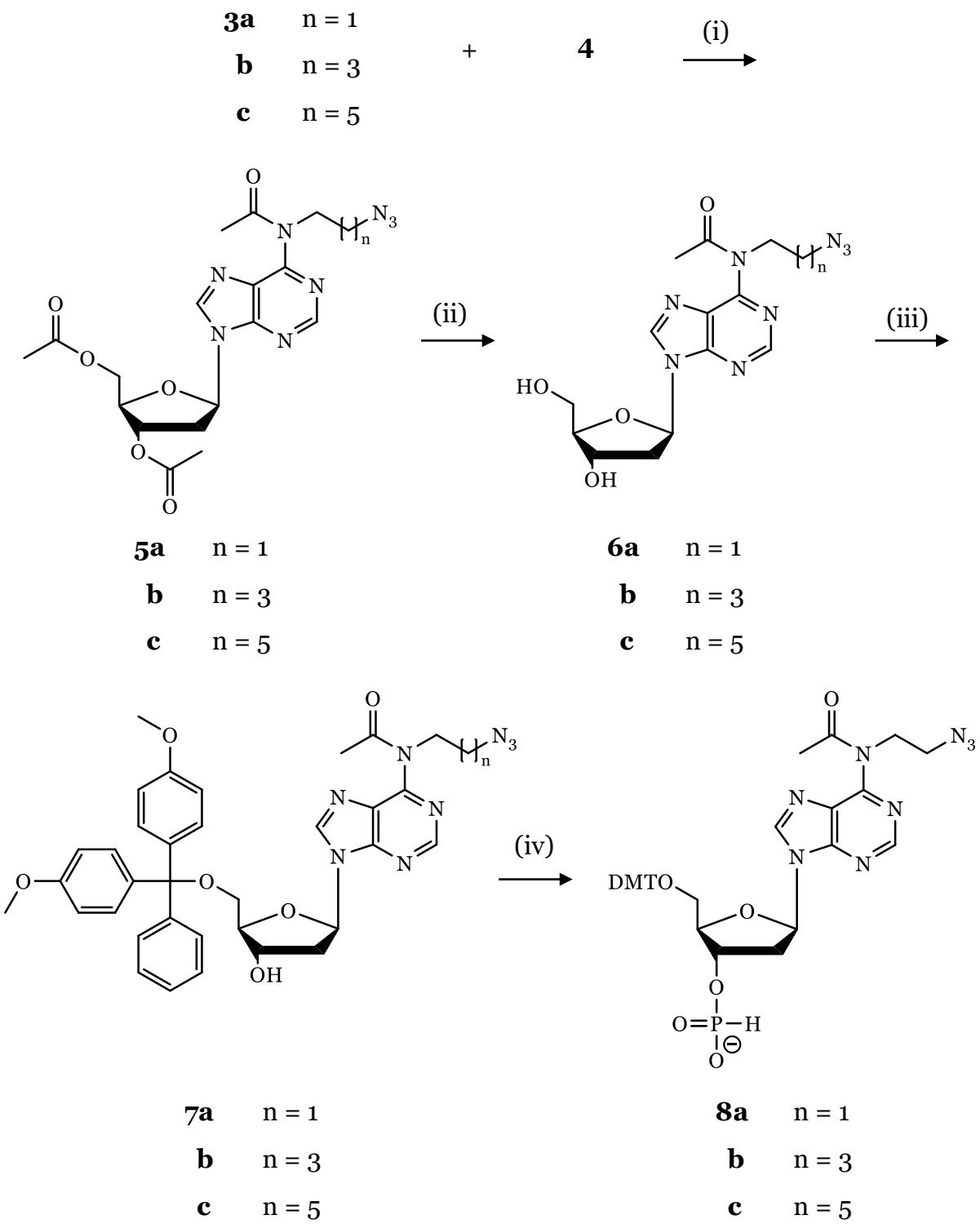


dA **4**

Scheme 5.3. Protection of 2′-deoxyadenosine. n = 1, 3, 5. (i) Acetic anhydride, overnight, r.t., followed by 4 hours, 60 °C, solvent: pyridine.

2′-Deoxyadenosine was protected with acetal groups by a reaction in pyridine with excess of acetic anhydride (Scheme 5.3 (i)). Hydroxyl groups were protected after stirring at room temperature overnight, as indicated by TLC (5% methanol/DCM), but it was necessary to heat the solution to 60°C for 3 – 4 hours to ensure protection of the less reactive aromatic amine. This step was monitored by TLC (5% methanol/DCM) to avoid formation of the tetra-acetylated product, with an additional acetyl group at the N^6-position, as this product could not be alkylated. The product was purified by silica gel column chromatography (0 – 10% MeOH/DCM) and characterised by ^{1}H and ^{13}C NMR, with the appearance of three ^{1}H NMR peaks around 2 ppm indicating the presence of the acetyl groups, in addition to peaks corresponding to the 2′-deoxyadenosine starting material.

Protected adenosine was dissolved in ACN with Cs_2CO_3 and stirred for 15 minutes before one of the previously synthesised tosylated azides was added (Scheme 5.4 (i)). The secondary amide of adenosine is relatively unreactive, so Cs_2CO_3 was added to activate this group prior to addition of tosylated azides. The



Scheme 5.4. Alkylation of N^6-position of adenosine with azide modification, followed by deprotection of 3′- and 5′-OH groups and synthesis of DMT-protected phosphoramidite for DNA synthesis; n = 1, 3, 5. (i) Cs_2CO_3, overnight, 60 °C, in argon atomosphere, solvent: ACN. (ii) TEA, 15 mins, r.t., solvent: 50/50 MeOH/H_2O. (iii) DMT-Cl, overnight, r.t., solvent: dry pyridine. (iv) Diphenyl H-phosphonate, 15 mins, r.t., solvent: pyridine.

solution was stirred overnight at 60°C under argon. When TLC (10% methanol/DCM) indicated the disappearance of starting materials, the reaction was quenched by adding water and the product was extracted and purified by

silica gel column chromatography (0 – 20% MeOH/DCM). 1D NMR was used to confirm addition of the modification, but 2D NMR was also necessary to ensure that the reaction occurred at N^6 rather than N^1. ^1H NMR contains peaks corresponding to both the protected adenosine and alkyl azide, but not the aromatic ring and CH_3 peaks observed after the tosylation. HMBC was used to confirm long range coupling across three bonds between protons on the first carbon of the modification and both the quaternary carbon of the acetyl group and the C^6 position of adenosine rather than the C^2 position (Figure 5.6). Coupling on the nitrogen at the 1-position would result in cross-peaks between H-1'' and both C-6 and C-2. However, cross-peaks were only observed with C-6 and the quaternary carbon of the acetyl protection group, indicating that coupling occurred at the N^6-position, as intended. In previous examples, N^6 alkylation of adenosine had been achieved first by alkylating the N^1 position, followed by Dimroth rearrangement to obtain the N^6-alkylated product.[124, 125] However, as Dr. Kurup reported previously[62, 126] the use of Cs_2CO_3 and toluenesulfonyl alkyl azides resulted in direct alkylation of the N^6 atom of adenosine.

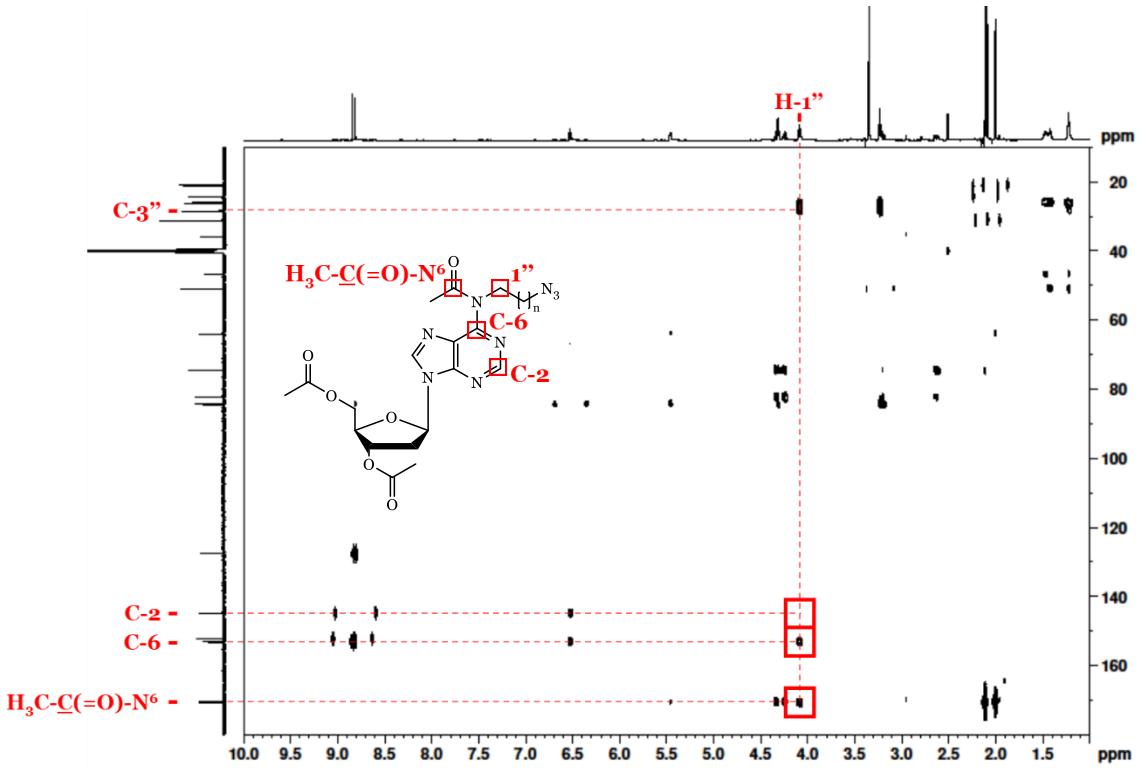


Figure 5.6. HMBC profile of alkylated adenosine, highlighting long-range J-coupling between H-1'' and both quaternary carbon of acetyl protecting group and C-6 of adenosine. Lack of cross-peak with C-2 provides additional evidence that alkylation has occurred at N^6-position rather than N^1-position.

The 3′- and 5′-OH were deprotected by mixing the previous products for approximately 15 mins in a 1:1 water/methanol solution in the presence of excess TEA (Scheme 5.4 (ii)). When disappearance of the starting material was observed by TLC (10% methanol/DCM) the products were extracted, purified by silica gel column chromatography (0 – 20% MeOH/DCM) and characterised by NMR spectroscopy to confirm the disappearance of two acetyl peaks, which had previously been observed at around 2 ppm in 1H NMR spectra.

Finally, the modified nucleosides were prepared for DNA synthesis by protecting the 5′-position with DMT and converting the 3′-OH into an H-phosphonate or phosphoramidite. Due to the hygroscopicity of the deprotected alkylated adenosine starting materials they were prepared for DMT protection by repeated evaporation with pyridine to remove residual water. They were then dissolved in freshly distilled pyridine, cooled to 0°C and 4,4-dimethoxytrityl chloride was added (Scheme 5.4 (iii)). The reaction was monitored by TLC (5% MeOH/DCM with 1% TEA). The product can be observed by the appearance of an additional orange spot when the TLC plate is heated. The reaction mixture was quenched with water when TLC indicated that no starting material remained. These products were purified by silica gel column chromatography (0 – 12% MeOH/DCM) after silica was neutralised with a 1% TEA/DCM solution to prevent the mildly acidic conditions of silica from cleaving DMT. The isolated products were characterised by 1H NMR spectroscopy. In addition to the peaks from the starting materials, the appearance of multiple peaks in the aromatic region (6.5 – 8 ppm) corresponding to trityl protons as well as a singlet at approximately 3.7 ppm corresponding to two dimethoxy groups indicates that the reactions were successful.

Initially the intention was to synthesise phosphoramidites, which are the more common reactants used in DNA synthesis. However, the phosphoramidite material synthesised by Dr. Kurup was unable to be concentrated or stored for any significant period due to the possibility of a Staudinger reaction occurring between the azide and the phosphoramidite. Longer carbon chains at the N6-position give increased flexibility to the chain attached to the azide, increasing the risk of product degradation. H-Phosphonates are an alternative method of DNA synthesis. They are less used because they may result in slightly lower synthetic

yields, but organic azides do not react with H-phosphonates allowing products to be concentrated and stored long term. Additionally, they are compatible with phosphoramidite synthesis meaning that individual bases can be added using nucleoside H-phosphonates while the remainder of the sequence uses nucleoside phosphoramidites. A protocol is described in Chapter 6 for the synthesis of a phosphoramidite with a 6-carbon alkyl azide modification. This test reaction was carried out to synthesise and characterise a small amount of material. This suggests that phosphoramidites of these materials are obtainable, but the quantity produced was insufficient for DNA synthesis. Following the synthesis of this material we chose to synthesise H-phosphonates instead as they are a more convenient method of introducing these types of modifications into DNA.

H-Phosphonates were produced by dissolving the DMT protected materials in pyridine and adding diphenyl H-phosphonate (Scheme 5.4 (iv)). The reaction proceeded quickly and after 15 minutes TLC indicated that no starting material remained. The reaction was quenched by adding water, followed by TEA. Like the previous reaction, the product was purified using a silica gel column neutralised with a 1% TEA/DCM solution (0 – 20% MeOH/DCM) and characterised by NMR. ^{31}P NMR clearly indicates an H-phosphonate with a peak around 0 ppm. Minimal change is observed in ^{1}H and ^{13}C NMR spectra from the previous products and the TEA salt dominates both the ^{13}C and ^{1}H spectra. Mass spectrometry confirmed the formation of the product, with the H-phosphonate appearing in the positive mode with a mass of 727.09 gmol^{-1}, and the TEA salt appearing in the negative mode with a mass of approximately 101. A small amount of non-acetyl protected material also appears with a mass of 685.09.

This method was used to successfully and reliably produce adenosine modified with a two-carbon alkyl azide at the N^{6} position. However, synthesis of four- and six-carbon modifications was less reliable. Adenosine containing the four-carbon alkyl azide was obtained with DMT protecting groups, but this material degraded significantly before H-phosphonate synthesis could be performed. Subsequent synthesis resulted in low yields for alkylation reactions and continued degradation of the DMT-protected material. A small quantity of adenosine containing the six-carbon alkyl azide modification was obtained and characterised. However, when this reaction was scaled up the TLC indicated

formation of the product but following purification on a silica gel column neither the product nor the starting materials were present in any of the collected fractions. In both cases the reactions were attempted several times with varying purification conditions, but sufficient quantities of product were not obtained for DNA synthesis.

Increased flexibility of the longer alkyl modifications may result in an increased risk of internal reactions and degradation. H-phosphonates were chosen to avoid reaction between the phosphoramidite and the azide, which we expected to be problematic for longer linkers. However, this does not appear to have been the problem. NMR spectra of the fractions collected during attempted purifications are also unclear. In most fractions, peaks corresponding to ribose sugars were not present. The most likely possibility, I would suggest, is a reaction between the organic azide and an aldehyde in the open-chain form of ribose, such as an intramolecular Schmidt reaction.[127] This would account for the degradation of the sugar, and is more likely to be possible both at higher concentrations and with longer linkers. This can potentially be addressed with more time spent investigating the products of these reactions and adjusting purification conditions accordingly. However, as discussed previously, the shorter linker was sufficient for testing on telomeric G4s while longer linkers would be necessary for other sequences. An adenosine H-phosphonate and phosphoramidite containing a 6-carbon linker was obtained in a small-scale, suggesting that this was possible, but, at this stage, obtaining sufficient quantities for DNA synthesis was not.

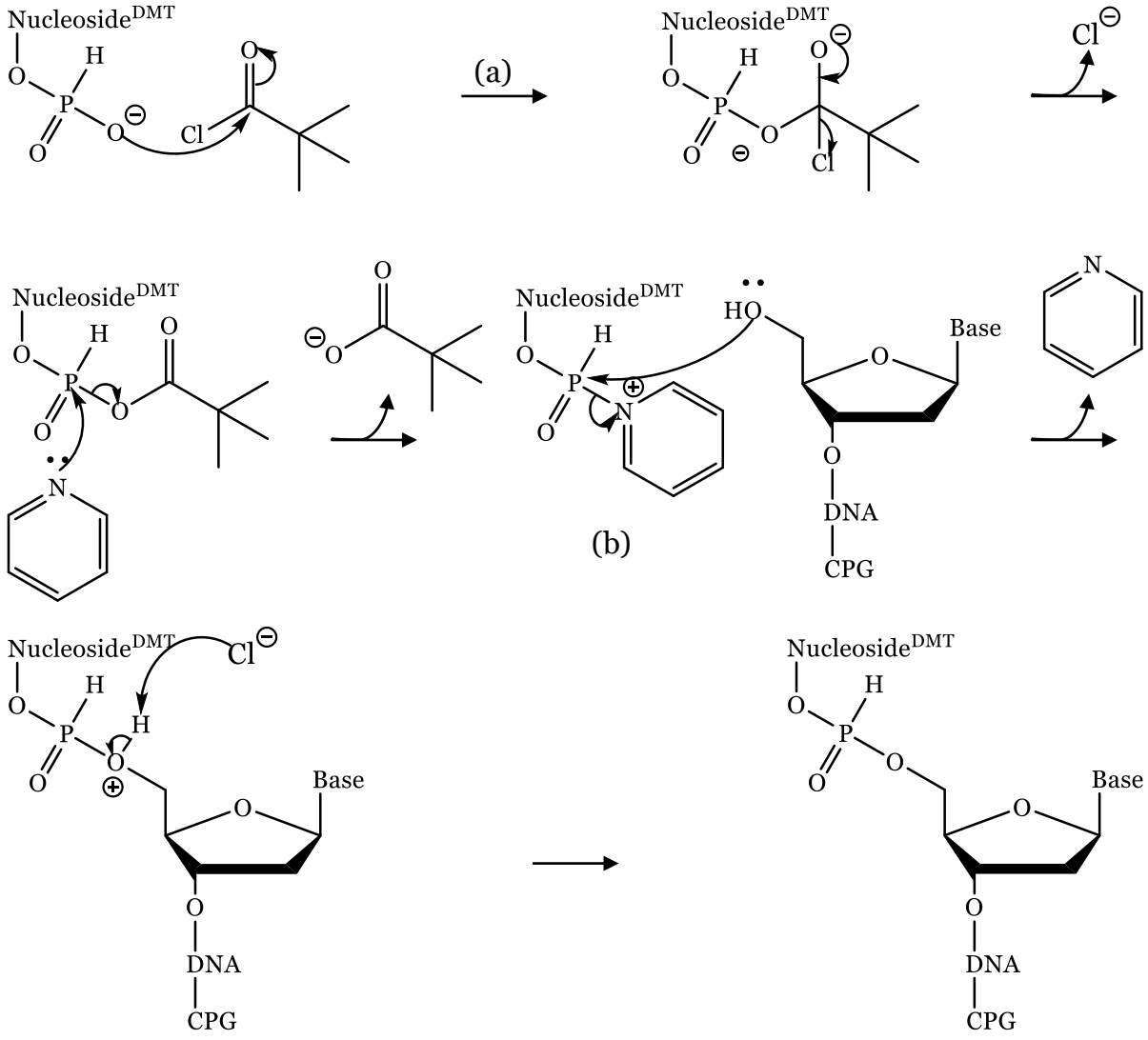
5.4. Cross-Linking of Modified G4-Forming Oligonucleotides

The DNA chain was assembled using standard phosphoramidite chemistry for unmodified nucleotides and 2′-O-propargylguanosine (see chapter 7.2.1 and 7.4.4). The coupling of nucleoside H-phosphonate was carried out by adding a 0.1 M solution of the H-phosphonate in a 30/70 mixture of dry pyridine and ACN to the column containing the DNA on a CPG-support in combination with a 0.5 M solution of pivaloyl chloride in dry ACN. All solvents were dried because the reaction's yield is affected by water similarly to phosphoramidite coupling. Additionally, yield of 2′-*O*-propargyl guanosine coupling was improved by performing three coupling steps, instead of two, increasing concentration of the phosphoramidite solution from 0.1 M to 0.2 M, and increasing coupling time

from two minutes to ten minutes. Once the synthesis was completed the product was deprotected and cleaved from the CPG with 28% aqueous ammonia solution for 12 hours at room temperature. This solution was filtered, freeze-dried, resuspended in Milli-Q water, and purified by reverse phase HPLC. The collected fractions were characterised by mass spectrometry to identify the product.

The intermediates formed during H-phosphonate coupling are not fully characterised, but the overall mechanism is reasonably straight-forward (Scheme 5.5). Initially, an S_N2 substitution reaction occurs when the O^- of the H-phosphonate attacks the δ+ carbon of the carbonyl in pivaloyl chloride. It is possible that an intermediate is formed by pyridine first attacking the carbonyl, but this step occurred too quickly for any other intermediate to be observed.[128] The nitrogen of pyridine acts as a nucleophile, attacking the phosphorus, with the pivaloyl anion as the leaving group. Several potential pyridinium adducts were proposed by Sigurdsson and Strömberg,[128] but Powles et al.,[129] who described H-phosphonate coupling in the context of oligonucleotides, used the intermediate shown in Scheme 5.5. They potentially identified this intermediate by NMR but were unable to confirm it with mass spectrometry. Finally, the free 5′-OH of the nucleotide attached to the CPG attacks the phosphorus with pyridine as the leaving group. The H-phosphonate is converted to a phosphate by a standard iodine-pyridine oxidation, although the duration of the oxidation step immediately after H-phosphonate coupling was doubled to 160 instead of 80 seconds to ensure the H-phosphonate was completely oxidised.

After synthesis, cleavage and purification, the oligonucleotide was dissolved in a volume of water determined by its OD_{260}, as shown in the table in the experimental section. 2 M Triethylammonium acetate buffer (pH 7.0) and freshly prepared aq. sodium ascorbate solution were added to the oligonucleotide dissolved in a known volume of Milli-Q water (see Chapter 7.4.4. and Table 7.6). This mixture was flushed with argon for several minutes before a premixed copper sulphate and Tris((1-hydroxy-propyl-1H-1,2,3-triazol-4-yl)methyl)amine (THPTA) solution was added, and the reaction mixture was sealed. This reaction was monitored using RP-HPLC, comparing the retention times of peaks to the previous analytical RP- HPLC (see Chapter 7.2.2. for RP-HPLC conditions). After approximately four hours, minimal starting material was observed (e.g. r_t = 15.38



Scheme 5.5. Suggested mechanism for H-phosphonate coupling during DNA synthesis. (a) complex with pivaloyl chloride formed too quickly for any other possible intermediates to be observed. (b) A range of possible pyridine adducts have been proposed. This appeared to be the most likely according to literature.

min, Figure 5.7A, Figure SI-30A) and a new peak was observed at a shorter retention time (e.g. r_t = 14.69 min, Figure 5.7B, Figure SI-30B). This peak was isolated using RP-HPLC and characterised by mass spectrometry, giving the same mass as the starting oligonucleotide, which is expected for the click-product.

Success of cross-link formation was determined primarily by RP-HPLC because no change in mass was observed, as expected (see Chapter 7.2.2). Previously reported click-chemistry in DNA showed a decrease in retention time of approximately one minute following successful coupling.[130] A similar shift in retention time was observed for both tel-G3A8 and tel-G5A8 after two hours of

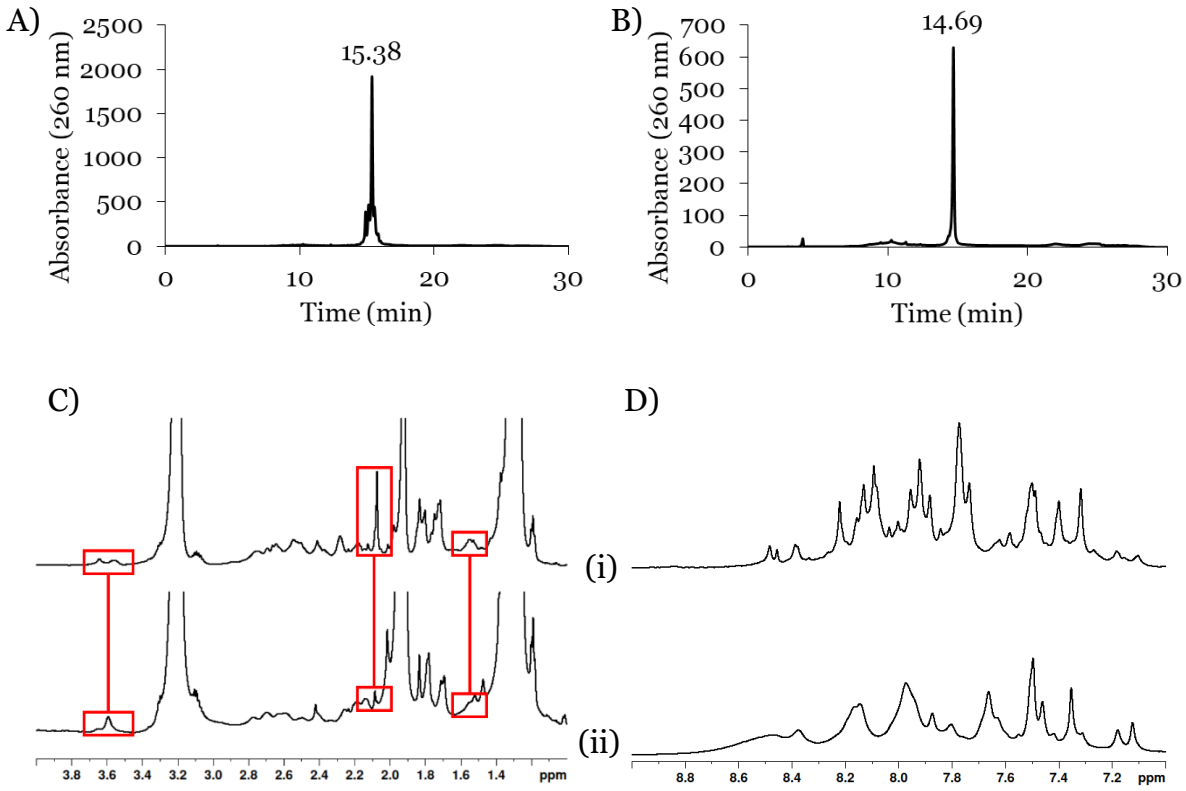


Figure 5.7. Comparisons of pre- and post-cross-linking tel-G3A8 sequences using RP-HPLC and ^{1}H NMR spectroscopy. A) RP-HPLC of tel-G3A8 and B) RP-HPLC profile of tel-G3A8-X, showing an approximately one-minute decrease in retention time. C) ^{1}H NMR of (i) tel-G3A8 and (ii) tel-G3A8-X. Highlighted regions, from left to right, indicate shift in protons corresponding to CH_2 group adjacent to azide in adenosine modification, disappearance of propargyl proton and slight shift in H-2′ of propargyl guanosine. D) ^{1}H NMR in the aromatic region of (i) tel-G3A8 and (ii) tel-G3A8-X. Triazole peak could be expected to be visible in this region, but this peak is obscured by the dominance of aromatic protons in this region (e.g., H-8 of guanosine or H-8 and H-2 of adenosine). However, significant changes in the chemical shift of several peaks in this region are observed.

coupling. ^{1}H NMR profiles comparing G4s before and after cross-linking also changed significantly but identifying specific expected changes in NMR or IR (e.g., the disappearance of an azide stretch) was difficult because of the multiple overlapping peaks and relatively low concentration of samples. However, several key features in ^{1}H NMR can be identified providing additional evidence for the formation of a cross-link. Specifically, a change in proton peaks around 3.6 ppm was observed which suggests a change in the shifts of CH_2 peaks in the carbon chain of the azide containing modification (Figure 5.7C, Figure SI-30C). This is expected due to the change from an azide to a triazole. We also observed the disappearance of a singlet around 2.1 ppm, potentially corresponding to the proton on the terminal carbon of the propargyl modification (Figure 5.7C). The reported chemical shift of this proton varies, but it typically is in the range 2.1 – 2.7 ppm and has high intensity relative to other peaks in nucleosides.[131] This region has a high density of peaks due to the overlapping protons of ribose sugars

and the methyl groups of thymidine, but the propargyl peak is particularly pronounced in the non-cross-linked sample and absent in the product. It is also possible to see a change in the 2′-proton on the same carbon as the propargyl/triazole group, which is expected to appear around 1.6 ppm (Figure 5.7C). Theoretically, triazole peaks should be visible around 7.8 ppm, but the large number of protons from the nucleobases, particularly the 8-position of multiple dG residues make it difficult to confirm (Figure 5.7D). The combination of mass spectrometry indicating no change in mass, a shift in RP-HPLC retention time and the corresponding changes to key peaks in ^{1}H NMR spectra support the conclusion that a triazole cross-link has been formed for both tel-G3A8 and tel-G5A8.

5.5. Investigation of Properties of Cross-Linked G4s

5.5.1. Formation and Topology of G4s Containing Cross-Links

The sequences were folded in a 20 mM sodium phosphate buffer and analysed. Later, 10 mM KCl was added for K^+-containing samples to observe differences between complexes formed in Na^+ and K^+. Cross-linked samples were compared both with an unmodified tel control and a modified, but not cross-linked control to ensure that any changes in properties were not purely the result of the addition of modified nucleotides.

The tel control remains unfolded in Na^+ buffer, shown by both CD and NMR spectroscopy (Figure 5.8A and C), but CD indicates formation of a parallel structure in K^+ buffer (Figure 5.8B). However, this is not clear in NMR as most peaks in the imino proton range were indistinguishable from the noise, although some low intensity peaks appeared between 13 and 14 ppm (Figure 5.8C). This suggests that unmodified tel is partially folded in K^+ buffer, and likely contains a mixture of primarily parallel topologies. tel-G3A8 was also mostly unfolded in Na^+ buffer. A slight shift in the λ_{max} towards 265 nm suggests formation of a parallel G4 structure. The ^{1}H NMR spectra showed broad peaks in the imino proton region (Figure 5.8C) in Na^+ buffer , but the low intensity of the CD peak (Figure 5.8A) and the lack of definition in ^{1}H NMR peaks suggests only partial formation and a mixture of topologies present.

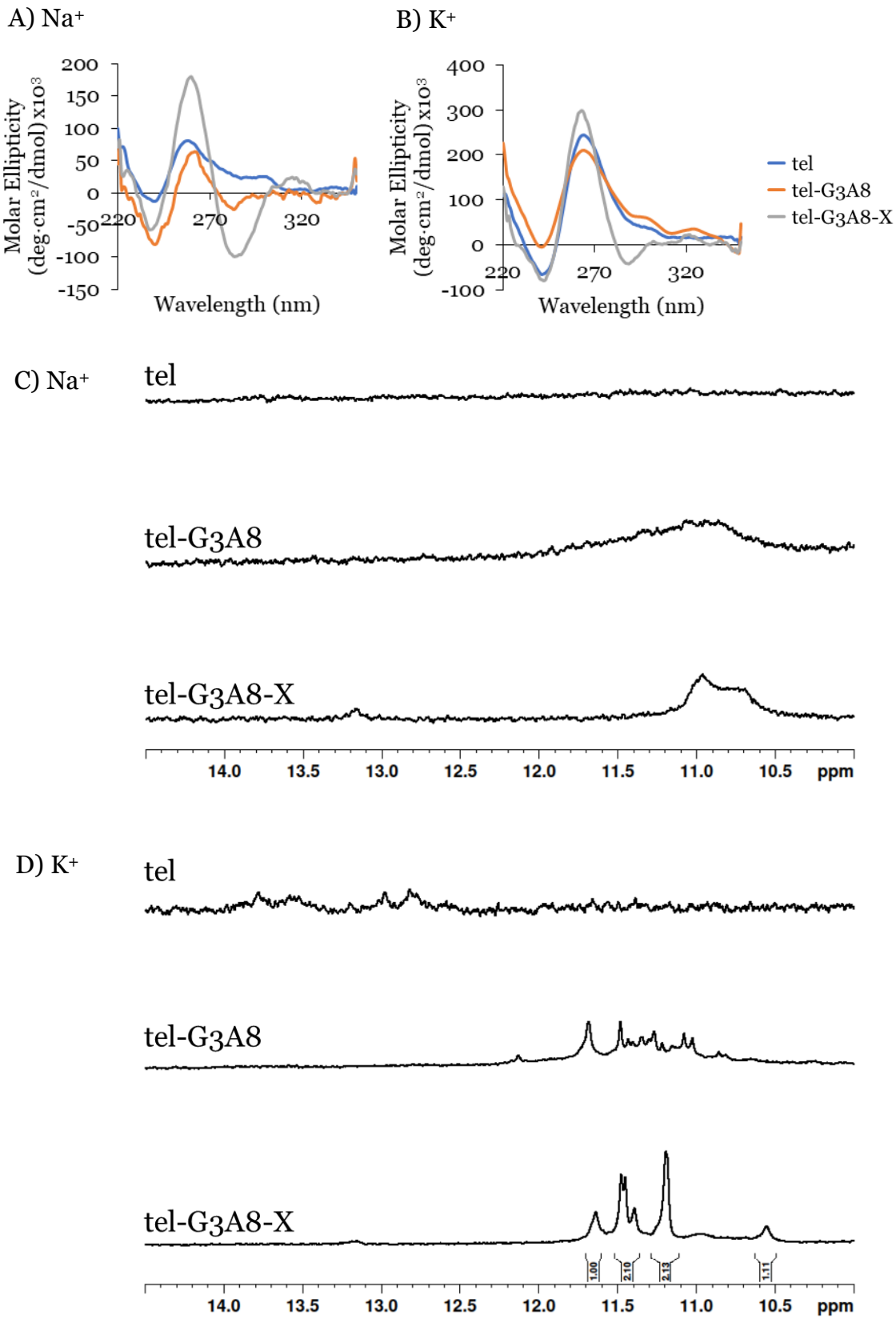


Figure 5.8. Analysis of topological changes induced by G3—A8 cross-links in a G4-forming sequence based on the human telomeric repeat. A) CD profile in Na^+ buffer. B) CD profile in K^+ buffer. C) NMR spectra in Na^+ buffer. D) NMR spectra in K^+ buffer. Conditions: 20 μM (CD) or 200 μM (NMR) strand concentration, 20 mM sodium phosphate, 10 mM KCl (K^+ samples only), 10% D_2O, 1% TSP, pH 7.0.

In K^+ buffer a distinctly parallel topology was observed, and more clearly defined peaks were visible in the imino proton region of the ^1H NMR spectra (Figure 5.8D). The presence of a shoulder at approximately 290 nm in CD and the number of peaks visible in NMR suggests that multiple G4 topologies are present in the tel-G3A8 sample. As expected, tel-G3A8-X forms a parallel G4 in both Na^+- and K^+-containing buffers according to CD spectra (Figure 5.8A and B). In Na^+ buffer the peak at 260 nm was more clearly defined and had higher intensity than either of the controls. ^1H NMR also showed undefined peaks similar to tel-G3A8 in Na^+ buffer. In K^+ buffer the intensity of CD peaks corresponding to the parallel G4 topology was comparable to the controls in CD, but the shoulder at around 290 nm is not present. ^1H NMR peaks were also more clearly defined (Figure 5.8C and D). Integration of imino proton peaks indicated that six imino protons of guanosine were involved in hydrogen-bonding (twelve if we assume diagonally opposed nucleotides are chemically equivalent). This suggests that a single parallel topology is formed rather than a mixture of topologies present in the non-cross-linked sample. As anticipated, the introduction of the cross-link encouraged the formation of a parallel G4. Increased intensity of peaks corresponding to parallel G4 formation in the CD spectrum indicates that the cross-linked sequence was significantly more structured in Na^+ buffer, compared to non-crosslinked sequences, suggesting that the presence of the cross-link encouraged G4 formation independent of the cations used, although these signals increased in intensity considerably in K^+ buffer. Both tel and tel-G3A8 had minimal G4 formation in Na^+ buffer, meaning the increase in G4 formation is attributable to the cross-link.

Results obtained for tel-G5A8-X were drastically different than for tel-G3A8-X. In Na^+-containing buffers, tel-G5A8 and tel-G5A8-X are both unfolded like the tel control (Figure 5.9A and C). The cross-link between G5 and A8 did not appear to encourage formation of the expected antiparallel G4 in Na^+ and K^+-containing buffers. The only noticeable difference is the decrease in intensity of the peak at approximately 300 nm after the cross-link was introduced. In K^+ buffer tel-G5A8 forms a G4 with a similar topology to the unmodified sequence (Figure 5.9B). However, the CD profile of tel-G5A8-X indicates that it remains unfolded in K^+ buffer, rather than forming an antiparallel G4 (Figure 5.9B). These results were

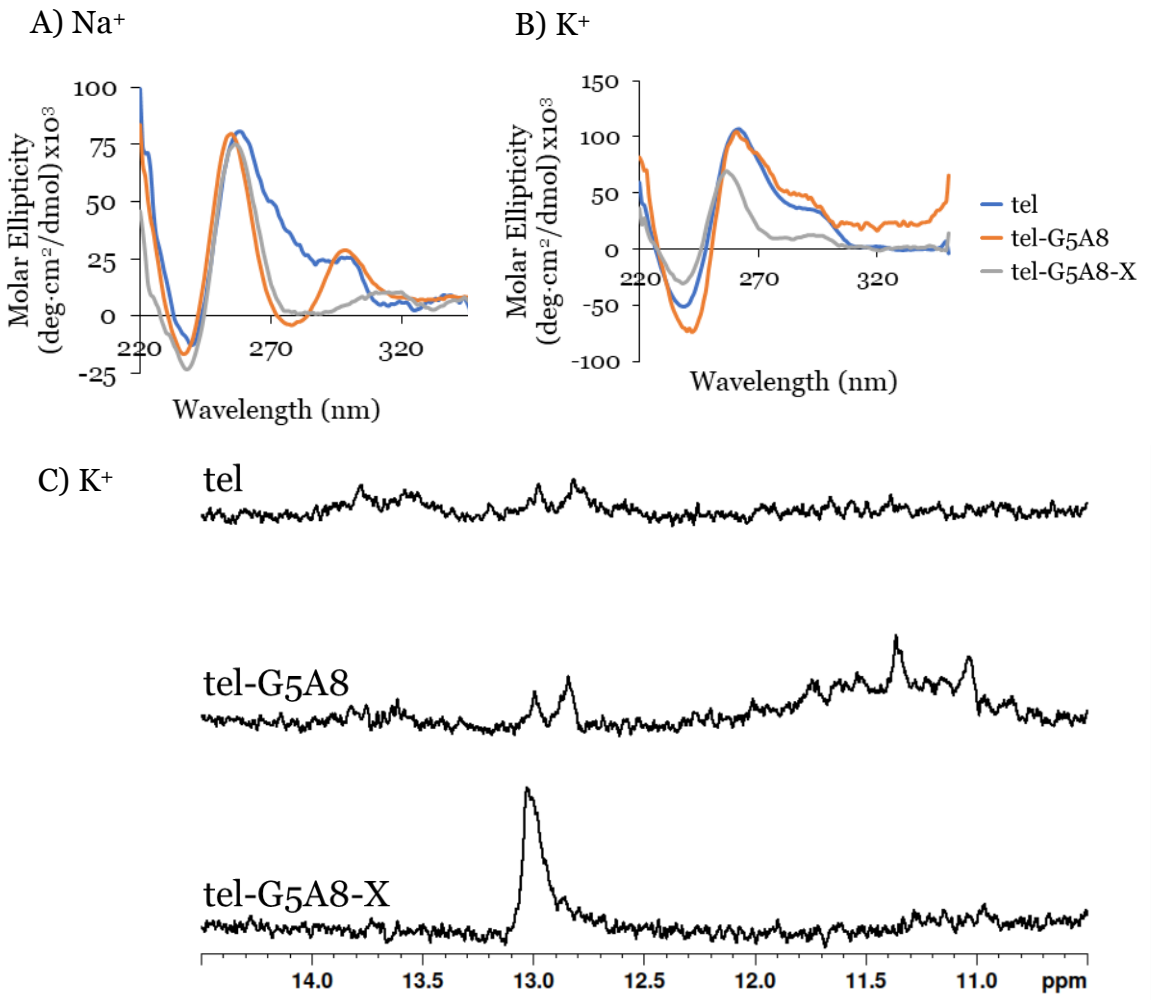


Figure 5.9. Analysis of topological changes induced by G5—A8 cross-links in a G4-forming sequence based on the human telomeric repeat. A) CD profile in Na^+ buffer. B) CD profile in K^+ buffer. C) NMR spectra in Na^+ buffer. D) NMR spectra in K^+ buffer. Conditions: 20 μM (CD) or 200 μM (NMR) strand concentration, 20 mM sodium phosphate, 10 mM KCl (K^+ samples only), 10% D_2O, 1% TSP, pH 7.0.

supported by ^1H NMR, as none of the sequences had significant imino proton peaks in Na^+ buffer, and both tel and tel-G5A8 had similar, low intensity peaks at approximately 11.5 ppm, indicating some G4 formation has occurred (Figure 5.9C and D). This suggests that rather than reinforcing an unfavourable antiparallel topology, cross-links between G5 and A8 only disrupted the normally favoured parallel G4. The presence of a significant peak at around 13 ppm (Figure 5.9D) could indicate formation of self-dimerisation through canonical A-T base pairing or potentially non-canonical G-T base pairing as a result of guanosines no longer being involved in G4 formation.

These results suggest some limitations of cross-linking modifications for encouraging the formation of unfavourable secondary structures. It is possible

that the length of the linker did not allow sufficient flexibility for antiparallel G4 formation, but longer linkers would also be more likely to allow formation of a parallel topology. Alternatively, the position of the cross-link may have forced bases into suboptimal orientations for G4 formation. This orientation of the modified sites was considered when analysing X-ray structures, but it is difficult to determine from these structures how cross-links will be positioned in practice. For future applications of this strategy, it would be desirable to test a range of linker lengths and positions. This obviously requires overcoming the synthetic challenges experienced while producing longer linkers. Additionally, while the G3–A8 cross-link matched our expectations for secondary structure formation based on existing X-ray crystallography and NMR structural data, the G5–A8 cross-link did not. This makes it clear that considering and testing a range of modification sites will be necessary when incorporating modifications into other sequences. Based on BLItz data obtained using modified tel, c-KIT and c-MYC sequences in Chapter 4, HP1α showed little or no affinity for tel sequences. Therefore, it will be necessary to expand this strategy to other sequences before testing with proteins can commence. However, the intended level of topological control was possible using cross-links, but only with careful selection of modification sites.

5.5.2. Thermal Stability of G4s Containing Cross-links

Assessing the impact of cross-linking on G4 stability, determined using CD melting experiments, is difficult. Only the cross-linked structure was formed in Na^+ buffer, but the low intensity of the peaks made it difficult to obtain a precise T_m value. tel-G3A8-X is only partially folded so the T_m value which can be calculated from the melting profile is inaccurate (Table 5.1, Figure SI-31). However, no structure was formed for unmodified sequences, suggesting that the cross-linked structure was more stable than the native structure. In K^+ buffer, the tel control gave a T_m of 57 °C (Table 5.1, Figure SI-32A). Both oligonucleotides with G3 and A8 modifications were folded into G4s in K^+ buffer giving a T_m of 42°C for tel-G3A8-X compared to a T_m of 58°C for tel-G3A8 (Table 5.1, Figure SI-32B and C). While the samples differ slightly in topology, the non-cross-linked structure appears to be more thermally stable than the structure supported by the cross-link. This suggests that the structure formed by the non-cross-linked

sample is more thermodynamically stable, and the cross-link hinders formation of the most thermodynamically stable structure. This is supported by the similarity in T_m of tel-G3A8 and the unmodified sequence. Potentially the reduction in polymorphism imposed by the cross-link results in less favourable hydrogen-bonding arrangements within the G-tetrads. However, while the G4 is less thermally stable, the cross-link promoted G4 formation, even in Na^+ buffer, meaning it may also disrupt duplex formation. Both tel-G5A8 and tel-G5A8-X were unfolded in Na^+ buffer, and only tel-G5A8 folded in K^+. The T_m of this sequence is only 50 °C, slightly lower than both tel and tel-G3A8. This suggests that the G5A8 modification was less effective at stabilising the G4 than G3A8 but did not completely disrupt G4 formation like the cross-linked counterpart.

5.5.3. Competition of Cross-Linked G4s with Canonical Duplexes

Duplex formation of these sequences was studied by challenging each G4 with complementary DNA and observing the changes in secondary structure from the imino proton region of the ^1H NMR spectra. The unmodified control forms a duplex immediately upon addition of the complementary strand and within one day is almost completely folded into a duplex (Figure 5.10A). tel-G3A8 also folded into a duplex but was only partially folded after three days and required heating before the duplex became the dominant structure (Figure 5.10B). This can be compared to experiments discussed in Chapter 1.3.1 and 4.1 with modifications such as 2′-fluoroarabinonucleic acids in G4s. Oligonucleotides containing non-native nucleotides were reported to form kinetically trapped G4s which remained stable for days or weeks.[64, 65] However, upon heating they rapidly unfolded and formed the canonical duplex upon cooling. Similar to 2′-fluoroarabinonucleic acids, 2′-O-propargylguanosine replaces the proton normally as the 2′-position in unmodified DNA with a larger functional group, and in both cases, we see duplex formation disrupted kinetically, but not thermodynamically. The G4 formed by the cross-linked tel-G3A8-X was still present after the sample was heated with no observation of ^1H NMR peaks corresponding to duplex formation (Figure 5.10C). Comparing the observations of tel-G3A8 to tel-G3A8-X indicated that these thermodynamically trapped G4s do not result solely from the incorporation of modified nucleotides. Duplex formation was only completely disrupted once the triazole cross-link was formed. Comparing this result to CD

melting experiments is interesting because this suggests that an increase in thermal stability is not necessary to disrupt duplex formation. By preventing the G4 structure from unfolding through the introduction of a chemical cross-link we were able to reinforce the G4 structure. This did not necessarily improve the G4 stability, but it did prevent the formation of the duplex.

The tel-G5A8 sequence behaved similarly to the tel control. Upon addition of complementary DNA, we observed the immediate appearance of peaks corresponding to the duplex (Figure 5.11). However, similar to tel-G3A8 the structure was not completely converted to duplex until the sample was heated and cooled to obtain the thermodynamic product. This suggests that this modification also results in kinetically trapped G4s, independently of the modification sites. However, as mentioned previously, tel-G5A8-X did not form a G-quadruplex structure according to NMR or CD (Figure 5.9). The initial ^{1}H NMR spectra after addition of the complementary strand suggests that some G4 formation might have occurred more slowly, but this is not a significant amount of G4 folding. With the complementary sequence present, we did not observe formation of a duplex either, indicating that this modification was sufficient to disrupt duplex formation in addition to G4 formation (Figure 5.11). Overall, this supports our hypothesis that G4s containing this type of cross-link will disrupt duplex formation, but it does suggest that selection of appropriate modification sites is crucial for success.

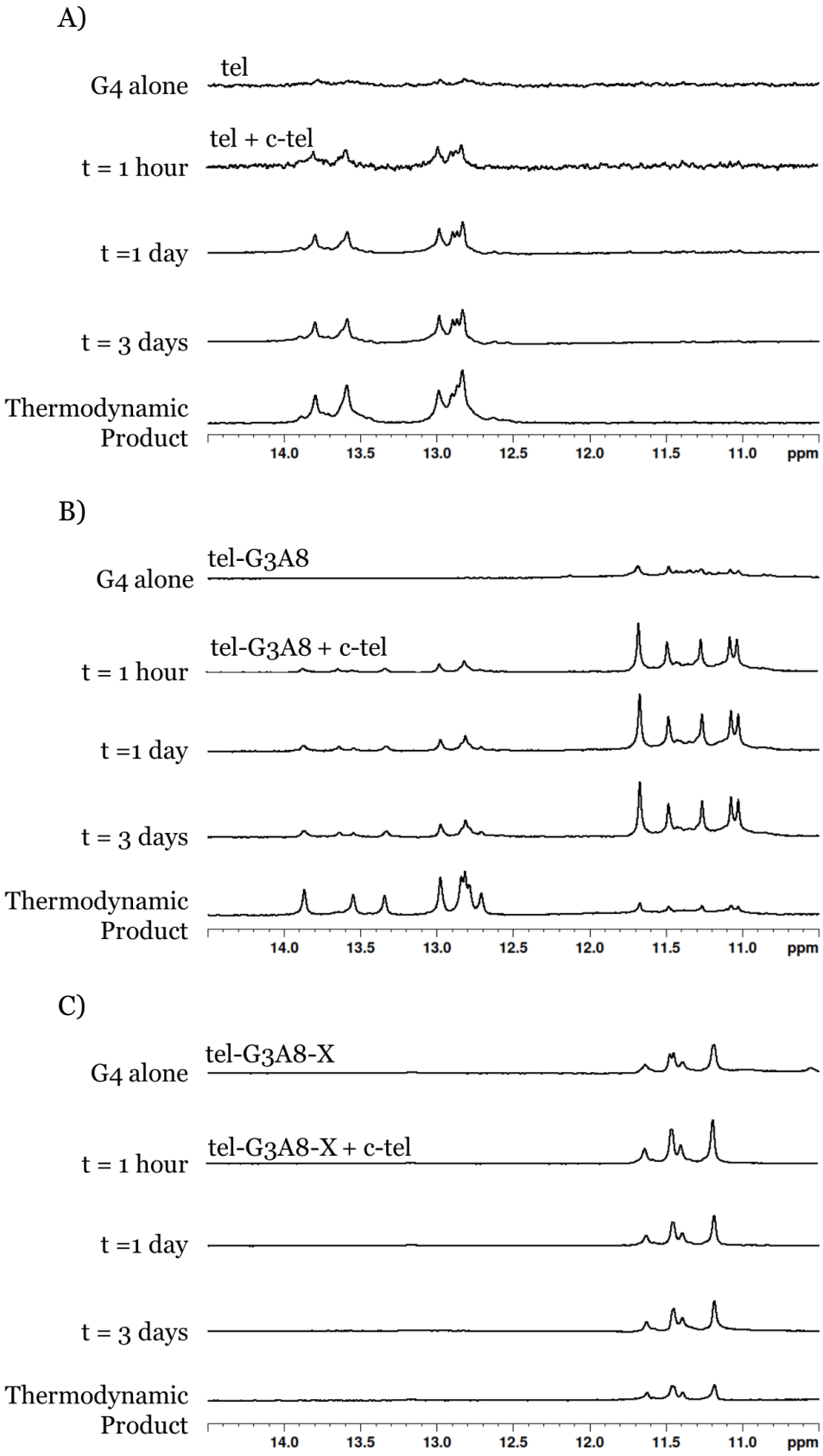


Figure 5.10. ^{1}H NMR of telomeric G4 sequences containing G3 to A8 CuAAC cross-links challenged with complementary DNA. A) Unmodified tel sequence + c-tel. B) tel-G3A8 + c-tel. C) tel-G3A8-X + c-tel. Conditions: 200 μM strand concentration, 20 mM sodium phosphate, 10 mM KCl, 10% D_2O, 1% TSP, pH 7.0.

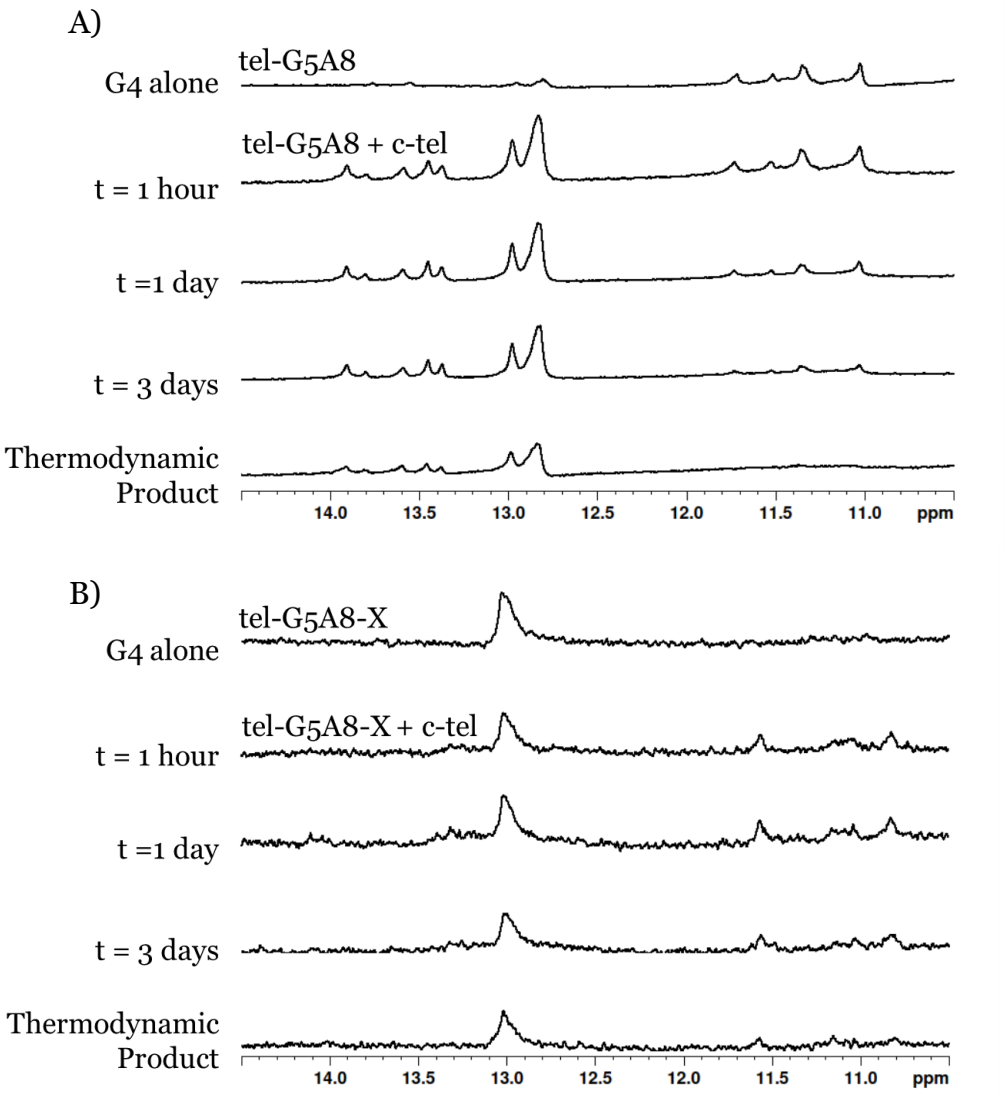


Figure 5.11. ^{1}H NMR of telomeric G4 sequences containing G5 to A8 CuAAC cross-links challenged with complementary DNA. A) tel-G5A8 + c-tel. B) tel-G5A8-X + c-tel. Conditions: 200 µM strand concentration, 20 mM sodium phosphate, 10 mM KCl, 10% D_2O, 1% TSP, pH 7.0.

5.5.4. Preliminary Testing of HP1α Binding to Cross-Linked G4s

The binding of tel sequences was tested with the his_6-tagged HP1α hinge discussed in Chapter 4.2.4. (Figure SI-20). BLI experiments (see Chapter 7.2.8.) were used to determine if these sequences bound to the hinge. HP1α showed minimal affinity for the tel control, relative to TERRA45 and no affinity for modified sequences (Figure 5.12A). The hinge showed comparable binding to tel, tel-G3A8, tel-G5A8 and the TERRA45 control. However, TERRA45 bound more quickly than any of the tel sequences. The hinge showed no affinity for either tel-G3A8-X or tel-G5A8-X. This is not surprising for tel-G5A8-X, which does not form a G4. However, tel-G3A8 formed a thermodynamically stable G4.

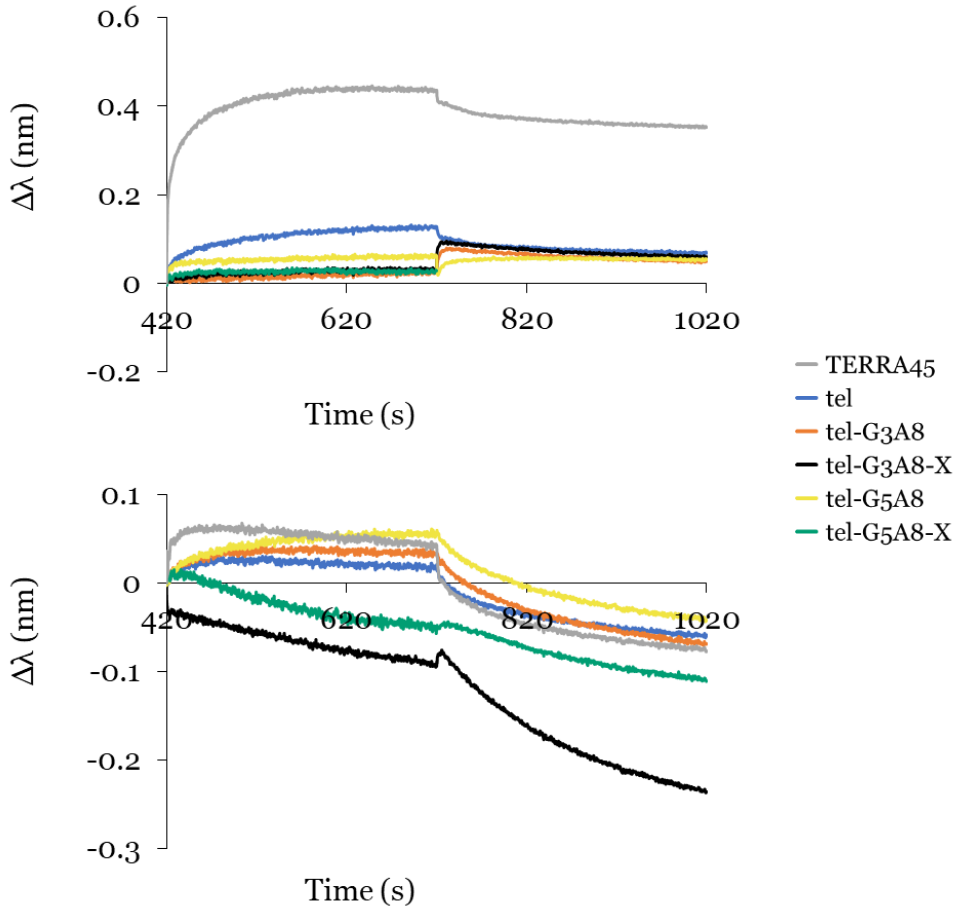


Figure 5.12. BLItz profile, showing binding of A) immobilised wt-HP1α and B) immobilised HP1α-hinge to tel sequences. Conditions: 100 μg/mL protein concentration, 2 μM strand concentration, 20 mM sodium phosphate, 100 mM KCl, 50 mM NaCl, pH 8.0.

The affinity shown for tel-G3A8 and tel-G5A8 suggests that the presence of the 2′-*O*-propargylguanosine and N^6-[2-azidoethyl]-2′-deoxyadenosine alone is not sufficient to disrupt protein binding. HP1α affinity only decreased when the cross-link was introduced. This result may indicate that the presence of the cross-

link in the groove of the G4 blocks binding of the protein. Where possible, it would be optimal to test G4s with cross-links outside of the grooves to avoid this issue. However, this could be challenging for parallel G4s, which typically only contain propeller loops, meaning they don't contain useful modification sites outside of the grooves. This could also be less of an issue for unimolecular G4s. In these cases, cross-links may only block some groove binding sites, which may decrease HP1α binding. Additionally, this may have less of an effect on the affinity of other proteins that do not bind in the groove. Furthermore, these proteins may bind to antiparallel G4s, where cross-links between lateral loops could be possible.

5.6. Conclusions and Perspectives

Several challenges were encountered synthesising nucleoside H-phosphonates containing longer alkyl azide linkers for cross-linking, but G4 sequences containing 2-carbon adenosine linkers were obtained using a previously described method.[62, 126] Azide-modified adenosine nucleotides were successfully cross-linked with 2'-*O*-propagyl guanosine nucleotides incorporated elsewhere in the sequence using copper(I)-catalysed azide-alkyne cycloaddition and produced two cross-linked sequences. Without cross-linking, both sequences containing G3 – A8 and G5 – A8 modifications produced G4s with similar structures and thermal stability to unmodified sequences, as evidenced by CD and ^{1}H NMR spectroscopy. When challenged with complementary DNA, the conversion to the canonical duplex was slowed, but not completely disrupted and the duplex was the thermodynamic product. After cross-linking, the tel-G3A8-X sequence was shown to form a G4 of parallel topology. This sequence appeared to form this G4 topology even in Na^+ buffer, but in K^+ it had decreased thermal stability compared to the unmodified and non-cross-linked examples. Despite this, the cross-linked G4 was thermodynamically stable and did not form a duplex in the presence of the complementary strand. Conversely, the tel-G5A8-X sequence, intended to create an antiparallel G4, did not form a G4 in either buffer, instead a ^{1}H NMR spectroscopy suggested formation of a hairpin. When challenged with complementary DNA, it also did not form a duplex.

Overall, these results suggest that cross-linking is an effective method of reinforcing an existing topology, in this case the parallel topology, but it was less effective at supporting an unfavourable topology. This could be more effective for

a sequence which is known to form both topologies more readily. Ultimately this makes this strategy ideal for our intended application, studying the interaction of HP1α with parallel G4s, since it did stabilise the naturally occurring G4 topologies, but it also illustrates a potential limitation which must be considered if this strategy is extended to drug design or the study of other proteins or DNA secondary structures.

Based on the results of these experiments, G4s with cross-links similar to G3—A8 would be an ideal candidate for further testing with HP1α. However, HP1α showed no affinity for the tel-G3A8-X sequence, unlike tel and tel-G3A8. This suggests that these cross-links may block the approach of HP1α to the G4 groove, preventing binding. This could be circumvented or minimised by selecting cross-linking sites in other G4s to avoid blocking potential binding sites and therefore disrupting protein interactions. This would be the next step in transferring this strategy to nucleosomal arrays. Both c-KIT and c-MYC contain adenosine nucleotides in the loop. Closer investigation of their crystal structures would determine if similar cross-links to the tel-G3A8 modification are possible. The results of the telomeric modification suggest that this approach would produce thermodynamically stable G4s capable of interacting with HP1α. An alternative would be to investigate a longer telomeric G4. TERRA45 is an RNA G4-forming sequence which has shown significant affinity to HP1α. The lack of affinity for the tel sequences is potentially a result of both the length of the sequences and the structure being bimolecular. As such, a longer, unimolecular G4 might produce more promising results. Additionally, this modification could be incorporated into longer DNA duplexes by DNA ligation, or sequences with duplex-forming tails similar to those produced in Chapter 4. These could be used to test interactions with other proteins. Ultimately, the best approach would appear to be incorporating similar cross-links into other G4-forming sequences. This could even potentially be applied to other secondary structures such as i-motifs, although this would potentially require use of other modified nucleotides.

Overall, modified nucleotides enforced G4 formation and disrupted duplex formation, based on the results obtained in both Chapters 4 and 5. In both cases, it was necessary to carefully select modification strategies to ensure the correct G4 topology was obtained, but when this was done effectively these strategies

produced G4s which closely mimic the unmodified sequences. Either strategy presents a potentially unique method of obtaining thermodynamically stable G4s in larger DNA structures. Going forward, either modification strategy presents a useful method of obtaining G4s with minimal change in topology. The next challenge is incorporating inverted-, α- and cross-linking modifications into oligonucleotides of sufficient length for DNA ligation. Previous discussion has indicated that this will require 40-70 nucleotides sequences, depending on the length of the G-rich strand, to allow for flexible regions between G4s and duplexes, duplex-forming regions on either side of the G4, and the overhangs required for ligation. Obtaining long DNA duplexes containing either cross-linked or inverted-/α-nucleotides will still require considerable optimisation of synthesis for longer oligonucleotides.

6. Conclusions and Final Perspectives

6.1. Investigation of Strategies from Literature

This research primarily focuses on understanding the role G4s play in the function of the cell, and in particular the specific interactions of G4s with proteins, such as HP1α. For this purpose, we intend to develop models containing mixtures of duplex- and G4-DNA which can be used to construct naturally occurring, but transient, structures which occur in cellular DNA, such as nucleosomes. Using current methods, it is difficult to ensure G4s are formed due to the preference of complementary DNA forming canonical duplexes instead. We require sequences in which the G4s are thermodynamically stable even in the presence of the complementary strand. Ideally, they would not affect any interactions with proteins. Stable G4s can then be incorporated into larger DNA structures and used to study DNA complexes with proteins. We expect that some modifications may also be applicable to other DNA secondary structures. This could allow for studies of similar specific interactions between i-motifs or triplexes and proteins.

Additionally, these modifications have potential as targeted drugs. G4s are often enriched in the promoter region of oncogenes[28-33] and therefore could be useful targets for anticancer drugs.[34] Elevated levels of G4s have also been observed in some types of cancers.[39] Proteins targeting G4s may bind preferentially to the thermodynamically stable G4s produced using our methods compared to G4s in cells which are in equilibrium with duplexes. In addition to HP1α (see Chapter 1.2), enzymes which have shown specific interactions with G4s include a helicase, Pif1, which unwinds G4 structures[37] and a DNA repair enzyme, DNA methyltransferase 3A.[35] The presence of Pif1 specifically has been linked to increased genomic stability.[132] It is currently unclear if any of the modifications described in this thesis have therapeutic applications, but allowing targeting of proteins involved in maintaining genomic integrity (including HP1α) suggests that stable G4 constructs could be a useful method of destabilising cancer cell genomes to a level that would trigger apoptosis. It is potentially worth testing modified sequences described in this sequence as inhibitors of G4-specific enzymes.

A number of existing methods have already been used to stabilise G4s, but they were determined to be unsuitable for our primary intended application. In many cases,

particularly small-molecule G4 ligands, these methods were designed with the intended therapeutic application of disrupting the function of proteins which process duplex DNA by stabilising G-quadruplexes.[61] In other cases, they were able to create kinetically trapped, but not thermodynamically stable, G4s through minor modifications of nucleotides.[64, 65] Other strategies were also developed more recently for G4 stabilisation. Monsen *et al.*[133] created a thermodynamically trapped G4 by replacing the c-rich region on the complementary strand with a polyT DNA sequence. This resulted in a G4 embedded within a duplex, which was characterised by electron microscopy, small-angle X-ray scattering and molecular dynamics. This is similar to the strategy discussed in Chapter 4. However, the complementary C-rich strand is often capable of forming i-motif structures and proteins may interact differently with G4s when C-rich DNA or i-motifs are nearby. For this reason, the strategies described in this thesis, which stabilise G4s even when unmodified complementary sequences are present, provide a significant improvement over the use of polyT sequences for duplex disruption. The concurrent development of stable G4s using this strategy, among others, does indicate that there remains interest in effective methods of creating thermodynamically stable G4s.

We investigated the effectiveness of small-molecule ligands using both the PDS ligand and several new, peptide-based ligands provided by Dr. Masayuki Fujii of Kindai University (Osaka, Japan). While we did find that these ligands typically induce an increase in thermal stability, they also caused a significant change in the imino proton chemical shifts of G4s in ^{1}H NMR (Figure 6.1A). PDS was also shown to partially disrupting duplex formation. Peptide-based ligands resulted in increased hydrophobicity and precipitation of G4s, most likely indicating that a complex was formed, but making further testing impractical.

In the context that these ligands are often developed (i.e., disrupting the function of proteins which process only duplex DNA), a change in G4 topology or increased aggregation is not necessarily problematic. However, when studying the types of interactions between proteins and G4s that occur in cells we require G4s that closely mimic the native G4 topologies. Additionally, while peaks corresponding to G4 formation were still present in ^{1}H NMR after the thermodynamic product was formed, we still observed some duplex formation, making this strategy less reliable for long-term stability of G4s. Finally, the binding of ligands to G4s has the potential to block

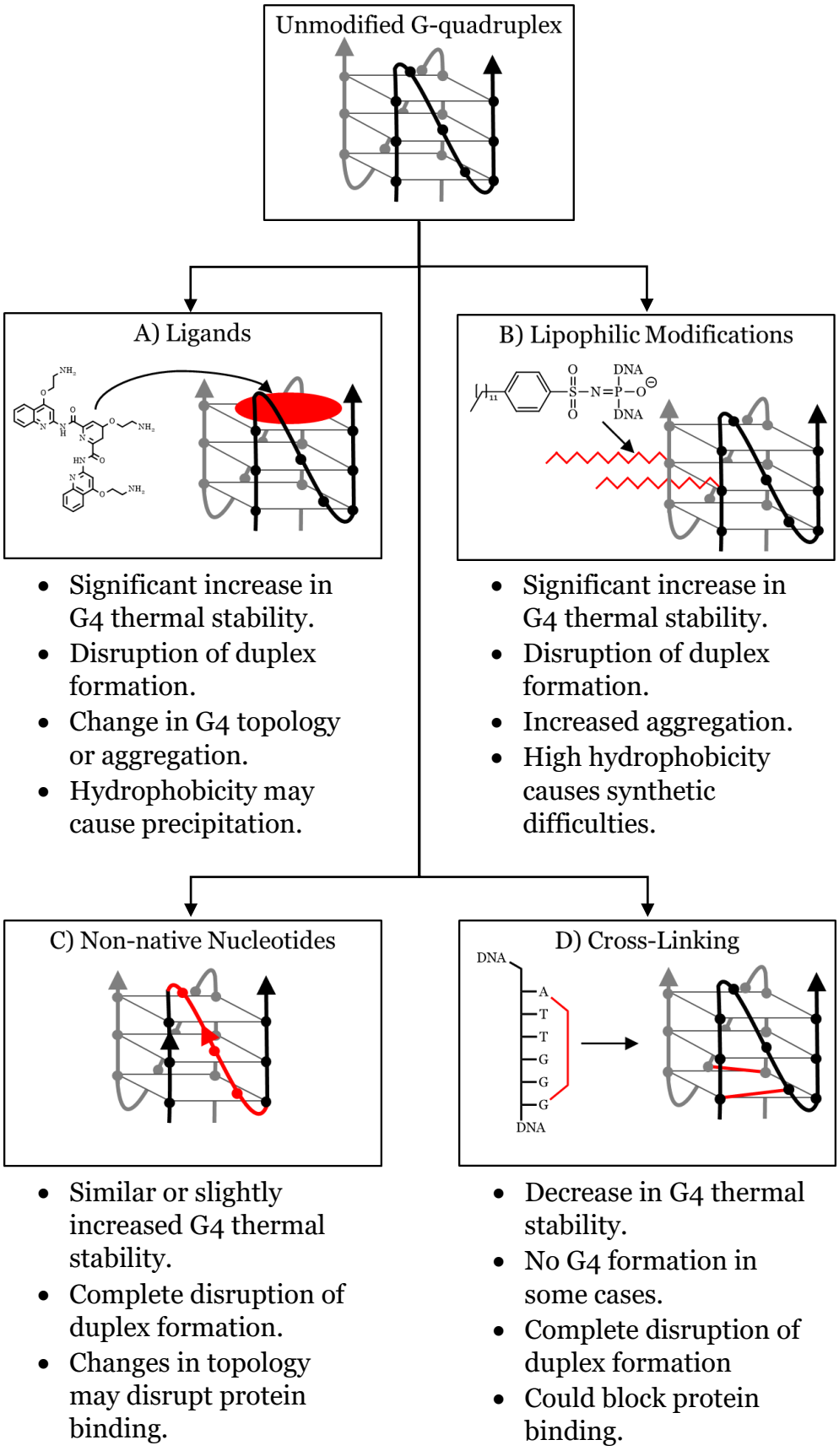
protein-binding entirely, meaning their application is dependent on the binding modes of the ligand and the protein. For this reason, the alternative strategies developed within this thesis are recommended alternative for studying G4-protein complexes.

6.2. G4s Containing Hydrophobic Modifications

Introduction of the lipophilic 4-dodecylbenzenesulfonyl imine phosphorus modification instead of a native phosphate, discussed in Chapter 3, showed significant increases in G4 thermal stability, comparable to small-molecule ligands, while also disrupting duplex formation (Figure 6.1B). The increased aggregation of strands due to the hydrophobic effect encouraged G4 formation. It is unclear if some changes observed in CD spectra of modified sequences are a result of the increased aggregation of G4s or changes in G4 topology, but the low mobility of modified sequences in native PAGE experiments suggested modified G4 structures may aggregate more than unmodified DNA. This was also supported by the initial issues encountered with G4 formation. To completely break down lipophilic aggregates heating at 90 °C for 1 hour was required, rather than approx. 5 mins for unmodified G4s.

However, the biggest issue encountered with this type of modification was during purification. While attempting synthesis of $GXG_3T_3G_3XG$ a significant amount of DNA containing only a single phosphate modification was obtained. Normally this could be removed during HPLC purification, but the hydrophobicity of these sequences meant they had similar retention times in both reverse-phase and ion-exchange HPLC. While this modification provided promising results using the tetra- and bi-molecular sequences we tested, single modifications in these sequences corresponds to two or four modifications in the resulting G4.

Increased G4 stability and disruption of duplexes in the presence of complementary strands suggests that these modifications could provide useful models of G4-protein complexes, provided the hydrophobic modifications do not disrupt protein interactions. However, testing interactions with proteins would require unimolecular G4s, such as Ttel, c-KIT, c-MYC or Pu39, which could be challenging to obtain. It may be possible to stabilise G4s with only two modifications, similar to $G_4T_4G_4$ modified sequences. At this stage of our investigation, other strategies presented more effective methods of building models of G4s in cells, and they therefore received greater focus.



- Significant increase in G4 thermal stability.
- Disruption of duplex formation.
- Change in G4 topology or aggregation.
- Hydrophobicity may cause precipitation.

- Significant increase in G4 thermal stability.
- Disruption of duplex formation.
- Increased aggregation.
- High hydrophobicity causes synthetic difficulties.

- Similar or slightly increased G4 thermal stability.
- Complete disruption of duplex formation.
- Changes in topology may disrupt protein binding.

- Decrease in G4 thermal stability.
- No G4 formation in some cases.
- Complete disruption of duplex formation
- Could block protein binding.

Figure 6.1. Summary of results from various strategies used to create thermodynamically trapped G4s.

Hydrophobic modifications may be worth reconsidering as an alternative to cross-linking. The tel sequence could receive modifications at similar positions (e.g., $TAXG_3TTXAG_3T$, would be similar to tel-G3A8), which would not introduce a cross-link, but could still encourage G4 formation. The tel-G5A8 cross-link completely disrupted both G4 and duplex formation, most likely due to cross-links restricting the movement of strands, preventing optimal hydrogen-bonding. Hydrophobic moieties could still encourage G4 formation, but with less restriction of strands. This type of sequence may only require two modifications, which was previously manageable.

Alternatively, short sequences (e.g., existing TG_4T or $G_4T_4G_4$ sequences) containing hydrophobic modifications could be tested as enzyme inhibitors, with aggregation of hydrophobic chains driving continued formation of stable G4 substrates. Enzymes such as Pif1 may be unable to effectively unfold G4s stabilised with lipophilic chains. This may be the preferable application of hydrophobic modifications based on the effectiveness of the other strategies we have explored.

I would suggest several changes to this modification strategy moving forward. Recent testing of phosphate modification with p-toluene-sulfonyl azide has shown greater coupling efficiency at 37 °C with concentrations of 0.5 M or even 1 M. Increasing concentration or temperature may allow for fewer partially modified sequences to be produced, simplifying purification and isolation of final products. Longer HPLC programmes with different conditions might be more successful in separating hydrophobic modifications. The modified sequences typically had retention times corresponding to high ACN concentration (i.e., 80 %), so better separation might be possible with a programme containing, for example, an additional step with a 70 – 80% ACN gradient. Fine-tuning of this programme could allow for more effective separation of G4s with multiple hydrophobic modifications, allowing for purification of unimolecular sequences like Ttel, c-MYC, c-KIT and Pu39 with 2 – 4 modifications. Alternatively, purification could be attempted using preparative dPAGE, which might allow for better separation of sequences with different numbers of hydrophobic modifications than HPLC.

Alternatively, less hydrophobic modification could be tested to achieve a balance between duplex disruption and hydrophobicity. Shorter lipophilic chains could be a valid alternative which encourages G4 formation without changing topology or introducing significant issues with purification. Some sulfonyl azides, such as 1-octanesulfonyl azide (Figure 6.2A), are commercially available, but a range of sulfonyl modifications can also be synthesised using hydrophobic carbon chains containing halides. 1-Bromooctane (Figure 6.2B) or 1-bromo-4-octylbenzene (Figure 6.2C) are both commercially available. They could be converted to a sulfonate though a reaction with sodium sulphite (Figure 6.2D (i))[134], followed by a conversion to a chloride with thionyl chloride (Figure 6.2D (ii))[135] and then an azide with sodium azide (Figure 6.2D (iii)).[135, 136] Many of the sulfonyl azides which could be obtained in this manner are not commercially available, but a wide range of synthetic precursors are. We are unsure what the optimal modification would be for balancing hydrophobicity, so testing multiple options would be ideal. This potentially expands the range of inhibitors which could be developed using hydrophobic modifications and allows for greater control of G4 stabilisation and the corresponding effect on interactions with proteins.

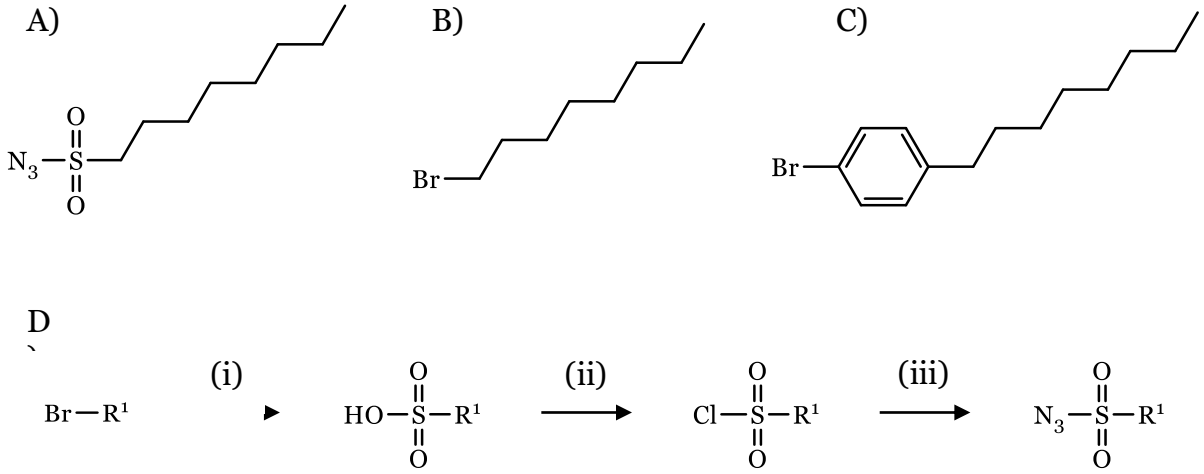


Figure 6.2. Alternative phosphate modification reagents and precursors. A) octylsulfonyl azide. B) 1-bromooctane. C) 1-bromo-4-octylbenzene. D) Reaction pathway converting halide to sulfonyl azide (i) sodium sulfite, (ii) thionyl chloride, (iii) sodium azide, solvent: water.

6.3. G4 Sequences Containing Inverted- or α-Nucleotides

Modifications containing inverted- and α-nucleotides appeared to be more applicable to developing the required type of thermodynamically trapped G4s. Modified sequences generally had higher thermal stability than unmodified G4s and did not form duplexes with complementary strands (Figure 6.1C). However, in preliminary experiments with HP1α, these sequences often showed affinity to the hinge of HP1α, but not for wild type-HP1α. This suggests that while they were effective

at preventing duplex formation, conserving the optimal G-tetrad H-bonding arrangement may result in changes to the loop configuration which are disruptive to the interactions with proteins. We saw evidence in experiments with loop and tetrad only modifications that this problem could be avoided, suggesting that by carefully selecting modification sites we can create synthetic G4 constructs which accurately mimic native G4 behaviour. In particular, HP1α often showed higher affinity for α-dG containing sequences than other modifications, or the unmodified controls. The development of nucleosomal arrays containing inverted or α-nucleotides would be the next step for these modifications.

In terms of drug development, these modifications show some advantages over hydrophobic modifications. Those modifications introduced long, bulky carbon chains with potential to disrupt protein binding. Inverted nucleotides in particular create a change in strand directionality but contain otherwise identical nucleotides to unmodified DNA. α-Nucleotides are similar, but anomers have more potential to interact differently with proteins since many proteins have chiral specificities. Overall, however, these modifications were the most effective at conserving the native G4 topology and therefore may be an optimal strategy for creating useful substrates and inhibitors which accurately represent unmodified DNA.

6.4. Chemical Cross-Linking in G4s

We also described an example of copper(I)-catalysed azide-alkyne cycloaddition used to introduce cross-links to G4s (Figure 6.1D). While this modification completely disrupted duplex formation, it also showed potential to disrupt the G4 if cross-links were not positioned optimally. Cross-links in the G4 groove also appeared to block the binding of the HP1α-hinge. This result provides useful insight into the binding of HP1α to G4s, but also highlighted the need for careful selection of modifications sites. These structural alterations are also limited with respect to encouraging the formation of thermodynamically unfavourable secondary structures. Overall, this means that inverted- and α-nucleotides may be preferable for developing models for HP1α binding. While they also disrupted protein binding in some cases, they allowed for more flexible modification, which could be used to circumvent this problem. Adjusting cross-linking positions is more challenging because the relative orientations of each position must be taken into account. To address some of these concerns it might be beneficial to incorporate other modified nucleotides (e.g., 5-ethynyl-2'-deoxyuridine,

which was developed by Dr. Kurup originally),[62] allowing for cross-links between positions in loops. However, this also requires sequences with ideally positioned nucleotides. Despite this, chemically cross-linked G4 constructs showed a similar ability to disrupt duplex formation, so it is worth exploring additional modifiable sequences to determine if better options can be found.

The cross-linked oligos developed by Dr. Kurup had also been used as inhibitors of the APOBEC3 enzyme.[126] Cross-links created an ideal shape for the protein active site, making binding to the inhibitor preferable, at which point cytosine analogues were used to prevent activity of the enzyme. A similar strategy could be attempted using G4 cross-links to target enzymes such as Pif1. Cross-links could prevent unfolding of the G4 structure and inhibit Pif1 in a similar manner to the analogues used for APOBEC3 inhibitors. As mentioned above, hydrophobic modifications may also be useful in this regard.

6.5. Development of Large DNA Structures Containing G4s

Formation of large DNA/protein complexes will first utilise a technique called DNA ligation. This technique requires phosphate-on sequences containing duplex tails with overhangs. During ligation a ligase enzyme joins two DNA sequences by creating a phosphodiester bond between the 3′-OH of one sequence and the 5′-phosphate of the other. DNA sequences used for ligation require overhangs, or sticky-ends, to allow for base pairing between the two ends intended to be ligated (Figure 6.3A), which controls where ligation occurs and is necessary for DNA ligase to function (Figure 6.3B). Additionally, the 5′-end must already have a phosphate group. DNA is normally synthesised with free OH-groups at both the 3′- and 5′-ends, requiring the use of a commercially available phosphate-on reagent (Figure 6.3C) to introduce a 5′-phosphate. This ultimately results in a G-rich sequence similar to the structure shown in Figure 6.3D. DNA ligation allows for the creation of longer synthetic DNA constructs which would be difficult or impossible to produce using chemical DNA synthesis alone due to their size (i.e., >200 nucleotides in length). DNA ligation is necessary to produce, for example, nucleosomal arrays containing thermodynamically stable G4s to study with HP1α.

The tel sequence used as a model for both of the previously discussed strategies provided useful evidence that within a canonical duplex, the modified G4s remained stable. However, HP1α did not have significant affinity for this structure, so

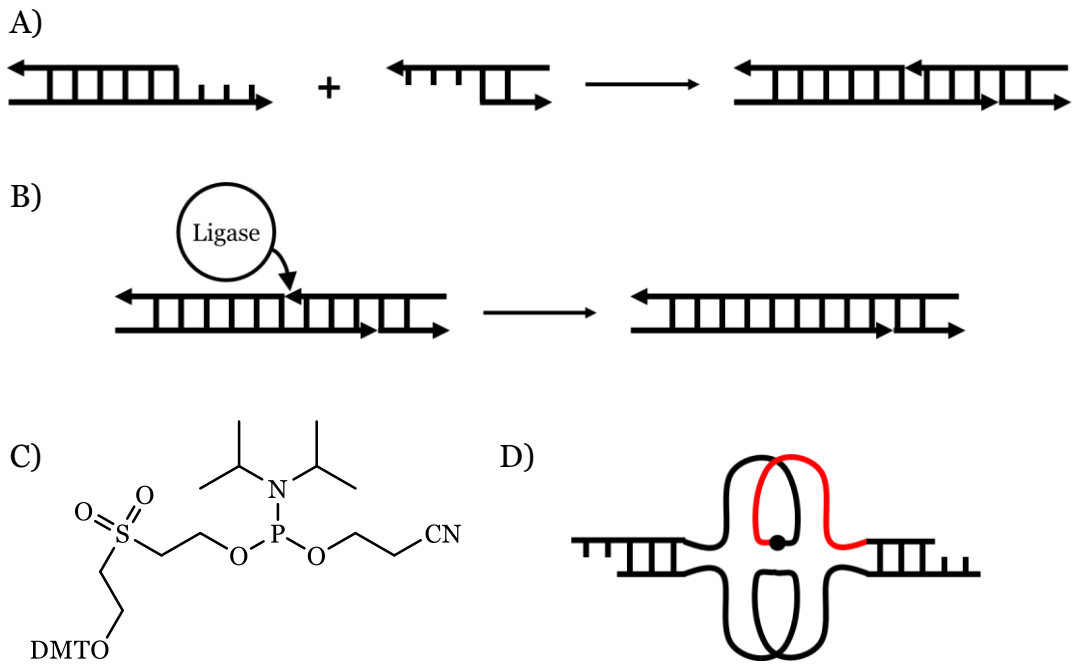


Figure 6.3. A) DNA overhangs or sticky end allow control of ligation sites by forming duplexes between different duplex-forming strands. B) DNA ligase processes duplex DNA to create a phosphodiester bond at the 5′-end between two sequences. This requires a 5′-phosphate on one strand. C) Phosphate-on reagent used to introduce a 5′-phosphate during chemical DNA synthesis. D) Target structure containing thermodynamically trapped G4 with overhangs to allow for DNA ligation.

investigating alternative sequences (e.g., c-MYC, c-KIT, Pu39, kras or src3, as discussed for previous modifications) will improve the applicability of these modifications for experiments testing protein binding. Additionally, obtaining H-phosphonates containing longer azide linkers will allow for a greater range of cross-linking sites, allowing incorporation into more sequences. The increased flexibility introduced by longer linkers could also avoid the decrease in thermal stability observed for the tel-G3A8-X cross-linked sequence.

c-KIT and c-MYC were used with inverted- and α-nucleotides because they had shown good affinity to HP1α and were easily synthesisable, making them useful as a proof of concept. However, HP1α has shown affinity for multiple other sequences. Pu39 is a sequence with considerable interest due to its location in the promoter of the BCL-2 oncogene promoter and would be an interesting target for modification. It was initially discarded because of its length. However, it is similar in length to the 5′inv-Ttel tail sequences, which were synthesised, and with evidence supporting the efficacy of these modifications in disrupting G4 formation, longer sequences are potentially more useful. Pu39 could include similar modifications to c-MYC or c-KIT (i.e., replacing all G-tetrad or loop nucleotides) but Pu39 contains six G-rich regions, and forms multiple G4 topologies.[72,137] Therefore, modifying only some G-rich regions may

encourage formation of some topologies over others, providing interesting insights into the specificity of HP1α. While Ttel, c-MYC, c-KIT and Pu39 are the main sequences studied in this thesis, the original paper demonstrating HP1α affinity for G4s contained several other sequences of interest.[38] Kras ($AG_3CG_2TGTG_3A_2GAG_3A_2GAG_5AG_2$) and src3 ($G_3AG_3AG_3CTG_5$) structures both show significant affinity to HP1α (K_D = 68.8 and 140 nM, respectively) and could be worth investigating. Our investigation was limited to c-MYC and c-KIT for synthetic reasons. In addition to developing precursors for insertion into duplex DNA, expanding these strategies to other G4-forming sequences will provide a wider range of sequences to consider when developing nucleosomal arrays. Similarly, synthesising sequences intended to form structures such as i-motifs or triplexes may provide a basis for large DNA structures intended to study the interactions of proteins with other secondary structures, which can also be formed in chromatin, and could also be worth investigating. A significant number of sequences could be considered for this, depending on the secondary structure chosen to stabilise.

6.6 Conclusion

The results presented in this thesis would suggest that inverted- and α-nucleotides and cross-linked G4s are both potentially useful methods for disrupting duplex formation. While their implementation in protein studies is dependent on further experimentation with sequence selection and modification sites, they can form the basis of future development of nucleosomal arrays and models for eventual studies of interactions with proteins. Additionally, there is the possibility of testing these strategies for developing anticancer drugs. In particular, inverted nucleotides may present a method of creating thermodynamically trapped G4 substrates with minimal change in nucleotide composition. Cross-linked oligonucleotides have already been shown to make effective inhibitors of APOBEC3, and while there is some risk of blocking protein-binding sites they may also prevent unfolding by enzymes such as Pif1, creating effective helicase inhibitors. It is unlikely that one strategy would result in effective stabilisation of G4 structures in all scenarios. The significant variation between both different G4s (including factors such as nucleotide composition, topology, and molecularity) and different proteins interacting with G4s means that different strategies need to be considered in each case. The same is also true for other secondary structures. However, all of strategies developed in this thesis provide new

tools for designing thermodynamically stable models of non-canonical DNA secondary structures.

Hydrophobic phosphate modifications could also present a useful method of creating thermodynamically stable G4s, but at this stage they seemed less effective than other options. This could change with adjustments to the strategy or the use of a different hydrophobic modification. Testing modified TG4T and G4T4G4 sequences for targeting proteins such as HP1α or DNA methyltransferase 3A may be worthwhile, but these proteins have not shown affinity for tetra- or bi-molecular DNA. This means that development of unimolecular G4s with this type of modification may be necessary for further experimentation.

7. Experimental

7.1. Reagents and Instruments

Solvents and reagents were obtained from a range of sources, certified ACS grade. Unmodified DNA sequences were obtained from IDT (USA), standard solid supports were obtained from Dnature Diagnostics and Research Ltd. (NZ), standard phosphoramidites were obtained from Innovassynth Technologies (I) Ltd (India) and modified phosphoramidites were obtained from Chemgenes Corporation (USA) or Glen Research (USA). Chemicals were used as received without further purification. Acetonitrile and pyridine were dried using standard distillation techniques and dried over 3 Å molecular sieves. 'Brine' refers to a saturated solution of NaCl in water.

All reactions were performed in oven-dried glassware. Moisture-sensitive reactions were performed under an inert atmosphere of argon, using syringe septum techniques. Column chromatography was performed using silica gel 60-120 and 100-200 mesh or using a Buchi automated column system (Cartridge PP 40/150, 25/150, 12/150, 12/75).

DNA synthesis was carried out using a Mermade 4 DNA/RNA automated synthesiser. Oligonucleotides were purified using Thermo Scientific UltiMate 3000 UHPLC with an Alltech 250 mm x 4.6 mm, 10 μm Hypersil Gold column (Reverse-Phase) and a TSKgel SuperQ-5PV 7.5 mm I.D. x 7.5 cm, 10 μm column (Ion-Exchange). DNA was desalted using NAP-5 size exclusion Columns. Circular dichroism spectra were recorded using a Chirascan CD spectrophotometer (150 W Xe arc) from Applied Photophysics with a Quantum Northwest TC125 temperature controller. UV-Vis spectrometry was performed with a Cary 100 Bio UV-Vis spectrophotometer. Biolayer interferometry was performed using a BLItz system (Fortebio, USA) and Ni-NTA biosensors (Fortebio).

Analytical thin-layer chromatography (TLC) was performed with MERCK precoated silica gel 60- F254 (0.5-mm) aluminium plates. Spots were visualised using UV or chemicals such as 2 M $KMnO_4$ that was used to identify oxidisable functional groups such as alcohols. TLC plates were submerged in $KMnO_4$ solution, then heated. The product is identified by the appearance of a brown spot, characteristic of MnO_2.

^{1}H and ^{13}C NMR spectra of small molecules were recorded using a Bruker 500 MHz spectrometer using tetramethylsilane (TMS) as an internal standard for ^{1}H and solvent peaks for ^{13}C. ^{1}H and ^{13}C NMR of DNA were recorded using a Bruker 700 MHz spectrometer with trimethylsilylpropanoic acid (TSP, 1%) as an internal standard. Chemical shifts are reported in parts per million (ppm) downfield of TMS or TSP. Spin multiplicities are described a: s (singlet), br.s. (broad singlet), d (doublet), dd (doublet of doublets), t (triplet), q (quartet), m (multiplet). Coupling constants are reported in Hertz (Hz). Mass spectra were recorded with methanol as the mobile phase using electrospray ionisation mass spectrometry with a Thermo Scientific Q Exactive Focus Hybrid Quadrupole-Orbitrap Mass Spectrometer. DNA samples were dissolved in Milli-Q water and 15% methanol was added to improve detection. Other samples were dissolved in methanol. Masses are reported in atomic mass units (a.m.u.).

7.2. General Methods

Multiple techniques were used in the synthesis and study of oligonucleotides. Brief descriptions of the principles involved in their use and their specific application to DNA chemistry are provided in this section.

7.2.1. Automated DNA Synthesis

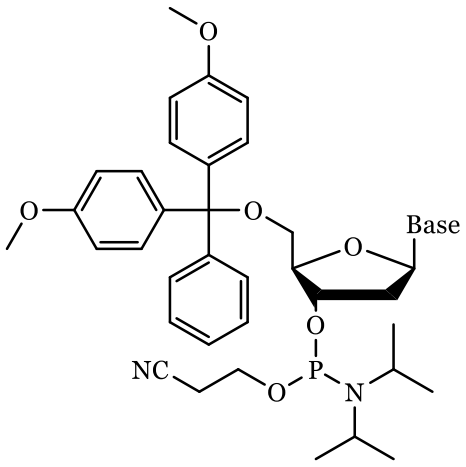


Figure 7.1. 5′-DMT-3′-O-phosphoramidite, the basic reagent in DNA *synthesis.*

DNA synthesis is an automated process composed of four steps, carried out consecutively as each base is added to the sequence. The reaction occurs on controlled pore glass, usually preloaded with the first nucleoside in the sequence. Sequences are conventionally written from the 5′-end to the 3′-end, but DNA synthesis begins from the 3-end. Each base is added using a 5′-DMT-3′-

phosphoramidite (Figure 7.1). The large number of reactions required to complete a sequence means that ideally the yields of each individual step are high (i.e., > 99.9%). The steps are as follows:

1. Deblocking (Figure 7.2 (i)): The 4,4′-dimethoxytrityl protecting group on the previously added nucleoside (or the preloaded base for the first addition when using preloaded CPG) is removed using an acidic solution (3% dichloroacetic acid in DCM)
2. Coupling (Figure 7.2 (ii)): The next nucleoside is added through a reaction of the DMT-protected 3′-*O*-phosphoramidite with the free 5′-OH. Solutions containing the phosphoramidite (0.1 mol L^{-1} phosphoramidite in dry ACN) and an activator (0.45 mol L^{-1} 5-ethylthio-*1H*-tetrazole in dry ACN) are mixed by alternating their addition to the column. These solutions must be in freshly dried ACN as this reaction is very sensitive to the presence of water. Coupling times are 120 seconds for A or T and 160 seconds for C or G.
    - H-Phosphonate Coupling: This reaction can also be accomplished using a 3′-H-phosphonate rather than a phosphoramidite. A 0.1 molL^{-1} solution of H-phosphonate in 30:70 dry pyridine/ACN is added to the column along with 0.5 M pivaloyl chloride in ACN. Coupling times were the same as for phosphoramidites.
3. Oxidation (Figure 7.2 (iii)): The trivalent phosphorus is oxidised to the pentavalent phosphorus seen in the final oligonucleotide using 0.02 mol L^{-1} I_2 in 88:10:2 mixture of tetrahydrofuran (THF):pyridine:water.
4. Capping (Figure 7.2 (iv)): Any unreacted 5′-OH groups are capped with acetyl groups to prevent the formation of side products using Cap A (85:15 mixture of THF and acetic anhydride) and Cap B (16% 1-methylimidazole in THF), which are delivered at the same time on the column.

The volume and duration of each step is dependent on the reagent and size of the column. These four steps are repeated for each base until synthesis is complete. Once synthesis is complete, oligonucleotides are cleaved from the solid

support by placing an oligonucleotide in a solution of 28% aq. ammonia at room temperature overnight or 55 °C for 12 hours (Figure 7.2 (v)).

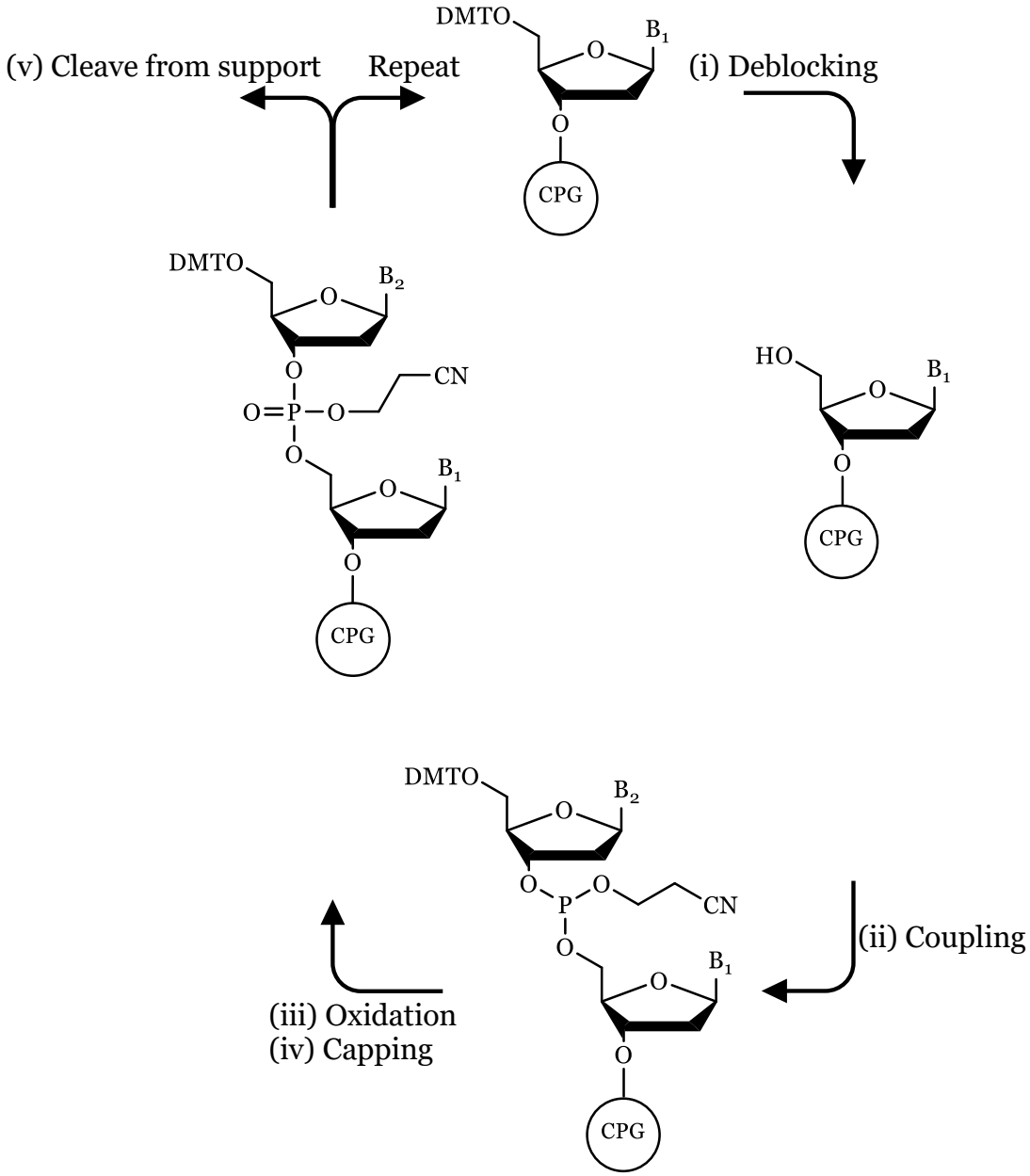


Figure 7.2. General protocol for DNA synthesis. (i) Dichloroacetic acid, solvent: DCM. (ii) Phosphoramidite, 5-ethylthio-1H-tetrazole, solvent: ACN. (iii) I_2, solvent: 80:10:2 THF:pyridine:H_2O. (iv) Acetic anhydride, 1-methylimidazole solvent: THF. (v) 28% aq. NH_3.

These steps represent a standard DNA synthetic protocol. Additional or alternative steps may be added for modified sequences. The duration and conditions of these steps are often adjusted to maximise yield. For details on the changes made to synthesis in these cases, see the relevant chapter or Chapter 7.4.

After purification and characterisation (see Chapter 7.2.2 and 7.2.4), DNA was typically folded into a secondary structure by mixing with the buffer, heating

to 90 °C for approx. 5 minutes, then cooling slowly to room temperature and storing at 4 °C. The time required varies between structures and sequences. Typically, unimolecular G4s would be folded within hours, but bimolecular DNA should remain at 4 °C overnight, and tetramolecular G4s may take several days to completely fold. Samples were usually folded at 100 μM strand concentration or higher and, if necessary, diluted to the desired concentration after folding.

7.2.2. High-Performance Liquid Chromatography (HPLC)

In HPLC a solution containing the analyte is passed through a pressurised column containing an absorbent material. Ion exchange (IE-HPLC) and reverse phase (RP-HPLC) columns are used for purification. Ion exchange separates molecules based on their charge, which is useful for DNA due to the presence of negatively charged phosphates in the backbone, meaning that sequences are separated by the number of nucleotides (i.e., phosphates) present. Reverse phase separates molecules based on their hydrophobicity. Different DNA sequences often have similar hydrophobicity, but RP-HPLC is often used for sequences containing a 5′-DMT group to separate them from the short sequences derived from incomplete coupling.

HPLC had two main applications for oligonucleotides: purification and analysis. Both RP-HPLC and IE-HPLC can be used for DNA purification, but RP-HPLC is used throughout this thesis. RP-HPLC used a gradient of 0.1 M triethylammonium acetate buffer and ACN (ACN Percentage: 0 – 20 min: 0-25% ACN, 20 – 22 mins: 25-80% ACN, 22 – 24 mins: 80% ACN, 24 – 25 mins: 80-0% ACN). This programme was also used to verify oligonucleotide purity following desalting and mass spectrometry. In some cases, G4 formation occuring in samples results in additional peaks appearing between 24 and 28 mins, during the washing step of the programme. These peaks were particularly prevalent when sequences were modified to stabilise G4s, meaning RP-HPLC profiles of pure modified oligonucleotides often contain additional peaks after 24 mins.

Size exclusion-HPLC (SE-HPLC), which separates molecules based on size, was used in some preliminary experiments to evaluate duplex formation. The initial results had given good indication that the modified telomeric sequences described in Chapter 4.2. did not form duplexes. However, they ultimately provided less quantifiable results than native gels or ^{1}H NMR. The native gel

experiments described in Chapter 4 were based on those initial size exclusion-HPLC experiments and produced nearly identical results. Additionally, gravity-fed Nap-5 size exclusion columns were used to remove TEAA salts from oligonucleotide samples following RP-HPLC purification. This is not an HPLC technique, but the principle is otherwise identical.

7.2.3. Nuclear Magnetic Resonance (NMR) Spectroscopy

NMR spectroscopy observes the spin relaxation time of protons and neutrons (nucleons) when they are excited within a magnetic field. Atoms will only be observable if they have net spin greater than zero, ideally resulting from an odd number of total nucleons. The chemical shift of each atom is determined by how shielded it is from the magnetic field, which reflects its proximity to electron-withdrawing groups. Atoms that are close to electron-withdrawing groups will be less shielded and consequently have higher chemical shift in an NMR spectrum. 2D NMR can show correlation between peaks in multiple spectra, indicating atoms which are bonded to each other or close together in space.

NMR spectroscopy is a common method of identifying small molecules and is used throughout this chapter to identify products of reactions. When possible, these results are also confirmed by mass spectrometry. ^{1}H and ^{13}C are used to identify the proton and carbon atoms in each molecule (both have spin $s = ½$). These results are confirmed most often by 2D techniques such as correlation spectroscopy (COSY), which identifies correlation between protons on adjacent carbon atoms and heteronuclear single-quantum correlation spectroscopy (HSQC) which identifies single bond correlations between protons and carbon atoms (i.e., C-H bonds). When necessary, heteronuclear multiple-bond correlation spectroscopy (HMBC) is also used to identify coupling between atoms three bonds apart.

NMR spectroscopy can also be useful for analysing oligonucleotides and the complexes formed. However, oligonucleotides have a significant number of overlapping peaks in ^{1}H NMR (for example, all thymine methyl groups appear at similar chemical shift) so fully characterising sequences using 1D NMR is difficult. It is possible with 2D NMR, but high resolution is required to distinguish peaks and observe low intensity peaks.

More often, we used ^{1}H NMR in the context of oligonucleotides to identify secondary structures. Different H-bonding arrangements in each secondary structure cause significant changes in the chemical shifts of imino protons in ^{1}H NMR. G4 imino-protons have chemical shifts of approximately 11 ppm, duplexes have shifts of 13-14 ppm and i-motifs have shifts of 15 ppm.[102] While full characterisation of these oligonucleotides is challenging, identifying formation of specific secondary structures is straightforward. This is the primary way NMR spectroscopy is employed for oligonucleotide analysis in this thesis.

2D NMR techniques can be used to fully assign DNA complexes. Correlations shown in nuclear Overhauser effect spectroscopy (NOESY), which identifies correlation between protons close together in space, can indicate protons of adjacent nucleobases (e.g., H-8 of guanosine or the methyl group of thymine). NOESY has also been used to identify which nucleobases are interacting through hydrogen-bonding, which can be used to determine overall structure more clearly.[138, 139] Our experiments focus on the 1D NMR techniques used to differentiate various secondary structures, but reported NMR structures of G4s using 2D NMR data were evaluated when developing strategies.

7.2.4. Mass Spectrometry

In mass spectrometry an analyte is ionised and passed through an electric field. This allows molecules to be separated based on mass and charge, and ultimately obtain the molecular weight of the analyte. When possible, this technique was used to confirm the composition of products of small molecule reactions in addition to NMR spectroscopy. It was also used to identify the composition of DNA sequences obtained after synthesis and purification.

Oligonucleotides often give high charges in mass spectrometry and often contain large quantities of Na^{+} and K^{+} adducts. As mentioned above, Nap-5 size exclusion columns were used to desalt DNA samples prior to mass spectrometry, which minimises formation of adducts. Typical oligonucleotide mass spectrometry profiles will include numerous low m/z peaks with higher charges, which can be used to confirm the mass. This can often be challenging without deconvolution software due to the large number of peaks associated with Na^{+} and K^{+} adducts. However, this is the most accurate method of obtaining oligonucleotide mass and determining if the target sequence was obtained. It is

also sometimes possible to view larger secondary structures using mass spectrometry, usually by adding ions such as NH_4OAc, but preventing these structures from breaking down during ionisation is challenging and can require careful adjustment of temperatures and voltage to preserve structures. For this reason, we favour native PAGE, size exclusion-HPLC or NMR over native mass spectrometry when possible.

Mass spectrometry samples of DNA were typically prepared by mixing 6 μL of methanol, 4 μL of the DNA stock solution and 30 μL of Milli-Q H_2O, resulting in a sample with 15% methanol. The precise concentration of DNA in samples used to confirm the oligonucleotide sequence was not controlled. For native mass spectrometry, samples were folded at 100 μM in pH 7.4 NH_4OAc buffer.

7.2.5. UV-Vis Spectroscopy

Light is passed through a sample and the absorbance of the sample is measured. This technique was primarily used to calculate DNA concentration using the Beer-Lambert Law (equation 6.1). Molar absorptivity (ε) of DNA is mostly dependent on the individual bases in the sequence meaning it can be calculated relatively easily with known absorptivity of individual nucleotides. In this thesis, the DNA sequence analyser tool provided by IDT[77] was used to obtain molar absorptivity unless otherwise specified. DNA has a λ_{max} at approximately 260 nm, so light is passed through the sample at this wavelength and the absorbance obtained is used to calculate DNA concentration. This technique can also be used for melting/kinetics experiments. CD was selected for thermal stability experiments because melting of G4 structures results in a larger observable shift in CD spectra than UV-Vis spectra. This meant that CD spectrometry gave more precise data for determining T_m of individual topologies and was less susceptible to fluctuations.

$$A = \varepsilon cl \qquad \text{(Equation 6.1)}$$

7.2.6. Circular Dichroism (CD) Spectroscopy

Circular Dichroism is one of the most important techniques specifically for investigating DNA secondary structures. Circularly polarised light is passed through the sample, with different stereoisomers rotating this light in different directions. This is also observed when circularly polarised light in passed through

Table 7.1. Characteristic CD peaks of DNA secondary structures

Secondary Structure	Characteristic peaks (nm)	Glycosidic Bond Conformation
Parallel G4	+265, -240	Anti
Antiparallel G	+290, -260	Syn
A-DNA	-210, +260 – 280	Anti
B-DNA	+260 – 280, -245[a]	Anti
Z-DNA	-290, +260, -205	G: Syn, C: Anti
i-motif	+290, - 260	Anti

The + and – values indicate the phase of the peak. (a) Duplex DNA CD peaks vary based on sequence composition, but often have multiple peaks, or vary in intensity compared to G-quadruplexes. Structures such triplexes can also be characterised using CD, but their characteristic peaks vary significantly depending on the sequence and topology.

G4 samples depending on the conformation of guanosine present. The stacking of different *syn*- and *anti*-conformers in DNA results in characteristic spectra for each secondary structure (Table 7.1). Parallel and antiparallel G4s give distinct CD spectra because they contain different guanosine conformers.[140-143] This technique can be limited in some cases because it only reflects the conformation of the bases, meaning that if the base conformation remains the same while structure changes, CD may not reflect the structural change (e.g., strand orientation). A previously reported example of this is discussed in Chapter 4 for internal polarity inversions within the thrombin binding aptamer (TBA). Identifying the secondary structure in this case required the use of NMR

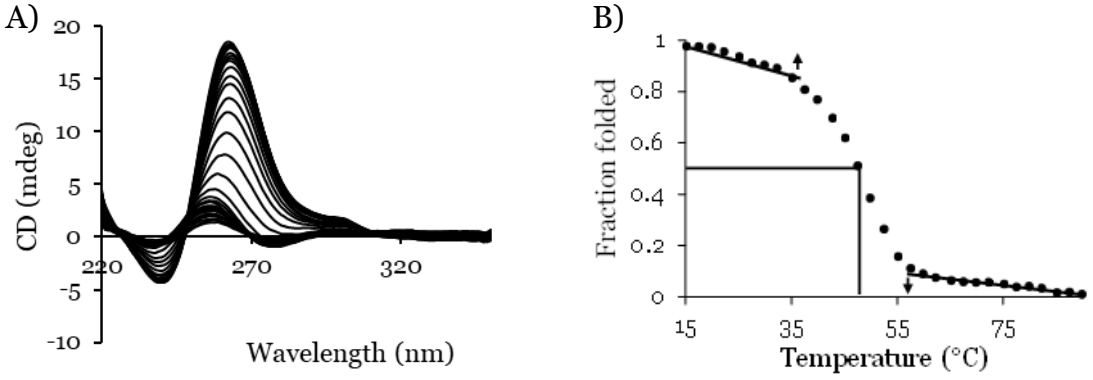


Figure 7.3. A) Example of melting spectra of tetramolecular G-quadruplex. B) Melting profile, showing adjustments made using Equation 6.2.

spectroscopy.

This technique is also used for melting experiments to determine thermal stability and rate of formation. Melting experiments are carried out by gradually heating the sample while taking regular scans to observe the decrease in the peaks as the structure breaks down. Samples were heated at 1°C/min (2.5 °C steps, with

spectra recorded every 150 seconds) between 15 and 90 °C (Figure 7.3A). CD signal at the wavelength corresponding to secondary structure formation was adjusted using equation 6.2 to obtain the fraction of DNA folded into the secondary structure at a given temperature (α).[70] The value θ_n represents the CD signal at λ_{max} at a given temperature (n), while θ_{min} and θ_{max} represent the CD signal when the sample is completely unfolded and completely folded, respectively. These values were obtained before and after the melting experiment was performed. The fraction folded values were plotted over temperature to obtain a melting profile, as shown in Figure 7.3. As shown in the example profile, these spectra were adjusted to account for variations in CD signal before and after melting. The T_m is defined as the inflection point, which represent the temperature at which half of the structure is unfolded.

$$\alpha = (\theta_n - \theta_{min}) / (\theta_{max} - \theta_{min}) \qquad \text{(Equation 6.2)}$$

We also initially attempted to use this technique to observe duplex formation, expecting peaks to shift from characteristic G4 peaks to the wider variety of duplex spectra observed. This was useful in some cases, but for many samples the changes in CD spectra were insignificant, with positive peaks corresponding to duplex formation appear around 260 – 280 nm, making the distinction from parallel G4 peaks at 265 nm difficult for some sequences. While some variation in peak intensity was observable, size exclusion-HPLC, native PAGE and, in particular, ^{1}H NMR spectroscopy were more effective methods of distinguishing secondary structures. However, CD provides an easy and reliable method of identifying G4 topologies and measuring thermal stability.

7.2.7. Polyacrylamide Gel Electrophoresis (PAGE)

This technique is used to separate molecules based on size, charge and shape. Samples were loaded on a 20% polyacrylamide gel and a voltage is applied across it. Charged molecules move through the gel based on their size and charge. DNA has a negatively charged backbone, allowing it to move through the gel when a voltage is applied. Shorter sequences move through the gel more quickly, allowing for a sequence to be separated based on length. We use a polythymidine ladder (usually T_5 – T_{50}) to provide a reference for sequence length. However, this is complicated by secondary structure formation, as folding into a secondary

structure often results in a more compact shape. Some G4s will therefore behave similarly on PAGE to significantly shorter polythymidine sequences.

In denaturing gels (dPAGE) the samples and gel contained urea (7 M), which disrupts DNA secondary structure. Some urea was added to the acrylamide solution instead of milli-Q water prior to degassing. This allows analysis of the unfolded structure. A native gel is carried out without urea and using a buffer in DNA folding and gel formation (the buffer used can vary and typically is selected to match other experiments such as CD or NMR), allowing analysis of the folded structure. In addition to the buffer, native gel samples also included 20% glycerol, which delayed the samples from diffusing out of the well before a voltage was applied. Denaturing PAGE is useful for analysing sequence length and purity of DNA. Native PAGE gave results similar to size exclusion-HPLC and were useful for observing duplex/G4 formation.

Electrophoretic mobility shift assay (EMSA) refers to a type of gel used to study protein-DNA interactions. A fluorescent or radio-labelled DNA sequence was loaded into the well along with a protein. The mobility of the labelled sequence changes if it binds to the protein. The ratio of labelled oligonucleotide:protein complex to the non-complexed oligonucleotide, relative to a control in the absence of the protein, indicates the affinity of the protein for the labelled sequence. Competitive assays are performed by including another DNA sequence in the sample. This allows comparison between the affinity of the labelled oligonucleotide to the protein in the absence and presence of competing sequence and therefore the relative affinities of the two sequences.

Analytical gels were prepared by filtering and degassing a 40 mL solution containing 40% acrylamide solution diluted with the running buffer (1x tris-borate-EDTA). All gels described in this thesis are 20% acrylamide unless specified otherwise. Urea was also added to denaturing gels at this stage. After degassing, 120 μL of freshly prepared 10% w/v ammonium persulfate solution was added followed by 12 μL of *N,N,N′,N′*-tetramethylethane-1,2,-diamine (TMED). This mixture was then loaded into a mould. The gel was formed after approximately 30 minutes, after which samples can be loaded into wells. The results were visualised using a stains-all dye solution (37 mg in 100 mL of 1:1 formamide: water solution).

7.2.8. Biolayer Interferometry (BLI)

BLI uses a probe with a biosensor tip upon which a tagged protein can be bound and immobilised. Our experiments use Ni-NTA (Nickel-charged tris-nitriloacetic acid) tips. The his-tag on the protein coordinates to the nickel ions, immobilising the protein on the tip. The interference pattern of white light passing through the probe is then analysed. A change in absorbance indicates the binding of the analyte to the protein, giving a profile representing the binding over time.

In our experiments it is used to indicate affinity of a His-tagged HP1α protein for a particular DNA sequence or structure. The tip is inserted into a blank buffer solution for 60 seconds, followed by a solution containing the labelled protein for 5 minutes to attach protein to the tip (Figure 7.4, association). The tip is then returned to the buffer solution for 60 seconds to observe protein dissociation from the tip. It is then inserted into a solution containing the sequence to be tested for 5 minutes. A control (5 minutes in buffer instead) is subtracted from this data to account for protein dissociation occurring during this step. The tip is then placed back into a buffer solution to observe the dissociation of the analyte (Figure 7.4, dissociation). The data can be used to calculate k_{on} (rate of association), k_{off} (rate of dissociation) and K_D (k_{off}/k_{on}, dissociation constant). High k_{on}, low k_{off} and low K_D values are used to indicate DNA with good affinity to the protein.

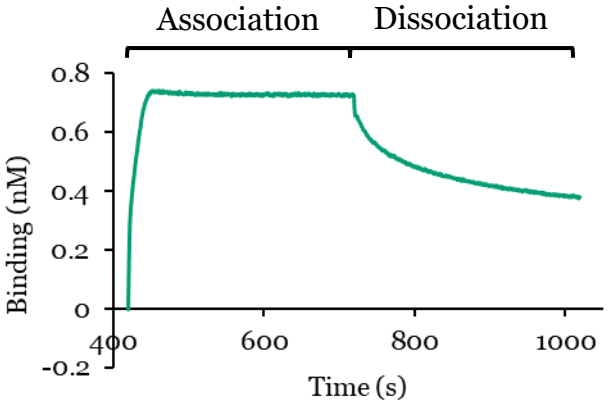
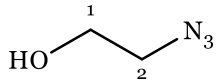


Figure 7.4. Example of BLItz binding profile showing the association and dissociation step of oligonucleotides binding to his-tagged HP1α immobilised on a Ni-NTA tip.

7.3. Synthesis of Adenosine H-Phosphonate Containing Azide

2-Azido-1-ethanol (2a)



2-Chloro-1-ethanol (**1a**, 20.1 g, 0.249 mol, 1 eq) was stirred in 250 mL of water and NaN_3 (48.75 g, 0.75 mol, 3 eq) was added. The reaction mixture was stirred at 60 °C for 72 hours. Mixture was cooled and extracted three times with 150 mL of diethyl ether. The organic phase was evaporated *in vacuo*. The product was obtained as a clear liquid and used in the following step without purification. Yield: 18.1 g (83%).

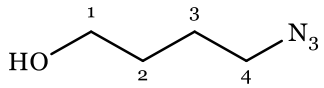
^{1}H NMR results agreed with previously synthesised material provided by Dr. Hari Kurup.[126]

^{1}H NMR (500 MHz, $CDCl_3$) δ = 3.72 (2H, q, J = 4.19 Hz, H-1), 3.36 (2H, t, J = 5.11 Hz, H-2).

Example of general protocol for azide substitution (**2b** and **c**)

4-Chloro-1-butanol (**1b**, 5.44 g, 50.1 mmol, 1 eq) was stirred in 50 mL of water, and NaN_3 (9 g, 0.145 mol, 3 eq) was added. The reaction mixture was refluxed overnight. The aqueous phase was extracted three times with 20 mL of ethyl acetate. The organic phase was dried using anhydrous sodium sulphate and evaporated *in vacuo*. The product was obtained as a clear liquid and used in the following step without purification.

4-Azido-1-butanol (2b)



Starting Material: 5.44 g, 50.1 mmol.

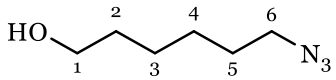
Yield: 2.09 g (38%).

NMR results were in agreement with Bates & Dewey (2009).[144]

^{1}H NMR (500 MHz, $CDCl_3$) δ = 3.57 (2H, t, J = 6.31 Hz, H-1), 3.26 (2H, t, J = 6.7 Hz, H-4), 1.65 – 1.52 (4H, m, H-2/3).

^{13}C NMR (125 MHz, $CDCl_3$) δ = 61.73 (C-1), 51.23 (C-4), 29.63 (C-2/3), 25.31 (C-2/3).

6-Azido-1-hexanol (2c)



Starting material: 0.43 g, 3.15 mmol.

Yield: 0.28 g (62%).

NMR results were in agreement with Zhang *et al.* (2010).[145]

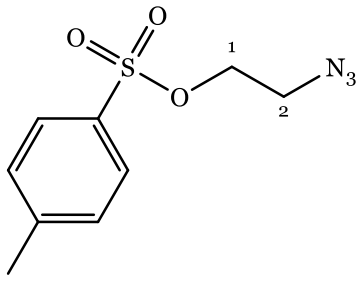
^{1}H NMR (500 MHz, $CDCl_3$) δ = 3.59 (2H, t, J = 6.7 Hz, H-1), 3.24 (2H, t, J = 6.9 Hz, H-6), 1.62 – 1.50 (4H, m, H-2/5), 1.40 – 1.33 (4H, m, H-3/4).

^{13}C NMR (125 MHz, $CDCl_3$) δ = 62.53 (C-1), 51.36 (C-6), 32.64, 32.49 (C-2), 28.78 (C-5), 26.50/25.32 (C-3/4).

Example of general protocol for tosyl substitution (3a, b and c)

Compound **2a** (18.1 g, 20.8 mmol, 1 eq) was dissolved in 100 mL of DCM and TEA (43.4 mL, 31.2 mmol, 1.5 eq) and p-tosyl chloride (59.5 g, 31.2 mmol, 1.5 eq) were added. The solution was stirred for 18 hours, monitored with TLC (10% ethyl acetate/hexane, visualised with $KMnO_4$). Additional p-tosyl chloride (5.95 g) and TEA (4.34 mL) were added until the starting material was no longer visible on the TLC. The reaction was quenched with water (50 mL) and extracted with 2 × 30 mL of DCM. The combined organic phase was dried over anhydrous sodium sulfate, filtered and evaporated *in vacuo*. Product was purified by silica gel column chromatography (0-20% ethyl acetate/hexane).

2-Azido-1-tosylethane (3a)



Starting material: 18.1 g, 20.8 mmol.

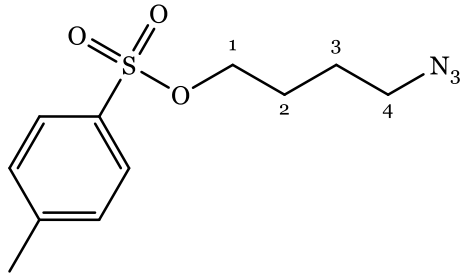
Yield: 22.4 g (44%).

NMR results agreed with the product obtained previously by Dr. Hari Kurup. [126]

^{1}H NMR (500 MHz, $CDCl_3$) δ = 7.81 (2H, d, J = 8.49 Hz, aromatic C<u>H</u>), 7.37 (2H, d, J = 8.05 Hz, aromatic C<u>H</u>), 4.15 (2H, t, J = 4.96 Hz, H-2), 2.48 (2H, t, J = 4.93 Hz, H-1), 2.46 (3H, s, C<u>H</u>$_3$).

^{13}C NMR (125 MHz, $CDCl_3$) δ = 144.92 (S–<u>C</u>$_{aromatic}$), 132.98 (CH_3–<u>C</u>$_{aromatic}$), 129.91 (<u>C</u>$_{aromatic}$H), 127.87 (<u>C</u>$_{aromatic}$H), 69.70 (C-4), 50.67 (C-1), 26.14 (C-3), 25.03 (C-2), 21.64 (CH_3).

4-Azido-1-tosylbutane (3b)



Starting Material: 1.5 g, 13.02 mmol.

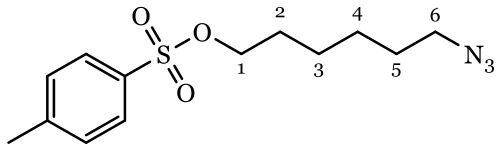
Yield: 2.48 g (71%).

NMR results were in agreement with Ding *et al.* (2006).[146]

^{1}H NMR (500 MHz, $CDCl_3$) δ = 7.81 (2H, d, J = 8.48 Hz, aromatic C<u>H</u>), 7.38 (2H, d, J = 8.03 Hz, aromatic C<u>H</u>), 4.07 (2H, t, J = 6.10 Hz, H-4), 3.28 (2H, t, J = 6.66 Hz, H-1), 2.47 (3H, s, C<u>H</u>$_3$), 1.79 – 1.72 (2H, m, H-3), 1.67 – 1.60 (2H, m, H-2).

^{13}C NMR (125 MHz, $CDCl_3$) δ = 144.92 (S–<u>C</u>$_{aromatic}$), 132.98 (CH_3–<u>C</u>$_{aromatic}$), 129.91 (<u>C</u>$_{aromatic}$H), 127.87 (<u>C</u>$_{aromatic}$H), 69.70 (C-4), 50.67 (C-1), 26.14 (C-3), 25.03 (C-2), 21.64 (CH_3).

6-Azido-1-tosylhexane (3c)



Starting material: 0.4 g, 2.79 mmol.

Yield: 0.54 g (65%).

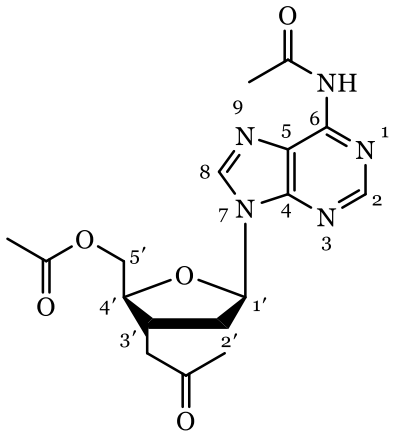
NMR results were in agreement with Li *et al.* (2011).[147]

^{1}H NMR (500 MHz, $CDCl_3$) δ = 7.77 (2H, d, J = 8.62 Hz, aromatic C<u>H</u>), 7.34 (2H, d, J= 8.62 Hz, aromatic C<u>H</u>), 4.01 (2H, t, J = 6.45 Hz, H-6), 3.21 (2H, t, J = 6.95

Hz, H-1), 2.44 (3H, s, C<u>H</u>$_3$), 1.64 (2H, quin, J= 6.93 Hz, H-5), 1.53 (2H, quin, J = 6.93 Hz, H-2), 1.379 – 1.258 (4H, m, H-3 and H-4).

^{13}C NMR (125 MHz, $CDCl_3$) δ = 144.77 (S–<u>C</u>$_{aromatic}$), 133.11 (CH_3–<u>C</u>$_{aromatic}$-CH_3), 129.85 (<u>C</u>$_{aromatic}$H), 127.83 (<u>C</u>$_{aromatic}$H), 70.39 (C-6), 51.20 (C-1), 28.65 (C-5), 28.58 (C-2), 26.14 (C-3 /4), 24.93 (C-3/4), 21.59 (CH_3).

3′-*O*,5′-*O*,*N*^{6}-Triacetyl-2′-deoxyadenosine (4)



2′-Deoxyadenosine (5 g, 19.94 mmol, 1 eq) was dissolved in 36 mL of pyridine and acetic anhydride (17.5 mL, 0.185 mol, 10 eq) was added. Reaction mixture was stirred overnight at room temperature, then heated to 60°C. Reaction mixture was monitored by TLC (5% MeOH/DCM) to observe the disappearance of 3′,5′-diacetyl-2′-deoxyadenosine and formation of tri- and tetra-acetylated products. Reaction was quenched with excess EtOH (50 mL) and allowed to cool to room temperature. Volatiles were evaporated *in vacuo* (pyridine was co-evaporated with successive additions of EtOH and MeOH). The resulting oil was diluted with 30 mL of ethyl acetate and washed with brine (4 × 30 mL). The organic phase was dried over anhydrous sodium sulfate, filtered and evaporated *in vacuo*. Product was purified by silica gel column chromatography (0-10% MeOH/DCM). ^{1}H NMR spectra is in agreement with that reported by Terrazas *et al.* (2004).[148]

Yield: 6.24 g (83%).

^{1}H NMR (500 MHz, d_6-DMSO) δ = 10.72 (1H, s, NH), 8.67 (2H, 2 x s, H-2/8), 6.49 (1H, t, J = 7.07 Hz, H-1′), 5.45 (1H, dt, J = 6.61 Hz, 2.46 Hz, H-3′), 4.35 – 4.20 (3H, m, H-4′/5′), 3.24 – 3.16 (1H, m, Hz, H-2$_a$′), 2.61 (1H, ddd, J = 14.4 Hz,

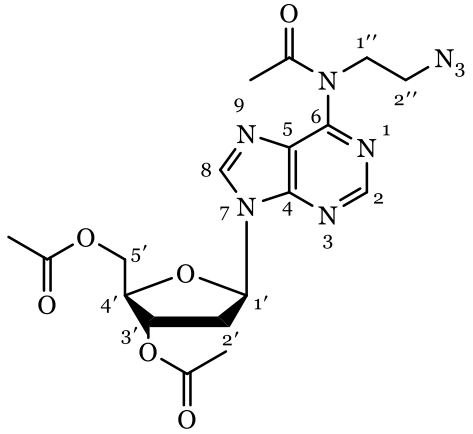
6.4 Hz, 2.69 Hz, H-2b′), 2.26 (3H, s, C(=O)-C<u>H</u>3), 2.11 (3H, s, C(=O)-C<u>H</u>3), 2.01 (3H, s, N-C(=O)-C<u>H</u>3).

ESI-MS (MeOH, positive): Calculated: 378.75 [M+H+], 400.74 [M+Na+]; Experimental: 378.10 [M+H+], 400.03 [M+Na+].

Example of general protocol for adenosine alkylation (**5a, b** and **c**)

Compound **4** (5.16 g, 13.67 mmol, 1 eq) was dissolved in 50 mL of ACN and Cs_2CO_3 (13.4 g, 41.02 mmol, 3 eq) was added. The mixture was stirred for 15 mins under argon, then Compound **3a** (19.79 g, 82.04 mmol, 6 eq) was added dropwise. The mixture was heated to 60°C and stirred overnight under argon. Disappearance of the starting material was monitored by TLC (5% MeOH/DCM). When the starting material had disappeared, the reaction mixture was diluted in ethyl acetate and washed with water (2 × 40 mL). The organic phase was dried over anhydrous sodium sulfate, filtered and evaporated *in vacuo*. Product was purified by silica gel column chromatography (0 – 10% MeOH/DCM).

3′-*O*,5′-*O*,*N*6-Triacetyl-N6-(2-azidoethane)-2′-deoxyadenosine (5a)



Starting material: 5.16 g, 13.67 mmol.

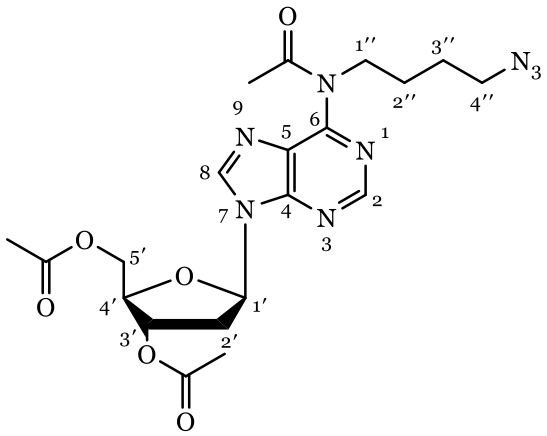
Yield: 3.36 g (55%).

1H NMR results agreed with previously synthesised material provided by Dr. Hari Kurup.[126]

1H NMR (500 MHz, d6-DMSO) δ = 8.86 (1H, s, H-8), 8.83 (1H, s, H-2), 6.54 (1H, t, J = 6.90 Hz, H-1′), 5.48 – 5.44 (1H, m, H-3′), 4.37 – 4.19 (5H, m, H-4′/5′/1′′), 3.54 (2H, t, J = 6.08 Hz, H-2′′), 3.26 – 3.13 (1H, m, H-2a′), 2.63 (1H, ddd, J =

14.29 Hz, 6.46 Hz, 2.81 Hz, H-2b′), 2.19 (3H, s, C(=O)-C<u>H</u>3), 2.12 (3H, s, C(=O)-C<u>H</u>3), 2.01 (3H, s, N-C(=O)-C<u>H</u>3).

3′-*O*,5′-*O*,*N*⁶-Triacetyl-N⁶-(4-azidobutane)-2′-deoxyadenosine (5b)



Starting material: 0.6 g, 1.59 mmol.

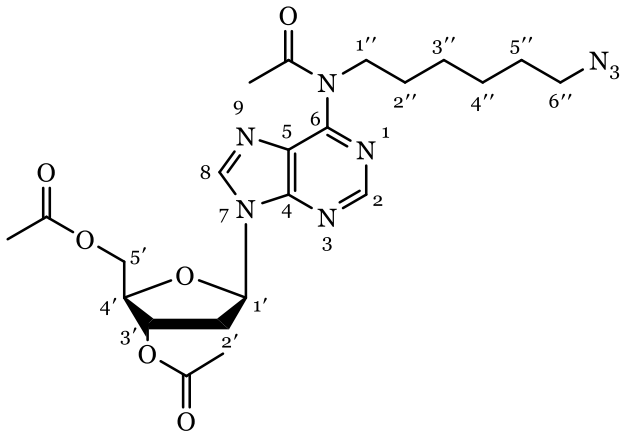
Yield: 0.55 g (73%)

¹H NMR (500 MHz, d6-DMSO) δ = 8.85 (1H, s, H-8), 8.82 (1H, s, H-2), 6.53 (1H, t, J = 7.18 Hz, H-1′), 5.46 (1H, dt, J = 6.59 Hz, 3.29 Hz, H-3′), 4.35 – 4.21 (3H, m, H-4′/5′), 4.12 (2H, t, J = 6.89 Hz, H-1′′), 3.27 (2H, t, J = 6.49 Hz, H-4′′), 3.24 – 2.17 (1H, m, H-2a′), 2.63 (1H, ddd, J = 14.45, 6.50, 2.86 Hz, H-2b′), 2.12 (6H, 2 x s, C(=O)-C<u>H</u>3), 2.01 (3H, s, N-C(=O)-C<u>H</u>3), 1.59 – 1.40 (4H, m, H-2′′/3′′).

¹³C NMR (125 MHz, d6-DMSO) δ = 206.94 (O=<u>C</u>-CH3), 170.87 (O=<u>C</u>-CH3), 170.57 (O=<u>C</u>-CH3), 170.52 (O=<u>C</u>-CH3), 153.10 (C-6), 153.04 (C-4), 152.26 (C-8), 145.05 (C-2), 127.67 (C-4/5), 84.40 (C-1′), 82.35 (C-4′), 74.60 (C-3′), 63.96 (C-5′), 50.70 (C-4′′), 46.29 (C-1′′), 35.89 (C-2′), 26.00/25.92 (C-2′′/3′′), 24.25 (N-C(=O)-<u>C</u>H3), 21.25/20.96 (O-C(=O)-<u>C</u>H3).

ESI-MS (MeOH, positive): Calculated: 475.2 [M+H⁺], 497.18 [M+Na⁺]; Experimental: 475.14 [M+H⁺], 497.07 [M+Na⁺].

3′-*O*, 5′-*O*, *N*^{6}-Triacetyl-N^6-(6-azidohexane)-2′-deoxyadenosine (5c)



Starting material: 0.76 g, 2.01 mmol.

Yield: 0.77 g (76%).

^{1}H NMR (500 MHz, d$_6$-DMSO) δ = 8.84 (1H, s, H-8), 8.81 (1H, s, H-2), 6.53 (1H, t, J = 7.02 Hz, H-1′), 5.46 (1H, dt, J = 6.57 Hz, 3.3 Hz, H-3′) 4.35 – 4.29 (2H, m, H-5′), 4.27 – 4.21 (1H, m, H-4′), 4.09 (2H, t, J = 7.3 Hz, H-1′′), 3.26 – 3.17 (3H, m, H-6′′/H-2$_a$′), 2.63 (1H, ddd, J = 14.29 Hz, 6.3 Hz, 2.94 Hz, H-2$_b$′), 2.11 (3H, s, C(=O)-C<u>H</u>$_3$), 2.09 (3H, s, C(=O)-C<u>H</u>$_3$), 2.01 (3H, s, N-C(=O)-C<u>H</u>$_3$), 1.51 – 1.39 (4H, m, H-2′′/5′′), 1.25 – 1.18 (4H, m, H-3′′/4′′).

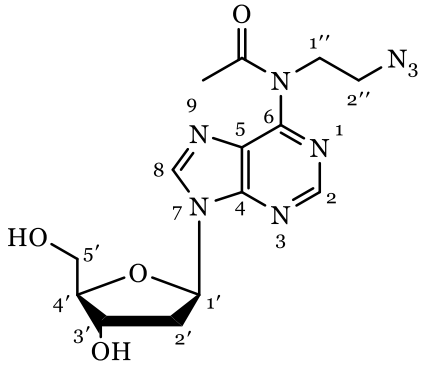
^{13}C NMR (125 MHz, d$_6$-DMSO) δ = 170.78 (O=<u>C</u>-CH$_3$), 170.57 (O=<u>C</u>-CH$_3$), 170.53 (O=<u>C</u>-CH$_3$), 152.27 (C-6), 152.99 (C-4), 152.22 (C-8), 144.95 (C-2), 127.60 (C-5), 84.38 (C-1′), 82.34 (C-4′), 74.60 (C-3′), 61.97 (C-5′), 50.92 (C-6′′), 46.73 (C-1′′), 35.86 (C-2′), 28.50/28.38 (C-2′′/5′′), 26.10/26.02 (C-3′′/4′′), 24.28 (N-C(=O)-<u>C</u>H$_3$), 21.25/20.95 (2x O-C(=O)-<u>C</u>H$_3$).

ESI-MS (MeOH, positive): Calculated: 502.53 [M+H$^+$], 524.53 [M+Na$^+$]; Experimental: 503.12 [M+H$^+$], 525.11 [M+Na$^+$].

Example of general protocol for acetyl deprotection (**6a, b** and **c**)

Compound **5a** (3.36 g, 7.5 mmol, 1 eq) was dissolved in 80 mL of a 50/50 MeOH/H$_2$O solution. Excess TEA (10.5 mL, 75 mmol, 10 eq) was added and the reaction was monitored by TLC (10% MeOH/DCM) until starting material disappeared after approx. 15 mins. Solution was evaporated under reduced pressure and the residue was purified using silica gel column chromatography (0 – 20% MeOH/DCM).

*N*⁶-Acetyl-*N*⁶-(2-azidoethane)-2′-deoxyadenosine (6a)



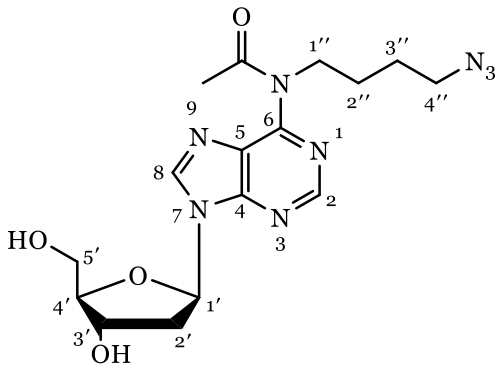
Starting material: 3.36 g, 7.5 mmol.

Yield: 2.14 g (79%).

^{1}H NMR results agreed with previously synthesised material provided by Dr. Hari Kurup.[126]

^{1}H NMR (500 MHz, d_6-DMSO) δ = 8.85 (2H, s, H-8/2), 6.50 (1H, t, 6.66 Hz, H-1′), 5.02 (1H, t, J = 5.47 Hz, H-3′), 4.30 (2H, t, 6.28 Hz, H-1′′), 3.92 (1H, q, J = 4.3 Hz, H-4′), 3.59 – 3.549 (4H, m, H-5′/2′′), 2.79 (1H, q, J = 6.61 Hz, H-2$_a$′), 2.38 (1H, ddd, J = 13.27 Hz, 6.36 Hz, 3.49 Hz, H-2$_b$′), 2.18 (3H, s, N-C(=O)-C$\underline{H}_3$).

*N*⁶-Acetyl-*N*⁶-(4-azidobutane)-2′-deoxyadenosine (6b)



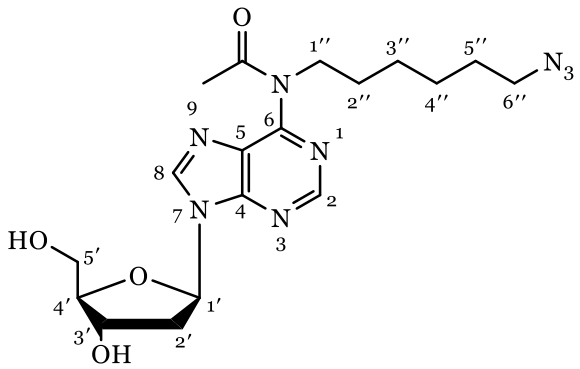
Starting material: 0.55 g, 1.16 mmol.

Yield: 0.36 g (80%).

^{1}H NMR (500 MHz, d_6-DMSO) δ = 8.84 (2H, s, H-8/2), 6.50 (1H, t, 6.72 Hz, H-1′), 5.38 (1H, br, OH), 5.01 (1H, br, OH), 4.46 (1H, dt, J = 8.77 Hz, 3.30 Hz, H-3′), 4.11 (2H, t, 6.91 Hz, H-1′′), 3.92 (1H, q, J = 4.3 Hz, H-4′), 3.67 – 3.52 (2H, m, H-5′), 3.27 (2H, t, 6.78 Hz, H-4′′), 2.79 (1H, m, H-2$_a$′), 2.38 (1H, ddd, J = 13.27 Hz, 6.36 Hz, 3.49 Hz, H-2$_b$′), 2.11 (3H, s, N-C(=O)-C$\underline{H}_3$), 1.56 – 1.44 (4H, m, H-2′′/3′′).

^{13}C NMR (125 MHz, d_6-DMSO) δ = 170.88 (O=<u>C</u>-CH_3), 152.99 (C-6), 152.95 (C-4), 152.14 (C-8), 144.88(C-2) 127.57 (C-5), 88.53 (C-1′), 84.34 (C-4′), 71.06 (C-3′), 61.97 (C-5′), 50.69 (C-4′′), 46.26 (C-1′′), 39.27 (C-2′), 25.99/25.79 (C-2′′/3′′), 24.26 (N-C(=O)-<u>C</u>H_3).

N^6-Acetyl-*N*^{6}-(6-azidohexane)-2′-deoxyadenosine (6c)



Starting material: 0.31 g, 0.61 mmol.

Yield: 0.14 g (55%).

^{1}H NMR (500 MHz, d_6-DMSO) δ = 8.83 (2H, s, H-2/8), 6.50 (1H, t, J = 6.83 Hz, H-1′), 5.39 (1H, br.s., 3′-OH), 5.01 (1H, br, 5′-OH), 4.46 (1H, dt, J = 9.2 Hz, 3.55 Hz, H-3′), 4.08 (2H, t, J = 7.4 Hz, H-1′′), 3.92 (1H, q, J = 4.34 Hz, H-4′), 3.66 – 3.50 (2H, m, H-5′), 3.24 (2H, t, J = 6.95 Hz, H-6′′), 2.82 – 2.75 (1H, m, 2_a′), 2.37 (1H. ddd, J = 13.42 Hz, 6.36 Hz, 3.42 Hz, 2_b′), 2.11 (3H, s, N-C(=O)-C<u>H</u>$_3$), 1.50 – 1.37 (4H, m, H-2′′/5′′), 1.25 – 1.18 (4H, m, H-3′′/4′′).

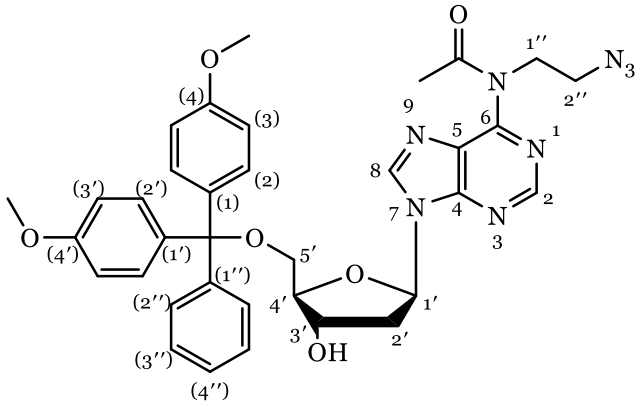
^{13}C NMR (125 MHz, d_6-DMSO) δ = 170.76 (O=<u>C</u>-CH_3), 152.11 (C-6), 152.95 (C-4), 152.09 (C-8), 144.77 (C-2), 127.48 (C-5), 88.53 (C-4′), 84.33 (C-1′), 71.08 (C-3′), 61.97 (C-5′), 50.92 (C-6′′), 46.70 (C-1′′), 35.89 (C-2′), 28.50 (C-2′′/5′′), 26.10 (C-3′′/4′′), 24.29 (N-C(=O)-<u>C</u>H_3).

ESI-MS (MeOH, positive): Calculated: 419.21 [M+H$^+$], 441.20 [M+Na$^+$]; Experimental: 419.21 [M+H$^+$], 441.26 [M+Na$^+$].

Example of general protocol for 5′-DMT protection (**7a, b** and **c**)

Compound **6a** (0.22 g, 0.61 mmol, 1 eq) was co-evaporated several times with HPLC grade pyridine then once with freshly distilled pyridine before it was dissolved in dry pyridine (10 mL) and cooled to 0°C under argon. DMT-Cl (0.27 g, 0.79 mmol, 1.3 eq) was added, and the solution was allowed to slowly warm to room temperature. The reaction was stirred overnight at room temperature, monitored by TLC (10% MeOH/DCM) and additional DMT-Cl (0.135 g, 0.40 mmol, 0.65 eq) was added if remaining starting material was observed. When no starting material was observed, the reaction was quenched with methanol (10 mL). Volatiles were evaporated under reduced pressure and the residue was dissolved in DCM (10 mL) and washed with sat. $NaHCO_3$ (10 mL). The organic phase was dried over anhydrous sodium sulphate, filtered, and evaporated *in vacuo*. Product was purified by silica gel column chromatography (0-20% Acetone/DCM, then 0-15% MeOH/DCM).

5′-*O*-(4,4′-Dimethoxytrityl)-*N*^{6}-acetyl-*N*^{6}-(2-azidoethane)-2′-deoxyadenosine (7a)



Starting material: 0.22 g, 0.61 mmol.

Yield: 0.29 g (72%).

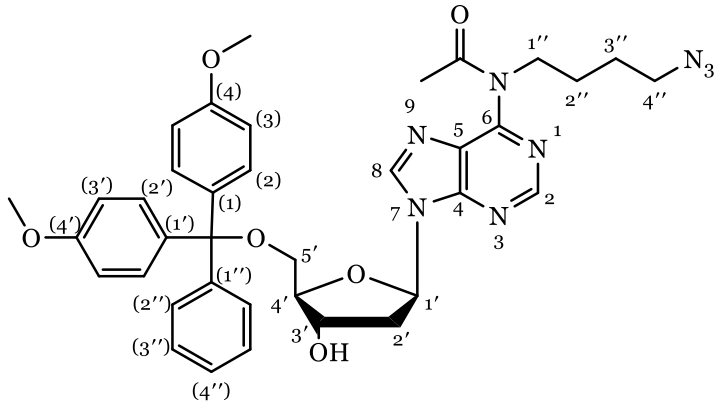
NMR results agreed with previously synthesised material provided by Dr. Hari Kurup.[126]

^{1}H NMR (500 MHz, d_6-DMSO) δ = 8.76 (1H, s, H-8), 8.73 (1H, s, H-2), 7.36 – 7.30 (2H, m, DMT aromatic C<u>H</u>), 7.24 – 7.14 (7H, m, DMT aromatic C<u>H</u> (4H + 2H +1H)), 6.83 – 6.77 (4H, m, DMT aromatic C<u>H</u>), 6.52 (1H, t, J = 6.52 Hz, H-1′), 5.31 (1H, br, 3′-OH), 4.55 – 4.50 (1H, m, H-3′), 4.28 (2H, t, J = 6.1 Hz, H-1′′), 4.06 – 4.01 (1H, m, H-4′), 3.72 (6H, s, 2x O-CH_3), 3.51 (2H, t J = 6.14 Hz, H-2′′)

3.25 – 3.16 (2H, m, H-5′), 2.93 (1H, quintet, J = 6.69 Hz, 2_a′), 2.45 – 2.36 (1H, m, 2_b′), 2.12 (3H, s, N-C(=O)-C$\underline{H}_3$).

^{13}C NMR (125 MHz, d_6-DMSO) δ = 171.28 (O=$\underline{C}$-CH_3), 158.47 (DMT-q), 152.69 (C-6), 151.95 (C-8), 145.29 (DMT-q), 145.11 (C-2), 136.07 (DMT-q), 135.93 (DMT-q), 130.13 (DMT $\underline{C}$H), 128.14 (DMT $\underline{C}$H), 128.10 (C-4) , 127.18 (C-5), 113.53 (DMT-$\underline{C}$H), 86.52 (DMT-q), 85.87 (C-4′), 84.36 (C-1′), 70.95 (C-3′), 64.50 (C-5′), 55.46 (O-CH_3 (DMT)), 52.47 (O-CH_3 (DMT)), 49.81 (C-2′′), 46.07 (C-1′′), 31.16 (C-2′), 24.25 (N-C(=O)-$\underline{C}H_3$).

5′-*O*-(4,4′-Dimethoxytrityl)-*N*^{6}-acetyl-*N*^{6}-(4-azidobutane)-2′-deoxyadenosine (7b)



Starting material: 0.36 g, 0.92 mmol

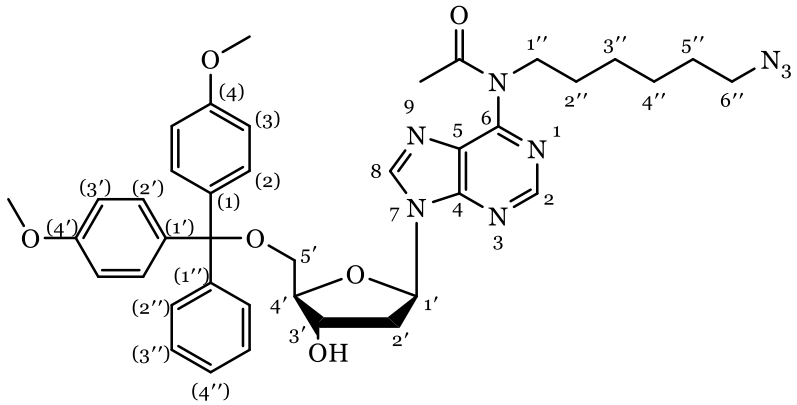
Yield: 0.52 g (81%).

^{1}H NMR (500 MHz, d_6-DMSO) δ = 8.78 (1H, s, H-8), 8.74 (1H, s, H-2), 7.33 – 7.29 (2H, m, DMT aromatic C$\underline{H}$), 7.22 – 7.15 (7H, m, DMT aromatic C$\underline{H}$ (4H + 2H +1H)), 6.80 (4H, t, J = 8.21 Hz, DMT aromatic C$\underline{H}$), 6.52 (1H, t, J = 6.65 Hz, H-1′), 5.45 (1H, br, 3′-OH), 4.56 (1H, dt, J = 10.44 Hz, 5.15 Hz, H-3′), 4.09 (2H, t, J = 6.65 Hz, H-1′′), 4.06 – 4.01 (1H, m, H-4′), 3.72 (6H, s, DMT-C$\underline{H}_3$), 3.21 – 3.14 (4H, m, H-5′/H-4′′), 2.96 (1H, m, 2_a′), 2.43 (1H, m, 2_b′), 2.06 (3H, s, N-C(=O)-C$\underline{H}_3$), 1.52 – 1.37 (4H, m, H-2′′/3′′).

^{13}C NMR (125 MHz, d_6-DMSO) δ = 170.85 (O=$\underline{C}$-CH_3), 158.47 (DMT-q), 152.97 (C-6), 152.08 (C-8), 145.29 (DMT-q), 145.19 (C-2), 136.06 (DMT-q), 135.95 (DMT-q), 130.12 (DMT $\underline{C}$H), 128.12 (DMT $\underline{C}$H), 127.68/127.03 (C-4/5), 113.50 (DMT-$\underline{C}$H), 86.87 (DMT-q), 85.85 (C-4′), 84.32 (C-1′), 70.82 (C-3′), 64.37 (C-5′), 62.49, 55.44 (O-CH_3 (DMT)), 55.36 (O-CH_3 (DMT)), 50.65 (C-4′′), 46.26 (C-1′′), 38.87 (C-2′), 25.95/25.76 (C-2′′/3′′), 24.18 (N-C(=O)-$\underline{C}H_3$).

ESI-MS (MeOH, positive): Calculated: 693.31 [M+H$^+$], 714.30 [M+Na$^+$]; Experimental: 693.22 [M+H$^+$], 714.61 [M+Na$^+$].

5′-*O*-(4,4′-Dimethoxytrityl)-*N*^{6}-acetyl-*N*^{6}-(6-azidohexane)-2′-deoxyadenosine (7c)



Starting material: 0.14 g, 0.33 mmol

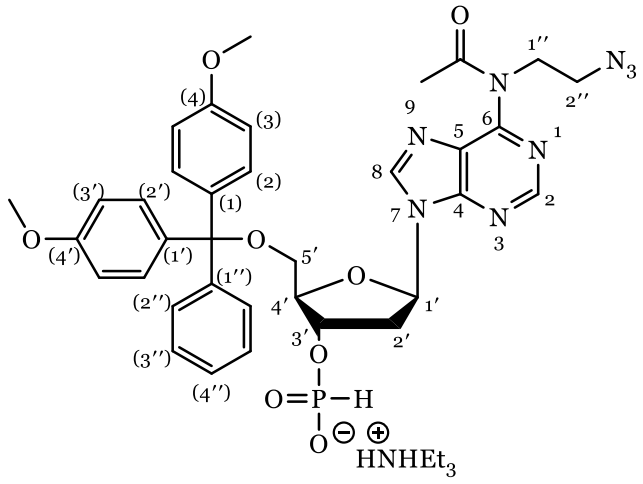
Yield: 0.18 g (77%).

^{1}H NMR (500 MHz, d$_6$-DMSO) δ = 8.77 (1H, s, H-8), 8.74 (1H, s, H-2), 7.32 – 7.28 (2H, m, DMT aromatic C<u>H</u>), 7.22 – 7.13 (7H, m, DMT aromatic C<u>H</u> (4H + 2H +1H)), 6.79 (4H, t, J = 8.71 Hz, DMT aromatic C<u>H</u>), 6.51 (1H, t, 6.30 Hz, H-1′), 5.44 (1H, br, 3′-OH), 4.56 (1H, dt, J = 5.54 Hz, 5.05 Hz, H-3′), 4.09 – 4.00 (4H, m, H-5′/H-1′′), 3.72 (6H, s, DMT O-C<u>H</u>$_3$), 3.21 – 3.15 (3H, m, H-4′/H-6′′), 2.98 (1H, m, 2$_a$′) 2.46 – 2.39 (1H, m, 2$_b$′), 2.05 (3H, s, N-C(=O)-C<u>H</u>$_3$), 1.46 – 1.33 (4H, m, H-2′′/5′′), 1.19 – 1.11 (4H, m, H-3′′/4′′).

^{13}C NMR (125 MHz, d$_6$-DMSO) δ = 170.73 (O=<u>C</u>-CH$_3$), 158.46, 153.12 (C-6), 152.94 (C-4), 152.05 (C-8), 145.28 (DMT-q), 145.09 (C-2), 136.08 (DMT-q), 135.95 (DMT-q), 130.12 (DMT <u>C</u>H), 130.09 (DMT <u>C</u>H), 128.11 (DMT-<u>C</u>H), 127.60 (C-5), 127.01 (DMT <u>C</u>H), 113.50 (DMT <u>C</u>H), 86.47 (C-4′), 85.85 (DMT-q), 84.30 (C-1′), 70.83 (C-3′), 64.73 (C-5′), 55.44 (O-CH$_3$ (DMT)), 55.43 (O-CH$_3$ (DMT)), 50.91 (C-6′′), 46.72 (C-1′′), 39.10 (C-2′), 28.47/28.35 (C-2′′/5′′), 26.07/26.01 (C-3′′/4′′), 24.23 (N-C(=O)-<u>C</u>H$_3$).

ESI-MS (MeOH, positive): calculated: 721.34 [M+H$^+$], 743.32 [M+Na$^+$]; Experimental: 743.07 [M+Na$^+$].

5′-*O*-(4,4′-Dimethoxytrityl)-*N*^{6}-acetyl-*N*^{6}-(2-azidoethane)-2′-deoxyadenosine 3′-H-Phosphonate TEA salt (8a)



Compound **7a** (0.29 g, 0.44 mmol, 1 eq) was dissolved in 5 mL of pyridine and diphenyl H-phosphonate (0.71 g, 3.1 mmol, 7 eq) was added. The mixture was stirred at room temperature for 15 mins, monitored by TLC (9:1 Chloroform/methanol). When TLC analysis indicated reaction was complete, 1 mL of water and 1 mL of TEA were added, and the mixture was left for 15 mins. Volatile solvent was evaporated under reduced pressure. 20 mL of DCM was added, and it was extracted three times with 20 mL of sat. $NaHCO_3$. Solution was dried over anhydrous sodium sulfate, filtered, and evaporated under reduced pressure. Product was purified by silica gel column chromatography (0 – 12% MeOH/DCM + 1% TEA) and obtained as a white foam. Yield: 0.31 g (85%).

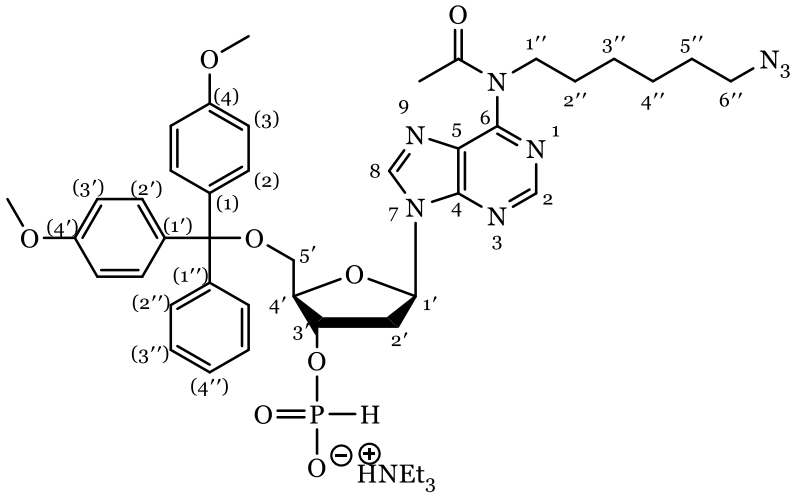
NMR spectra of this product indicated the presence of a nucleotide, but characterisation was not possible due to the TEA salt present. Formation of the correct product was confirmed by mass spectrometry.

^{1}H NMR (500 MHz, d_6-DMSO) δ = 8.73 (1H, s, H-8), 8.71 (1H, s, H-2), 7.36 – 7.32 (2H, m, DMT aromatic C<u>H</u>), 7.24 – 7.15 (7H, m, DMT aromatic C<u>H</u> (4H + 2H +1H)), 6.84 – 6.78 (4H, m, DMT aromatic C<u>H</u>), 6.50 (1H, t, J = 6.88 Hz, H-1′), 4.86 – 4.70 (1H, m, H-3′), 4.29 (2H, t, J = 6.05 Hz, H-1′′), 4.24 – 4.19 (1H, m, H-4′), 3.73 (6H, s, 2x O-CH_3), 3.52 (2H, t, J = 6.17 Hz, H-2′′) 3.23 – 3.15 (approx. 2H (conflicts with HOD peak in DMSO), m, H-5′), 3.01 (1H, quintet, J = 6.78 Hz, 2_a′), 2.58 – 2.54 (1H, m, 2_b′), 2.12 (3H, s, N-C(=O)-C<u>H</u>$_3$).

^{31}P NMR (125 MHz, $CDCl_3$) δ = -0.03108 (H-phosphonate).

ESI-MS (MeOH, positive): Calculated: 727.24 [M+H⁺]. Experimental: 685.09 [M – N⁶-Acetyl], 727.09 [M+H⁺].

5′-*O*-(4,4′-Dimethoxytrityl)-*N*⁶-acetyl-*N*⁶-(6-azidohexane)-2′-deoxyadenosine 3′-H-Phosphonate TEA salt (8c)



A quantity of this material, sufficient for NMR, was obtained using the same protocol as compound **8a**. However, when scaled up no product was obtained.

¹H NMR (500 MHz, d₆-DMSO) δ = 8.73 (1H, s, H-8), 8.71 (1H, s, H-2), 7.36 – 7.30 (2H, m, DMT aromatic C<u>H</u>), 7.20 – 7.13 (7H, m, DMT aromatic C<u>H</u> (4H + 2H +1H)), 6.83 – 6.76 (4H, m, DMT aromatic C<u>H</u>), 6.51 (1H, t, J = 6.65 Hz, H-1′), 4.91 – 4.83 (1H, m, H-3′), 4.25 – 4.19 (1H, m, H-4′), 4.07 (2H, t, J = 7.42 Hz, H-1′′), 3.73 (6H, s, 2x O-CH_3), 3.29 – 3.17 (4H, m, H-5′/6′′), 3.01 – 3.04 (1H, m, H-2_a′, 2.61 – 2.54 (1H*, m, H-2_b′), 2.45 – 2.36 (1H, m, 2_b′), 2.09 (3H, s, N-C(=O)-C<u>H</u>$_3$), 1.50 – 1.42 (2H, m, H-2′′), 1.42 – 1.36 (2H, m, H-5′′), 1.27 – 1.20 (4H*, m, H-3′′/4′′).

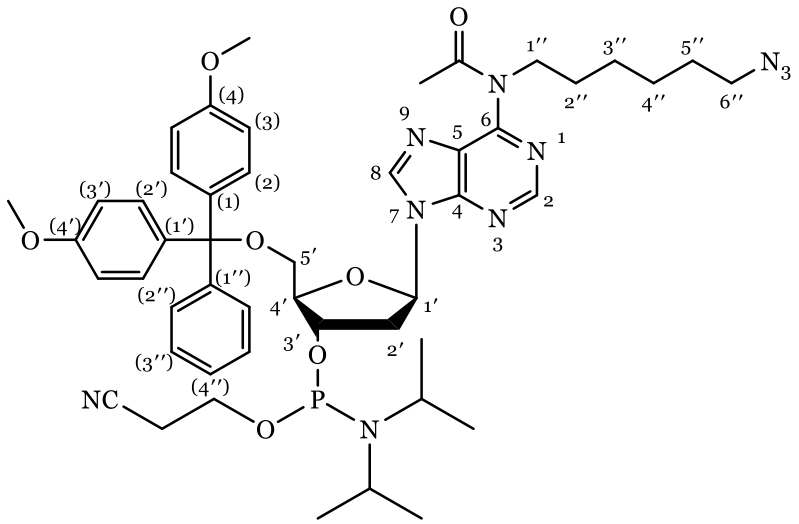
*These peaks are partially covered by peaks corresponding to the TEA salt

¹³C NMR (125 MHz, d₆-DMSO) δ = 170.76 (O=<u>C</u>-CH_3), 158.48 (DMT-q), 152.97 (C-6), 152.03 (C-8), 145.26 (DMT-q), 145.99 (C-2), 136.03 (DMT-q), 135.92 (DMT-q), 130.14 (DMT <u>C</u>H), 128.13 (DMT <u>C</u>H), 127.64 (C-4) , 127.04 (C-5), 113.52 (DMT-<u>C</u>H), 85.94 (DMT-q), 85.75 (C-4′), 84.68 (C-1′), 73.02 (C-3′), 64.27 (C-5′), 55.44 (O-CH_3 (DMT)), 50.91 (O-CH_3 (DMT)), 46.74 (C-6′′), 45.70 (C-1′′), 34.95 (C-2′), 28.48/28.37 (C-2′′/5′′), 26.09/26.03 (C-3′′/4′′), 24.27 (N-C(=O)-<u>C</u>H_3).

³¹P NMR (125 MHz, $CDCl_3$) δ = -0.0907 (H-phosphonate).

ESI-MS (MeOH, positive): Calculated: 783.30 [M+H$^+$]. Experimental: 783.09 [M+H$^+$].

5′-*O*-(4,4′-Dimethoxytrityl)-*N*^{6}-acetyl-*N*^{6}-(6-azidohexane)-3′-*O*-([2-cyanoethyl],[*N*,*N*-di*iso*propylamine] phosphoryl)-2′-deoxyadenosine (9c)



Compound **7c** (0.325 g, 0.45 mmol, 1 eq) was dissolved in 10 mL of freshly dried DCM under argon. TEA (82 µL, 0.59 mmol, 1.3 eq) followed by 2-cyanoethyl *N*,*N*-diisopropylchlorophosphoramidite (117 µL, 0.52 mmol, 1.15 eq) were added and the mixture was stirred at room temperature. The reaction was monitored by TLC (10% MeOH/DCM) until starting material disappeared (approximately 15 mins). The reaction mixture was washed with saturated $NaHCO_3$ solution (3 × 5 mL), followed by brine (2 × 5 mL). The organic layer was then passed through a column filled with anhydrous sodium sulfate. The solution was then cooled in a freezer to freeze any residual water and used without further purification in DNA synthesis. The solution was poured over activated molecular sieves approximately 30 mins before coupling.

The product was characterised by carrying out a small-scale reaction in freshly dried $CDCl_3$ instead of DCM and using the crude solution for ^{1}H, ^{13}C and ^{31}P NMR. This results in an NMR with multiple extra peaks due to the presence of impurities and diastereomers. The assignment below is the most accurate assignment possible of the resulting spectra, with TLC and ^{31}P NMR being the most definitive method of characterising the product.

^{1}H NMR (500 MHz, $CDCl_3$) δ = 8.69 (1H, s, H-8), 8.22 (1H, s, H-2), 7.38 – 7.31 (2H, m, DMT aromatic C<u>H</u>), 7.27 – 7.11 (7H, m, DMT aromatic C<u>H</u> (4H + 2H +1H)), 6.75 (4H, t, J = 6.38 Hz, DMT aromatic C<u>H</u>), 6.52 – 6.46 (1H, m, H-1′), 4.86 – 4.74 (1H, m, H-3′), 3.34 – 4.24 (1H, m, H-4′), 4.22 – 4.03 (4H, m, H-5′/H-1′′), 3.73 (6H, s, DMT O-C<u>H</u>$_3$), 3.65 – 3.25 (6H, m, CE-C<u>H</u>$_2$), 3.10 (2H, t, 6.95 Hz, H-6′′), 2.75 – 2.67 (2H, m, *iso*propyl C<u>H</u>), 2.52 – 2.44 (12H, m, *iso*propyl C<u>H</u>$_3$), 2.17 – 2.11 (3H, m, N-C(=O)-C<u>H</u>$_3$), 1.56 – 1.40 (4H, m, H-2′′/5′′), 1.12 – 1.06 (4H, m, H-3′′/4′′).

^{13}C NMR (125 MHz, d_6-DMSO) δ = 171.22 (O=<u>C</u>-CH_3), 154.23/153.49 (C-6) 152.72/152.65 (C-4), 151.91/149.42 (C-8), 144.46 (C-2), 142.55 (DMT-q), 135.55 (DMT-q), 130.04 (DMT-<u>C</u>H), 130.01 (DMT-<u>C</u>H), 129.99 (DMT-<u>C</u>H), 128.07 (DMT-<u>C</u>H), 128.03 (DMT-<u>C</u>H), 127.79 (DMT-<u>C</u>H), 127.66/127.61 (C-5), 126.87 (DMT-<u>C</u>H), 126.83 (DMT-<u>C</u>H), 117.60/117.45 (DMT-<u>C</u>H), 113.09 (DMT-<u>C</u>H), 106.53 (C-1′), 86.45 (DMT-q), 85.96/85.78 (C-4′), 84.73/84.66 (DMT-<u>C</u>H), 73.98/73.84 (C-3′), 58.36, 58.22 (C-5′), 55.15/55.13 (O-CH_3 (DMT)), 51.21 (C-6′′), 47.13, 46.12 (C-1′′), 45.29, 45.24, 43.29, 43.18, 39.25/38.93 (C-2′), 28.54/28.33 (C-2′′/5′′), 26.18/26.10 (C-3′′/4′′), 24.12 (N-C(=O)-<u>C</u>H_3).

^{31}P NMR (125 MHz, $CDCl_3$) δ = 148.77 (phosphoramidite diastereomer$_a$), 148.75 (phosphoramidite diastereomer$_b$), 14.12 (H-phosphonate).

7.4. Synthesis of DNA

7.4.1. Hydrophobic Modification of Phosphates

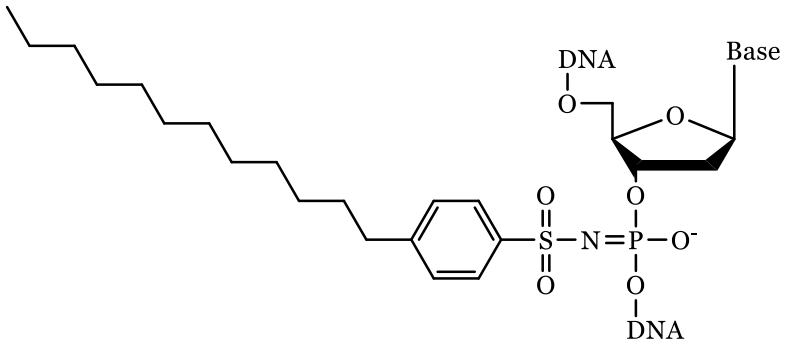


Figure 7.5. Deprotected 4′-dodecylsulfonyl phosphate modification.

For 5 μmol synthesis, 5 μmol of appropriate preloaded CPG (e.g., 64.1 mg of 78 μmol/g CPG for dT CPG) was loaded into the column and modified sequences were synthesised using standard DNA synthesis protocol, as described in the Chapter 7.2.1. For each desired modification (Table 7.2, Table 3.1 for full sequences), the oxidation step was replaced by thiolation in the programme. The standard thiolation solution was replaced with a 0.2 mol L^{-1} 4-dodecylbenzenesulfonyl azide solution in ACN, which was injected three times (700 μL) with a coupling time of 6 mins between each injection. This resulted in the cyanoethyl-protected, modified pentavalent phosphorus. Cyanoethyl deprotection occurred post-synthetically when DNA was deprotected and cleaved from the solid support with 28% aq. NH_4OH, resulting in the modified phosphate shown in Figure 7.5. The oligonucleotides were purified by RP-HPLC (0.1 M TEAA Buffer/ACN, see Chapter 7.2.2) and analysed with ESI-MS to ensure the correct species were present.

Table 7.2. 4-Dodecylbenzenesulfonyl azide modified TG_4T sequences

Sequence	RP-HPLC R_t (min)	Expected M_w (g mol^{-1})	Experimental M_w (g mol^{-1})
TXG_4T	24.46	2155	2155.500
TG_2XG_2T	24.27	2155	2155.527
TG_4XT	24.48	2155	2155.505
$G_4T_4G_3XG$	24.36	4094	4093.840
$GXG_3T_4G_4$	24.28	4094	4093.844
$GXG_3T_4G_3XG$	24.33	4401	4401.043

X indicates position of phosphate modification between bases.

7.4.2. Synthesis of Sequences Containing Inverted- and α-Nucleotides

For 5 μmol synthesis, 5 μmol of appropriate preloaded CPG was loaded into a column and sequences were synthesised using standard DNA protocols, with the following exceptions: Sequences in Table 7.3 (see Table 4.1, Table 4.3Table 4.3 and Table 4.6 for full sequences) used increased coupling times for modified phosphoramidites only (5 minutes for all modified phosphoramidites). Sequences in Table 7.4 (see Table 4.6 for full sequences) further increased coupling times of modified phosphoramidites, as well as increased coupling times of standard phosphoramidites. This was due to a noticeable decrease in trityl levels following coupling of different types of nucleotides (i.e., α- with β- or inverted- with standard-nucleotides), and it was necessary to increase coupling to maximise yield at inversion sites. Sequences with only G-tetrad modifications, had dG coupling time of 10 minutes. For these sequences, A, C and T used standard coupling times. Sequences with loop modifications, had coupling times of 10 mins for all modified phosphoramidites (i.e., A, C and T) and dG coupling time (i.e., unmodified bases) of 5 minutes. This was necessary due to the large number of dG nucleotides in each sequence.

Oligonucleotides were deprotected and cleaved in 28% aq. ammonia and purified by RP-HPLC.

Table 7.3. Modified sequences synthesised using 5-minute coupling times for modified bases only

Sequence	RP-HPLC r_t (min)	Expected M_w ($gmol^{-1}$)	Experimental M_w ($gmol^{-1}$)
5′inv-Ttel	14.19	8154	8246 (4x Na^+ adduct)
5′inv-Ttel-5′-tail	13.94	12514	12514.865
5′inv-Ttel-3′-tail	14.47	12157	12156.940
c-KIT-inv	18.95	5998	5998.000

Table 7.4. Modified sequences synthesised using 10-minute coupling times for modified bases and 5-minute coupling time for dG.

Sequence	RP-HPLC r_t (min)	Expected M_w ($gmol^{-1}$)	Experimental M_w ($gmol^{-1}$)
c-KIT-inv-loops	12.00	6327.9	6327. 1
c-KIT-α-loops	12.44	6327.9	6327.1
c-KIT-inv-dG	13.45	6632.0	6631.1
c-KIT-α-dG	24.10	6327.9	6327.1
c-MYC-inv-loops	11.58	5112.8	5111.9
c-MYC-α-loops	12.47	5112.8	5111.9
c-MYC-inv-dG	11.32	5416.8	5415.9
c-MYC-α-dG	24.17	5112.8	5111.9

7.4.3. Synthesis of Azide- and Alkyne-Containing Oligonucleotides

For 5 μmol synthesis, 64.1 mg (78 μmol/g) of preloaded dT CPG was loaded into the column. Modified sequences (Table 7.5) were synthesised using standard DNA synthesis protocol, as described in the methodology section. Modified bases were added through two methods:

(i) H-phosphonates (0.1 M) were dissolved in a dry pyridine/ACN (30/70) and attached to an unused phosphoramidite port. Pivaloyl chloride (0.5 M) was dissolved in dry ACN and attached to a different unused phosphoramidite port. The solutions were mixed in the column using a wobble base protocol ensure adequate mixing of the H-phosphonate with the activator. Coupling time was the same as for standard phosphoramidites. Oxidation time for this step was doubled to ensure the H-phosphonate was fully oxidised before subsequent additions were carried out.

Table 7.5: Telomeric sequences synthesised for cross-linking.

Name	Sequence	RP-HPLC R_t (min)	Expected M_w ($gmol^{-1}$)	Experimental M_w ($gmol^{-1}$)
tel-G3A8	TAYGGTTZGGGT	15.77	3877	3877.69
tel-G3A8-X	TAYGGTTZGGGT	14.55	3877	3877.68
tel-G5A8	TAGGYTTZGGGT	15.02	3877	3877.70
Tel-G5A8-X	TAGGYTTZGGGT	13.94	3877	3877.64

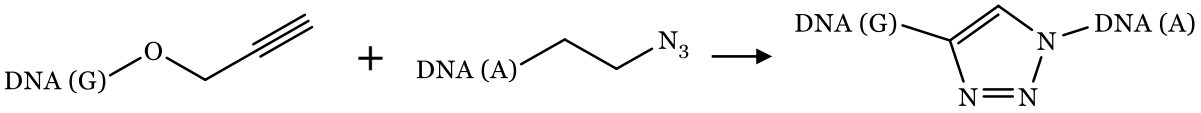
Y indicates position of 2′-O-propargyl guanosine and Z indicates position of N^6-ethylazide-2′-deoxyadenosine. X- indicates that Cu(I)-catalysed azide-alkyne cycloaddition has been performed.

(ii) 5′-DMT-2′-*O*-Propargyl guanosine 3′-*O*-phosphoramidite was added through semi-manual addition. After detritylation, 70 mg of this

phosphoramidite was added directly to the column. 400 µL of the standard activator solution was added and mixed with a pipette to dissolve phosphoramidite, forming a 0.2 M solution. This solution was passed through the column using the same methods as standard coupling, except coupling times were increased to 10 minutes.

Modified sequences were deprotected and cleaved from the solid support by 28% aq. NH_4OH at 25 °C for 24 hours. The solution was filtered and freeze dried before purification using RP-HPLC (0.1 M TEAA Buffer/ACN). Collected fractions were freeze dried again and desalted using NAP-5 columns. Standard oligonucleotides were cleaved at 55 °C and concentrated at 45 °C under reduced pressure (speed vac) but modified samples were deprotected at 30 °C for 24 hours and freeze dried to prevent formation of undesirable 1,5-triazole linkages.

7.4.4. Copper-catalysed Azide-Alkyne Cycloaddition (CuAAC)



Reagents were mixed according to Table 7.6, based on a protocol originally obtained from Lumiprobe Ltd. (Hong Kong).[149] Oligonucleotide was dissolved in 50 µL of Milli-Q water and concentration was measured using UV-Vis spectroscopy (7508.1 µM). OD_{260} is the amount of light absorbed by the oligonucleotide when it is suspended in 1 mL of water and measured in a 1 cm cuvette. The value can be calculated by multiplying initial concentration, molar absorptivity, and initial volume of oligonucleotide (e.g., 7508.1 µM × 122800 L mol^{-1} cm^{-1} × 25 µL gives an OD_{260} of 23.05). DNA was then diluted to obtain the reaction concentration (e.g., 25 µL DNA with 125 µL Milli-Q water = 150 µL oligonucleotide volume) and the corresponding volume of 2 M TEAA Buffer (250 µL) was added. The solution of 10 mM aq. sodium ascorbate was freshly prepared and added to oligonucleotide (25 µL). Reaction mixture was flushed with argon before addition of combined $CuSO_4$ and tris-hydroxypropyltriazolylmethylamine (THPTA) reagent (50 µl).

Every two hours, 20 µL of reaction mixture was assessed by analytical RP-HPLC using the same programme as previous preparative RP-HPLC. A new peak

with shorter retention time was observed. After 4 hours no starting material was observed and the reaction mixture was purified by preparative RP-HPLC, collected fractions were concentrated by speed-vac and desalted using Nap-5 size exclusion columns.

Table 7.6: Reagent volume (µL) required for various oligonucleotide OD_{260} values

		Oligo OD_{260}						
		< 5	**5 – 10**	**10 – 15**	**15 – 20**	**20 – 25**	**25 – 30**	**30+**
Reagent Volume (µL)	**Oligo**	30	60	90	120	150	180	+ 30 µL
	2M TEAA Buffer	50	100	150	200	250	300	+ 50 µL
	$CuSO_4$[a]	5	10	15	20	25	30	+ 5 µL
	THPTA[a]	5	10	15	20	25	30	+ 5 µL
	10 mM Sodium Ascorbate	5	10	15	20	25	30	+5 µL

(a) THPTA and $CuSO_4$ are added as a combined activator solution containing a 10 mM $CuSO_4$/THPTA complex.

References

1. Levene, P. A.; Jacobs, W. A. Über Inosinsäure. *Berichte der deutschen chemischen Gesellschaft* **2006**, *42* (1), 1198-1203.

2. Levene, P. A. The Structure of Yeast Nucleic Acid. *J Biol Chem* **1919**, *40* (2), 415-424.

3. Griffith, F. The Significance of Pneumococcal Types. *J Hyg (Lond)* **1928**, *27* (2), 113-159.

4. Avery, O. T.; Macleod, C. M.; McCarty, M. Studies on the Chemical Nature of the Substance Inducing Transformation of Pneumococcal Types : Induction of Transformation by a Desoxyribonucleic Acid Fraction Isolated from Pneumococcus Type Iii. *J Exp Med* **1944**, *79* (2), 137-158.

5. Chargaff, E. Chemical Specificity of Nucleic Acids and Mechanism of their Enzymatic Degradation. *Experientia* **1950**, *6*, 201-209.

6. Franklin, R. E.; Gosling, R. G. Molecular Configuration in Sodium Thymonucleate. *Nature* **1953**, *171* (4356), 740-741.

7. Watson, J. D.; Crick, F. H. C. Molecular Structure of Nucleic Acids - a Structure for Deoxyribose Nucleic Acid. *Nature* **1953**, *171* (4356), 737-738.

8. Richmond, T. J.; Davey, C. A. The Structure of DNA in the Nucleosome Core. *Nature* **2003**, *423* (6936), 145-150.

9. Volk, D. E.; Yang, X.; Fennewald, S. M.; King, D. J.; Bassett, S. E.; Venkitachalam, S.; Herzog, N.; Luxon, B. A.; Gorenstein, D. G. Solution Structure and Design of Dithiophosphate Backbone Aptamers Targeting Transcription Factor NF-Kappab. *Bioorg Chem* **2002**, *30* (6), 396-419.

10. Gao, Y. G.; Robinson, H.; Wang, A. H. High-Resolution α-DNA Crystal Structures of d(AGGGGCCCCT). An α-DNA Model of Poly(dG) x Poly(dC). *Eur J Biochem* **1999**, *261* (2), 413-420.

11. Drozdzal, P.; Gilski, M.; Jaskolski, M. Crystal Structure of Z-DNA in Complex with the Polyamine Putrescine and Potassium Cations at Ultra-High Resolution. *Acta Crystallogr B Struct Sci Cryst Eng Mater* **2021**, *77* (Pt 3), 331-338.

12. Whelan, D. R.; Hiscox, T. J.; Rood, J. I.; Bambery, K. R.; McNaughton, D.; Wood, B. R. Detection of an en Masse and Reversible β- To α-Dna Conformational Transition in Prokaryotes in Response to Desiccation. *J R Soc Interface* **2014**, *11* (97), 20140454.

13. Doluca, O.; Withers, J. M.; Filichev, V. V. Molecular Engineering of Guanine-rich Sequences: Z-DNA, DNA triplexes, and G-quadruplexes. *Chem. Rev.* **2013**, *113* (5), 3044-3083.

14. Wittig, B.; Dorbic, T.; Rich, A. Transcription is Associated with Z-DNA Formation in Metabolically Active Permeabilized Mammalian Cell Nuclei. *Proc. Natl. Acad. Sci.* **1991**, *88* (6), 2259-2263.

15. Champ, P. C.; Maurice, S.; Vargason, J. M.; Camp, T.; Ho, P. S. Distributions of Z-DNA and Nuclear Factor I in Human Chromosome 22: a Model for Coupled Transcriptional Regulation. *Nucleic Acids Res* **2004**, *32* (22), 6501-6510.

16. Ghosh, A.; Bansal, M. A Glossary of DNA structures from A to Z. *Acta Crystallogr D Biol Crystallogr* **2003**, *59* (Pt 4), 620-626.

17. Bansal, M. DNA Structure: Revisiting the Watson–Crick Double Helix. *Current Science* **2003**, *85* (11), 1556-1563.

18. Parkinson, G. N.; Lee, M. P. H.; Neidle, S. Crystal Structure of Parallel Quadruplexes from Human Telomeric DNA. *Nature* **2002**, *417* (6891), 876-880.

19. Harrell, W. A.; Neidle, S.; Balasubramanian, S. *Quadruplex Nucleic Acids*; RSC Pub., 2006.

20. Choi, J.; Majima, T. Conformational Changes of non-B DNA. *Chem Soc Rev* **2011**, *40* (12), 5893-5909.

21. Huppert, J. L.; Balasubramanian, S. Prevalence of Quadruplexes in the Human Genome. *Nucleic Acids Res* **2005**, *33* (9), 2908-2916.

22. Kikin, O.; D'Antonio, L.; Bagga, P. S. QGRS Mapper: a Web-Based Server for Predicting G-Quadruplexes in Nucleotide Sequences. *Nucleic Acids Res* **2006**, *34* (Web Server issue), W676-682.

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

24. Chambers, V. S.; Marsico, G.; Boutell, J. M.; Di Antonio, M.; Smith, G. P.; Balasubramanian, S. High-Throughput Sequencing of DNA G-Quadruplex Structures in the Human Genome. *Nat Biotechnol* **2015**, *33* (8), 877-881.

25. Jiang, H.; Ju, Z.; Rudolph, K. L. Telomere Shortening and Ageing. *Z Gerontol Geriatr* **2007**, *40* (5), 314-324.

26. Runge, K. W.; Zakian, V. A. Introduction of Extra Telomeric DNA Sequences into Saccharomyces Cerevisiae Results in Telomere Elongation. *Mol Cell Biol* **1989**, *9* (4), 1488-1497.

27. Zhou, J.; Monson, E. K.; Teng, S. C.; Schulz, V. P.; Zakian, V. A. Pif1p Helicase, a Catalytic Inhibitor of Telomerase in Yeast. *Science* **2000**, *289* (5480), 771-774.

28. Dexheimer, T. S.; Sun, D.; Hurley, L. H. Deconvoluting the Structural and Drug-Recognition Complexity of the G-Quadruplex-Forming Region Upstream of the Bcl-2 P1 promoter. *J Am Chem Soc* **2006**, *128* (16), 5404-5415.

29. Shirude, P. S.; Okumus, B.; Ying, L. M.; Ha, T.; Balasubramanian, S. Single-Molecule Conformational Analysis of G-Quadruplex Formation in the Promoter

DNA Duplex of the Proto-Oncogene C-Kit. *J Am Chem Soc* **2007**, *129* (24), 7484-7485.

30. Seenisamy, J.; Rezler, E. M.; Powell, T. J.; Tye, D.; Gokhale, V.; Joshi, C. S.; Siddiqui-Jain, A.; Hurley, L. H. The Dynamic Character of the G-quadruplex Element in the c-MYC Promoter and Modification by TMPyP4. *J Am Chem Soc* **2004**, *126* (28), 8702-8709.

31. Cogoi, S.; Paramasivam, M.; Spolaore, B.; Xodo, L. E. Structural Polymorphism within a Regulatory Element of the Human KRAS Promoter: Formation of G4-DNA Recognized by Nuclear Proteins. *Nucleic Acids Res* **2008**, *36* (11), 3765-3780.

32. Cogoi, S.; Xodo, L. E. G-Quadruplex Formation within the Promoter of the KRAS Proto-Oncogene and its Effect on Transcription. *Nucleic Acids Res* **2006**, *34* (9), 2536-2549.

33. Paramasivam, M.; Membrino, A.; Cogoi, S.; Fukuda, H.; Nakagama, H.; Xodo, L. E. Protein hnRNP A1 and its Derivative Up1 Unfold Quadruplex DNA in the Human KRAS Promoter: Implications for Transcription. *Nucleic Acids Res* **2009**, *37* (9), 2841-2853.

34. Balasubramanian, S.; Hurley, L. H.; Neidle, S. Targeting G-Quadruplexes in Gene Promoters: a Novel Anticancer Strategy? *Nat Rev Drug Discov* **2011**, *10* (4), 261-275.

35. Cree, S. L.; Fredericks, R.; Miller, A.; Pearce, F. G.; Filichev, V.; Fee, C.; Kennedy, M. A. DNA G-quadruplexes Show Strong Interaction with DNA Methyltransferases in Vitro. *Febs Lett* **2016**, *590* (17), 2870-2883.

36. Mao, S. Q.; Ghanbarian, A. T.; Spiegel, J.; Cuesta, S. M.; Beraldi, D.; Di Antonio, M.; Marsico, G.; Hansel-Hertsch, R.; Tannahill, D.; Balasubramanian, S. DNA G-quadruplex Structures Mold the DNA Methylome. *Nat Struct Mol Biol* **2018**, *25* (10), 951-957.

37. Hou, X. M.; Wu, W. Q.; Duan, X. L.; Liu, N. N.; Li, H. H.; Fu, J.; Dou, S. X.; Li, M.; Xi, X. G. Molecular Mechanism of G-Quadruplex Unwinding Helicase: Sequential and Repetitive Unfolding of G-Quadruplex by Pif1 Helicase. *Biochem J* **2015**, *466*, 189-199.

38. Roach, R. J.; Garavis, M.; Gonzalez, C.; Jameson, G. B.; Filichev, V. V.; Hale, T. K. Heterochromatin Protein 1α Interacts with Parallel RNA and DNA G-Quadruplexes. *Nucleic Acids Res* **2020**, *48* (2), 682-693.

39. Biffi, G.; Tannahill, D.; Miller, J.; Howat, W. J.; Balasubramanian, S. Elevated Levels of G-quadruplex Formation in Human Stomach and Liver Cancer Tissues. *PLoS One* **2014**, *9* (7), e102711.

40. Asensio, J. L.; Brown, T.; Lane, A. N. Solution Conformation of a Parallel DNA Triple Helix with 5' and 3' Triplex-Duplex Junctions. *Structure* **1999**, *7* (1), 1-11.

41. Wang, G.; Vasquez, K. M. Impact of Alternative DNA Structures on DNA Damage, DNA Repair, and Genetic Instability. *DNA Repair (Amst)* **2014**, *19*, 143-151.

42. Wang, G.; Vasquez, K. M. Effects of Replication and Transcription on DNA Structure-Related Genetic Instability. *Genes (Basel)* **2017**, *8* (1).

43. Snoussi, K.; Nonin-Lecomte, S.; Leroy, J. L. The RNA i-Motif. *J Mol Biol* **2001**, *309* (1), 139-153.

44. Benabou, S.; Aviñó, A.; Eritja, R.; González, C.; Gargallo, R. Fundamental Aspects of the Nucleic Acid i-Motif Structures. *RSC Adv.* **2014**, *4* (51), 26956-26980.

45. Wright, E. P.; Huppert, J. L.; Waller, Z. A. E. Identification of Multiple Genomic DNA Sequences which Form i-Motif Structures at Neutral pH. *Nucleic Acids Res* **2017**, *45* (22).

46. Phan, A. T.; Mergny, J. L. Human Telomeric DNA: G-quadruplex, i-motif and Watson-Crick Double Helix. *Nucleic Acids Res* **2002**, *30* (21), 4618-4625.

47. Li, W.; Wu, P.; Ohmichi, T.; Sugimoto, N. Characterization and Thermodynamic Properties of Quadruplex/Duplex Competition. *Febs Lett* **2002**, *526* (1-3), 77-81.

48. Miyoshi, D.; Fujimoto, T.; Sugimoto, N. Molecular Crowding and Hydration Regulating of G-Quadruplex Formation. *Top Curr Chem* **2013**, *330*, 87-110.

49. Cui, Y.; Kong, D.; Ghimire, C.; Xu, C.; Mao, H. Mutually Exclusive Formation of G-Quadruplex and i-Motif Is a General Phenomenon Governed by Steric Hindrance in Duplex DNA. *Biochemistry* **2016**, *55* (15), 2291-2299.

50. King, J. J.; Irving, K. L.; Evans, C. W.; Chikhale, R. V.; Becker, R.; Morris, C. J.; Pena Martinez, C. D.; Schofield, P.; Christ, D.; Hurley, L. H.; et al. DNA G-Quadruplex and i-Motif Structure Formation Is Interdependent in Human Cells. *J Am Chem Soc* **2020**, *142* (49), 20600-20604.

51. Brown, R. V.; Danford, F. L.; Gokhale, V.; Hurley, L. H.; Brooks, T. A. Demonstration that Drug-Targeted Down-Regulation of MYC in non-Hodgkins Lymphoma is Directly Mediated through the Promoter G-Quadruplex. *J Biol Chem* **2011**, *286* (47), 41018-41027.

52. Janssen, A.; Colmenares, S. U.; Karpen, G. H. Heterochromatin: Guardian of the Genome. *Annu Rev Cell Dev Biol* **2018**, *34*, 265-288.

53. Lander, E. S.; Linton, L. M.; Birren, B.; Nusbaum, C.; Zody, M. C.; Baldwin, J.; Devon, K.; Dewar, K.; Doyle, M.; FitzHugh, W.; et al. Initial Sequencing and Analysis of the Human Genome. *Nature* **2001**, *409* (6822), 860-921.

54. Vicient, C. M.; Casacuberta, J. M. Impact of Transposable Elements on Polyploid Plant Genomes. *Ann Bot* **2017**, *120* (2), 195-207.

55. Jenuwein, T.; Allis, C. D. Translating the Histone Code. *Science* **2001**, *293* (5532), 1074-1080.

56. Bannister, A. J.; Zegerman, P.; Partridge, J. F.; Miska, E. A.; Thomas, J. O.; Allshire, R. C.; Kouzarides, T. Selective Recognition of Methylated Lysine 9 on Histone H3 by the HP1 Chromo Domain. *Nature* **2001**, *410* (6824), 120-124.

57. Lachner, M.; O'Carroll, D.; Rea, S.; Mechtler, K.; Jenuwein, T. Methylation of Histone H3 lysine 9 Creates a Binding Site for HP1 Proteins. *Nature* **2001**, *410* (6824), 116-120.

58. Meehan, R. R.; Kao, C. F.; Pennings, S. HP1 Binding to Native Chromatin in vitro is Determined by the Hinge Region and not by the Chromodomain. *EMBO J* **2003**, *22* (12), 3164-3174.

59. Maison, C.; Bailly, D.; Peters, A. H.; Quivy, J. P.; Roche, D.; Taddei, A.; Lachner, M.; Jenuwein, T.; Almouzni, G. Higher-Order Structure in Pericentric Heterochromatin Involves a Distinct Pattern of Histone Modification and an RNA Component. *Nat Genet* **2002**, *30* (3), 329-334.

60. Muchardt, C.; Guilleme, M.; Seeler, J. S.; Trouche, D.; Dejean, A.; Yaniv, M. Coordinated Methyl and RNA Binding is Required for Heterochromatin Localization of Mammalian HP1α. *EMBO Rep* **2002**, *3* (10), 975-981.

61. Santos, T.; Salgado, G. F.; Cabrita, E. J.; Cruz, C. G-Quadruplexes and Their Ligands: Biophysical Methods to Unravel G-Quadruplex/Ligand Interactions. *Pharmaceuticals (Basel)*. **2021**, *14* (8).

62. Kurup, H. M.; Kvach, M. V.; Harjes, S.; Barzak, F. M.; Jameson, G. B.; Harjes, E.; Filichev, V. V. Design, Synthesis, and Evaluation of a Cross-Linked Oligonucleotide as the First Nanomolar Inhibitor of APOBEC3A. *Biochemistry* **2022**, *61* (22), 2568-2578.

63. Prestinari, C.; Richert, C. Intrastrand Locks Increase Duplex Stability and Base Pairing Selectivity. *Chem. Commun.* **2011**, *47* (38), 10824-10826, Article.

64. Martin-Pintado, N.; Yahyaee-Anzahaee, M.; Deleavey, G. F.; Portella, G.; Orozco, M.; Damha, M. J.; Gonzalez, C. Dramatic Effect of Furanose C2' Substitution on Structure and Stability: Directing the Folding of the Human Telomeric Quadruplex with a Single Fluorine Atom. *J Am Chem Soc* **2013**, *135* (14), 5344-5347.

65. Abou Assi, H.; El-Khoury, R.; Gonzalez, C.; Damha, M. J. 2'-Fluoroarabinonucleic Acid Modification Traps G-Quadruplex and I-Motif Structures in Human Telomeric DNA. *Nucleic Acids Res* **2017**, *45* (20), 11535-11546.

66. Gonzalez, C. Instituto de Química Física Rocasolano, CSIC, Madrid, Spain. Personal Communication, **2021.**

67. Zhou, J.; Rosu, F.; Amrane, S.; Korkut, D. N.; Gabelica, V.; Mergny, J. L. Assembly of Chemically Modified G-rich Sequences into Tetramolecular DNA G-Quadruplexes and Higher Order Structures. *Methods* **2014**, *67* (2), 159-168.

68. Merkina, E. E.; Fox, K. R. Kinetic Stability of Intermolecular DNA Quadruplexes. *Biophys J* **2005**, *89* (1), 365-373.

69. Tran, P. L.; De Cian, A.; Gros, J.; Moriyama, R.; Mergny, J. L. Tetramolecular Quadruplex Stability and Assembly. *Top Curr Chem* **2013**, *330*, 243-273.

70. Mergny, J. L.; De Cian, A.; Ghelab, A.; Sacca, B.; Lacroix, L. Kinetics of Tetramolecular Quadruplexes. *Nucleic Acids Res* **2005**, *33* (1), 81-94.

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

72. Agrawal, P.; Lin, C.; Mathad, R. I.; Carver, M.; Yang, D. The Major G-Quadruplex Formed in the Human BCL-2 Proximal Promoter Adopts a Parallel Structure with a 13-nt Loop in K+ Solution. *J Am Chem Soc* **2014**, *136* (5), 1750-1753.

73. Carvalho, J.; Pereira, E.; Marquevielle, J.; Campello, M. P. C.; Mergny, J. L.; Paulo, A.; Salgado, G. F.; Queiroz, J. A.; Cruz, C. Fluorescent Light-up Acridine Orange Derivatives Bind and Stabilize KRAS-22RT G-Quadruplex. *Biochimie* **2018**, *144*, 144-152.

74. Hu, M. H.; Zhou, J.; Luo, W. H.; Chen, S. B.; Huang, Z. S.; Wu, R.; Tan, J. H. Development of a Smart Fluorescent Sensor That Specifically Recognizes the c-MYC G-Quadruplex. *Anal Chem* **2019**, *91* (3), 2480-2487.

75. Rodriguez, R.; Muller, S.; Yeoman, J. A.; Trentaseux, C.; Riou, J.-F.; Balasubramanian, S. A Novel Small Molecule That Alters Shelterin Integrity and Triggers a DNA-Damage Response at Telomeres. *J. Am. Chem. Soc* **2008**, *130* (47), 15758-15759.

76. Feng, Y.; Yang, D.; Chen, H.; Cheng, W.; Wang, L.; Sun, H.; Tang, Y. Stabilization of G-quadruplex DNA and Inhibition of Bcl-2 Expression by a Pyridostatin Analog. *Bioorg Med Chem Lett* **2016**, *26* (7), 1660-1663.

77. *IDT Oligoanalyzer Tool*. Integrated DNA Technologies, Inc., https://sg.idtdna.com/calc/analyzer (accessed 2023-05-10).

78. Fujii, M.; Hasegawa, T.; Niidome, T.; Aoyagi, H. Enhancement of Stability of Double and Triple Stranded DNA by Cationic Amphiphilic α-Helix Peptide. *Nucleosides and Nucleotides* **2006**, *18* (6-7), 1623-1624.

79. Yokoyama, K.; Kubo, T.; Fujii, M. Amphiphilic β-Sheet Peptides can Bind to Double and Triple stranded DNA. *Nucleosides Nucleotides Nucleic Acids* **2001**, *20* (4-7), 1317-1320.

80. Fujii, M.; Hasegawa, T.; Niidome, T.; Aoyagi, H. Thermal Stability of Double and Triple Stranded DNA in the Presence of Cationic Amphiphilic α-Helix Peptide. *Pept. Sci.* **1999**, *1998*, 297-300.

81. Kubo, T.; Yokoyama, K.; Ueki, R.; Abe, S.; Goto, K.; Niidome, T.; Aoyagi, H.; Iwakuma, K.; Ando, S.; Ono, S.; et al. Design, Synthesis and Characterization of DNA Binding Peptides. *Pept. Sci.* **2001**, *2000*, 109-112.

82. Shinkai, Y.; Kashihara, S.; Minematsu, G.; Fujii, H.; Naemura, M.; Kotake, Y.; Morita, Y.; Ohnuki, K.; Fokina, A. A.; Stetsenko, D. A.; et al. Silencing of BCR/ABL Chimeric Gene in Human Chronic Myelogenous Leukemia Cell Line K562 by siRNA-Nuclear Export Signal Peptide Conjugates. *Nucleic Acid Ther.* **2017**, *27* (3), 168-175.

83. Eckstein, F. Nucleoside Phosphorothioates. *Annu Rev Biochem* **1985**, *54*, 367-402.

84. Eckstein, F. Phosphorothioates, Essential Components of Therapeutic Oligonucleotides. *Nucleic Acid Ther* **2014**, *24* (6), 374-387.

85. Su, Y.; Bayarjargal, M.; Hale, T. K.; Filichev, V. V. DNA with Zwitterionic and Negatively Charged Phosphate Modifications: Formation of DNA Triplexes, Duplexes and Cell Uptake Studies. *Beilstein J Org Chem* **2021**, *17*, 749-761.

86. Su, Y.; Fujii, H.; Burakova, E. A.; Chelobanov, B. P.; Fujii, M.; Stetsenko, D. A.; Filichev, V. V. Neutral and Negatively Charged Phosphate Modifications Altering Thermal Stability, Kinetics of Formation and Monovalent Ion Dependence of DNA G-Quadruplexes. *Chem Asian J* **2019**, *14* (8), 1212-1220.

87. Stetsenko, D. A. Novosibirsk State University, Novosibirsk, Russia. Personal Communication, **2019.**

88. Doluca, O.; Withers, J. M.; Loo, T. S.; Edwards, P. J.; Gonzalez, C.; Filichev, V. V. Interdependence of Pyrene Interactions and Tetramolecular G4-DNA Assembly. *Org Biomol Chem* **2015**, *13* (12), 3742-3748.

89. Morvan, F.; Debart, F.; Vasseur, J.-J. From Anionic to Cationic α-Anomeric Oligodeoxynucleotides. *Chem Biodivers* **2010**, *7*, 494-535.

90. Morvan, F.; Rayner, B.; Imbach, J. L.; Lee, M.; Hartley, J. A.; Chang, D. K.; Lown, J. W. α-DNA-V - Parallel Annealing, Handedness and Conformation of the Duplex of the Unnatural α-Hexadeoxyribonucleotide α-[D(CpApTpGpCpG)] with Its β-Complement β-[D(GpTpApCpGpC)] Deduced from High-Field H-1-Nmr. *Nucleic Acids Res* **1987**, *15* (17), 7027-7044.

91. Swarna Latha, Y.; Yathindra, N. Stereochemical Studies on Nucleic Acid Analogues. I. Conformations of α-Nucleosides and α-Nucleotides: Interconversion of Sugar Puckers via 04'-exo. *Biopolymers* **1992**, *32*, 249-269.

92. Sequin, U. Nucleosides and Nucleotides. 5. The Stereochemistry of Oligonucleotides Consisting of 2'-Deoxy-α-D-ribosides, a Study with Drieding Stereomodels. *Experientia* **1973**, *29* (9), 1059-1062.

93. Germann, M. W.; Aramini, J. M.; Kalisch, B. W.; van de Sande, J. H. Structural, Dynamic, and Enzymatic Properties of Mixed α/β-Oligonucleotides Containing Polarity Reversals. *Nucleosides Nucleotides Nucleic Acids* **2001**, *20* (4-7), 493-499.

94. Lancelot, G.; Guesnet, J. L.; Roig, V.; Thuong, N. T. 2D-Nmr Studies of the Unnatural Duplex α-D(TCTAAAC)-β-D(AGATTTG). *Nucleic Acids Res* **1987**, *15* (18), 7531-7547.

95. Filitcheva, J.; Edwards, P. J. B.; Norris, G. E.; Filichev, V. V. α-2'-Deoxyguanosine can Switch DNA G-Quadruplex Topologies from Antiparallel to Parallel. *Org. Biomol. Chem.* **2019**, 4031-4042.

96. Boutorine, A. S.; Venyaminova, A. G.; Repkova, M. N.; Sergueyeva, Z. A.; Pyshnyi, D. V. Effect of Derivatization of Ribophosphate Backbone and Terminal Ribophosphate Groups in Oligoribonucleotides on their Stability and Interaction with Eukaryotic Cells. *Biochimie* **1994**, *76* (1), 23-32.

97. Novopashina, D.; Kuznetsova, M.; Venyaminova, A. 2'-O-Modified Oligoribonucleotides with Terminal 3'-3'-Internucleotide Linkage and their Derivatives. *Nucleosides Nucleotides Nucleic Acids* **2001**, *20* (4-7), 903-907.

98. Virgilio, A.; Esposito, V.; Filosa, R.; Mayol, L.; Galeone, A. The Introduction of Inversion of Polarity Sites in DNA G-Quadruplex Structures: Effects and Perspectives. *Mini Rev Med Chem* **2016**, *16* (7), 509-523.

99. Sket, P.; Virgilio, A.; Esposito, V.; Galeone, A.; Plavec, J. Strand Directionality Affects Cation Binding and Movement within Tetramolecular G-Quadruplexes. *Nucleic Acids Res* **2012**, *40* (21), 11047-11057.

100. Esposito, V.; Virgilio, A.; Pepe, A.; Oliviero, G.; Mayol, L.; Galeone, A. Effects of the Introduction of Inversion of Polarity Sites in the Quadruplex Forming Oligonucleotide TGGGT. *Bioorg Med Chem* **2009**, *17* (5), 1997-2001.

101. Virgilio, A.; Esposito, V.; Mayol, L.; Galeone, A. More than one Non-Canonical Phosphodiester Bond in the G-Tract: Formation of Unusual Parallel G-Quadruplex Structures. *Org Biomol Chem* **2014**, *12* (3), 534-540.

102. Phan, A. T.; Gueron, M.; Leroy, J. L. Investigation of Unusual DNA Motifs. *Method Enzymol* **2001**, *338*, 341-371.

103. Lowary, P. T.; Widom, J. New DNA Sequence Rules for High Affinity Binding to Histone Octamer and Sequence-Directed Nucleosome Positioning. *J Mol Biol* **1998**, *276* (1), 19-42.

104. Hsu, S. T.; Varnai, P.; Bugaut, A.; Reszka, A. P.; Neidle, S.; Balasubramanian, S. A G-rich Sequence within the c-kit Oncogene Promoter forms a Parallel G-Quadruplex Having Asymmetric G-tetrad Dynamics. *J Am Chem Soc* **2009**, *131* (37), 13399-13409.

105. Ambrus, A.; Chen, D.; Dai, J.; Jones, R. A.; Yang, D. Solution Structure of the Biologically Relevant G-Quadruplex Wlement in the Human c-MYC

Promoter. Implications for G-Quadruplex Stabilization. *Biochemistry* **2005**, *44* (6), 2048-2058.

106. Vaux, D. L.; Cory, S.; Adams, J. M. Bcl-2 Gene Promotes Haemopoietic Cell Survival and Cooperates with c-myc to Ommortalize pre-B Cells. *Nature* **1988**, *335* (6189), 440-442.

107. Hockenbery, D.; Nunez, G.; Milliman, C.; Schreiber, R. D.; Korsmeyer, S. J. Bcl-2 is an Inner Mitochondrial Membrane Protein that Blocks Programmed Cell Death. *Nature* **1990**, *348* (6299), 334-336.

108. Campbell, K. J.; Tait, S. W. G. Targeting BCL-2 Regulated Apoptosis in Cancer. *Open Biol* **2018**, *8* (5).

109. Blackwell, H. E.; Grubbs, R. H. Highly Efficient Synthesis of Covalently Cross-Linked Peptide Helices by Ring-Closing Metathesis. *Angew Chem Int Ed Engl* **1998**, *37* (23), 3281-3284.

110. Walensky, L. D.; Bird, G. H. Hydrocarbon-Stapled Peptides: Principles, Practice, and Progress. *J Med Chem* **2014**, *57* (15), 6275-6288.

111. Schafmeister, C. E.; Po, J.; Verdine, G. L. An All-Hydrocarbon Cross-Linking System for Enhancing the Helicity and Metabolic Stability of Peptides. *J. Am. Chem. Soc* **2000**, *122*, 5891-5892.

112. Barthe, P.; Rochette, S.; Vita, C. Synthesis and NMR Solution Structure of an α-Helical Hairpin Stapled with two Disulfide Bridges. *Protein Science* **2000**, *9*, 942-955.

113. Chilton, B.; Filichev, V. V. Development of Cross-linking Strategies for DNA G-quadruplexes. Masters Thesis, Massey University, New Zealand, 2019.

114. Ramadan, M.; Bremner-Hay, N. K.; Carlson, S. A.; Comstock, L. R. Synthesis and Evaluation of N6-Substituted Azide- and Alkyne-Bearing N-Mustard Analogs of S-Adenosyl-l-methionine. *Tetrahedron* **2014**, *70* (34), 5291-5297.

115. Zhu, L.; Brassard, C. J.; Zhang, X.; Guha, P. M.; Clark, R. J. On the Mechanism of Copper(I)-Catalyzed Azide-Alkyne Cycloaddition. *Chem Rec* **2016**, *16* (3), 1501-1517.

116. Rostovtsev, V. V.; Green, L. G.; Fokin, V. V.; Sharpless, K. B. A Stepwise Huisgen Cycloaddition Process: Copper(I)-Catalyzed Regioselective "Ligation" of Azides and Terminal Alkynes. *Angew Chem Int Ed Engl* **2002**, *41* (14), 2596-2599.

117. El-Sagheer, A. H.; Brown, T. Click Chemistry with DNA. *Chem Soc Rev* **2010**, *39* (4), 1388-1405.

118. Gierlich, J.; Burley, G. A.; Gramlich, P. M. E.; Hammond, D. M.; Carell, T. Click Chemistry as a Reliable Method for the High-Density Postsynthetic Functionalization of Alkyne-Modified DNA. *Org Lett* **2006**, *8* (17), 3639-3642.

119. Bouillon, C.; Meyer, A.; Vidal, S.; Jochum, A.; Chevolot, Y.; Cloarec, J. P.; Praly, J. P.; Vasseur, J. J.; Morvan, F. Microwave Assisted "Click" Chemistry for the Synthesis of Multiple Labeled-Carbohydrate Oligonucleotides on Solid Support. *J Org Chem* **2006**, *71* (12), 4700-4702.

120. Geci, I.; Filichev, V. V.; Pedersen, E. B. Stabilization of Parallel Triplexes by Twisted Intercalating Nucleic Acids (TINAs) Incorporating 1,2,3-Triazole Units and Prepared by Microwave-Accelerated Click Chemistry. *Chemistry-Eur* **2007**, *13* (22), 6379-6386.

121. Wenge, U.; Ehrenschwender, T.; Wagenknecht, H. A. Synthesis of 2'-O-Propargyl Nucleoside Triphosphates for Enzymatic Oligonucleotide Preparation and "Click" Modification of DNA with Nile Red as Fluorescent Probe. *Bioconjug Chem* **2013**, *24* (3), 301-304.

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

123. Lim, K. W.; Ng, V. C.; Martin-Pintado, N.; Heddi, B.; Phan, A. T. Structure of the Human Telomere in Na+ Solution: an Antiparallel (2+2) G-Quadruplex Scaffold Reveals Additional Diversity. *Nucleic Acids Res* **2013**, *41* (22), 10556-10562.

124. Tararov, V. I.; Kolyachkina, S. V.; Alexeev, C. S.; Mikhailov, S. N. N6-Acetyl-2′,3′,5′-tri-O-acetyladenosine; A Convenient, 'Missed Out' Substrate for Regioselective N6-Alkylations. *Synthesis* **2011**, *2011* (15), 2483-2489.

125. Ottria, R.; Casati, S.; Baldoli, E.; Maier, J. A.; Ciuffreda, P. N(6)-Alkyladenosines: Synthesis and Evaluation of in vitro Anticancer Aactivity. *Bioorg Med Chem* **2010**, *18* (23), 8396-8402.

126. Kurup, H. M. Design, Synthesis, and Evaluation of Cross-linked Single-stranded DNAs as Inhibitors of APOBEC3 Enzymes. PhD Thesis, Massey University, New Zealand, 2021.

127. Aube, J.; Milligan, G. L. Intramolecular Schmidt Reaction of Alkyl Azides. *J Am Chem Soc* **1991**, *113* (23), 8965-8966.

128. Sigurdsson, S.; Strömberg, R. The H-Phosphonate Approach to Oligonucleotide Synthesis. An Investigation on the Mechanism of the Coupling Step. *J. Chem. Soc., Perkin Trans. 2* **2002**, (10), 1682-1688.

129. Powles, N.; Atherton, J.; Page, M. I. Reactive Ontermediates in the H-Phosphonate Synthesis of Oligonucleotides. *Org Biomol Chem* **2012**, *10* (30), 5940-5947.

130. Pourceau, G.; Meyer, A.; Vasseur, J. J.; Morvan, F. Azide Solid Support for 3'-Conjugation of Oligonucleotides and their Circularization by Click Chemistry. *J Org Chem* **2009**, *74* (17), 6837-6842.

131. Srivastava, S. C.; Raza, S. K. Synthesis of Propargyl Modified Nucleosides and Phosphoramidites and their Incorporation into Defined Seqeunce Oligonucleotides. WO 1995018139 A1, **1995**.

132. Paeschke, K.; Bochman, M. L.; Garcia, P. D.; Cejka, P.; Friedman, K. L.; Kowalczykowski, S. C.; Zakian, V. A. Pif1 Family Helicases Suppress Genome Instability at G-Quadruplex Motifs. *Nature* **2013**, *497* (7450), 458-462.

133. Monsen, R. C.; Chua, E. Y. D.; Hopkins, J. B.; Chaires, J. B.; Trent, J. O. Structure of a 28.5 kDa Duplex-Embedded G-Quadruplex System Resolved to 7.4 A Resolution with Cryo-EM. *Nucleic Acids Res* **2023**, *51* (4), 1943-1959.

134. Reed, R. M.; Tartar, H. V. The Preparation of Sodium Alkyl Sulfonates. *J. Am. Chem. Soc* **1935**, *57* (3), 570-571.

135. Hirota, M.; Haga, T.; Watabe, T. Preparation of Sulfonyl Chlorides from Sulfonic Acids or their Salts. JP2005179325, **2005**.

136. Matano, Y.; Ohkubo, H.; Honsho, Y.; Saito, A.; Seki, S.; Imahori, H. Synthesis and Charge-Carrier Transport Properties of Poly(phosphole P-alkanesulfonylimide)s. *Org Lett* **2013**, *15* (4), 932-935.

137. Onel, B.; Carver, M.; Wu, G.; Timonina, D.; Kalarn, S.; Larriva, M.; Yang, D. A New G-Quadruplex with Hairpin Loop Immediately Upstream of the Human BCL2 P1 Promoter Modulates Transcription. *J Am Chem Soc* **2016**, *138* (8), 2563-2570.

138. Novakovic, M.; Kupce, E.; Scherf, T.; Oxenfarth, A.; Schnieders, R.; Grun, J. T.; Wirmer-Bartoschek, J.; Richter, C.; Schwalbe, H.; Frydman, L. Magnetization Transfer to Enhance NOE Cross-Peaks among Labile Protons: Applications to Imino-Imino Sequential Walks in SARS-CoV-2-Derived RNAs. *Angew Chem Int Ed Engl* **2021**, *60* (21), 11884-11891.

139. Webba da Silva, M. NMR Methods for Studying Quadruplex Nucleic Acids. *Methods* **2007**, *43* (4), 264-277.

140. Małgowska, M.; Gudanis, D.; Teubert, A.; Dominiak, G.; Gdaniec, Z. How to Study G-quadruplex Structures. *BioTechnologia* **2012**, *4*, 381-390.

141. Karsisiotis, A. I.; Hessari, N. M.; Novellino, E.; Spada, G. P.; Randazzo, A.; Webba da Silva, M. Topological Characterization of Nucleic Acid G-quadruplexes by UV Absorption and Circular Dichroism. *Angew Chem Int Ed Engl* **2011**, *50* (45), 10645-10648.

142. Randazzo, A.; Spada, G. P.; da Silva, M. W. Circular Dichroism of Quadruplex Structures. *Top Curr Chem* **2013**, *330*, 67-86.

143. Vorlickova, M.; Kejnovska, I.; Sagi, J.; Renciuk, D.; Bednarova, K.; Motlova, J.; Kypr, J. Circular Dichroism and Guanine Quadruplexes. *Methods* **2012**, *57* (1), 64-75.

144. Bates, R. W.; Dewey, M. R. A Formal Synthesis of Swainsonine by Gold-Catalyzed Allene Cyclization. *Org Lett* **2009**, *11* (16), 3706-3708.

145. Zhang, P.; Ng, K.; Ling, C. C. Total Synthesis of Le(A)-LacNAc Pentasaccharide as a Ligand for Clostridium Difficile Toxin A. *Org Biomol Chem* **2010**, *8* (1), 128-136.

146. Ding, T. J.; Zhou, L.; Cao, X. P. A Facile and Green Synthesis of Sulforaphane. *Chinese Chem Lett* **2006**, *17* (9), 1152-1154.

147. Li, H.; Fahrenbach, A. C.; Coskun, A.; Zhu, Z.; Barin, G.; Zhao, Y. L.; Botros, Y. Y.; Sauvage, J. P.; Stoddart, J. F. A Light-Sstimulated Molecular Switch Driven by Radical-Radical Interactions in Water. *Angew Chem Int Ed Engl* **2011**, *50* (30), 6782-6788.

148. Terrazas, M.; Ariza, X.; Farras, J.; Guisado-Yang, J. M.; Vilarrasa, J. A Direct, Efficient Method for the Preparation of N6-protected 15N-labeled Adenosines. *J Org Chem* **2004**, *69* (16), 5473-5475.

149. *Click Chemistry Labeling of Oligonucleotides and DNA*. Lumiprobe Limited, https://www.lumiprobe.com/protocols/click-chemistry-dna-labeling (accessed 2023-05-10).

Appendix A. Supplementary Information

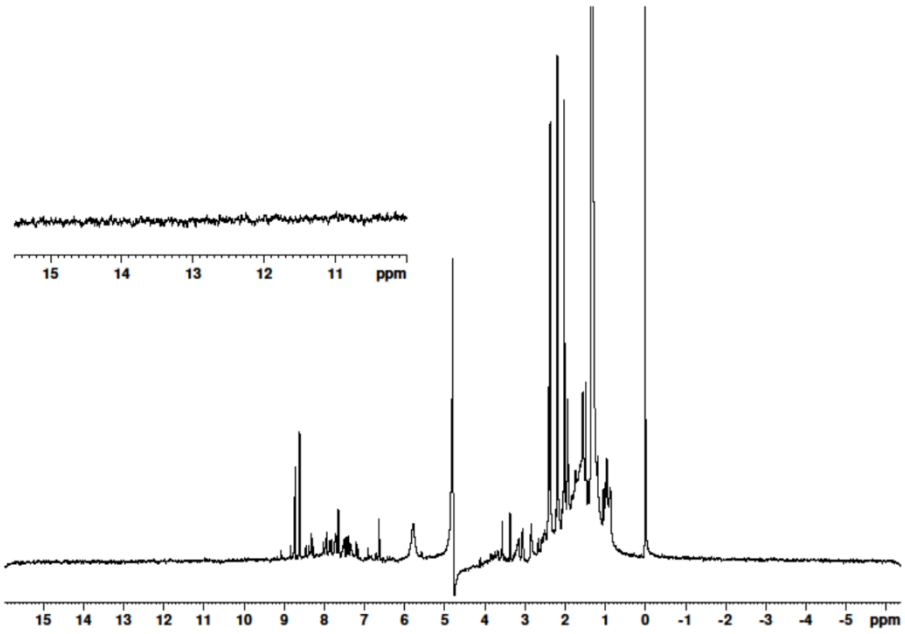


Figure SI-1. ^1H NMR spectra of pyridostatin. Insert is a 10 – 15.5 ppm region with no peaks in the imino proton region of G4 and DNA duplexes. Conditions: 200 μM pyridostatin, 20 mM sodium phosphate, 10 mM KCl, 10% D_2O, 1% TSP, pH 7.0, 25 °C.

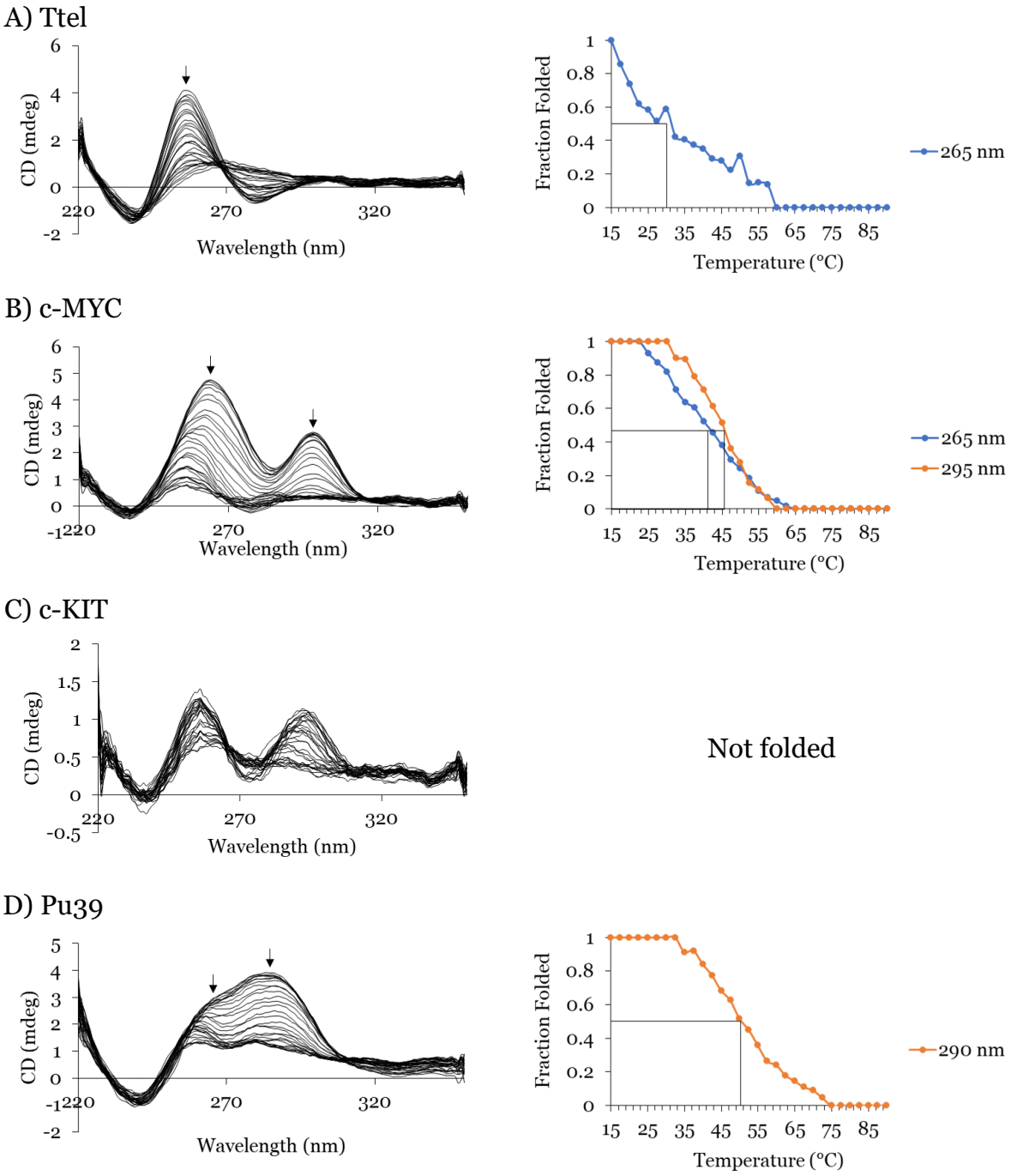


Figure SI-2. CD melting profiles of A) Ttel, B) c-MYC, C) c-KIT, and D) Pu39 in Na^+ buffer. Conditions: 10 μM strand concentration, 20 mM sodium phosphate, pH 7.0. 15 – 90 °C, 1 °C/min. Arrows indicate direction of change in peak intensity.

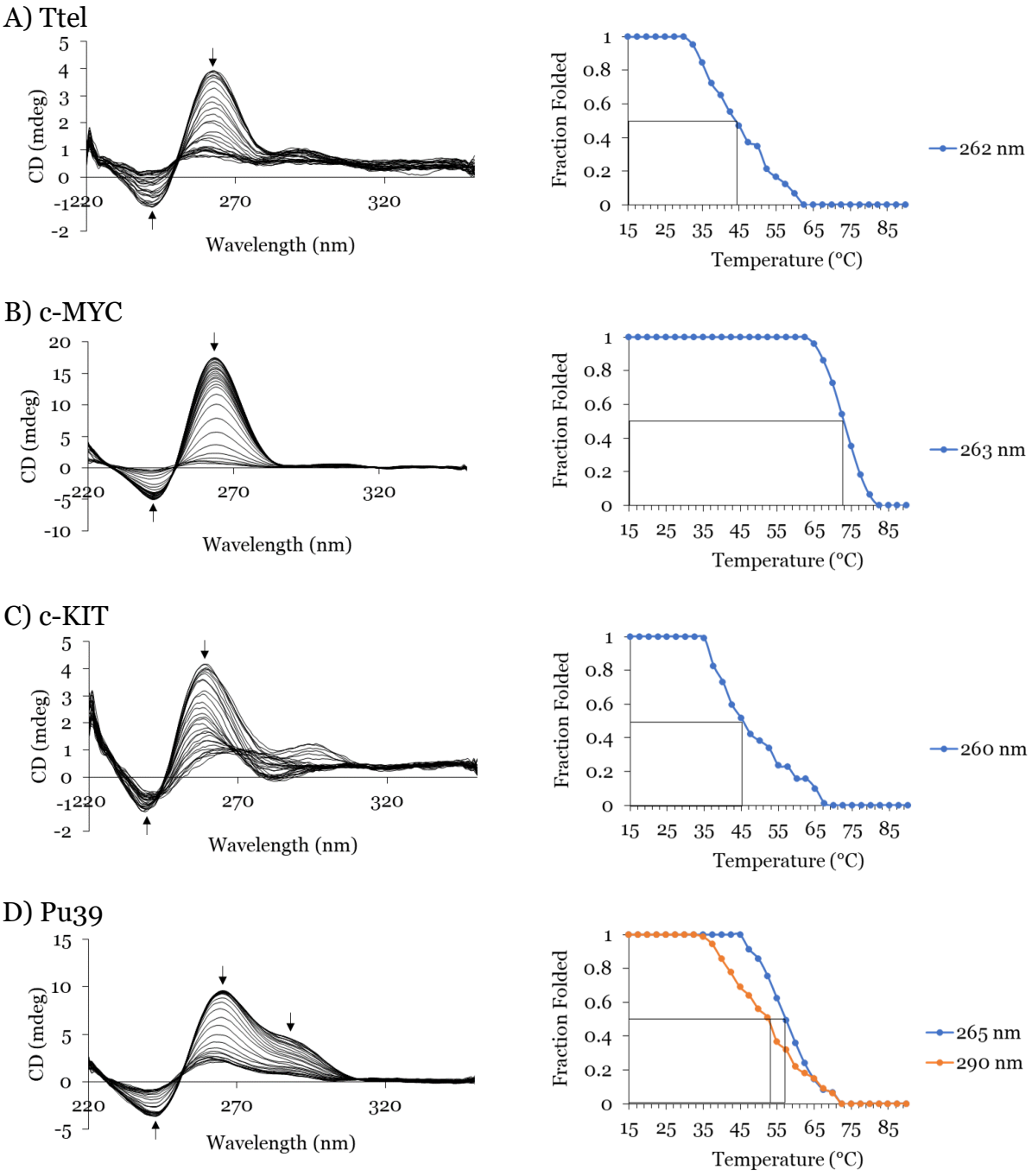


Figure SI-3. CD melting profiles of A) Ttel, B) c-MYC, C) c-KIT, and D) Pu39 in K^+ buffer. Conditions: 10 μM strand concentration, 20 mM sodium phosphate, 10 mM KCl, pH 7.0. 15 – 90 °C, 1 °C/min. Arrows indicate direction of change in peak intensity.

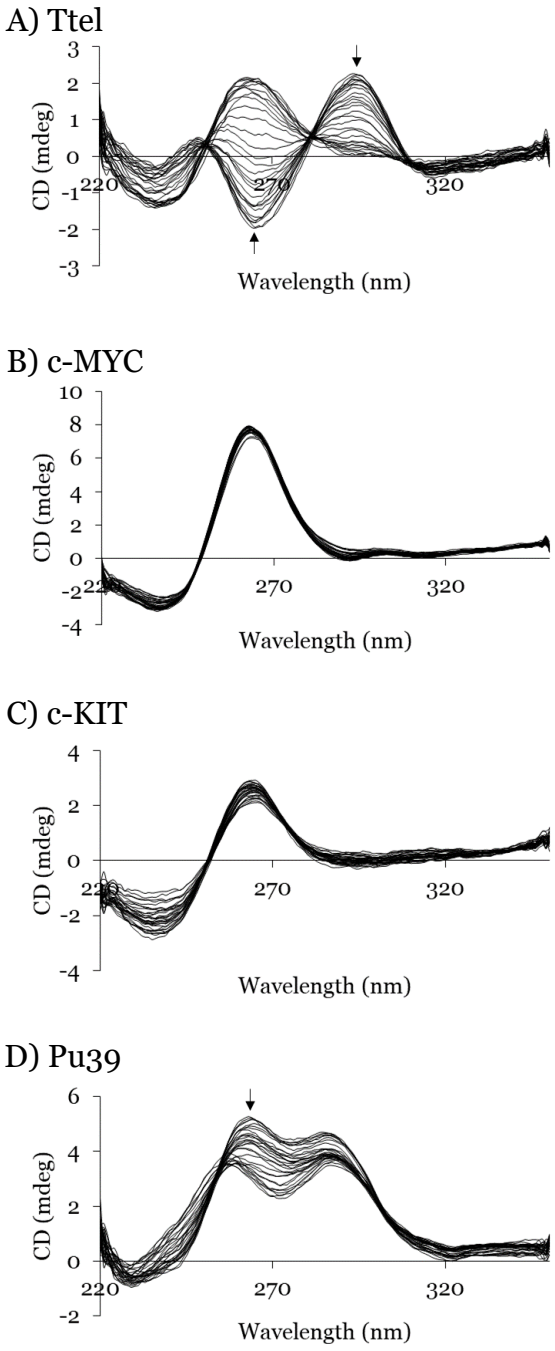


Figure SI-4. CD melting profiles of A) Ttel, B) c-MYC, C) c-KIT, and D) Pu39 in K^+ buffer, with the addition of the pyridostatin ligand. Ttel initially appears to form an antiparallel G4 but peak at 265 nm changes from negative to positive. This structure is stable to 90 °C. All other samples show no transition. Conditions: 10 µM strand concentration, 20 µM pyridostatin, 20 mM sodium phosphate, 10 mM KCl, pH 7.0. 15 – 90 °C, 1 °C/min. Arrows indicate direction of change in peak intensity.

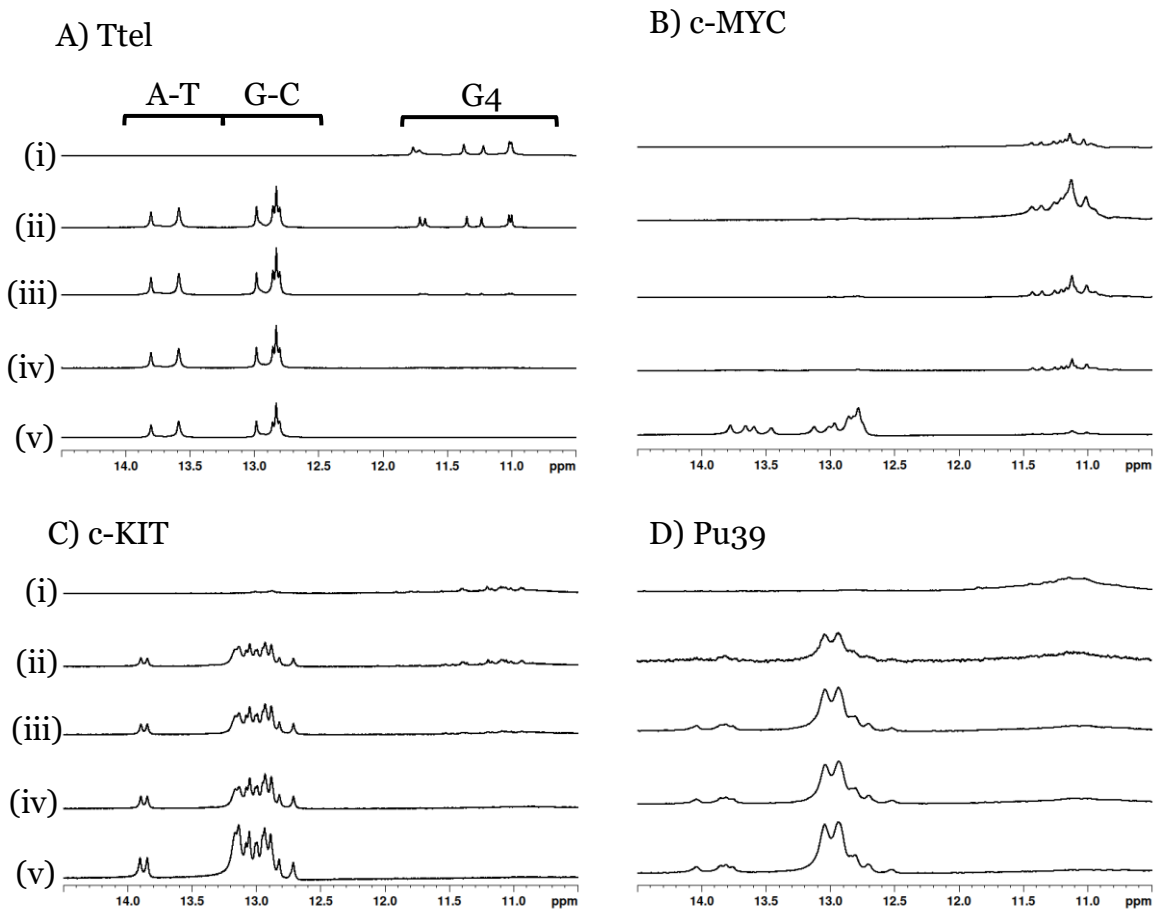


Figure SI-5. ^{1}H NMR spectra of control duplex formation experiments for A) Ttel, B) c-MYC, C) c-KIT and D) Pu39 with their respective complementary strands (i-Ttel, i-c-MYC, i-c-KIT and i-Pu39). (i) G4 alone, (ii) 1 hour, (iii) 1 day, (iv) 3 days, (v) thermodynamic product. Conditions: 200 µM strand concentration, 20 mM sodium phosphate, 10 mM KCl, 10% D_2O, 1% TSP, pH 7.0, 25 °C.

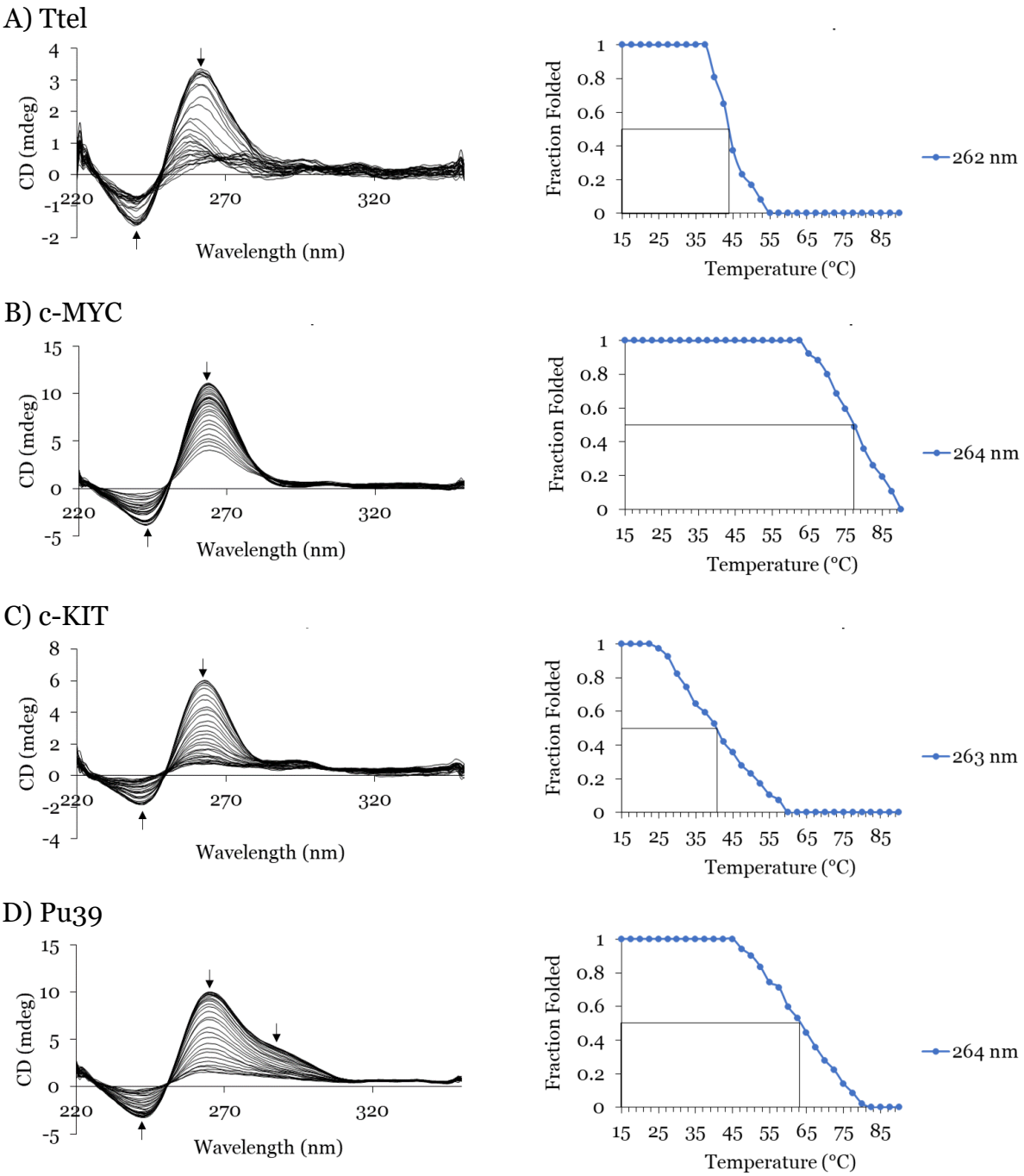


Figure SI-6. CD melting profiles of A) Ttel, B) c-MYC, C) c-KIT, and D) Pu39 in K^+ buffer, with the addition of the RL4 ligand. Conditions: 10 μM strand concentration, 20 μM Rl4, 20 mM sodium phosphate, 10 mM KCl, pH 7.0. 15 – 90 °C, 1 °C/min. Arrows indicate direction of change in peak intensity.

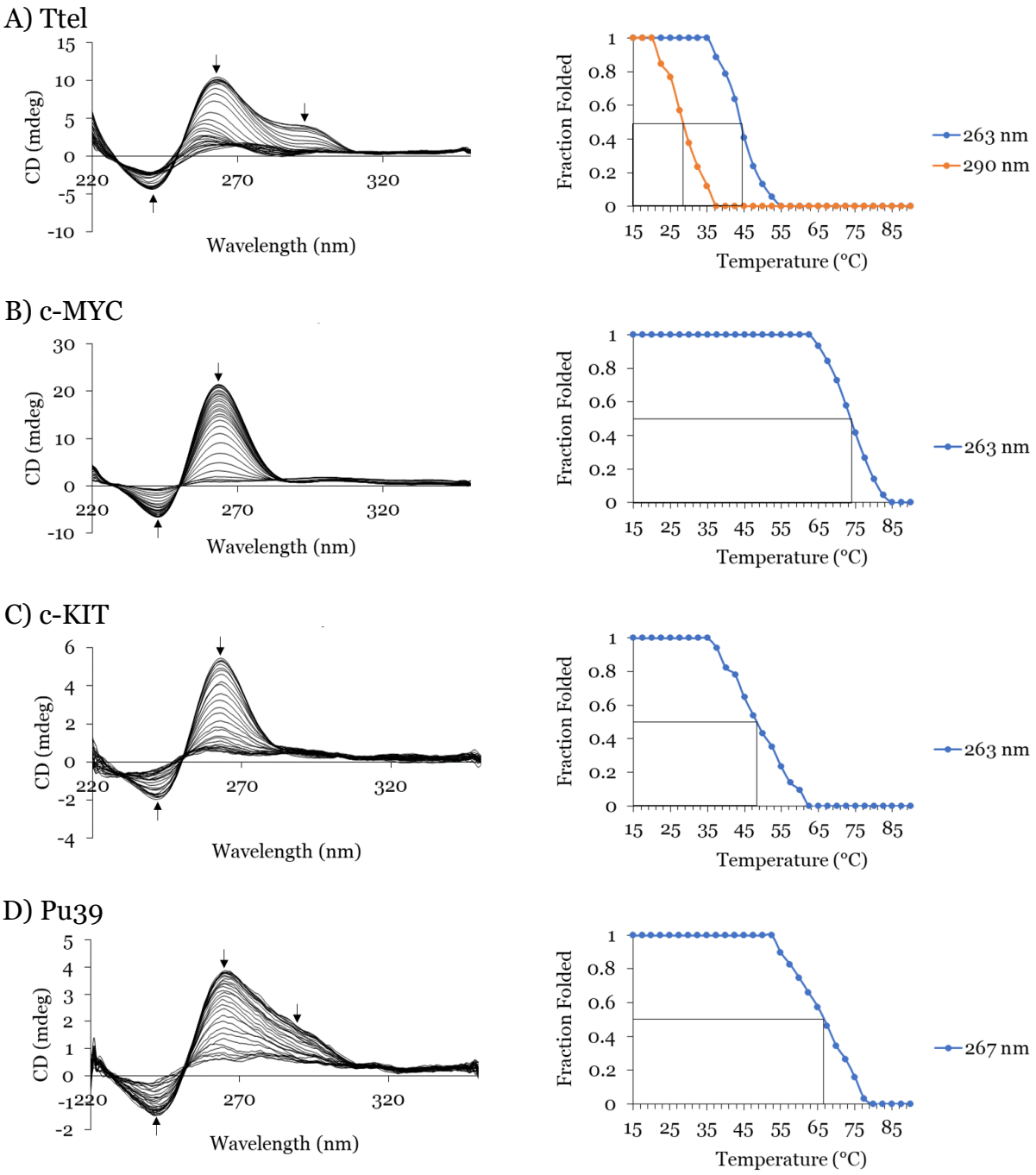


Figure SI-7. CD melting profiles of A) c-KIT, B) c-MYC, C) Pu39, and D) Ttel in K^+ buffer, with the addition of the RS4 ligand. Conditions: 10 µM strand concentration, 20 µM RS4, 20 mM sodium phosphate, 10 mM KCl, pH 7.0. 15 – 90 °C, 1 °C/min. Arrows indicate direction of change in peak intensity.

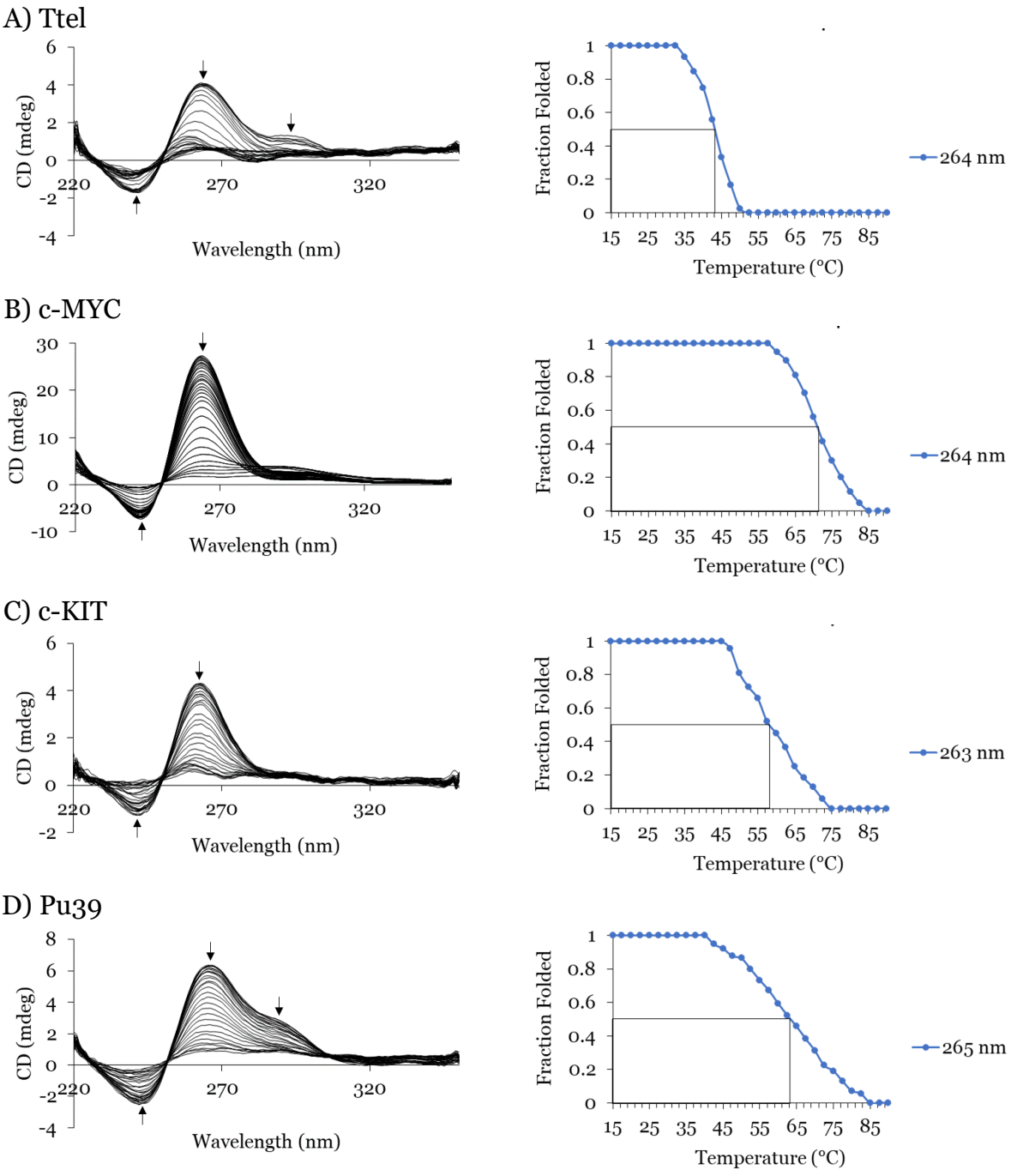


Figure SI-8. CD melting profiles of A) c-KIT, B) c-MYC, C) Pu39, and D) Ttel in K^+ buffer, with the addition of the RG4 ligand. Conditions: 10 µM strand concentration, 20 µM RG4, 20 mM sodium phosphate, 10 mM KCl, pH 7.0. 15 – 90 °C, 1 °C/min. Arrows indicate direction of change in peak intensity.

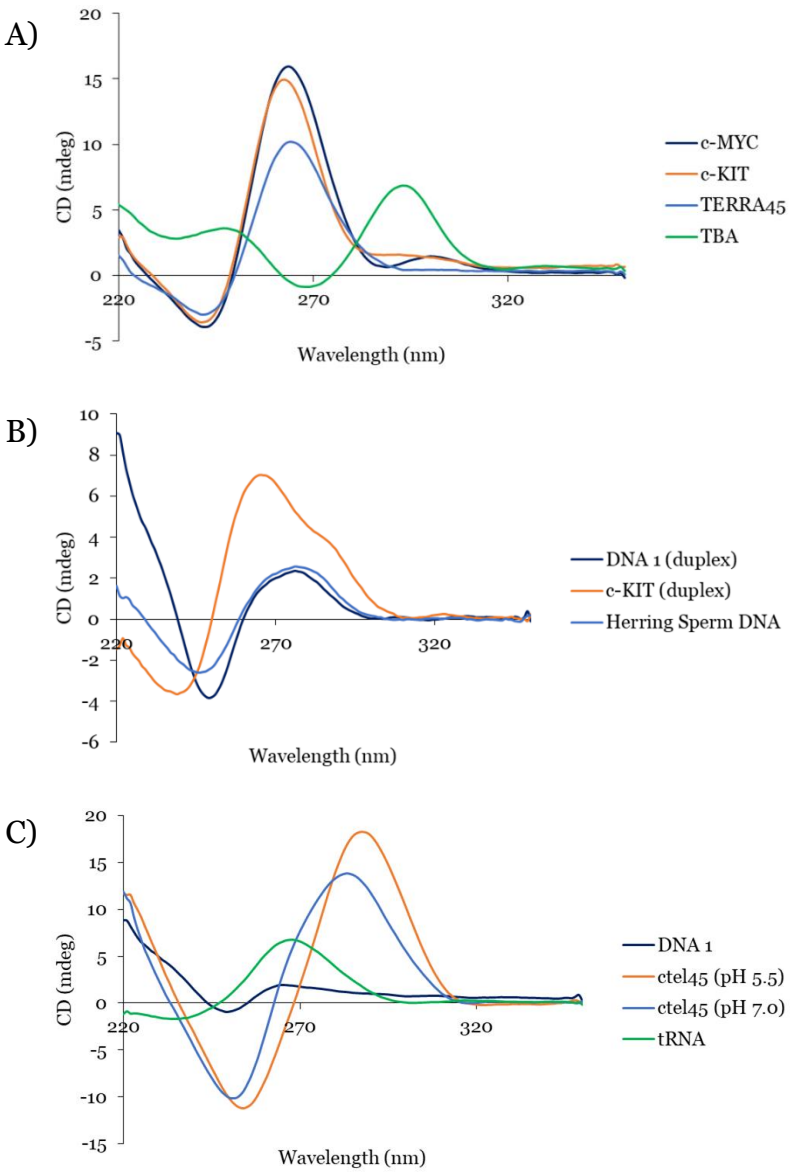


Figure SI-9. CD profiles of sequences forming a variety of DNA secondary structures. A) G4-forming sequences: c-MYC, c-KIT, TERRA45, parallel G4s, and TBA, antiparallel G4. B) Duplex-forming sequences: DNA 1, mixture of two short complementary oligonucleotides, c-KIT, mixture of the c-KIT and i-c-KIT sequences and herring sperm DNA, a long DNA duplex extracted from herring sperm. C) other structures: DNA 1, single short oligonucleotide from the DNA 1 duplex in B), ctel45 (pH 5.5), an i-motif formed in a c-rich strand, ctel45 (pH 7.0), the same sequence but single-stranded due to the higher pH and tRNA, a long single-stranded RNA oligonucleotide. Conditions: Approximately 20 μM strand concentration (significantly lower concentrations, e.g., 2μM or less, were required for TERRA45, tRNA and herring sperm DNA), 20 mM sodium phosphate buffer, 10 mM KCl, pH 7.0 (buffer adjusted to pH 5.5 for i-motif formation by addition of HCl).

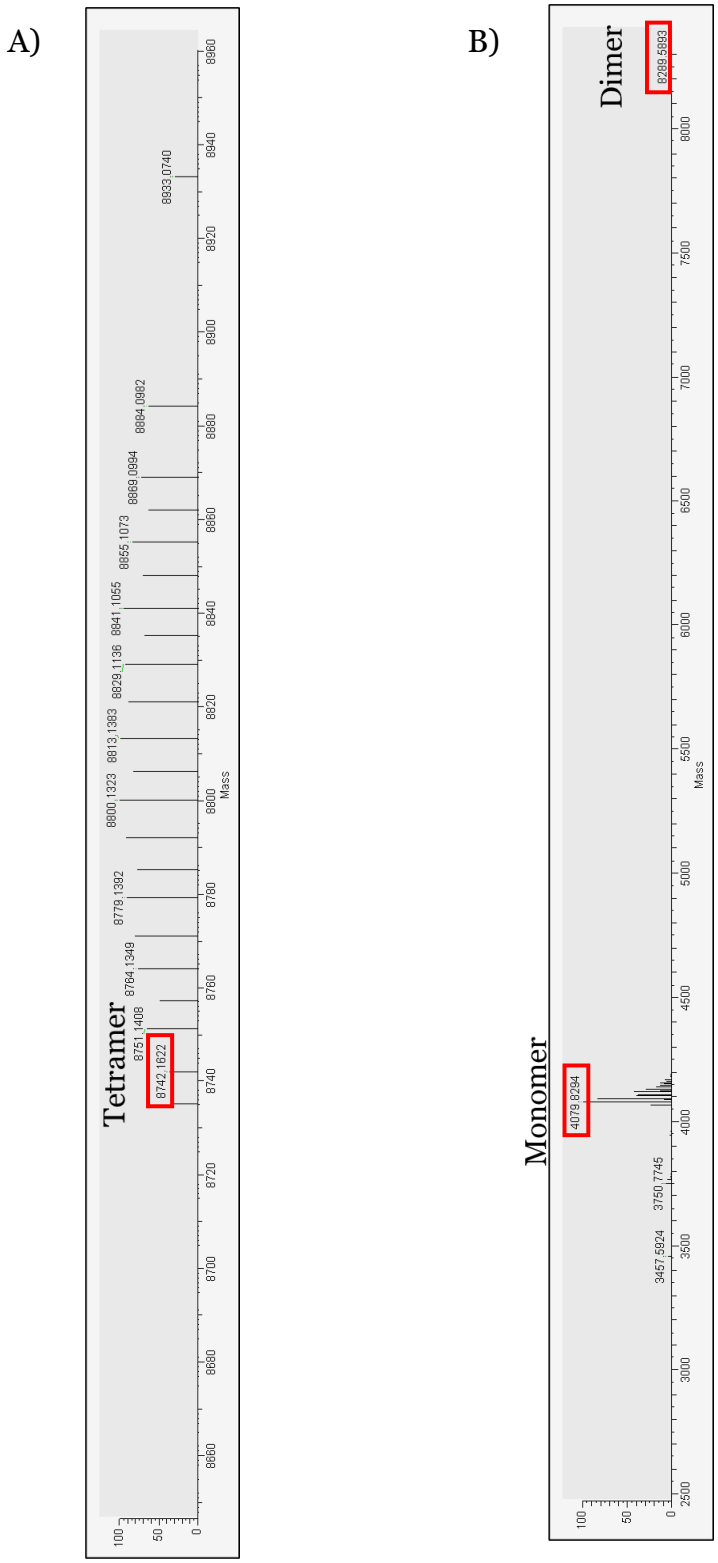
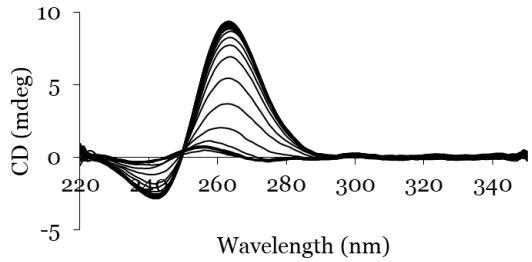
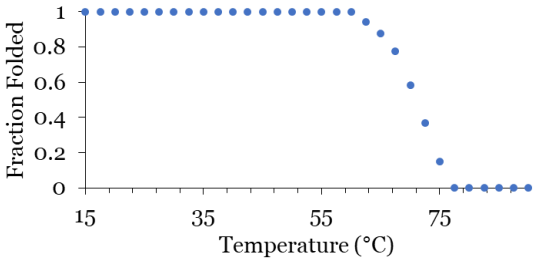


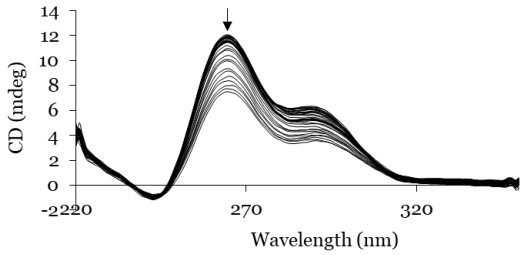
Figure SI-10. Native ESI-MS of A) TG_4XT G4 and B) $GXG_3T_4G_4$ G4. In addition to observing monomeric sequences, it is also possible to identify the tetramolecular and bimolecular complexes formed. Conditions: 100 μM strand concentration, 15% aq. MeOH, 15 mM, pH 7.4.

A) TG_4T



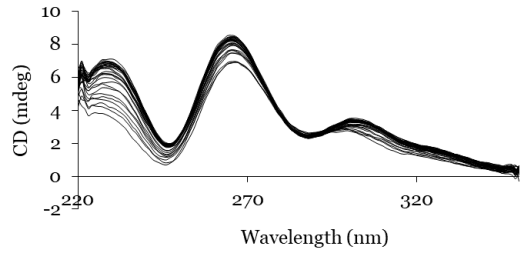


B) TG_4XT



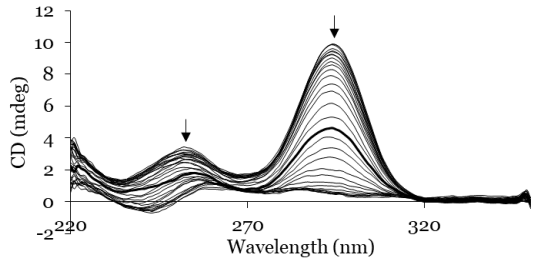
Not completely unfolded

C) TXG_4T



Not completely unfolded

D) TG_2XG_2T



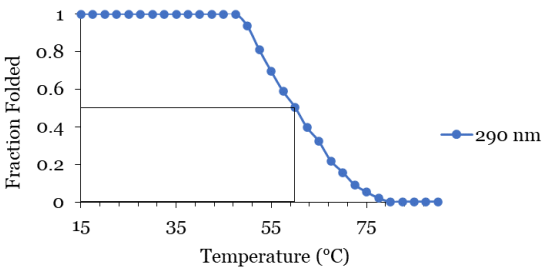
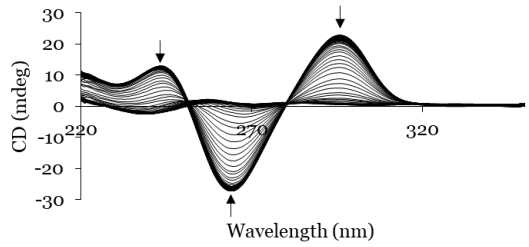
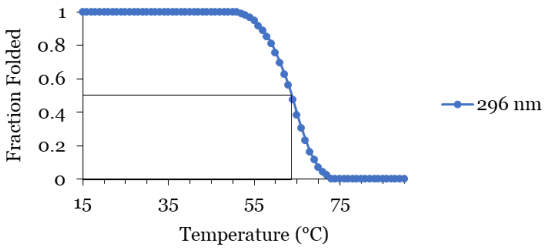


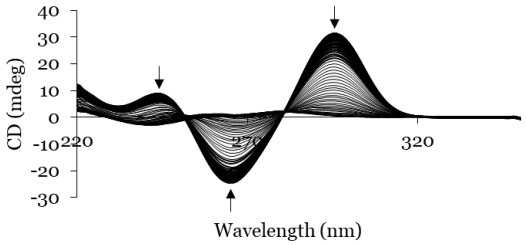
Figure SI-11. CD melting profiles of A) TG_4T, B) TG_4XT, C) TXG_4T, and D) TG_2XG_2T in Na^+ buffer. Conditions: 20 μM strand concentration, 10 mM lithium cacodylate, 100 mM NaCl, pH 7.2. 15 – 90 °C, 1 °C/min. Arrows indicate direction of change in peak intensity.

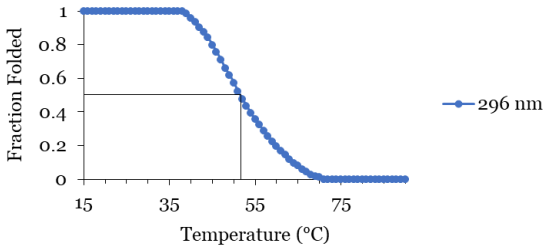
A) $G_4T_4G_4$



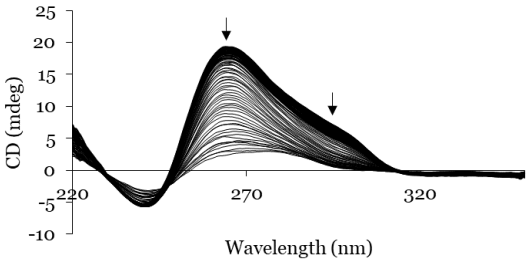


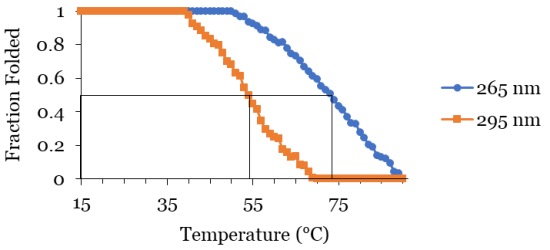
B) $GXG_3T_4G_4$



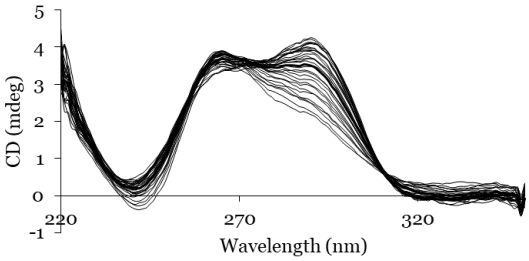


C) $G_4T_4G_3XG$





D) $GXG_3T_4G_3XG$



Not completely unfolded

Figure SI-12. CD melting profiles of A) $G_4T_4G_4$, B) $GXG_4T_4G_4$, C) $G_4T_4G_3XG$, and D) $GXG_4T_4\ G_3XG$ in Na^+ buffer. Conditions: 20 µM strand concentration, 10 mM lithium cacodylate, 100 mM NaCl, pH 7.2. 15 – 90 °C, 1 °C/min. Arrows indicate direction of change in peak intensity.

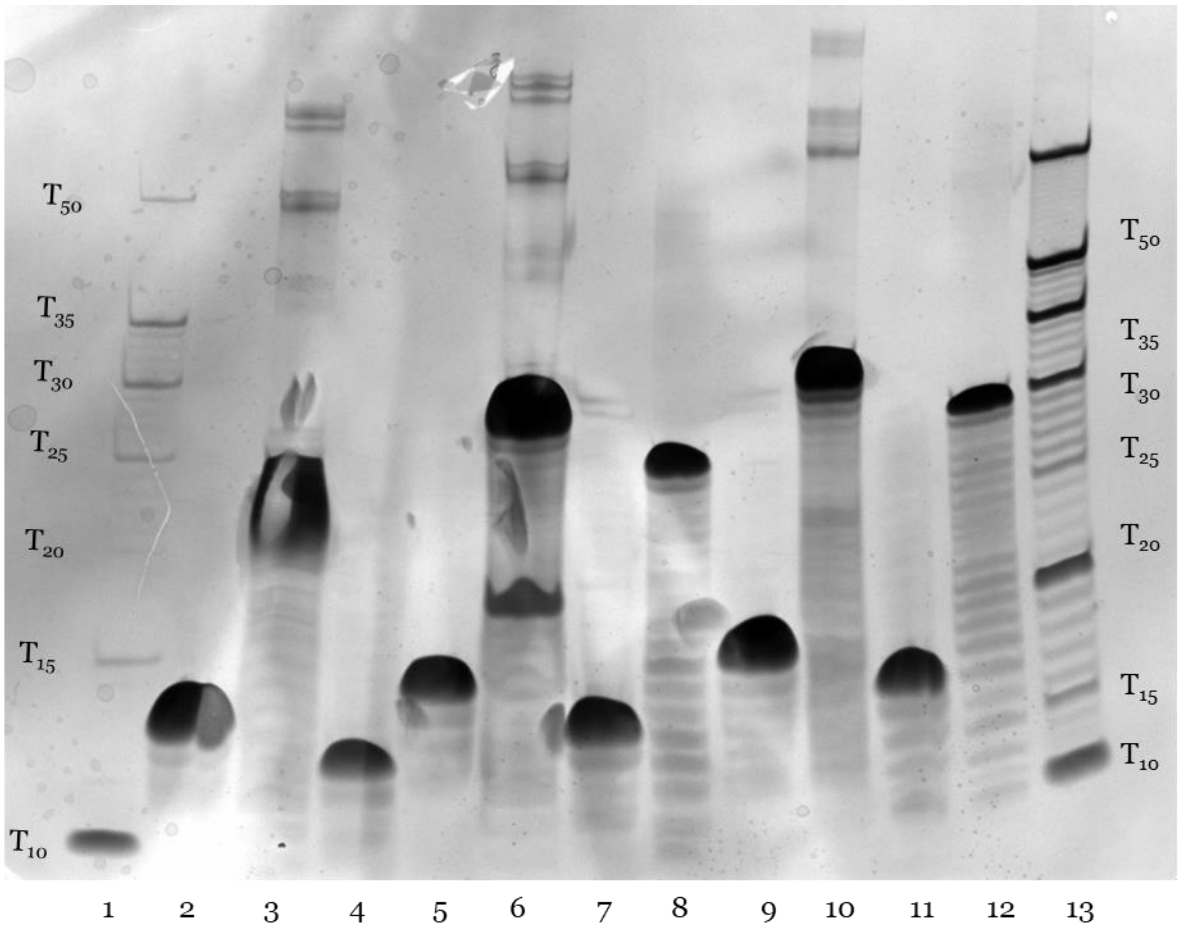


Figure SI-13. dPAGE of unmodified and inverted telomeric oligonucleotides. Lane 1: $T_5 – T_{50}$ ladder, lane 2: tel, lane 3: 5′inv-tel, lane 4: c-tel, lane 5: Ttel, lane 6: 5′inv-Ttel, lane 7: c-Ttel, lane 8: c-5′inv-Ttel, lane 9: telTA, lane 10: 5′inv-telTA, lane 11: c-telTA, lane 12: c-5′inv-telTA, lane 13: $T_5 – T_{50}$ ladder. Conditions: 100 μM strand concentration, 20% polyacrylamide gel, 7M Urea.

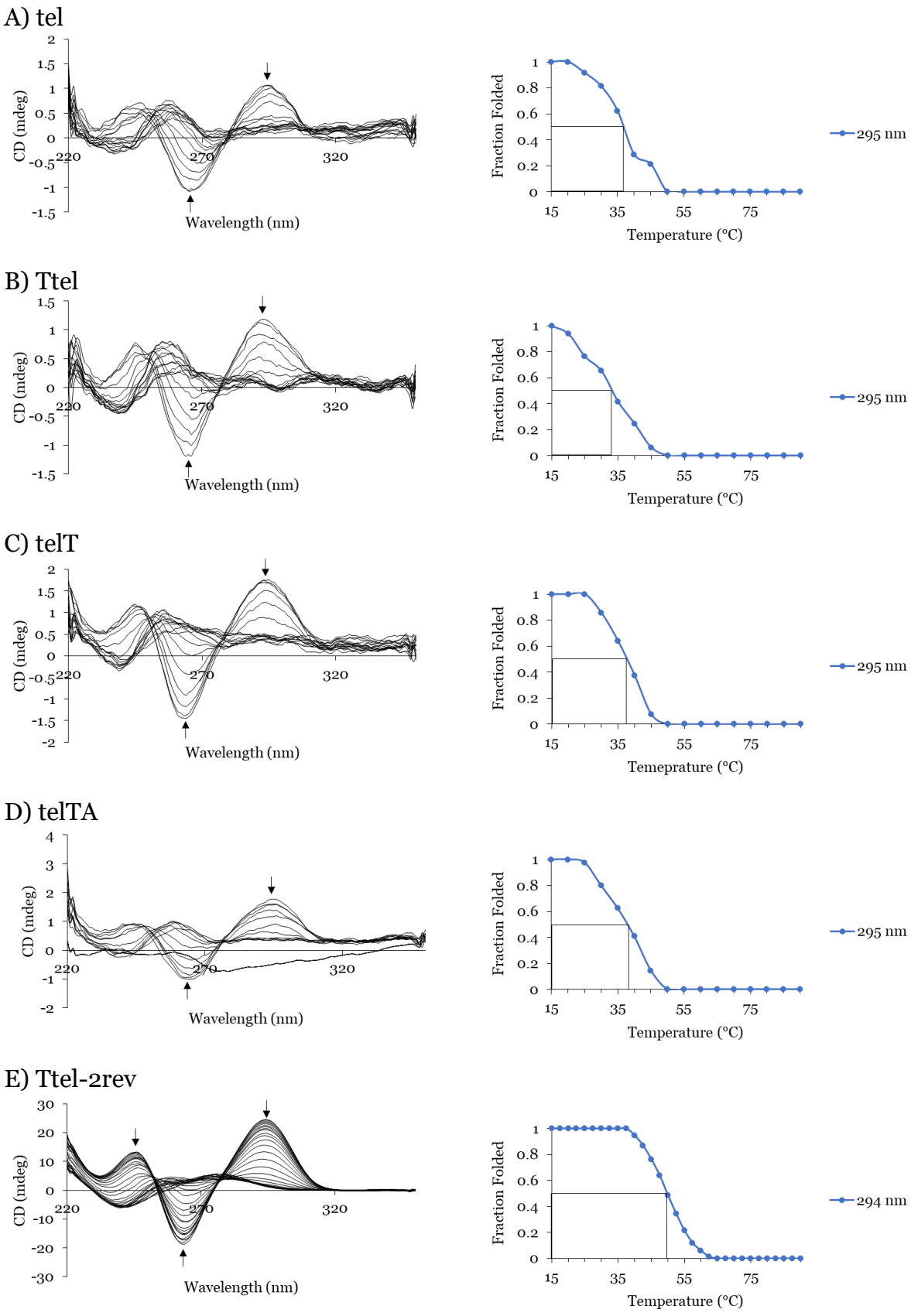
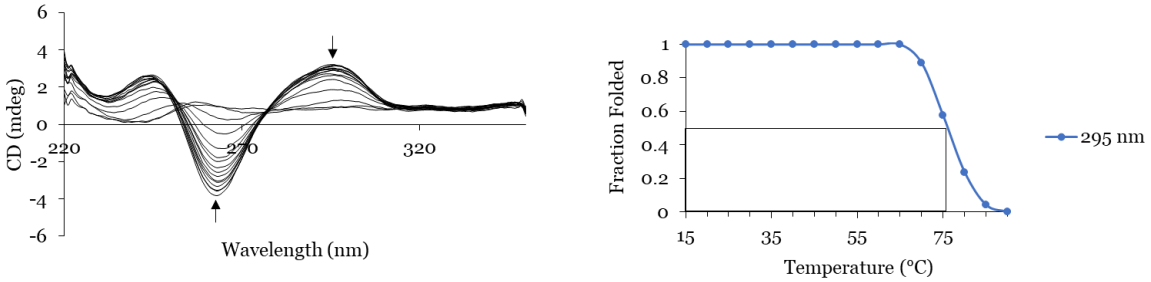
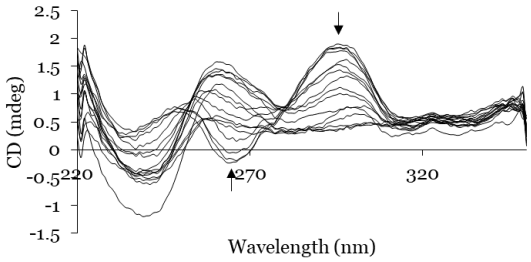


Figure SI-14. CD melting profiles of A) tel, B) Ttel, C) telT, D) telTA and E) Ttel-2rev in Na^+ buffer. Conditions: 20 µM strand concentration, 10 mM lithium cacodylate, 100 mM NaCl, pH 7.2. 15 – 90 °C, 1 °C/min. Arrows indicate direction of change in peak intensity.

A) 5′inv-tel

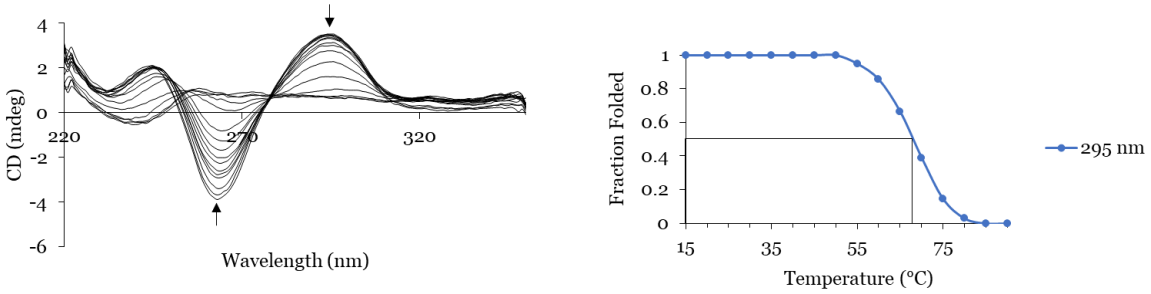


B) 5′inv-Ttel



Not completely unfolded

C) 5′inv-telT



D) 5′inv-telTA

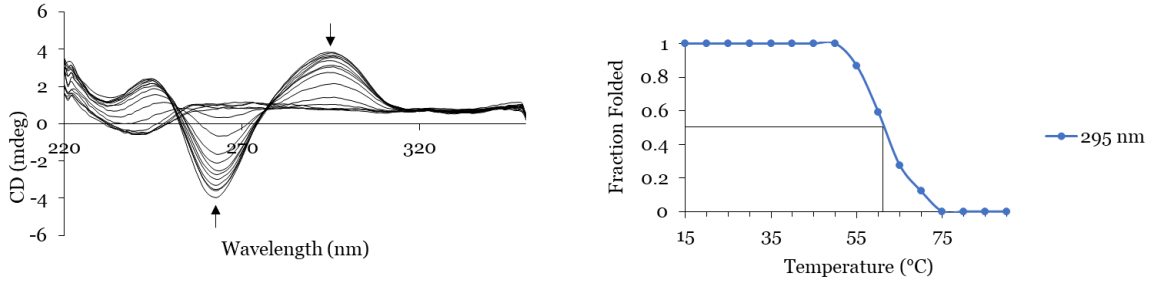


Figure SI-15. CD melting profiles of A) 5′inv-tel, B) 5′inv-Ttel, C) 5′inv-telT, and D) 5′inv-telTA in Na^+ buffer. Conditions: 20 μM strand concentration, 10 mM lithium cacodylate, 100 mM NaCl, pH 7.2. 15 – 90 °C, 1 °C/min. Arrows indicate direction of change in peak intensity.

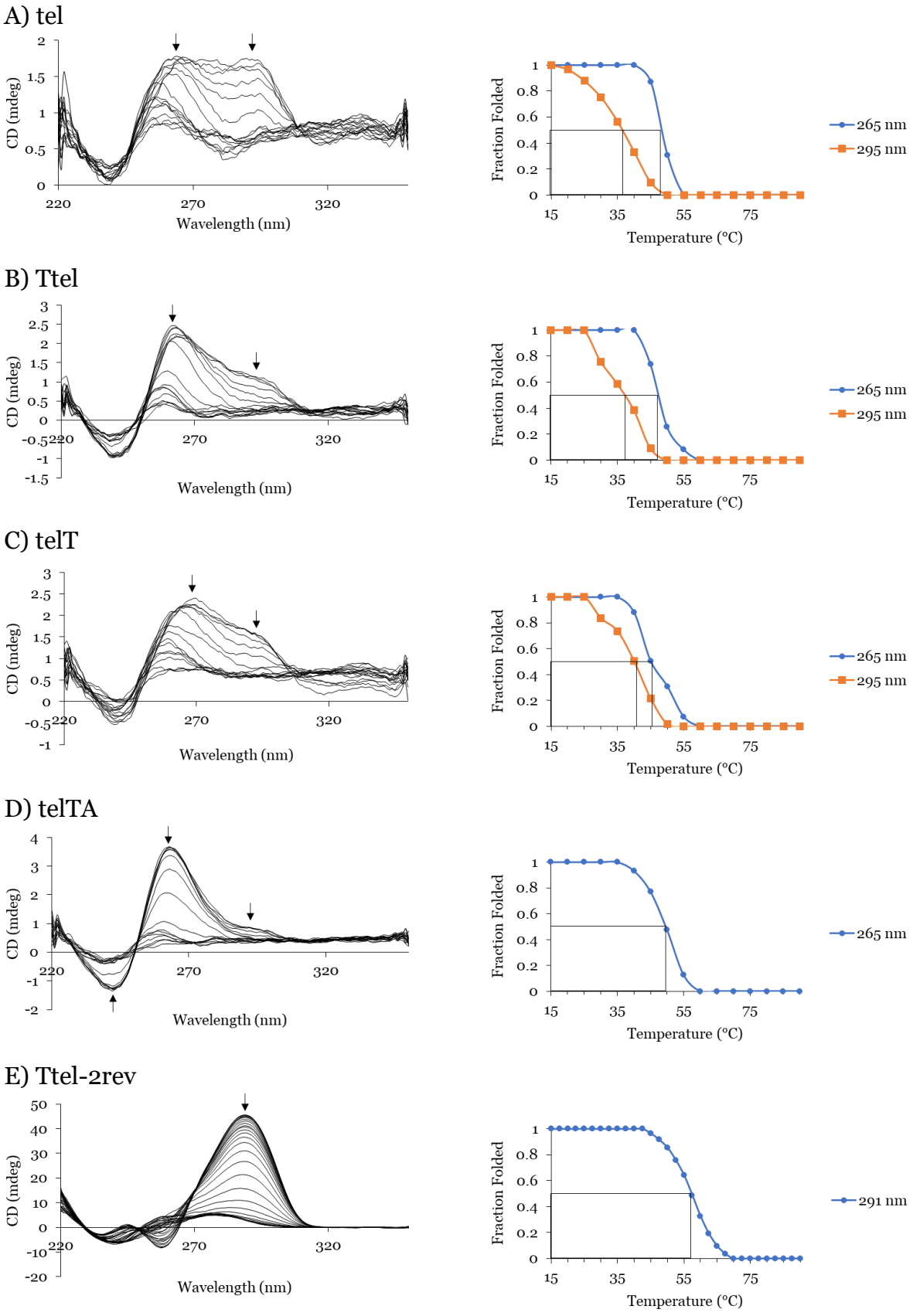
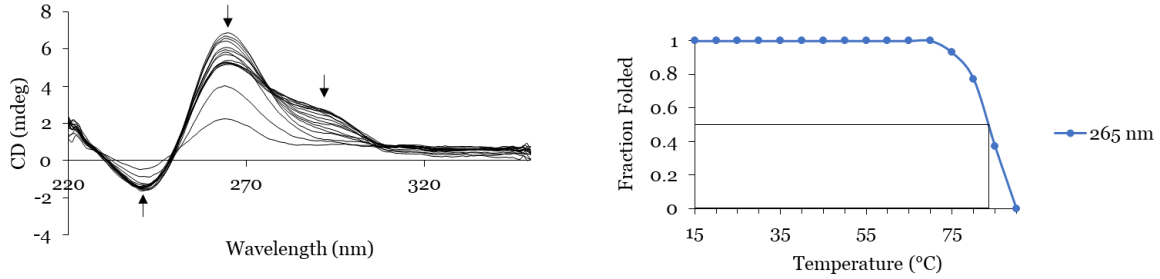
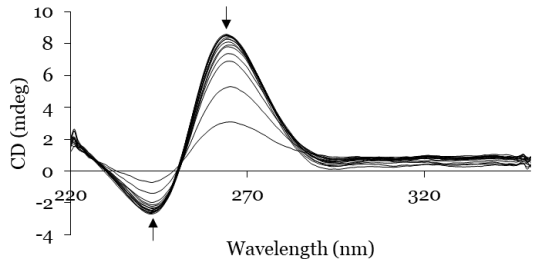


Figure SI-16. CD melting profiles of A) tel, B) Ttel, C) telT, D) telTA and E) Ttel-2rev in K⁺ buffer. Conditions: 20 μM strand concentration, 10 mM lithium cacodylate, 100 mM KCl, pH 7.2. 15 – 90 °C, 1 °C/min. Arrows indicate direction of change in peak intensity.

A) 5′inv-tel

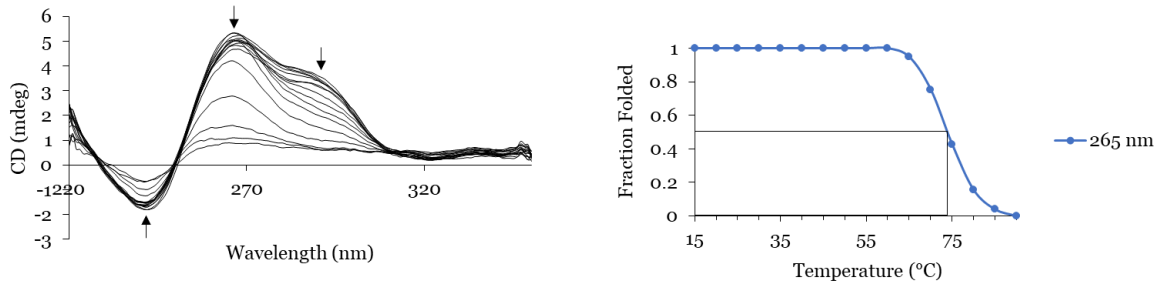


B) 5′inv-Ttel



Not completely unfolded

C) 5′inv-telT



D) 5′inv-telTA

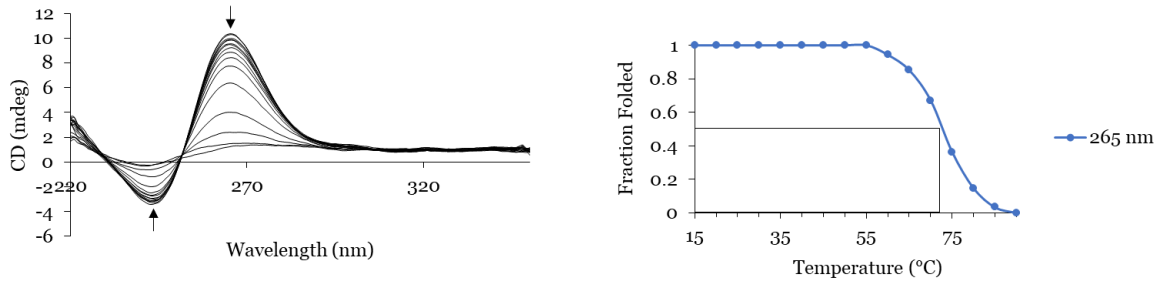
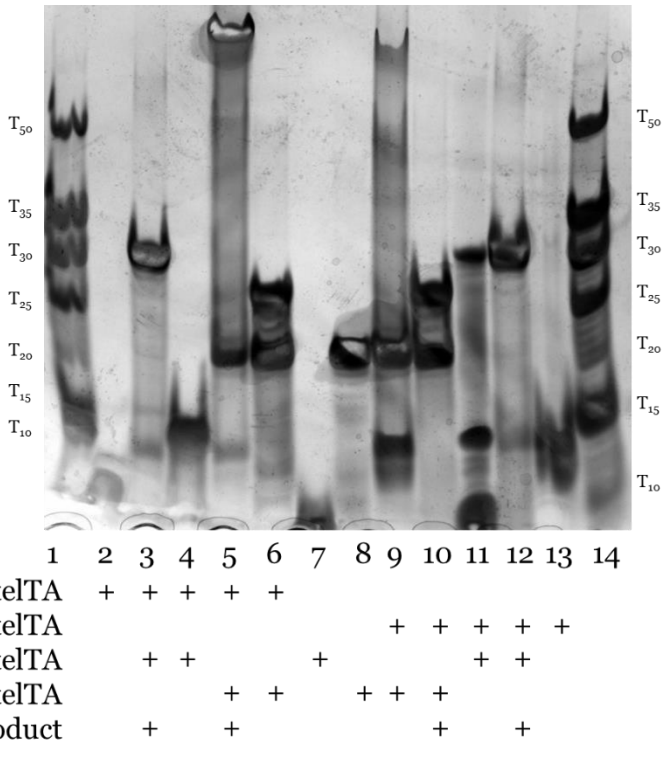


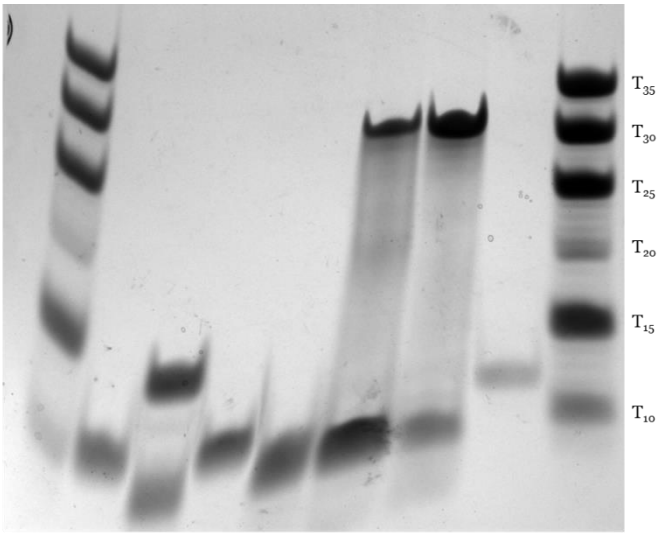
Figure SI-17. CD melting profiles of A) 5′inv-tel, B) 5′inv-Ttel, C) 5′inv-telT, and D) 5′inv-telTA in K^+ buffer. Conditions: 20 μM strand concentration, 10 mM lithium cacodylate, 100 mM KCl, pH 7.2. 15 – 90 °C, 1 °C/min. Arrows indicate direction of change in peak intensity.

A)



	1	2	3	4	5	6	7	8	9	10	11	12	13	14
telTA		+	+	+	+	+								
3'inv-telTA									+	+	+	+	+	
c-telTA			+	+			+				+	+		
c-3'inv-telTA					+	+		+	+	+				
Thermodynamic Product			+		+					+		+		

B)



	1	2	3	4	5	6	7	8	9
tel		+	+	+					
5'inv-tel						+	+	+	
c-tel			+	+	+	+	+		
Thermodynamic Product			+				+		

Figure SI-18. Native gels of telomeric G4s (A) telTA and 3'inv-telTA and B) tel and 5'inv-tel) with their respective complementary sequences (Table 4.1). Conditions: 100 µM strand concentration, 20% polyacrylamide gel, 10 mM lithium cacodylate, 100 mM KCl, pH 7.2.

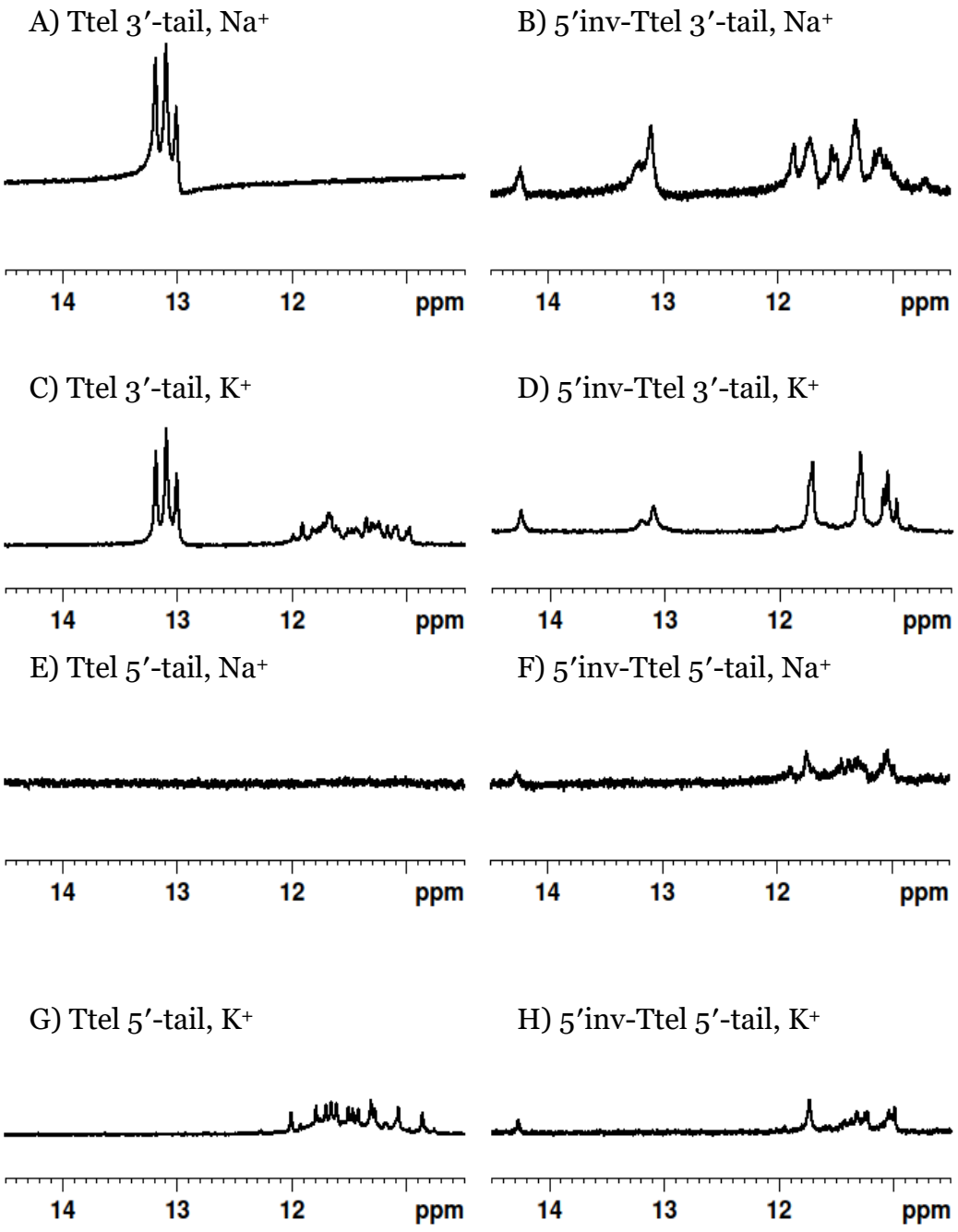


Figure SI-19. ^{1}H NMR of telomeric sequences with duplex-forming tails in different buffers. (A) Ttel 3′-tail, Na^+, (B) 5′inv-Ttel 3′-tail, Na^+, (C) Ttel 3′-tail, K^+, (D) 5′inv-Ttel 3′-tail, K^+, (E) Ttel 5′-tail, Na^+, (F) 5′inv-Ttel 5′-tail, Na^+, (G) Ttel 5′-tail, K^+, (H) 5′inv-Ttel 5′-tail, K^+. Conditions: 300 μM strand concentration, 20 mM sodium phosphate, 10% D_2O, 1% TSP, 10 mM KCl added (E – H), pH 7.0.

HP1α Hinge + His-tag (underlined)

MHHHHHHDYDIPTTENLYFQGKKYKKMKEGENNKPREKSEGNKRKSSFSNSADDIKSKKKREQSN
DIAR

HP1α + His-tag (underlined) – Hinge in bold

MHHHHHHDYDIPTTENLYFQGAMGSGKKTKRTADSSSSEDEEEYVVEKVLDRRMVKGQ
VEYLLKWKGFSEEHNTWEPEKNLDCPELISEFM**KKYKKMKEGENNKPREKSEGNKRKS**
SFSNSADDIKSKKKREQSNDIARGFERGLEPEKIIGATDSCGDLMFLMKWKDTDEADL
VLAKEANVKCPQIVIAFYEERLTWHAYPEDAENKEKESAKS

Figure SI-20. Full peptide sequences of HP1α + His-tag and HP1α hinge region + His-tag, both used in BLItz experiments.

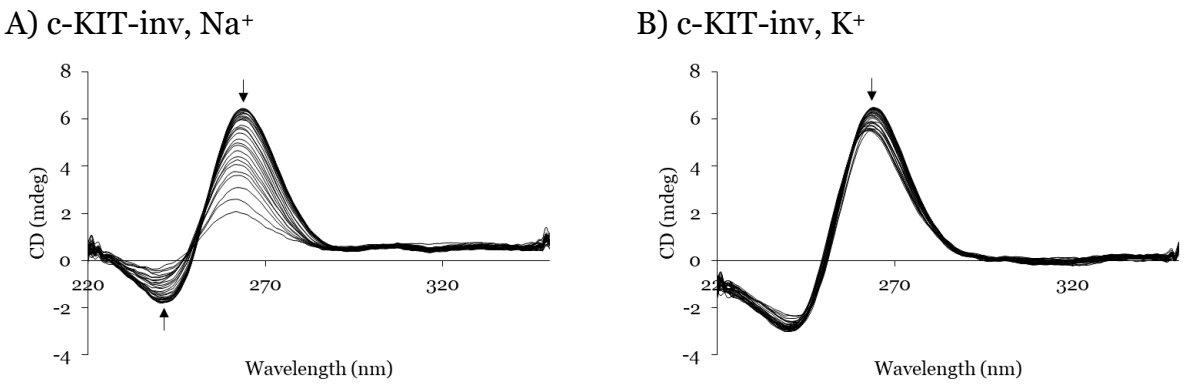


Figure SI-21. CD melting profiles of c-KIT-inv in A) Na^+ and B) K^+ buffer. G4s are not completely unfolded at 90 °C in either buffer. Conditions: 20 μM strand concentration, 20 mM sodium phosphate, 10 mM KCl (in K^+ buffer only), pH 7.0.

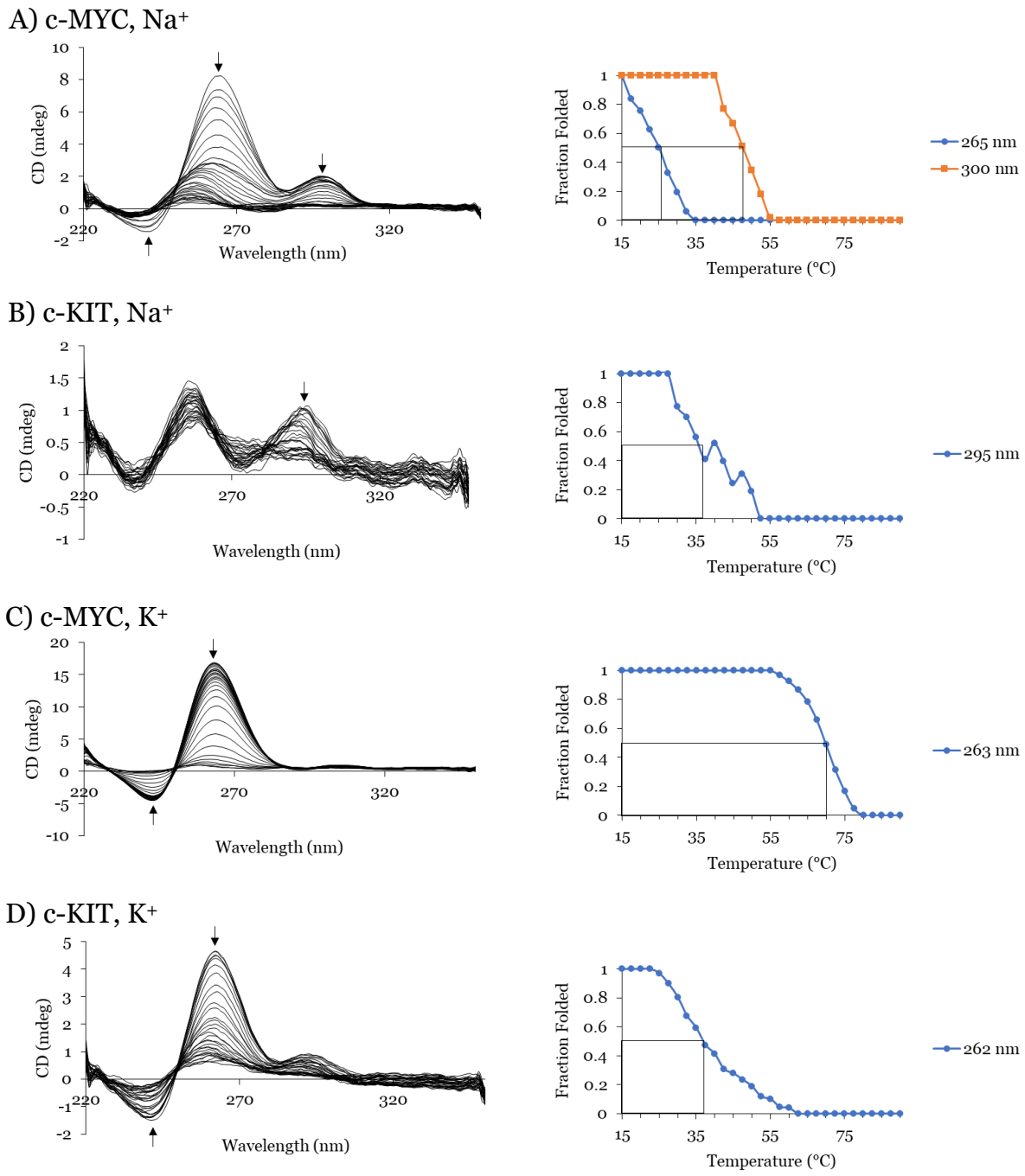


Figure SI-22. CD melting profiles of A) c-MYC and B) c-KIT in Na^+ buffer and melting profiles of C) c-MYC and D) c-KIT in K^+ buffer. Conditions: 20 μM strand concentration, 20 mM sodium phosphate, 10 mM KCl (in K^+ buffer only), pH 7.2. 15 – 90 °C, 1 °C/min. Arrows indicate direction of change in peak intensity.

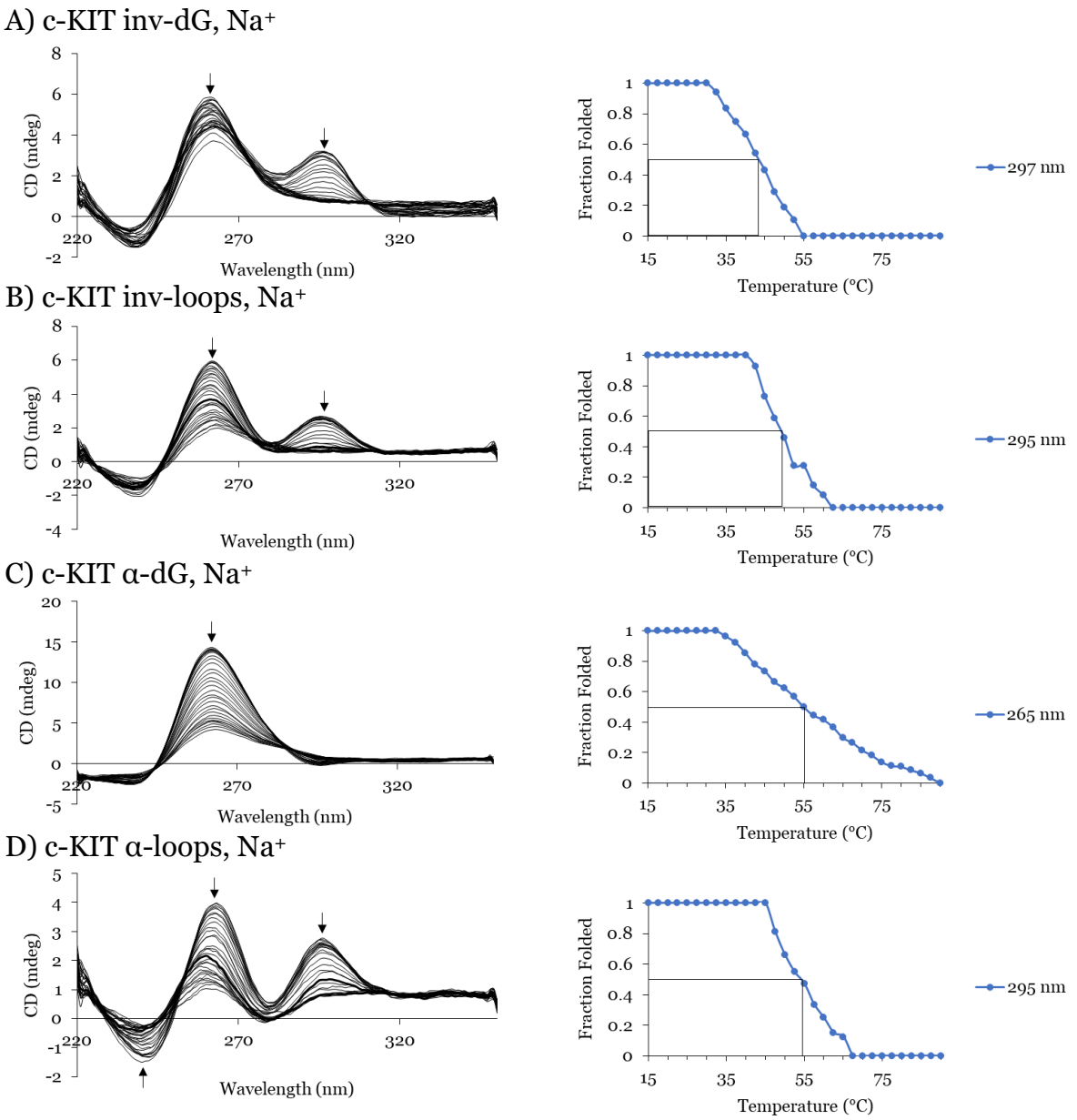


Figure SI-23. CD melting profiles of c-KIT modifications in Na+ buffer. A) inv-dG. B) inv-loops. C) α-dG. D) α-loops. Conditions: 20 μM strand concentration, 20 mM sodium phosphate, pH 7.2. 15 – 90 °C, 1 °C/min. Arrows indicate direction of change in peak intensity.

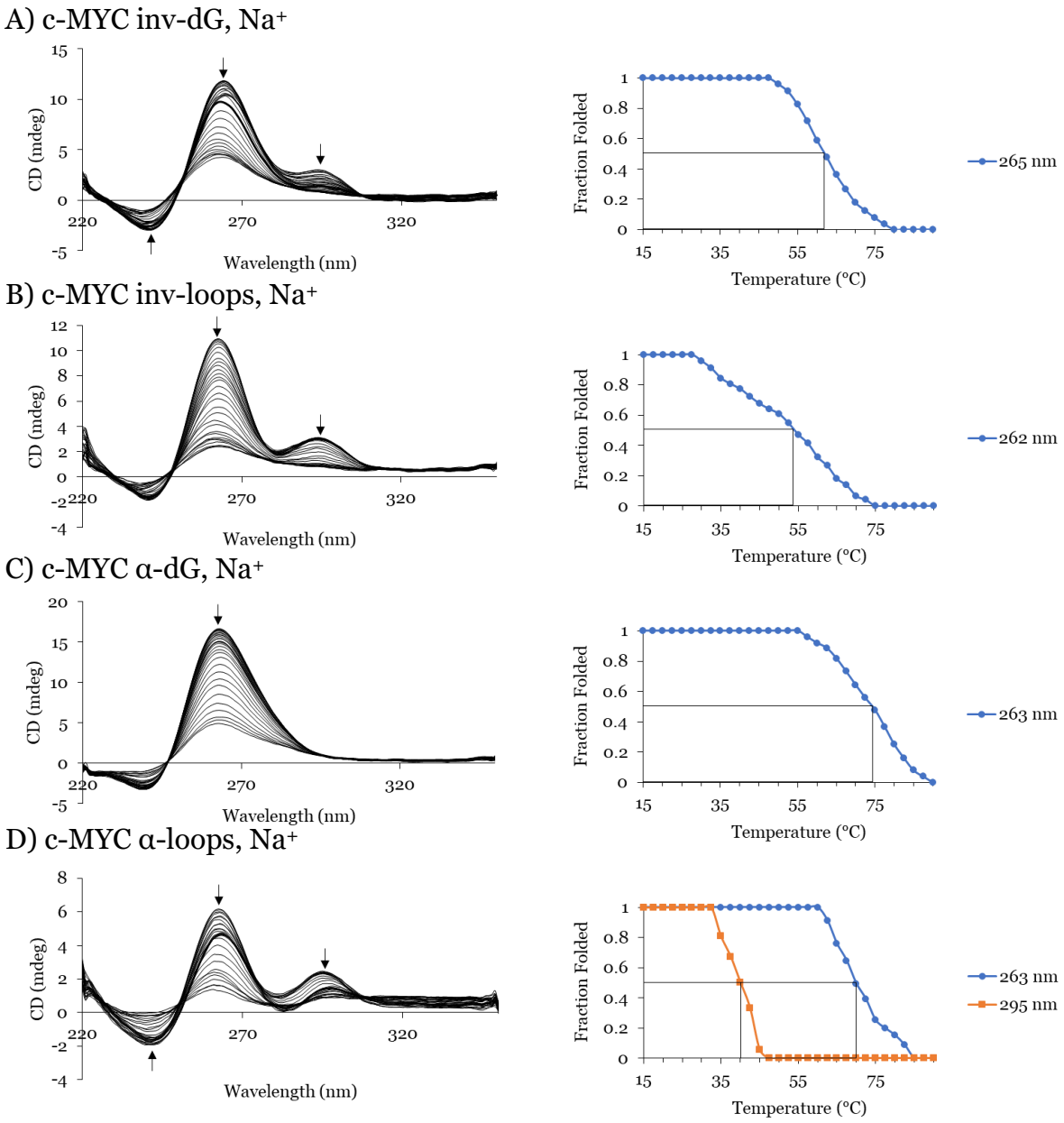


Figure SI-24. CD melting profiles of c-KIT modifications in K⁺ buffer. A) inv-dG. B) inv-loops. C) α-dG. D) α-loops. Conditions: 20 µM strand concentration, 20 mM sodium phosphate, 10 mM KCl, pH 7.2. 15 – 90 °C, 1 °C/min. Arrows indicate direction of change in peak intensity.

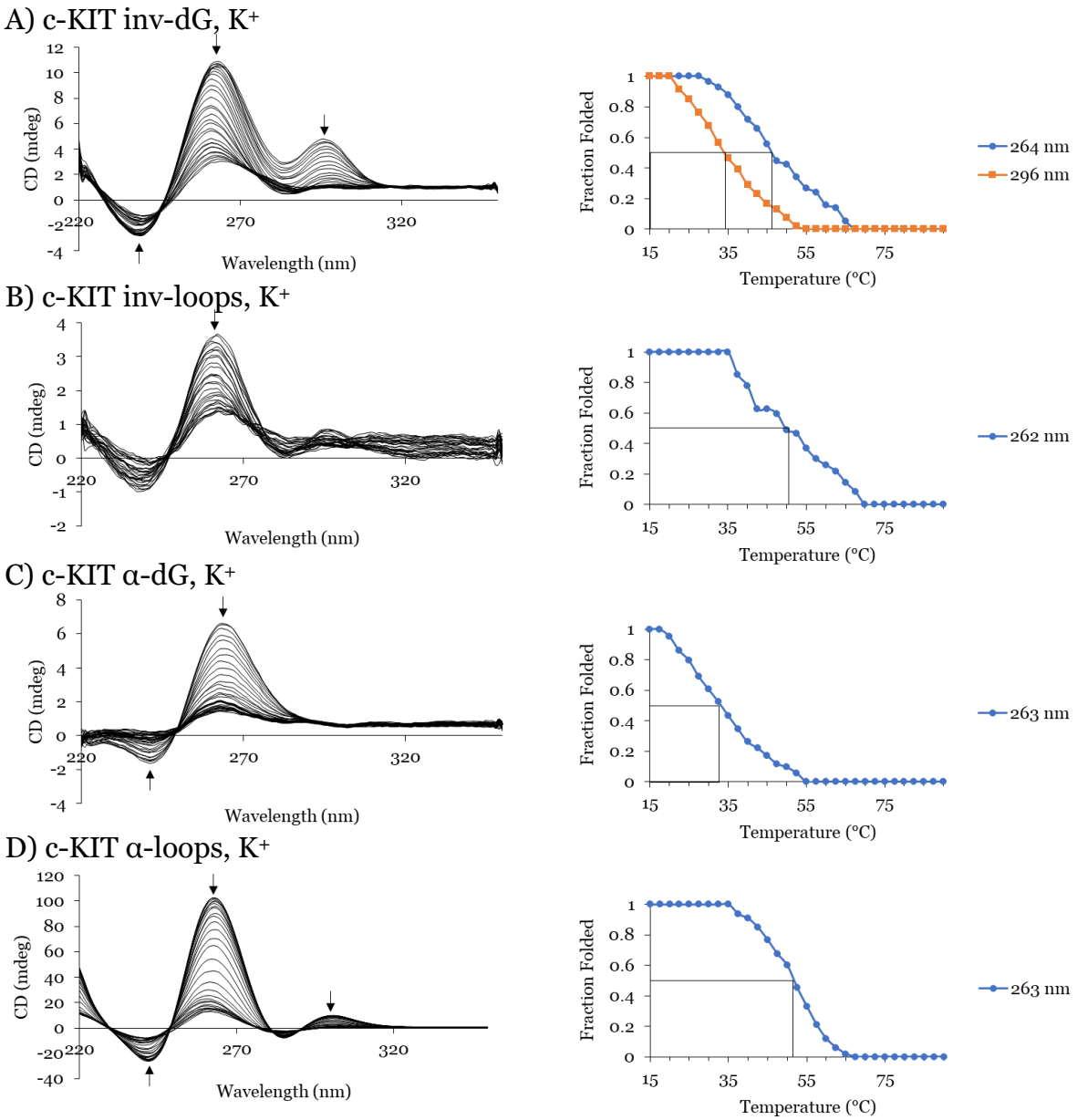


Figure SI-25. CD melting profiles of c-MYC modifications in Na^+ buffer. A) inv-dG. B) inv-loops. C) α-dG. D) α-loops. Conditions: 20 μM strand concentration, 20 mM sodium phosphate, pH 7.2. 15 – 90 °C, 1 °C/min. Arrows indicate direction of change in peak intensity.

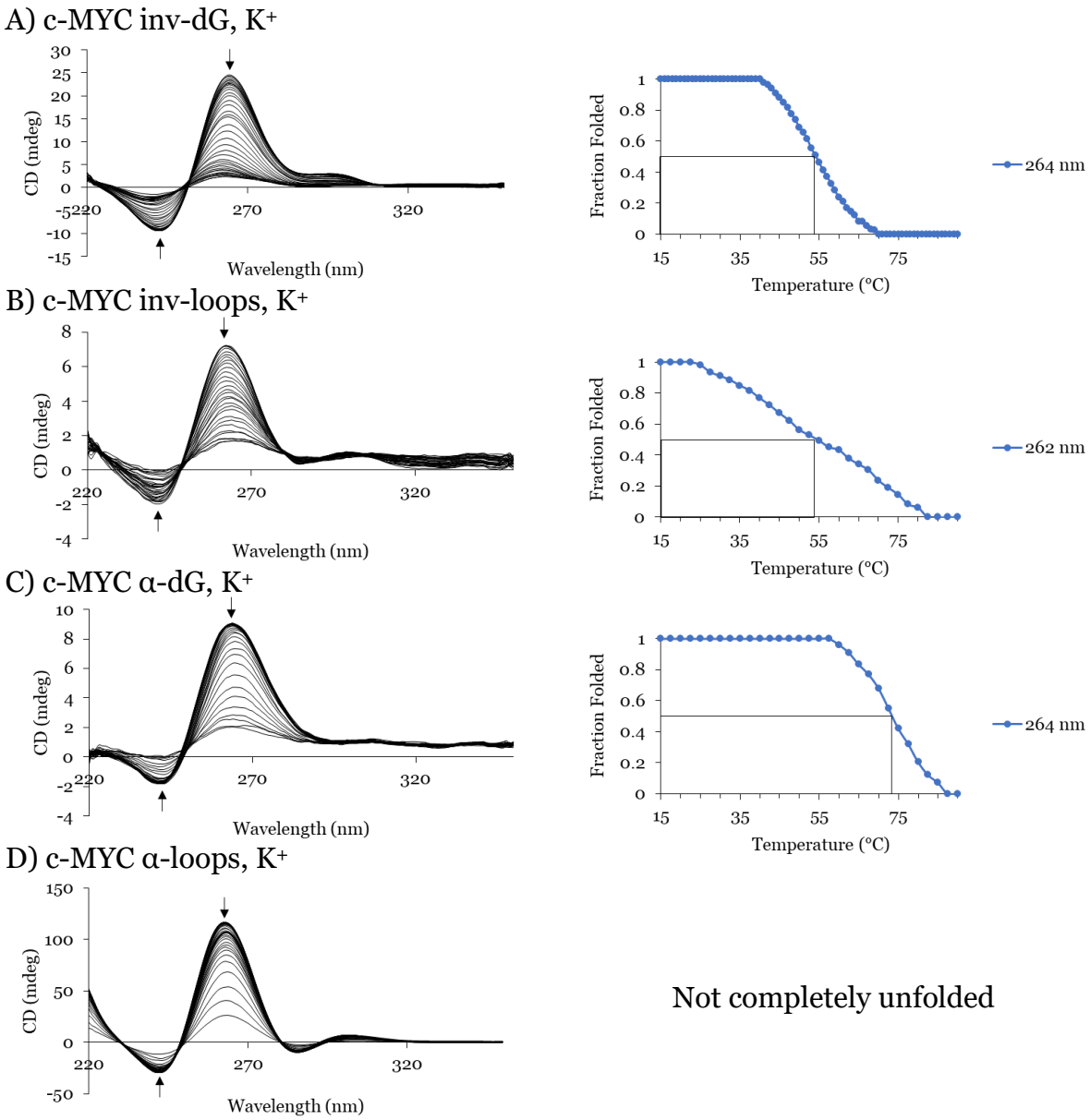


Figure SI-26. CD melting profiles of c-MYC modifications in K+ buffer. A) inv-dG. B) inv-loops. C) α-dG. D) α-loops. Conditions: 20 µM strand concentration, 20 mM sodium phosphate, 10 mM KCl, pH 7.2. 15 – 90 °C, 1 °C/min. Arrows indicate direction of change in peak intensity.

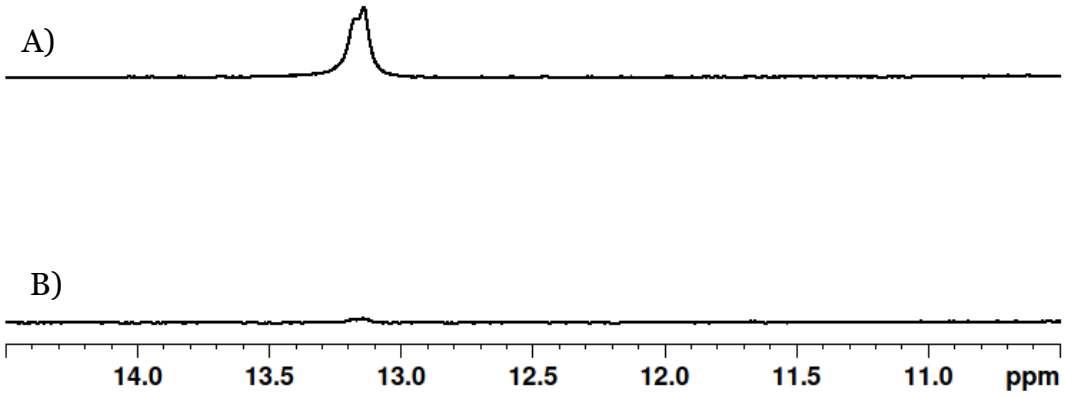


Figure SI-28. ^{1}H NMR profiles of A) i-c-KIT and B) i-c-MYC controls, in particular highlighting peaks resulting from DNA hairpin formation in single-stranded i-c-KIT. Conditions: 300 µM strand concentration, 20 mM sodium phosphate, 10 mM KCl, 10% D_2O, 1% TSP, pH 7.0.

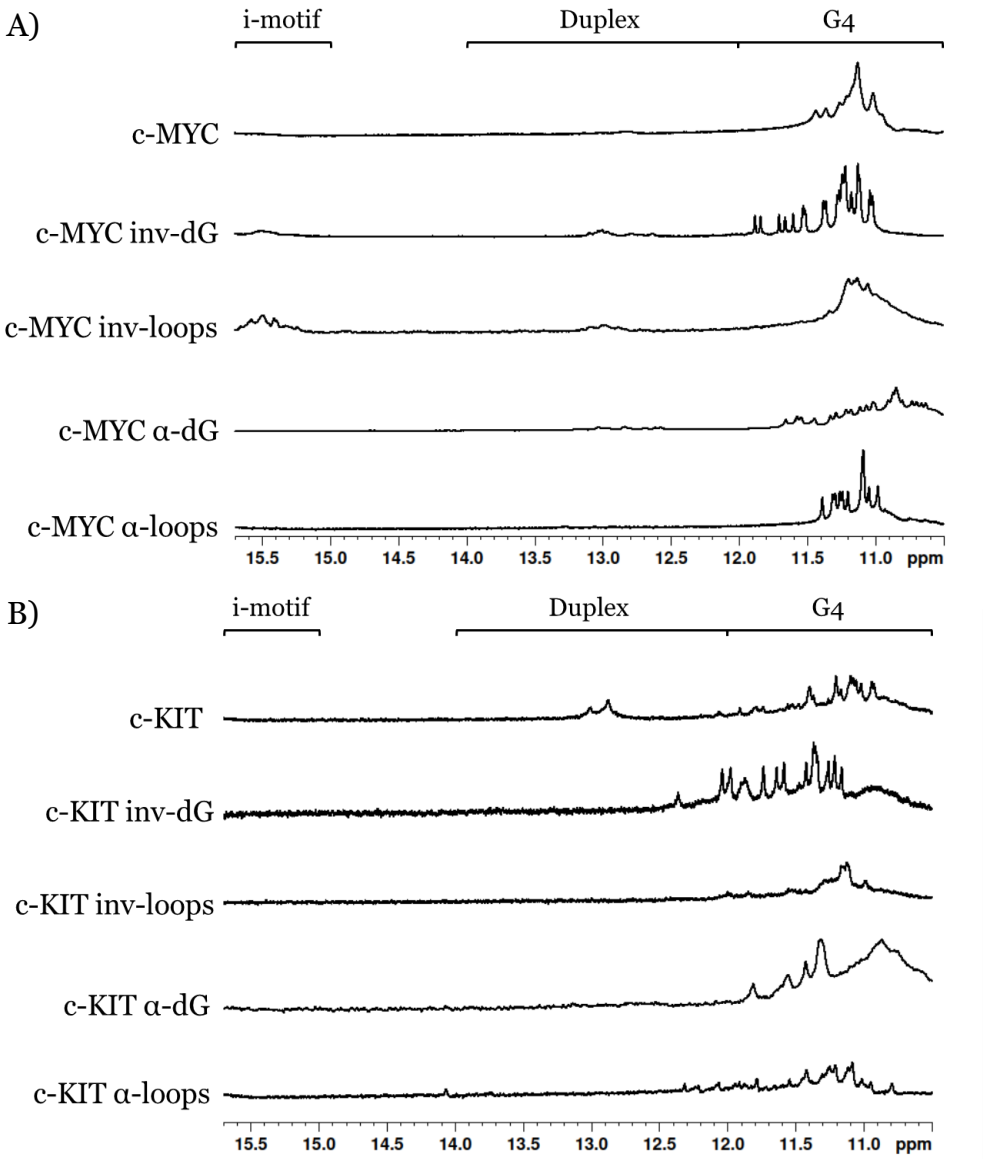


Figure SI-27. Comparison of ^{1}H NMR spectra of all A) c-MYC and B) c-KIT sequences in the presence of their complementary strand. I-motif formation, indicated by peaks around 15 – 16 ppm is observed in some c-MYC sequences (c-MYC-inv-dG and c-MYC-inv-loops in particular). Conditions: approx. 200 µM strand concentration, 20 mM sodium phosphate, 10 mM KCl, 10% D_2O, 1% TSP, pH 7.0.

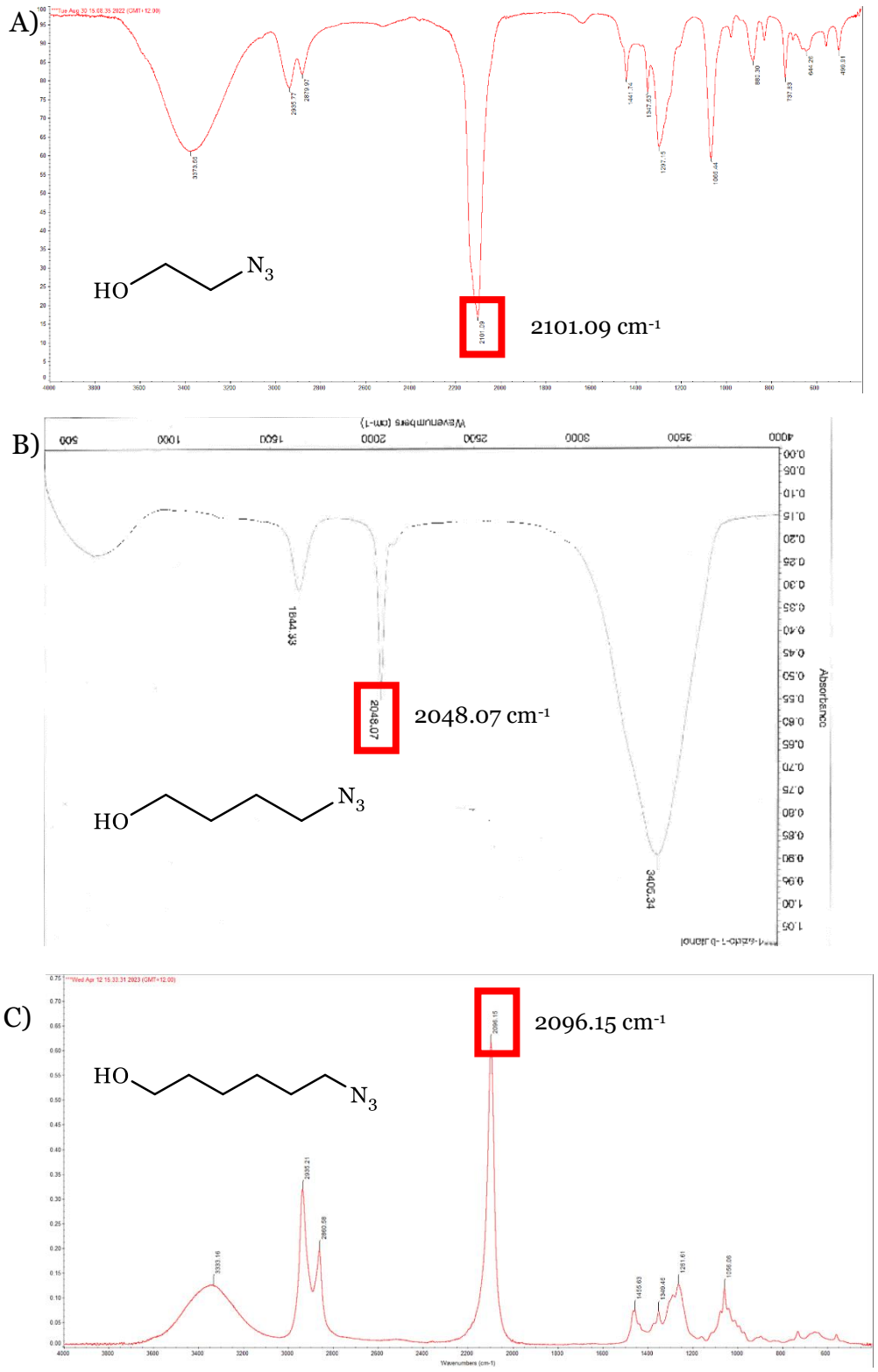


*Figure SI-29. IR spectra of A) 2-azido-1-ethanol (**1a**), B) 4-azido-1-butanol (**1b**) and C) 6-azido-1hexanol (**1c**). Presence of azide stretch at approx. 2100 cm⁻¹ indicates that product has been formed.*

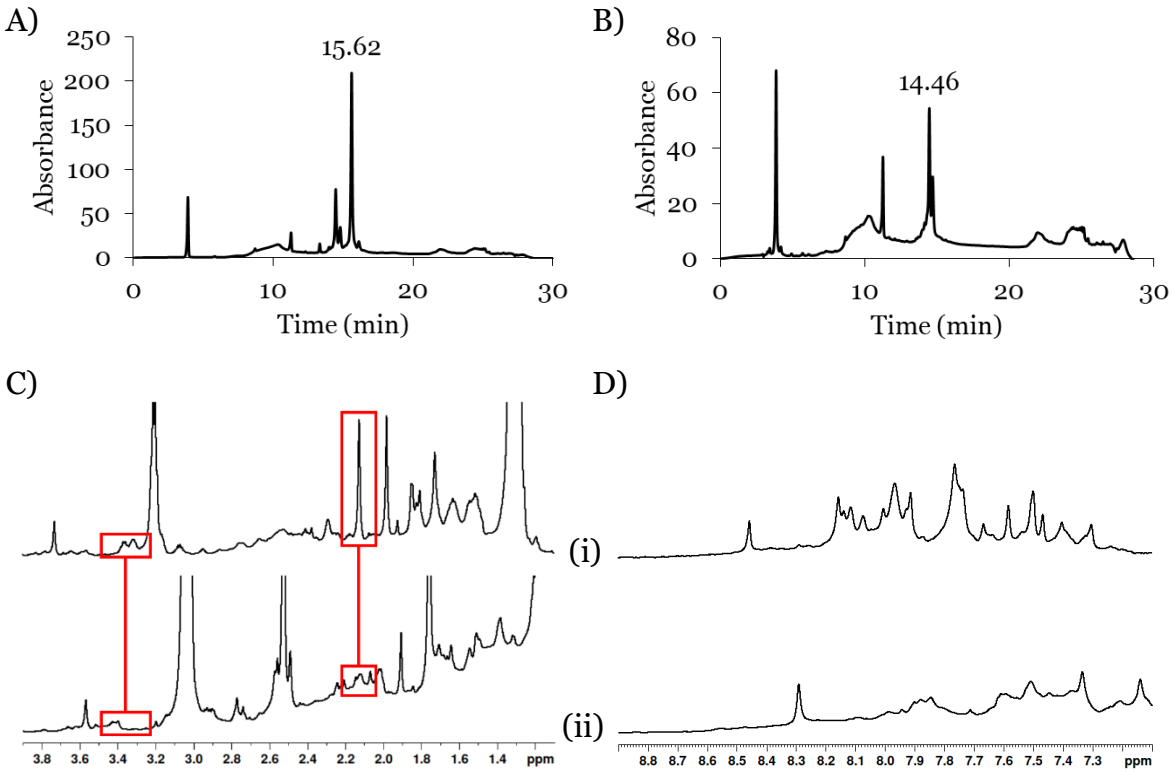


Figure SI-30. Comparisons of pre- and post-cross-linking tel-G5A8 sequences using RP-HPLC and ^{1}H NMR. A) RP-HPLC of tel-G5A8 and B) RP-HPLC profile of tel-G5A8-X, showing an approximately one-minute decrease in retention time. C) ^{1}H NMR of (i) tel-G5A8 and (ii) tel-G5A8-X. Highlighted regions, from left to right, indicate shift in protons corresponding to CH_2 group adjacent to azide in adenosine modification and disappearance of propargyl proton. Slight shift in H-2′ of propargyl guanosine in less clear in this figure than in Figure 5.6. D) ^{1}H NMR of (i) tel-G5A8 and (ii) tel-G5A8-X. Triazole peak could be expected to be visible in this region, but this peak is obscured by the dominance of aromatic protons in this region (e.g., H-8 of guanosine or H-8 and H-2 of adenosine). However, significant changes in chemical shift in this region are obvious. Conditions: approx. 200 µM strand concentration, 20 mM sodium phosphate, 10 mM KCl, 10% D_2O, 1% TSP, pH 7.0.

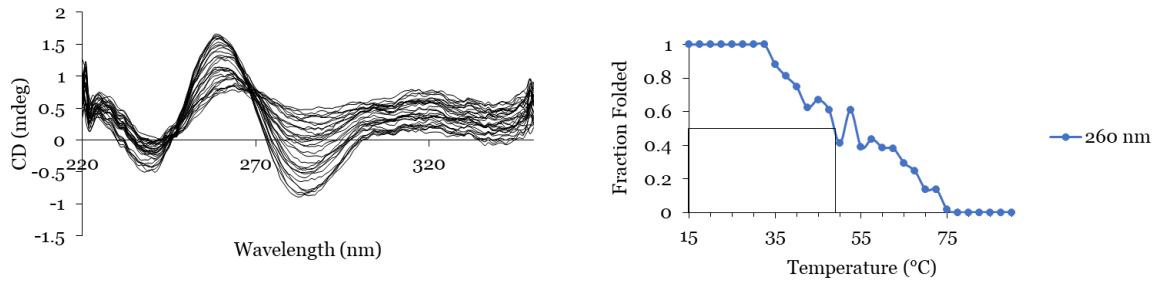


Figure SI-31. CD melting profile of tel-G3A8-X in Na^+ buffer. This in the only sequence to form a G4 in Na^+ buffer, but T_m value is approximate because it appears to only be partially folded. Conditions: 20 µM strand concentration, 20 mM sodium phosphate, pH 7.0.

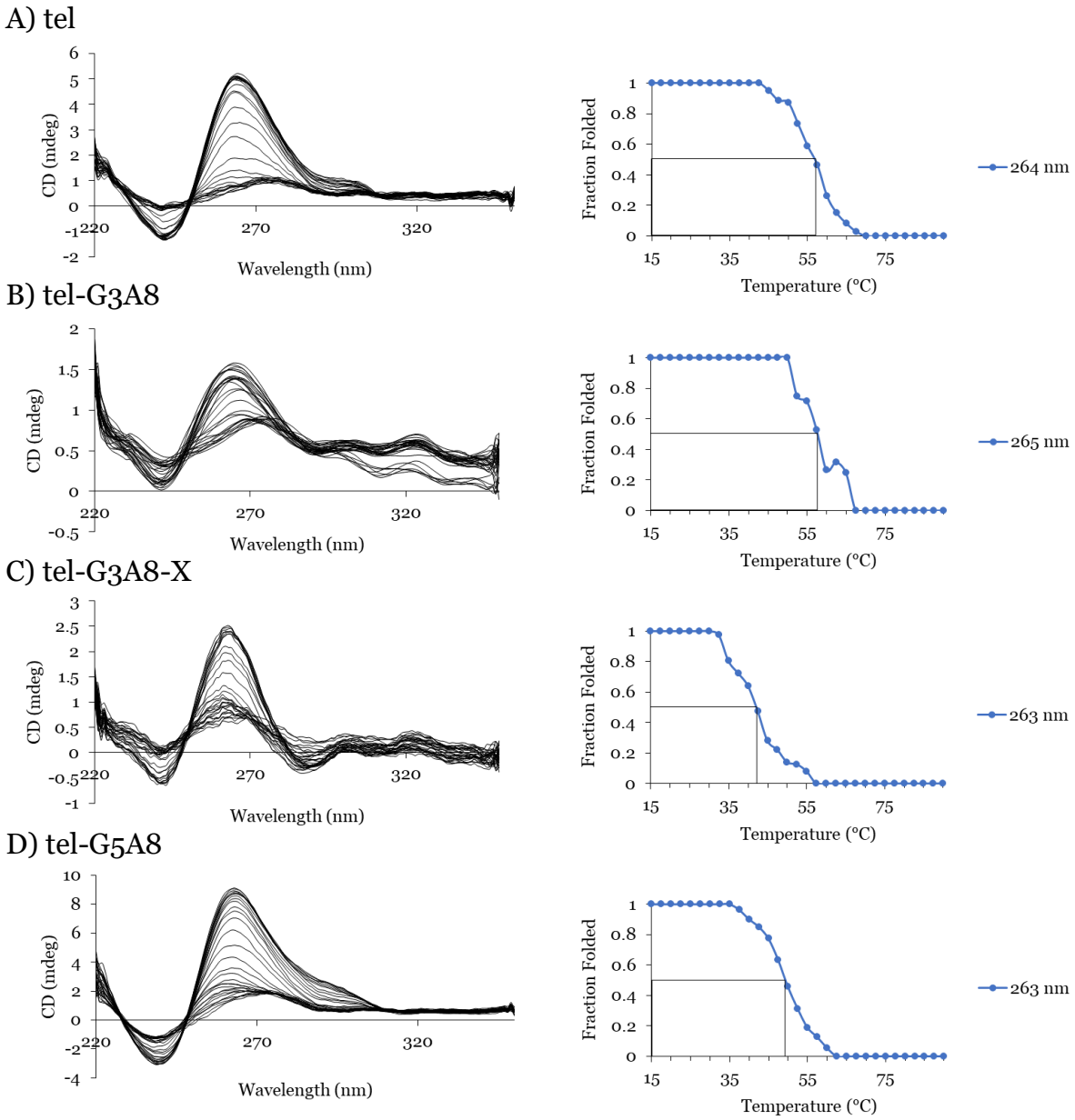


Figure SI-32. CD melting profiles of telomeric G4-forming sequences with cross-links in K^+ buffer. A) tel, B) tel-G3A8, C) tel-G3A8-X and D) tel-G5A8. Conditions: 20 µM strand concentration, 20 mM sodium phosphate, 10 mM KCl, pH 7.2. 15 – 90 °C, 1 °C/min. Arrows indicate direction of change in peak intensity.

Appendix B: NMR Spectroscopy, HPLC and Mass Spectrometry

RP-HPLC and mass spectra in Figure SI-33 – Figure SI-38 correspond to Table 7.2:

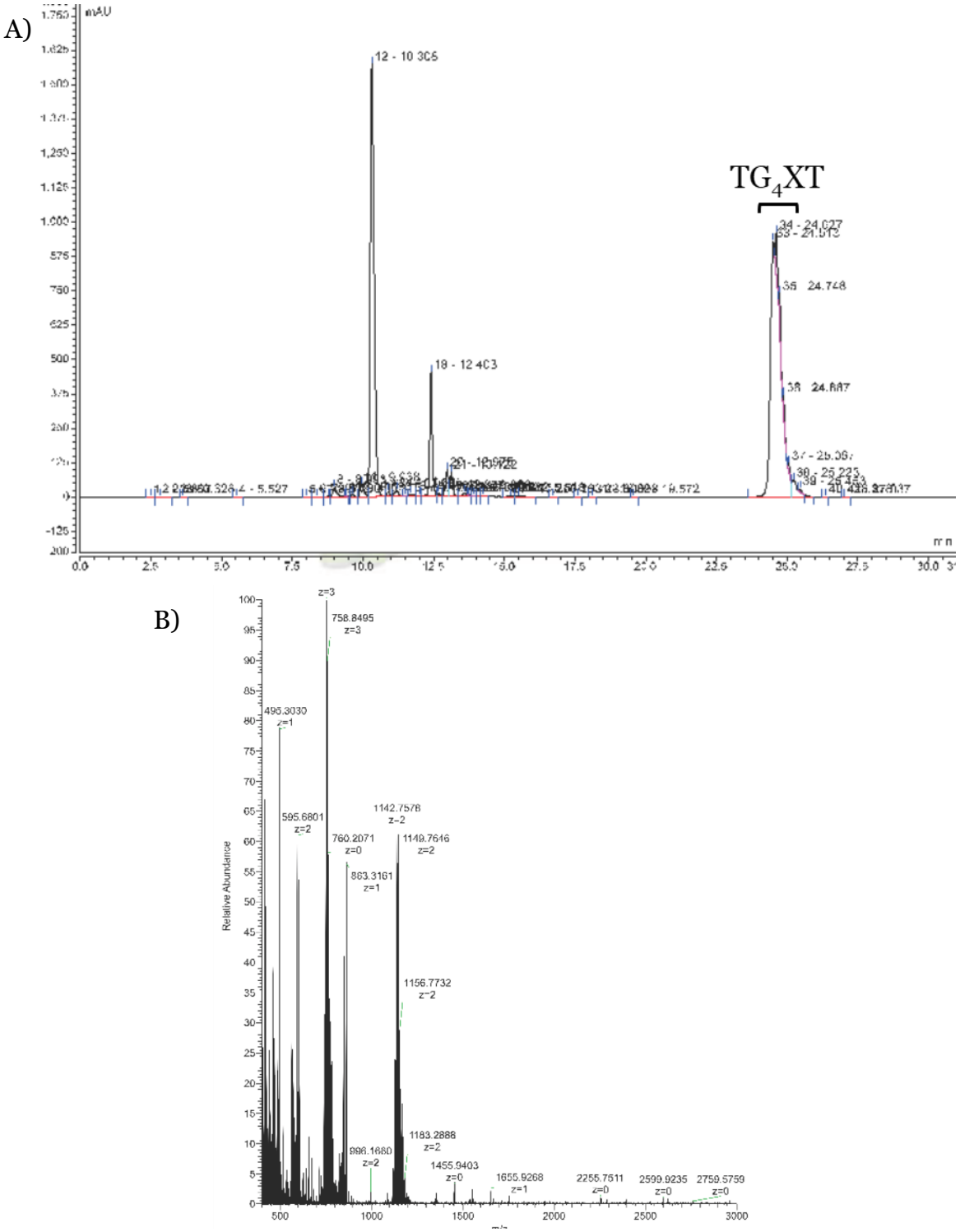
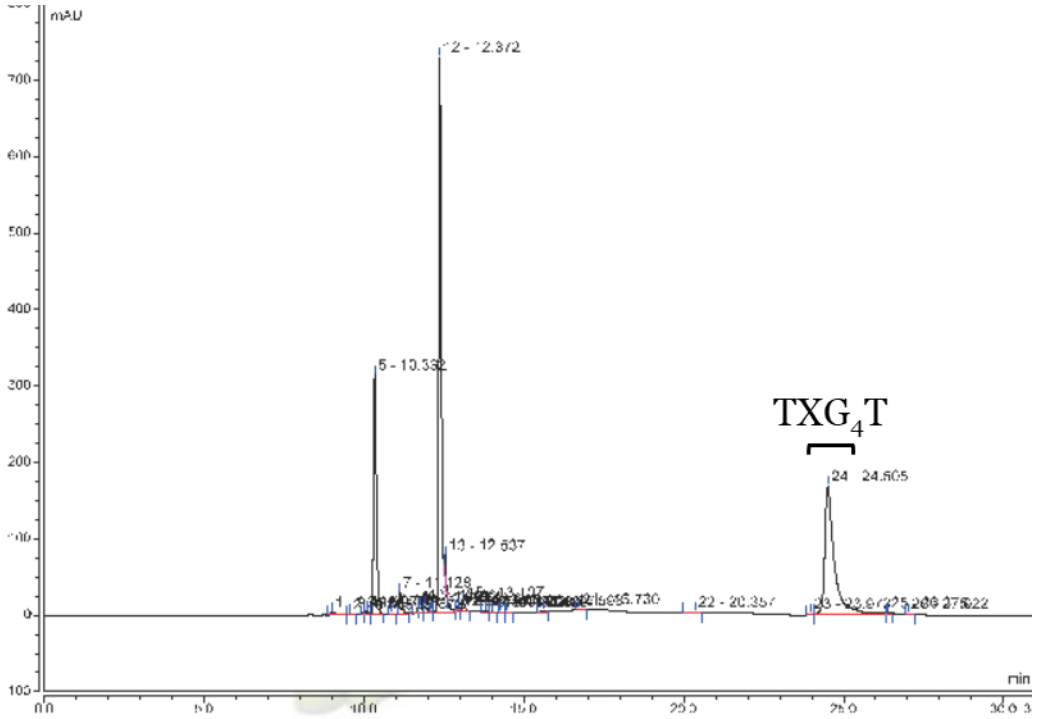


Figure SI-33. A) RP-HPLC profile of the TG_4XT sequence and B) corresponding ESI-MS confirming synthesis was successful.



B)

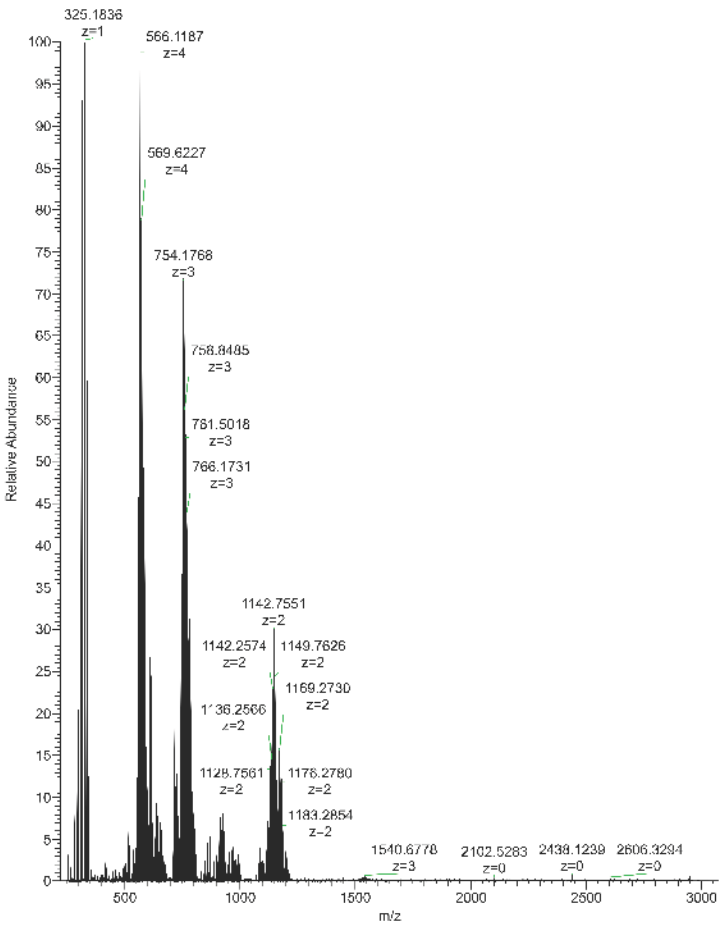


Figure SI-34. RP-HPLC profile of the TXG$_4$T sequence and B) corresponding ESI-MS confirming synthesis was successful.

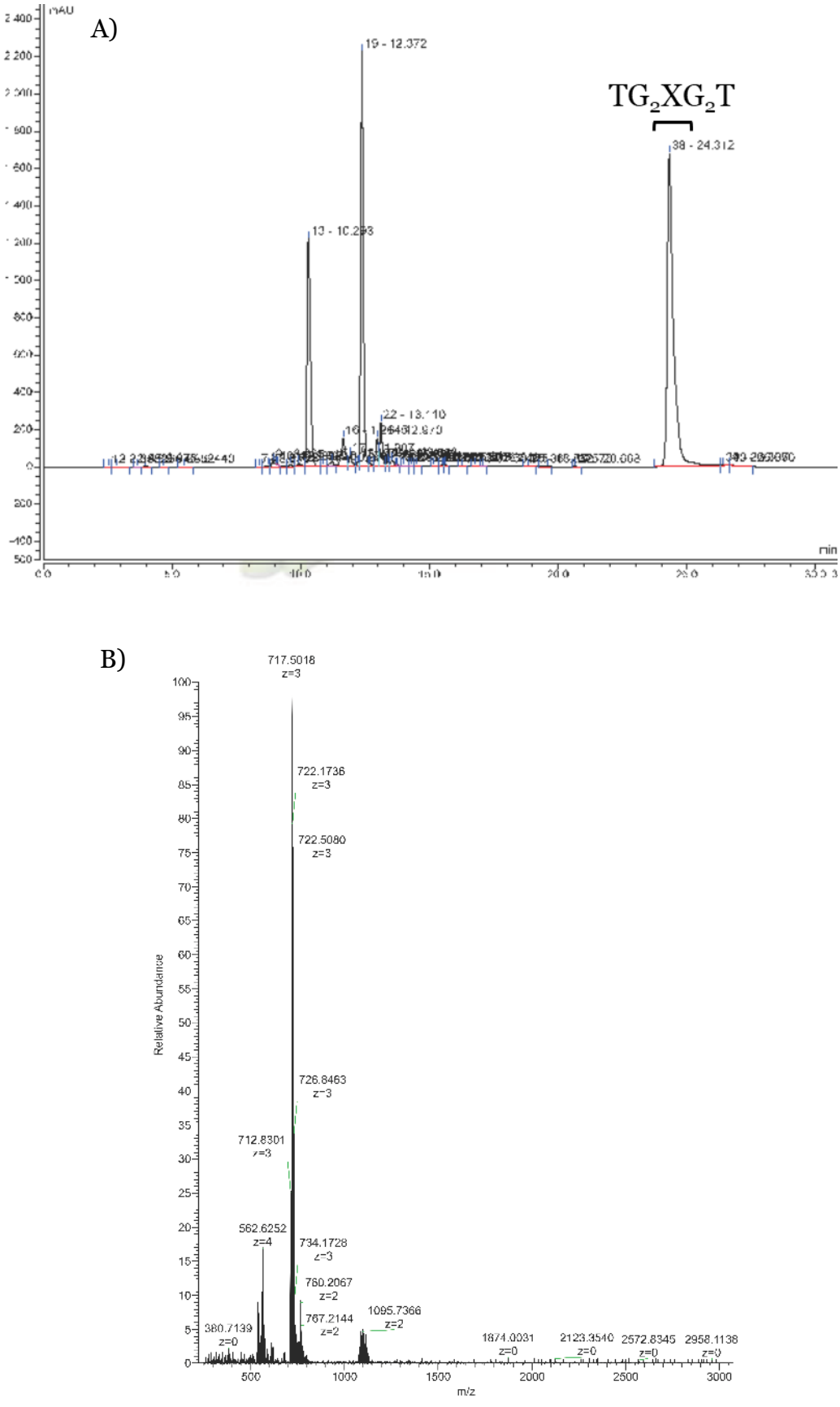


Figure SI-35. RP-HPLC profile of the TG_2XG_2T sequence and B) corresponding ESI-MS confirming synthesis was successful.

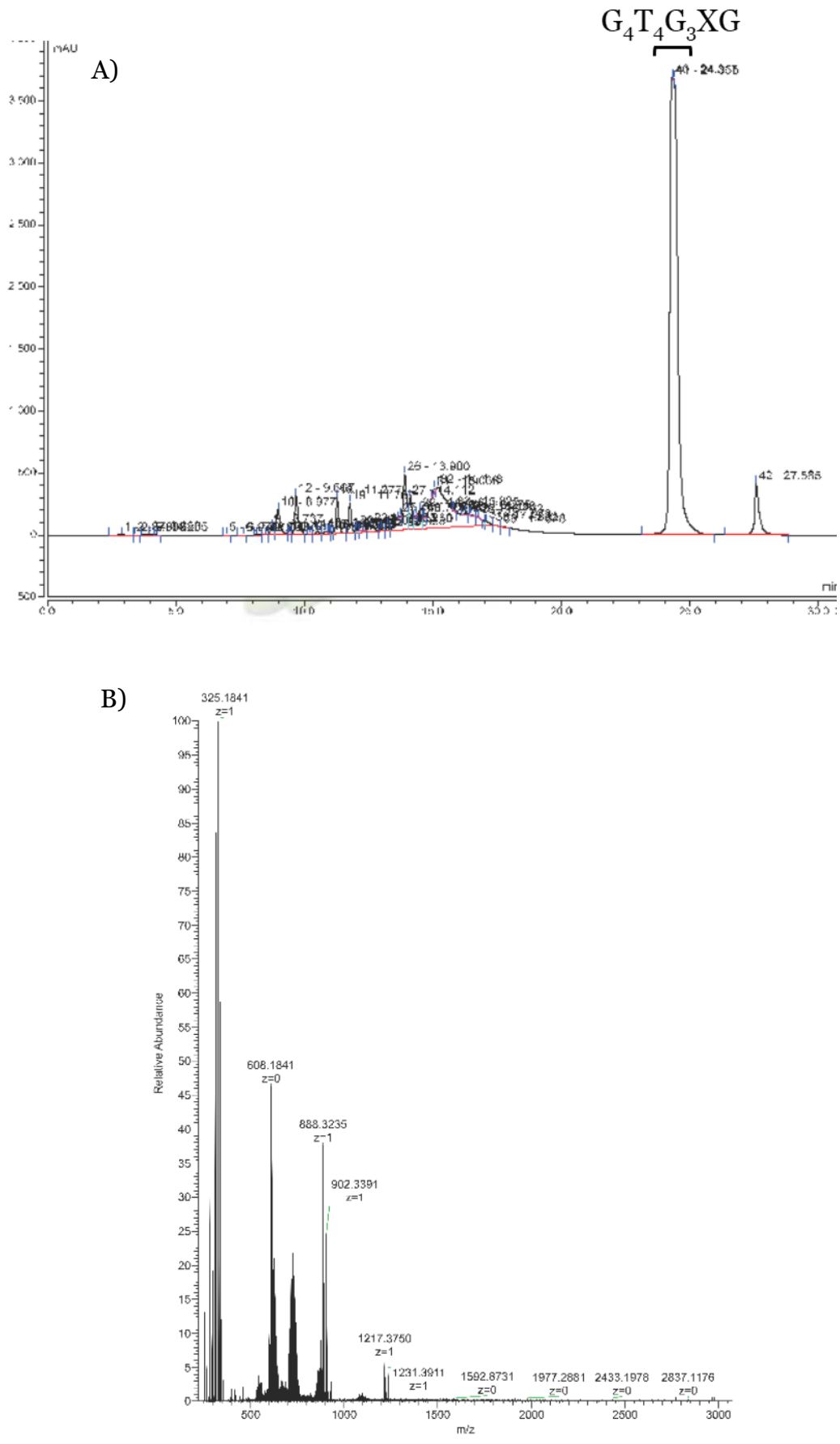


Figure SI-36. RP-HPLC profile of the $G_4T_4G_3XG$ sequence and B) corresponding ESI-MS confirming synthesis was successful.

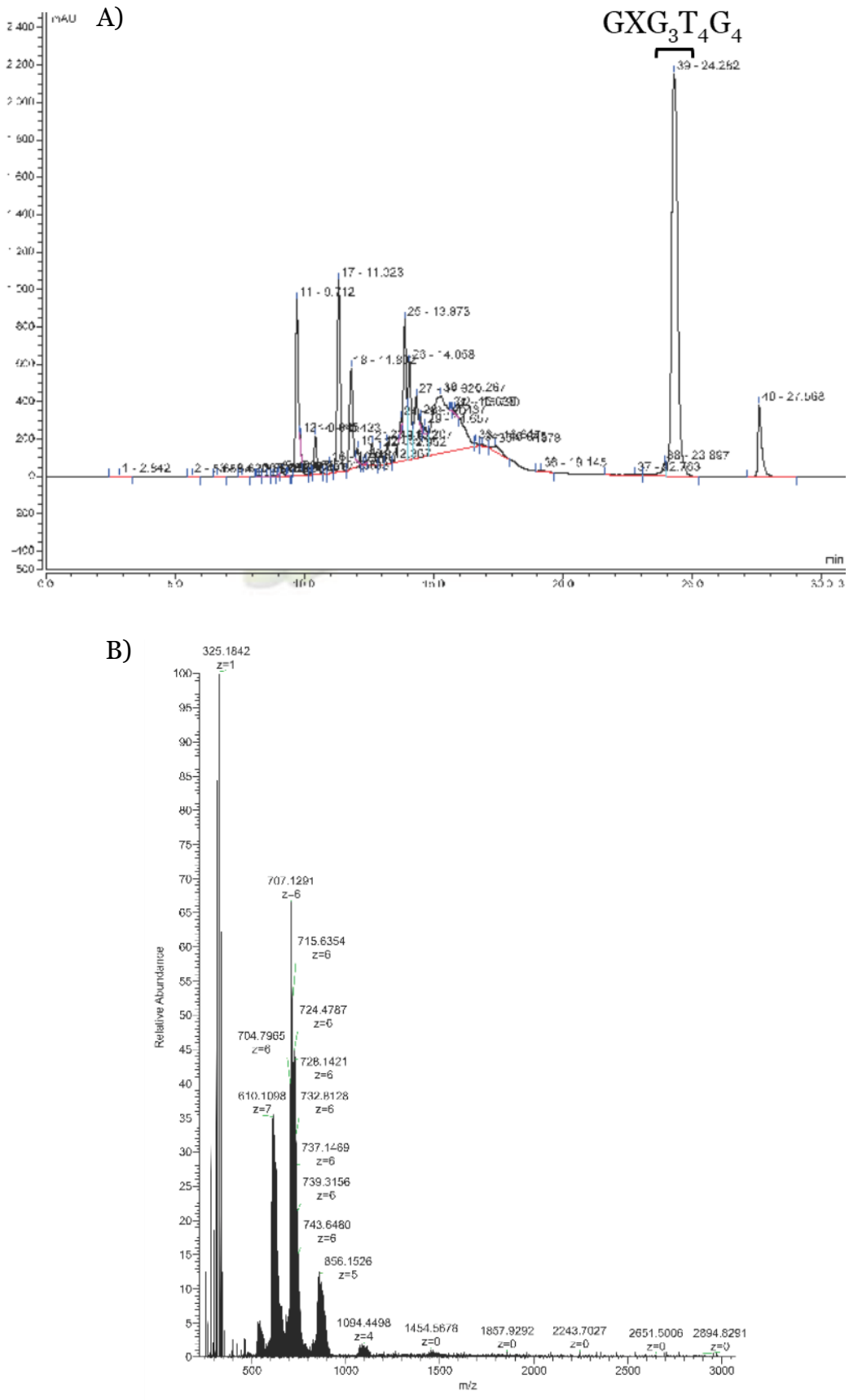


Figure SI-37. RP-HPLC profile of the $GXG_3T_4G_4$ sequence and B) corresponding ESI-MS confirming synthesis was successful.

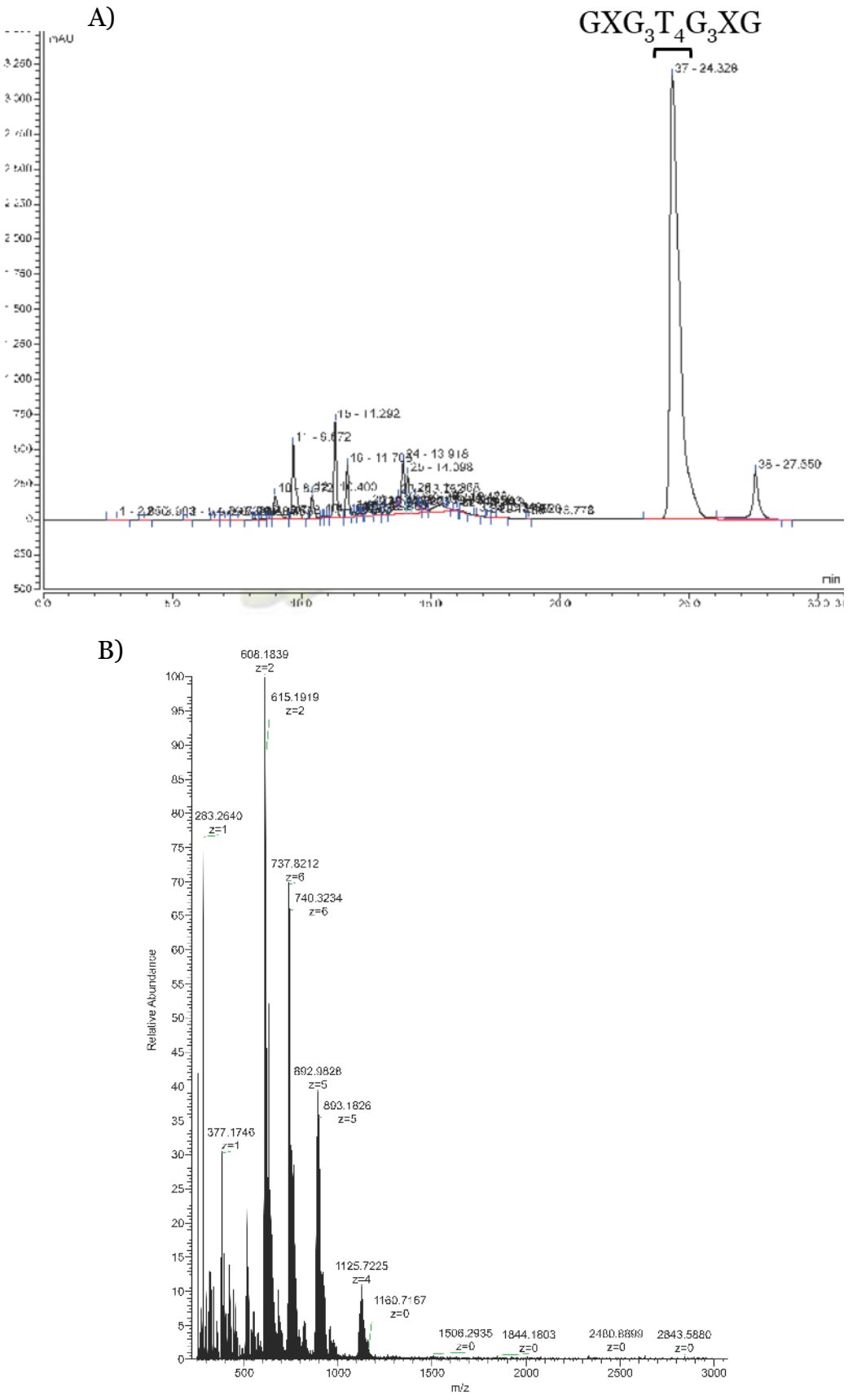


Figure SI-38. RP-HPLC profile of the $GXG_3T_4G_3XG$ sequence and B) corresponding ESI-MS confirming synthesis was successful.

RP-HPLC and mass spectra in Figure SI-39 – Figure SI-42 correspond to Table 7.3:

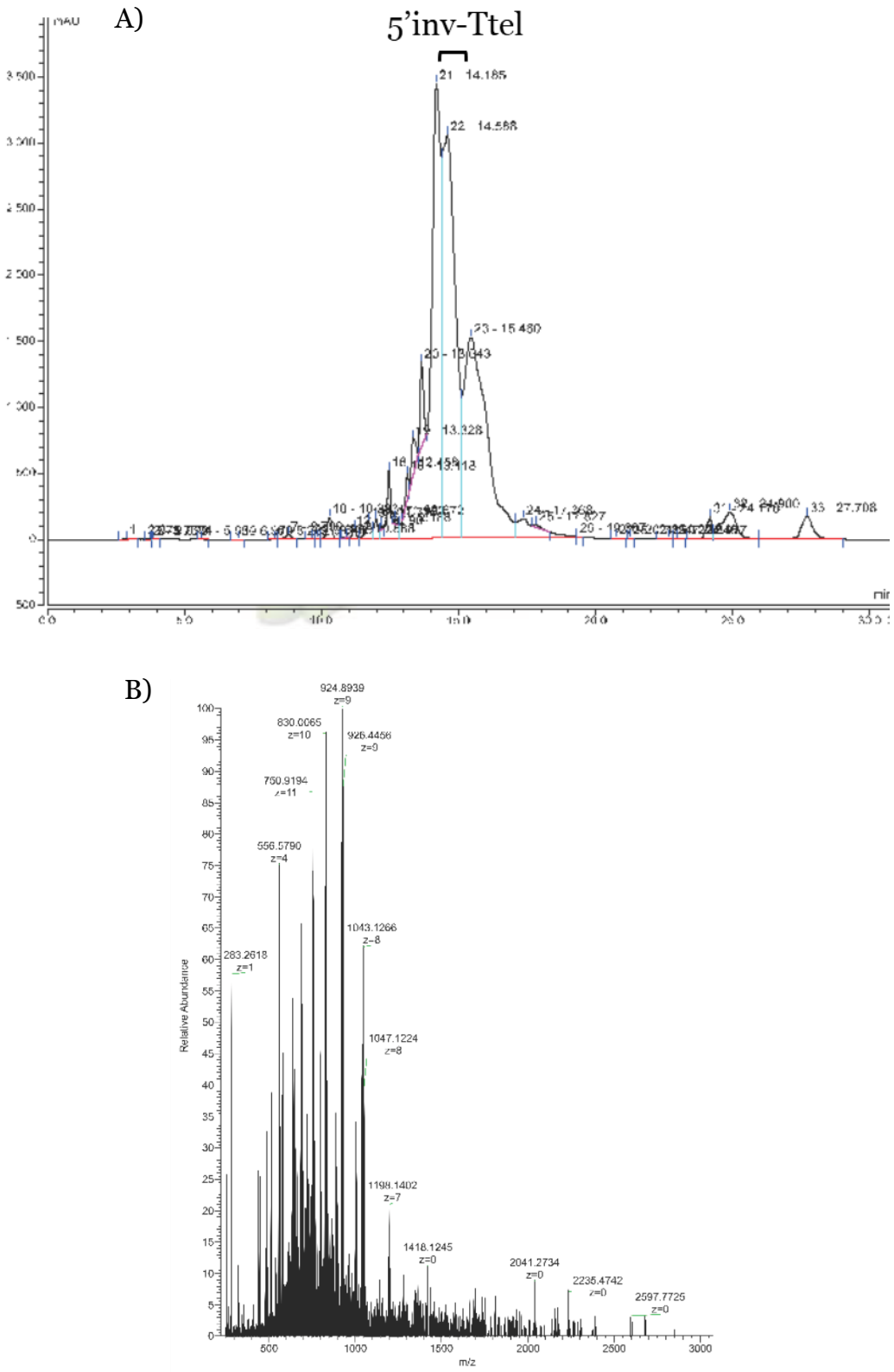


Figure SI-39. RP-HPLC profile of 5'inv-Ttel and B) corresponding ESI-MS confirming synthesis was successful.

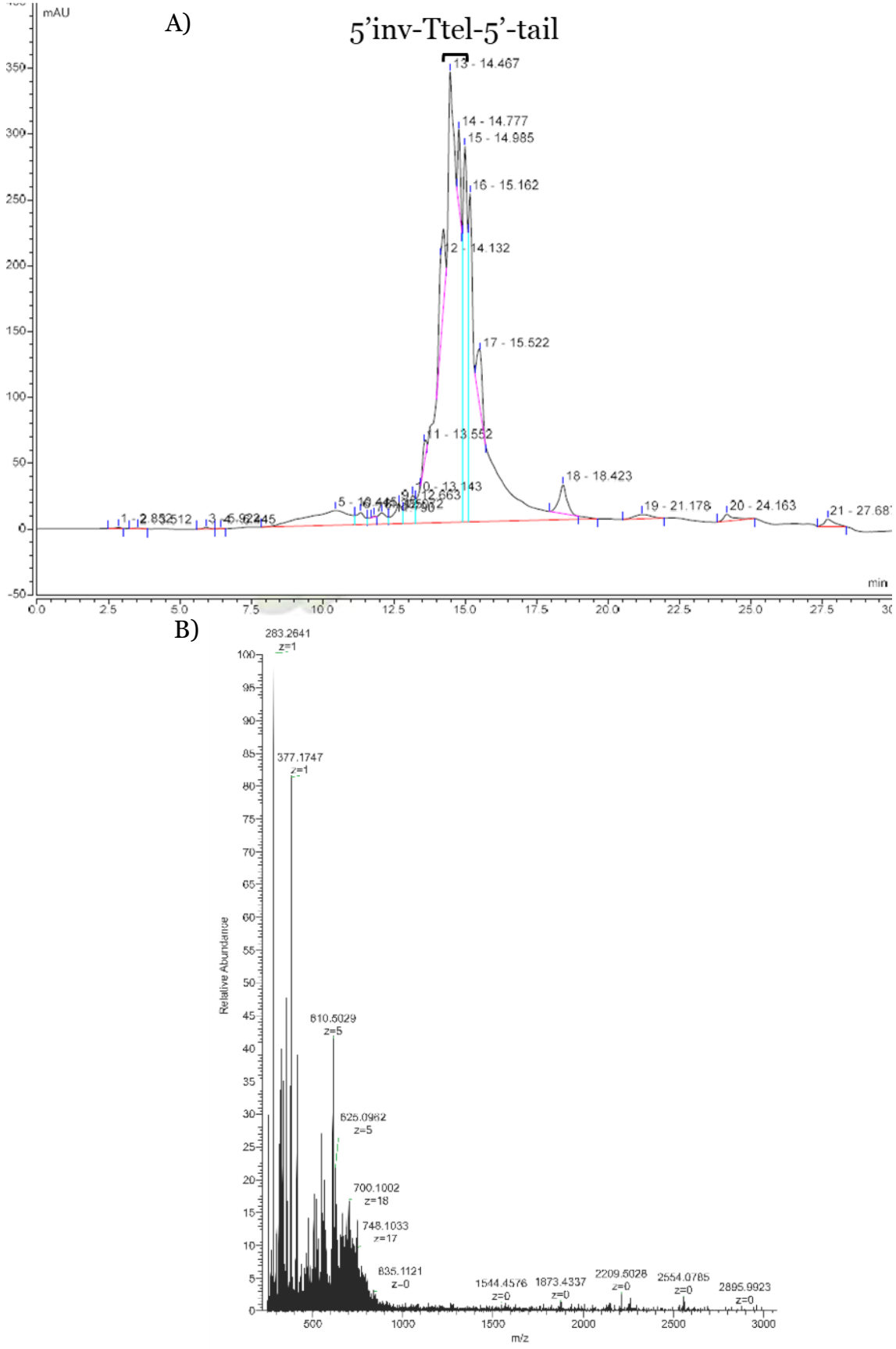


Figure SI-40. RP-HPLC profile of 5'inv-Ttel-5'-tail and B) corresponding ESI-MS confirming synthesis was successful.

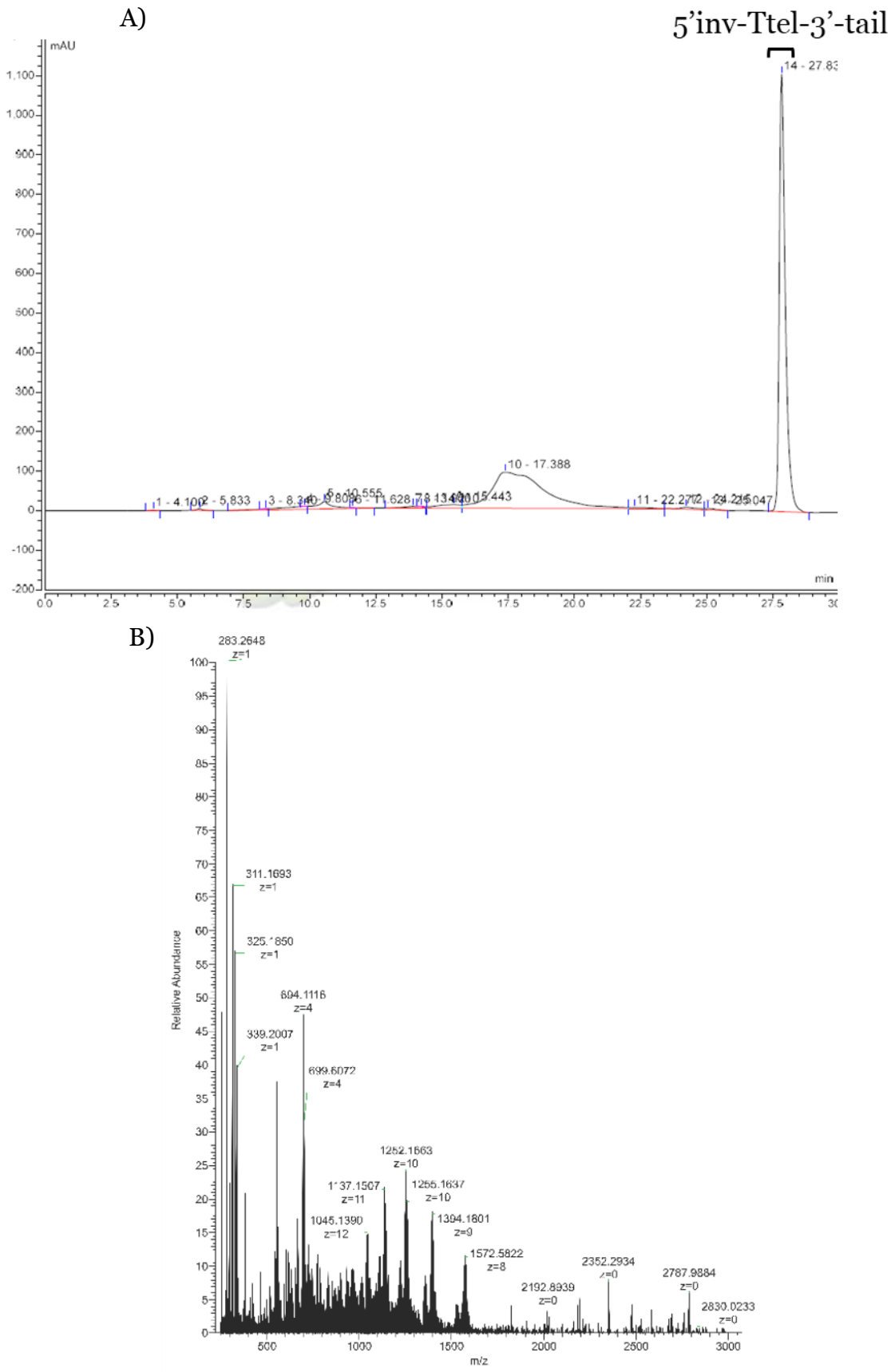


Figure SI-41. RP-HPLC profile of 5'inv-Ttel-3'-tail and B) corresponding ESI-MS confirming synthesis was successful.

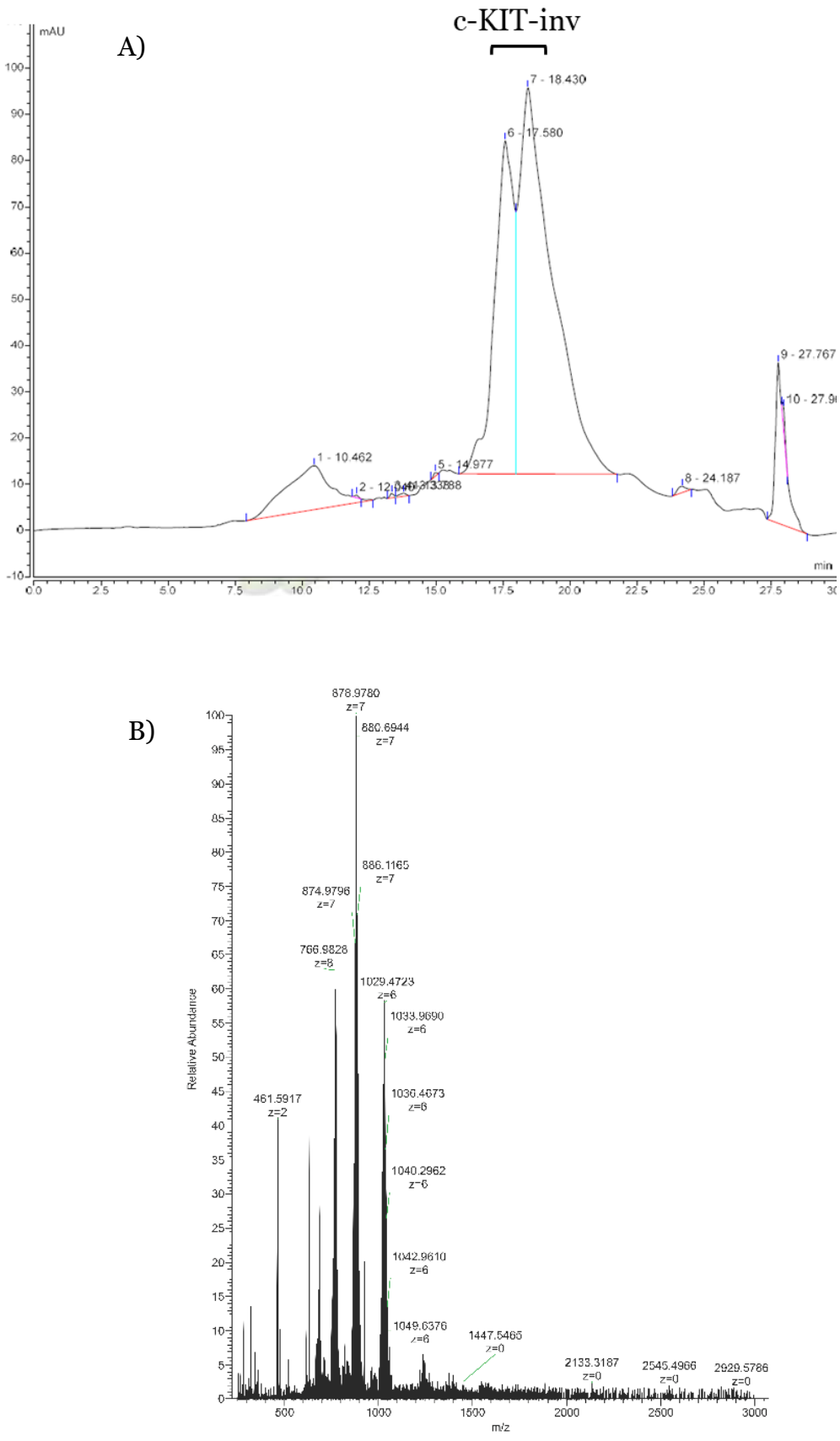


Figure SI-42. RP-HPLC profile of c-KIT-inv and B) corresponding ESI-MS confirming synthesis was successful.

RP-HPLC and mass spectra shown in Figure SI-43 – Figure SI-50 correspond to Table 7.4:

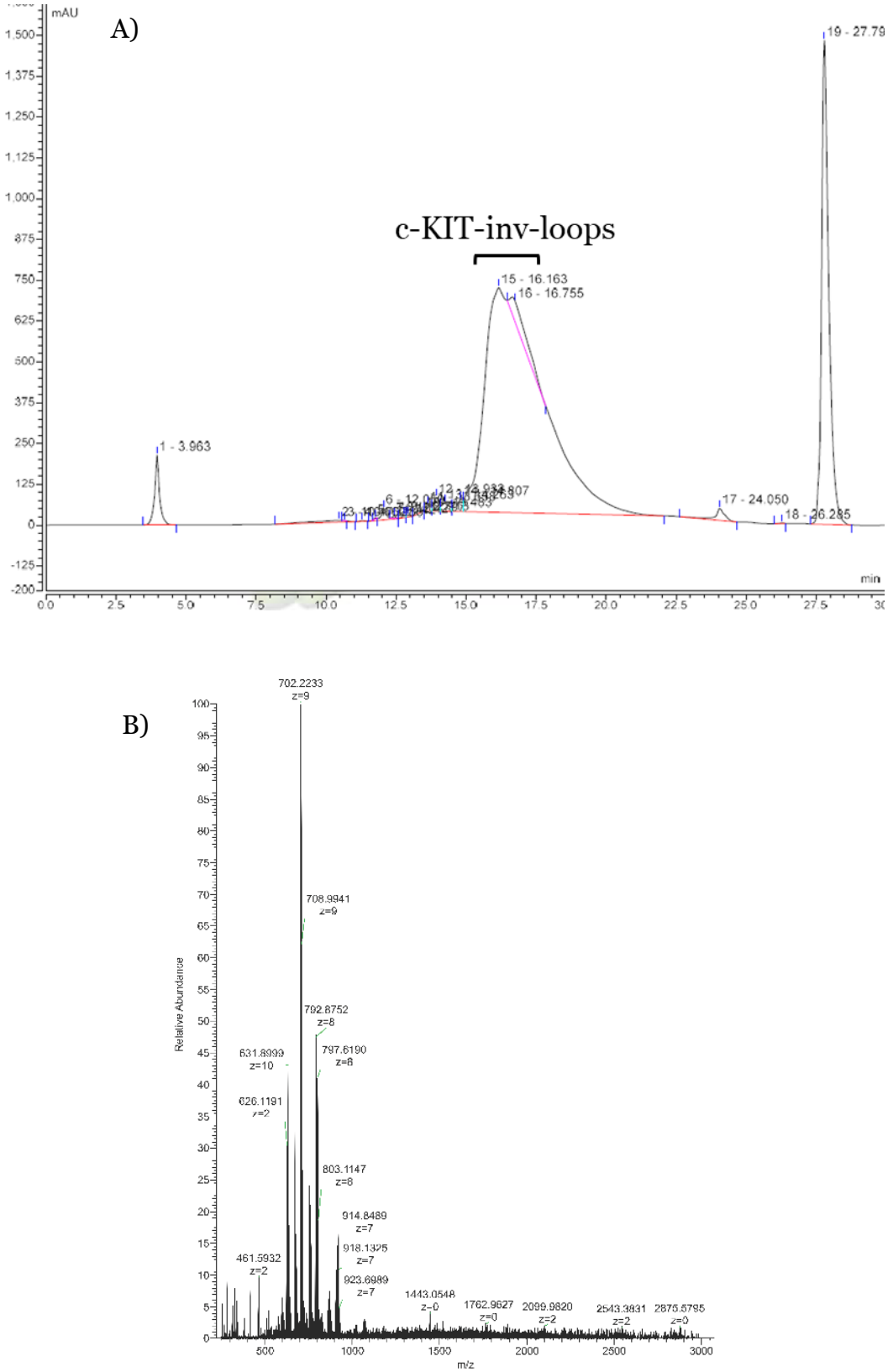


Figure SI-43. RP-HPLC profile of c-KIT-inv-loops and B) corresponding ESI-MS confirming synthesis was successful.

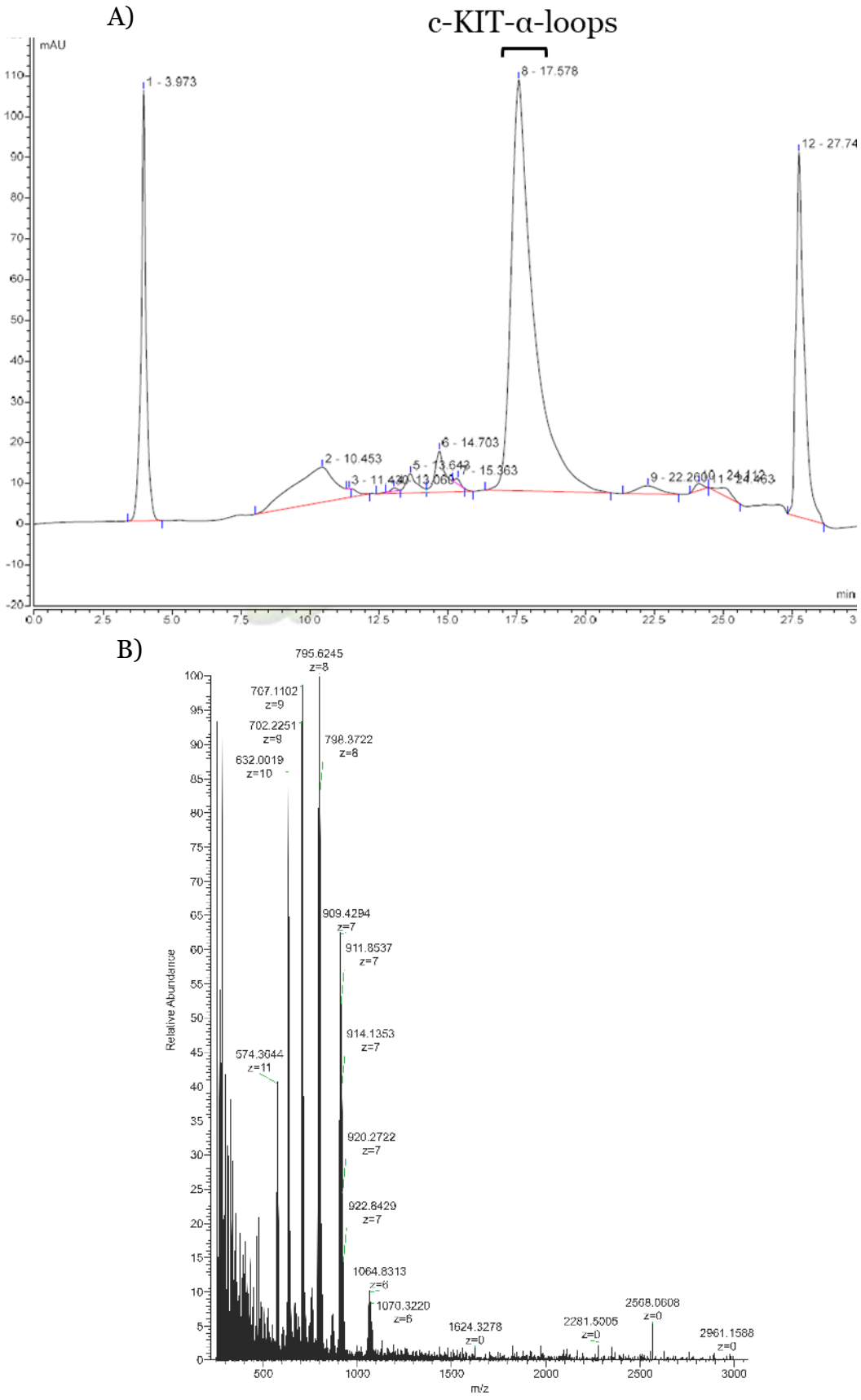


Figure SI-44. RP-HPLC profile of c-KIT-α-loops and B) corresponding ESI-MS confirming synthesis was successful.

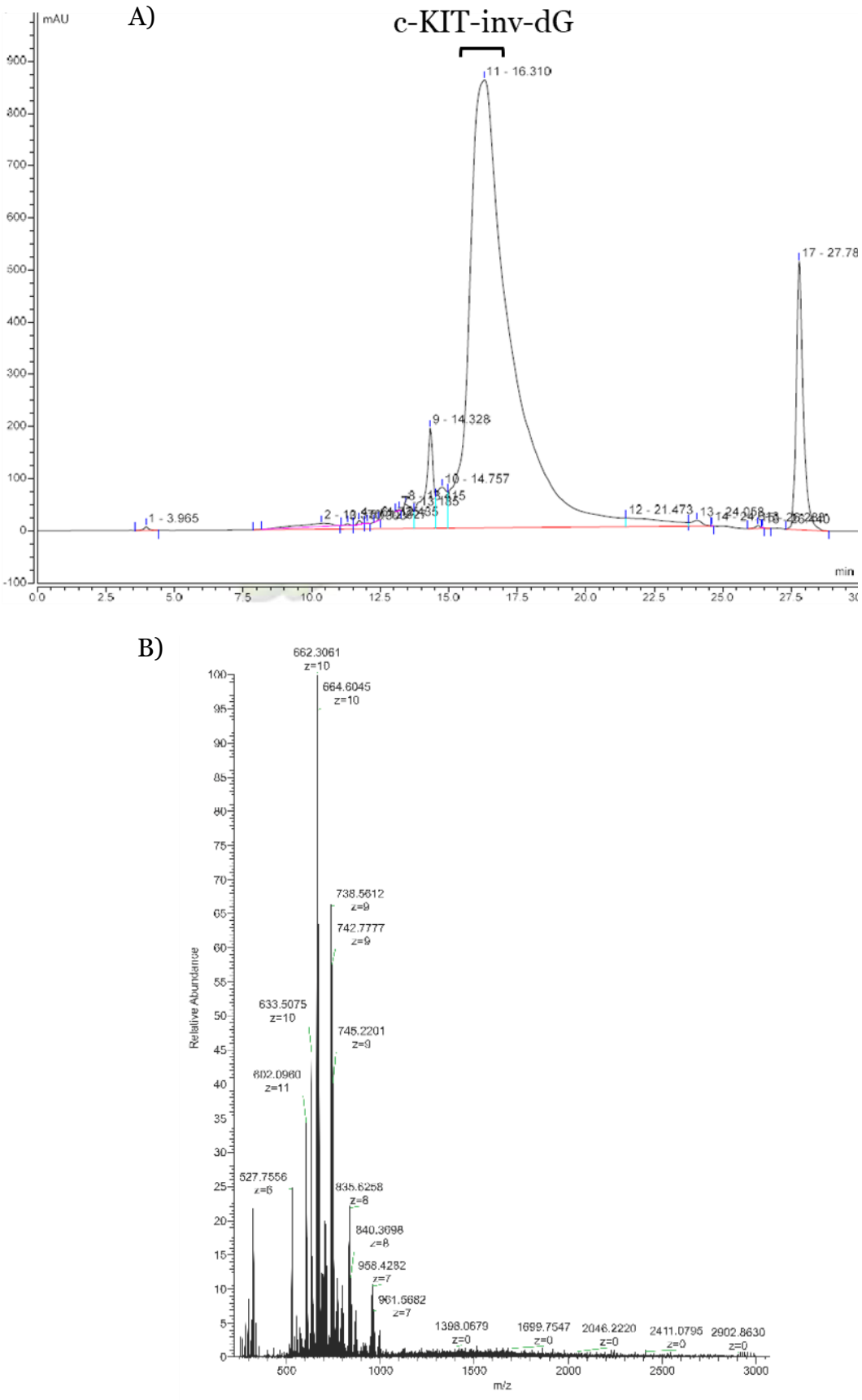


Figure SI-45. RP-HPLC profile of c-KIT-inv-dG and B) corresponding ESI-MS confirming synthesis was successful.

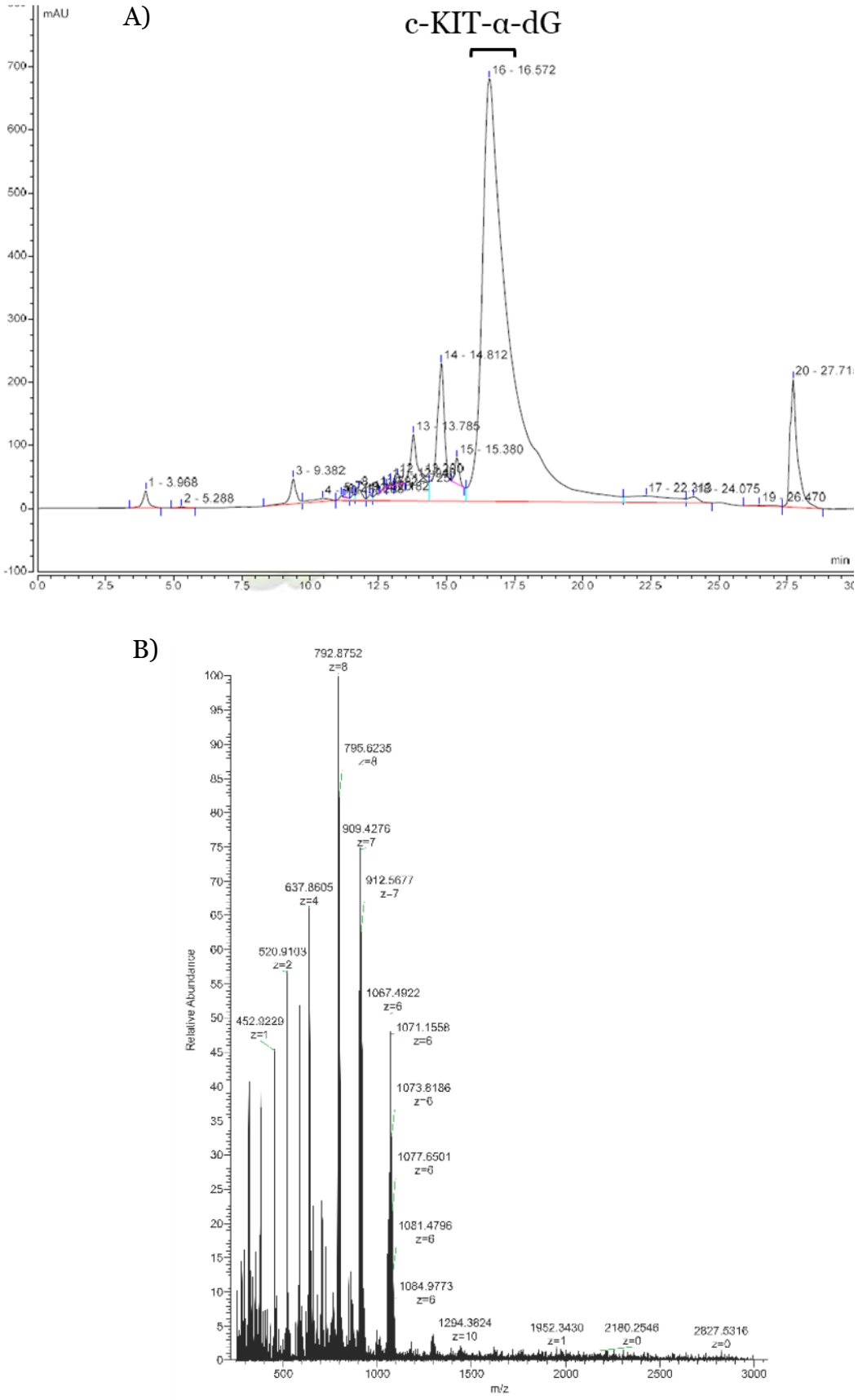


Figure SI-46. RP-HPLC profile of c-KIT-α-dG and B) corresponding ESI-MS confirming synthesis was successful.

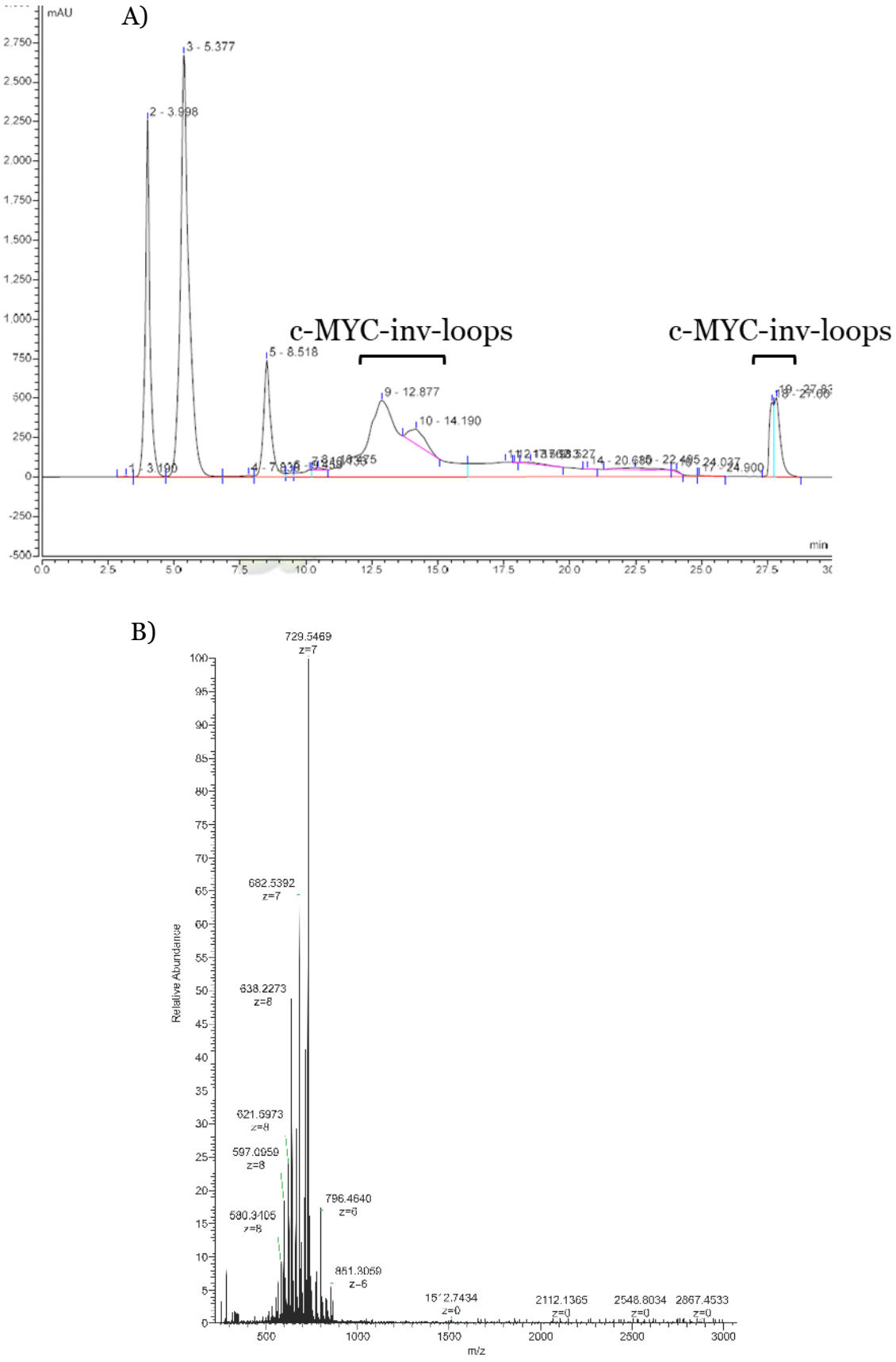


Figure SI-47. RP-HPLC profile of c-MYC-inv-loops and B) corresponding ESI-MS confirming synthesis was successful. Initial peak in preparative RP-HPLC appeared at approx. 14 min, but product also appears at approx. 27 min due to G4 formation in the sample.

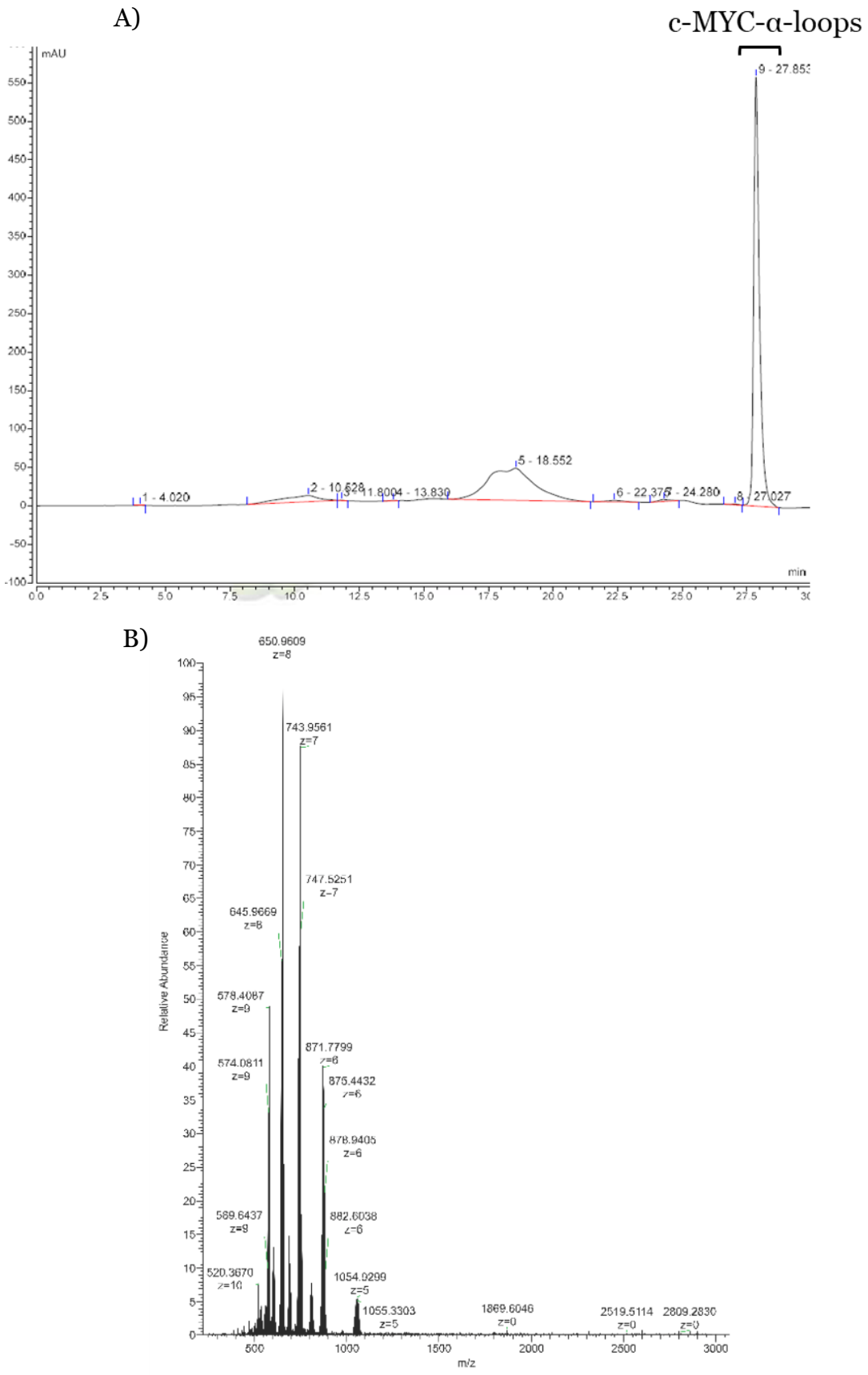


Figure SI-48. RP-HPLC profile of c-MYC-α-loops and B) corresponding ESI-MS confirming synthesis was successful.

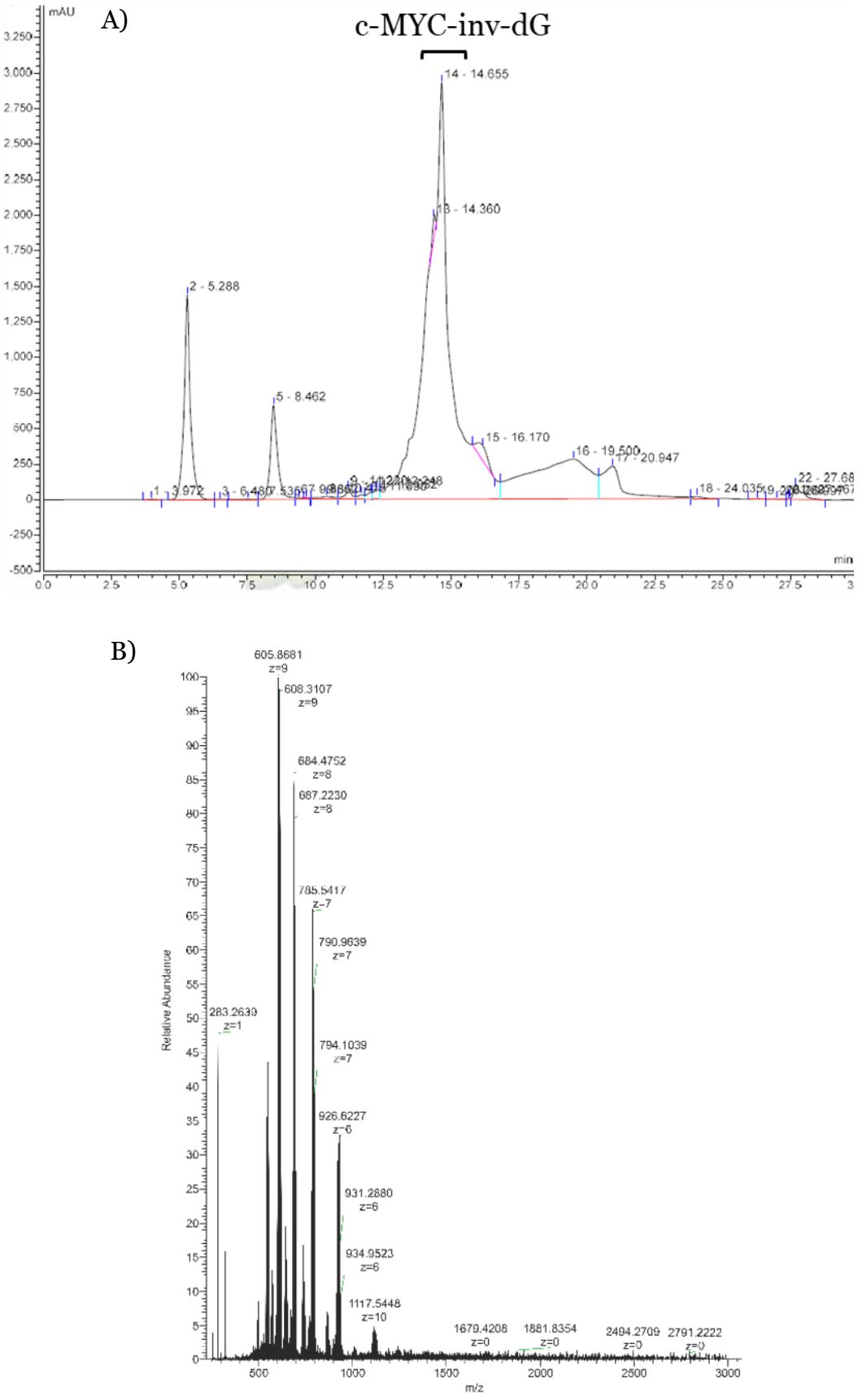


Figure SI-49. RP-HPLC profile of c-MYC-inv-dG and B) corresponding ESI-MS confirming synthesis was successful.

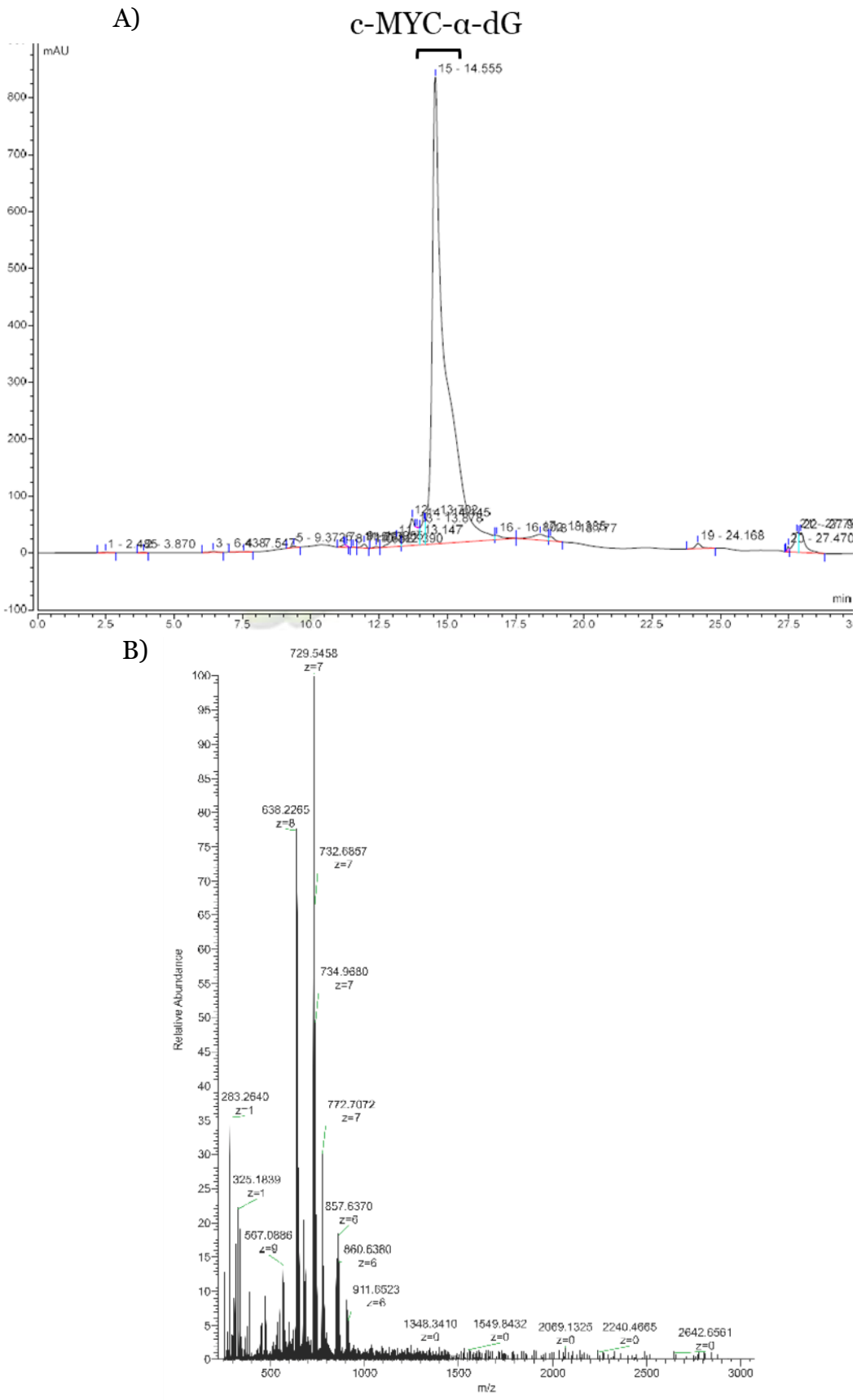


Figure SI-50. RP-HPLC profile of c-MYC-α-dG and B) corresponding ESI-MS confirming synthesis was successful.

See Table 7.5 for Figure SI-52 and Figure SI-51:

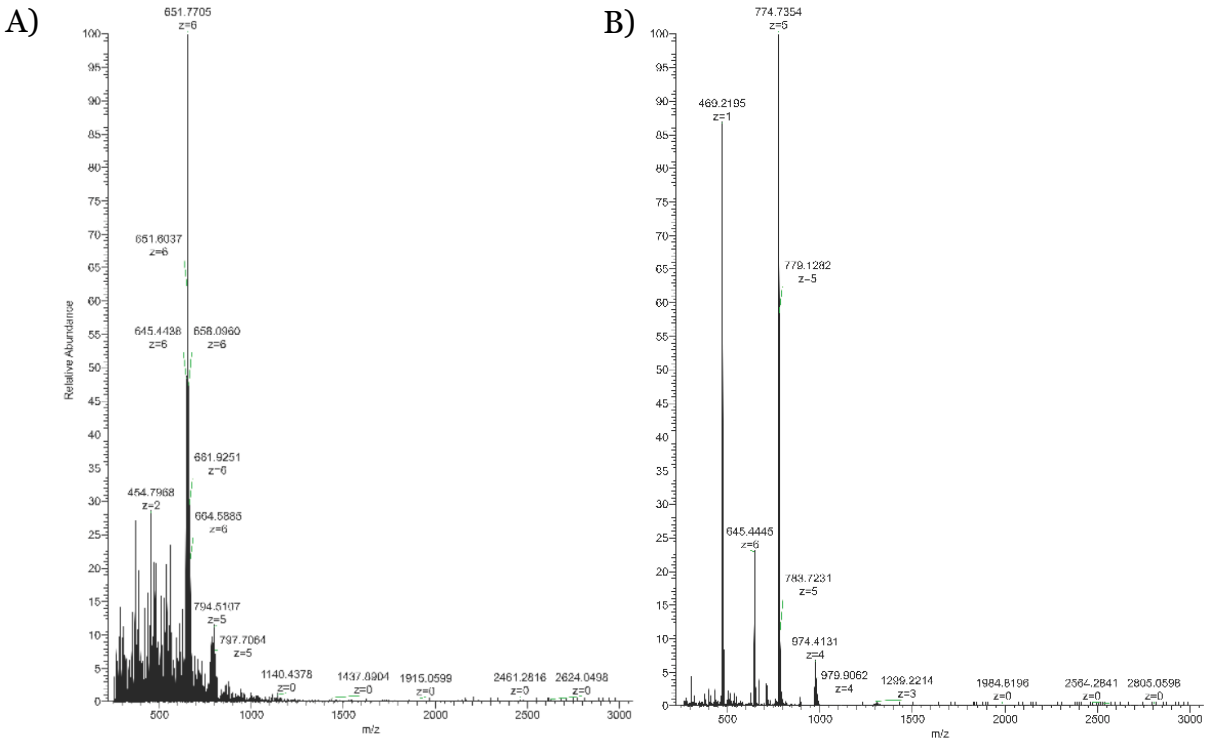


Figure SI-52. Comparison of ESI-MS spectra of A) tel-G3A8 and B) tel-G3A8-X. Both spectra show the same mass, indicating that the change in r_t observed in RP-HPLC is not due to a difference in sequence.

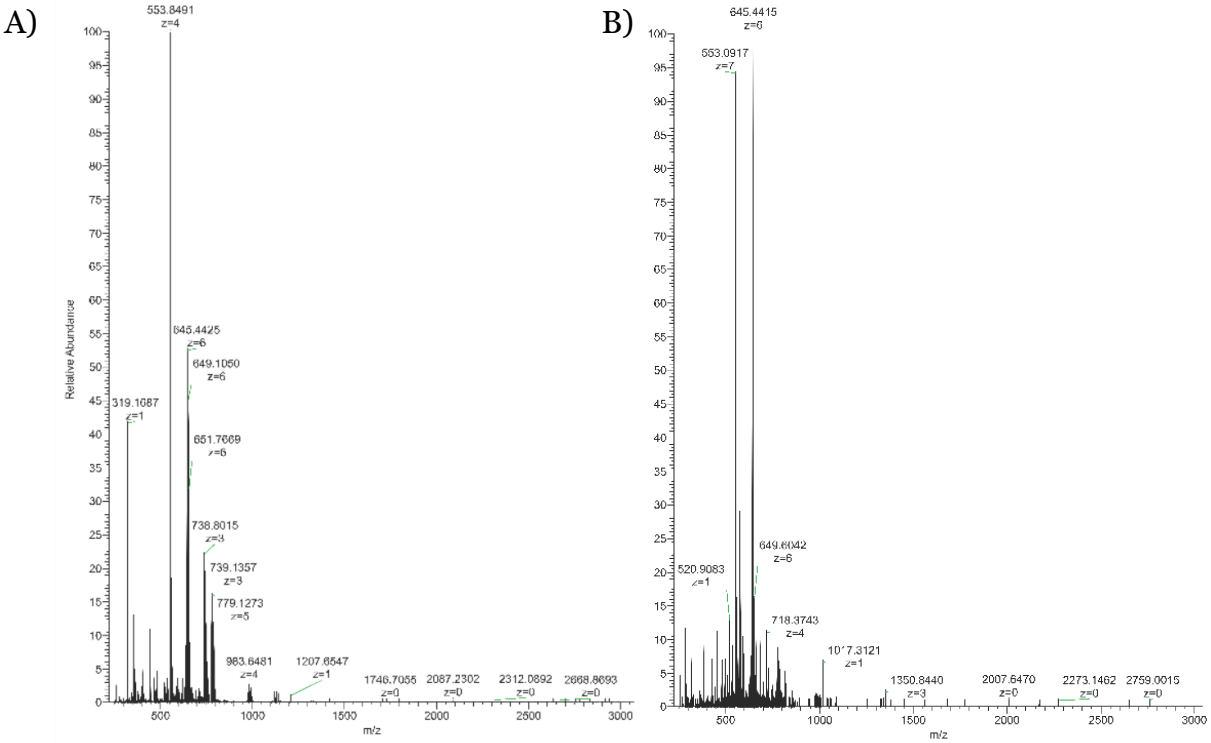


Figure SI-51. Comparison of ESI-MS spectra of A) tel-G5A8 and B) tel-G5A8-X. Both spectra show the same mass, indicating that the change in r_t observed in RP-HPLC is not due to a difference in sequence.

Full characterisation of the following spectra is included in Chapter 7.3.

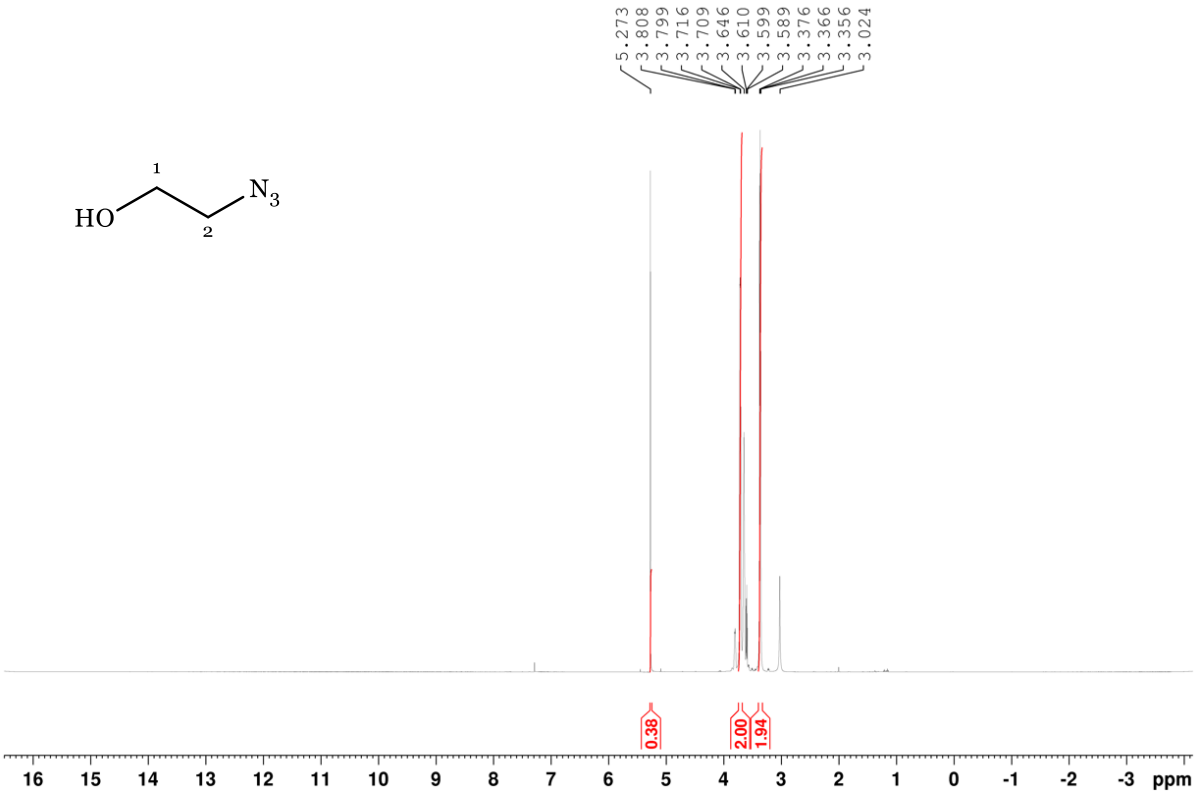


*Figure SI-53. ^{1}H NMR spectrum of 2-azido-1-ethanol (compound **2a**).*

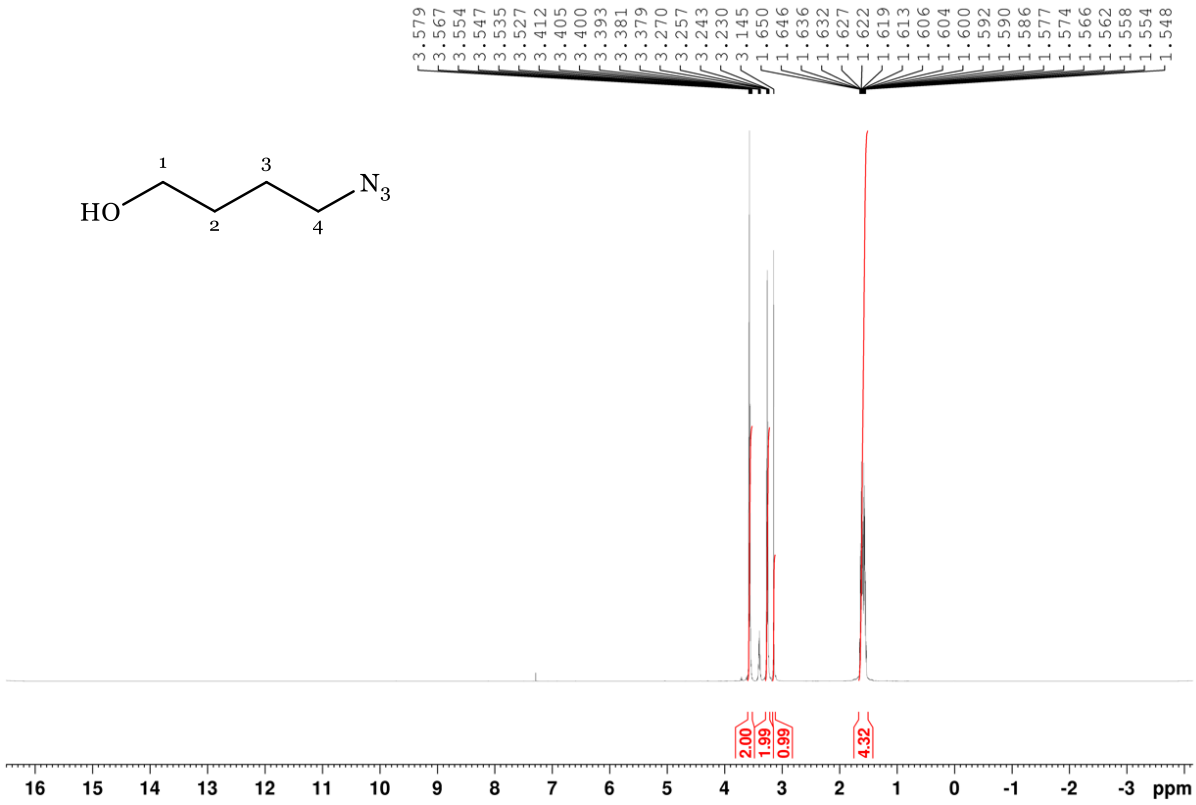


*Figure SI-54. ^{1}H NMR spectrum of 4-azido-1-butanol (compound **2b**).*

*Figure SI-55. [13]C NMR spectrum of 4-azido-1-butanol (compound **2b**).*

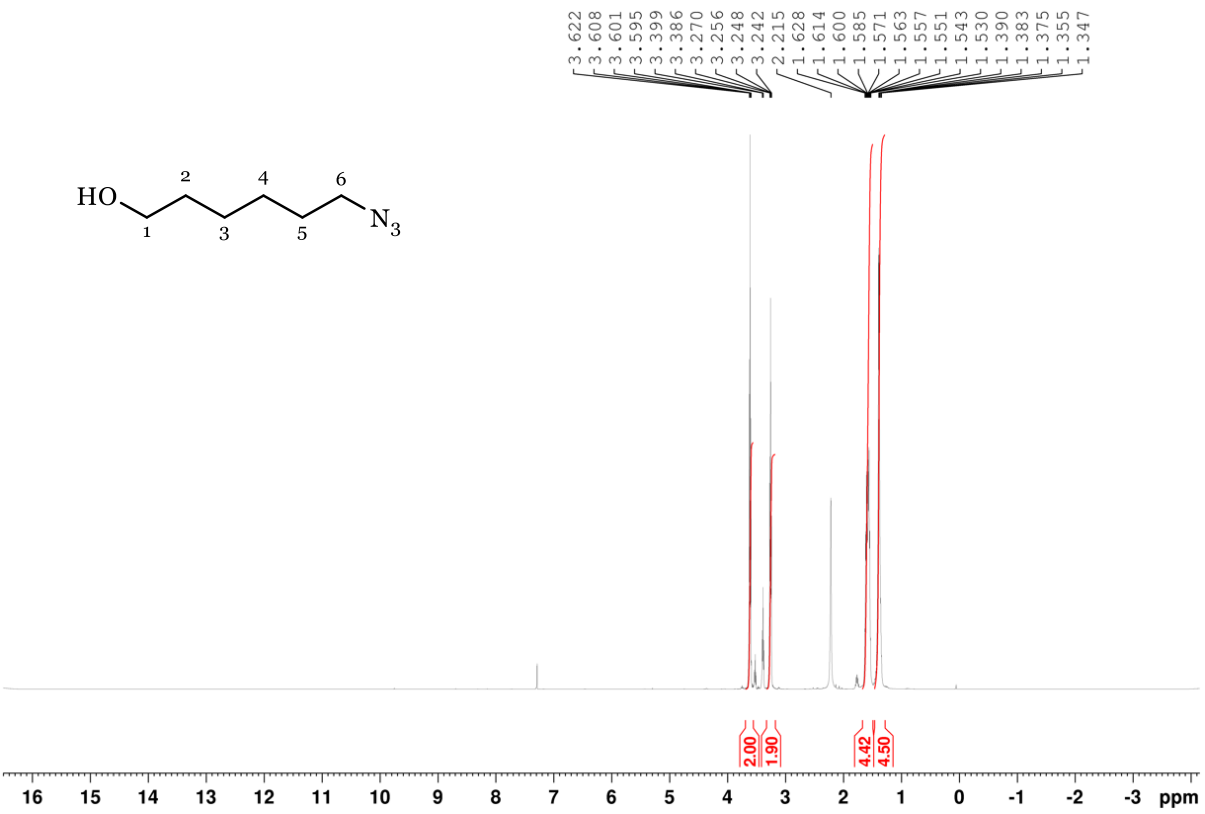


*Figure SI-56. [1]H NMR spectrum of 6-azido-1-hexanol (compound **2c**).*

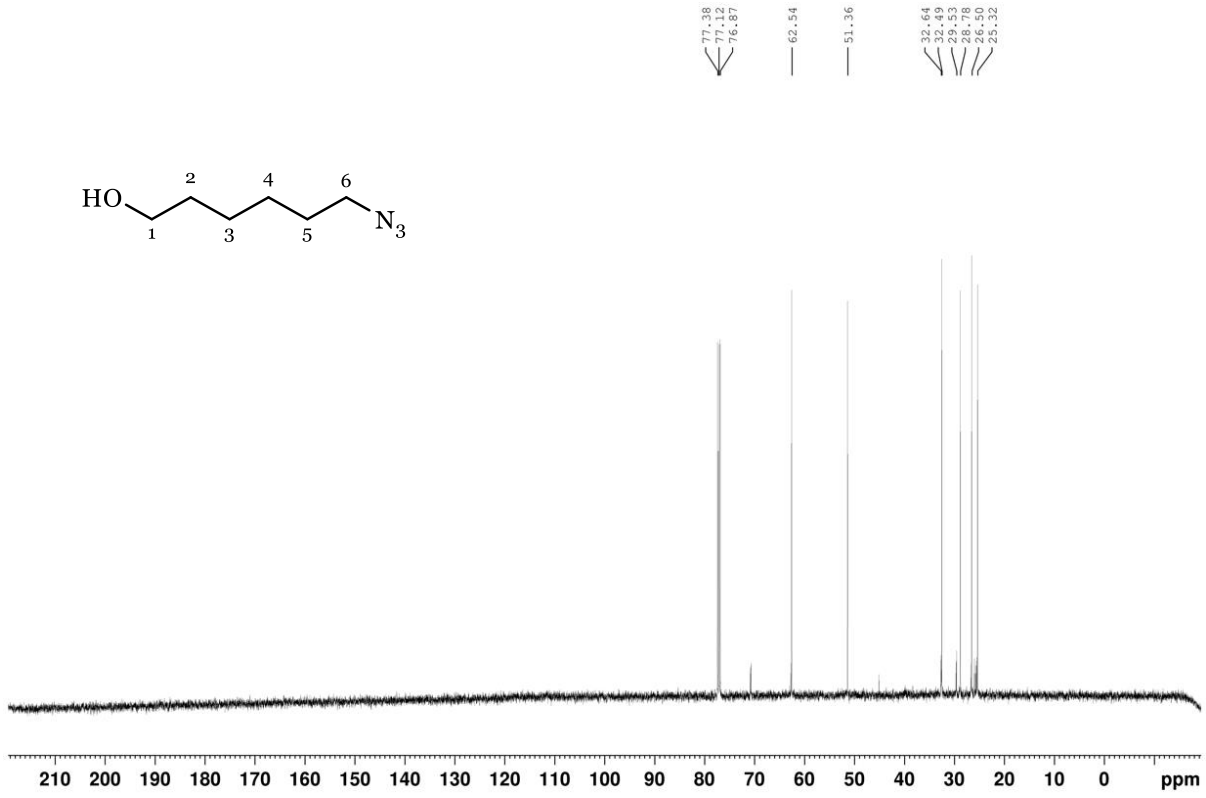


*Figure SI-57. ^{13}C NMR spectrum of 6-azido-1-hexanol (compound **2c**).*

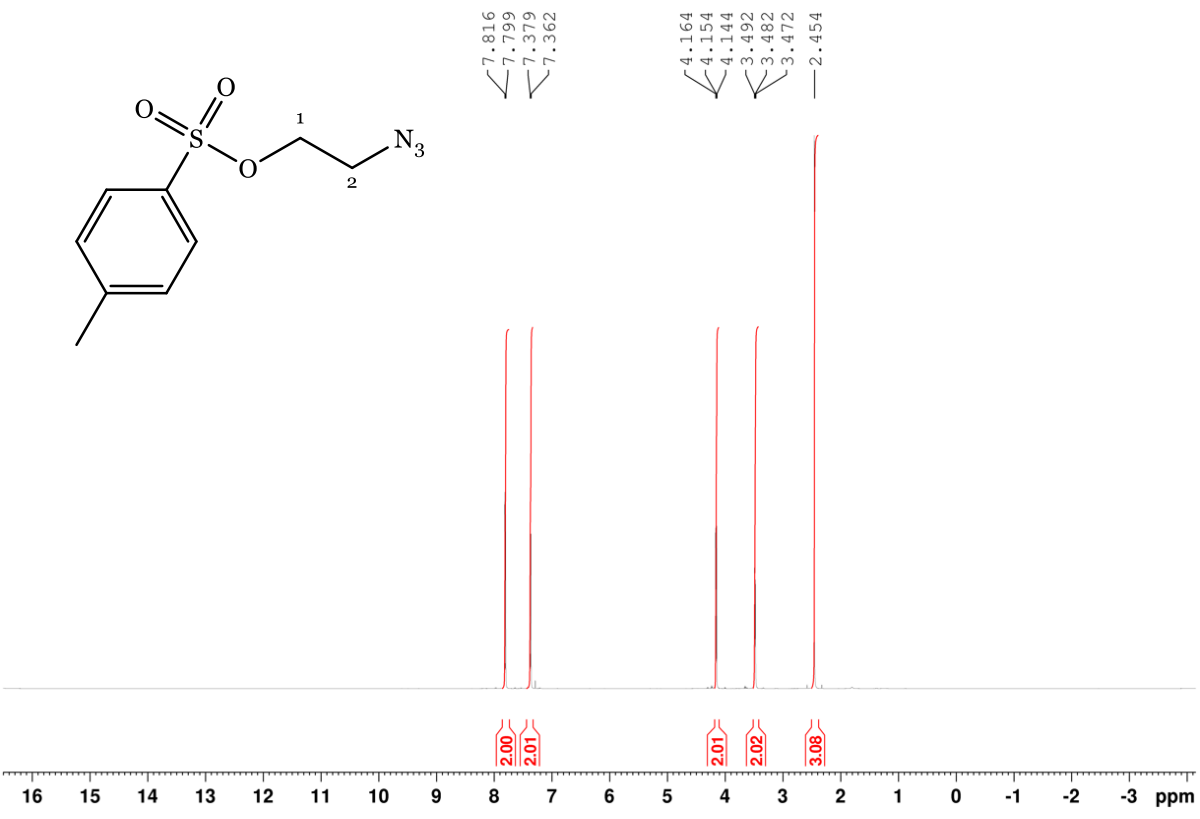


*Figure SI-58. ^{1}H NMR spectrum of 2-azido-1-toluenesufonyl ethane (compound **3a**).*

*Figure SI-59. [13]C NMR spectrum of 2-azido-1-toluenesufonyl ethane (compound **3a**).*

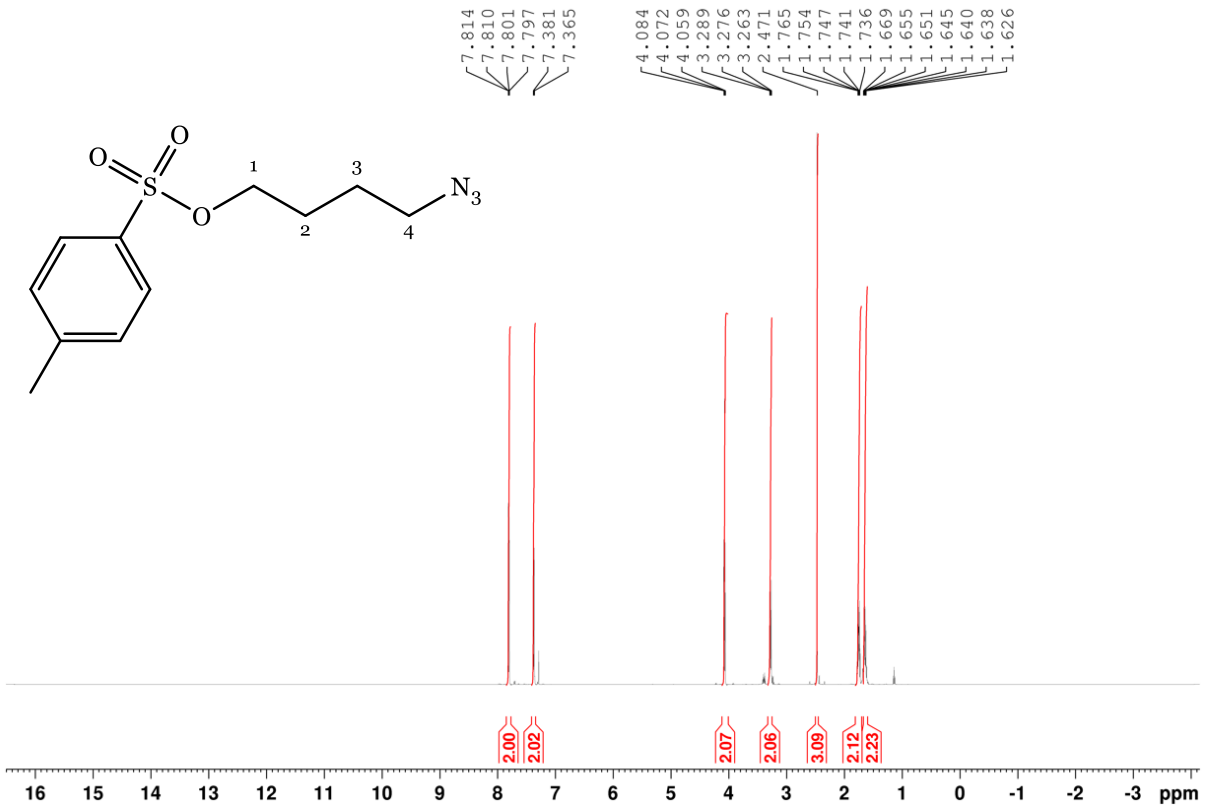


*Figure SI-60. [1]H NMR spectrum of 4-azido-1-toluenesufonyl butane (compound **3b**).*

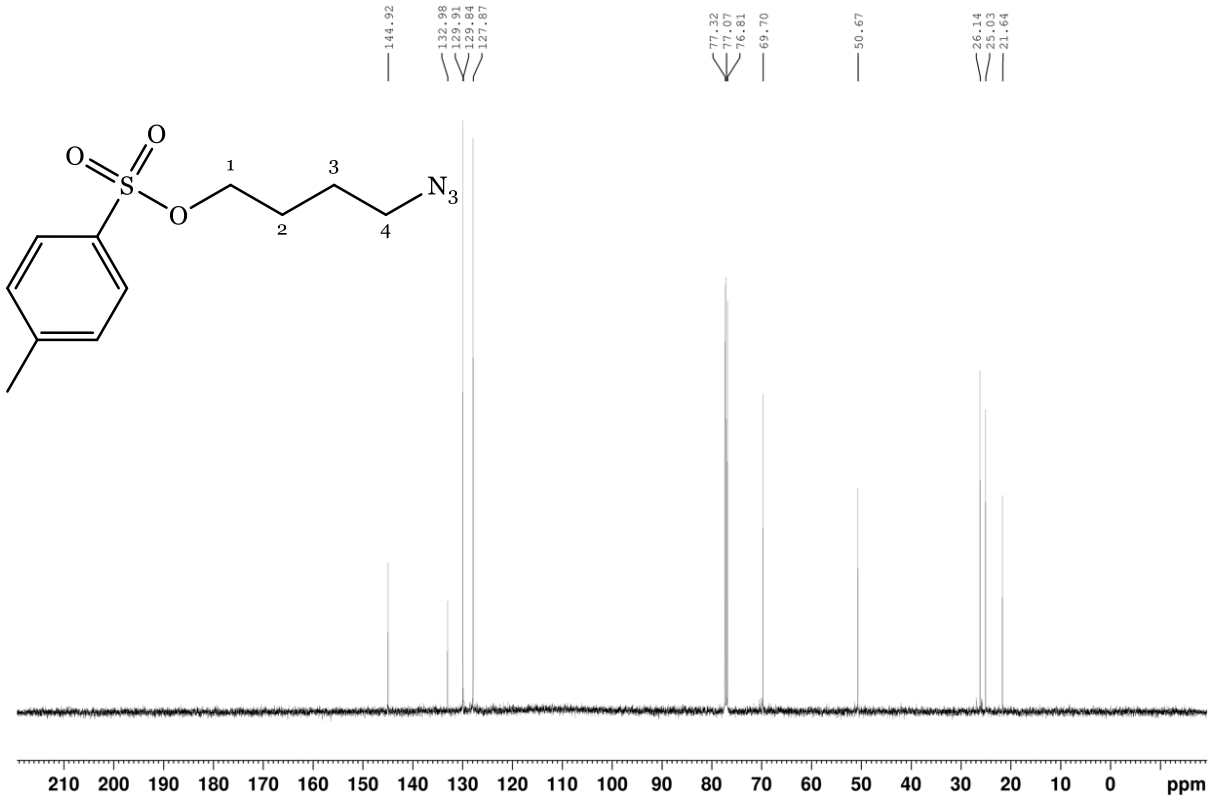


*Figure SI-61. ^{13}C NMR spectrum of 4-azido-1-toluenesufonyl butane (compound **3b**).*

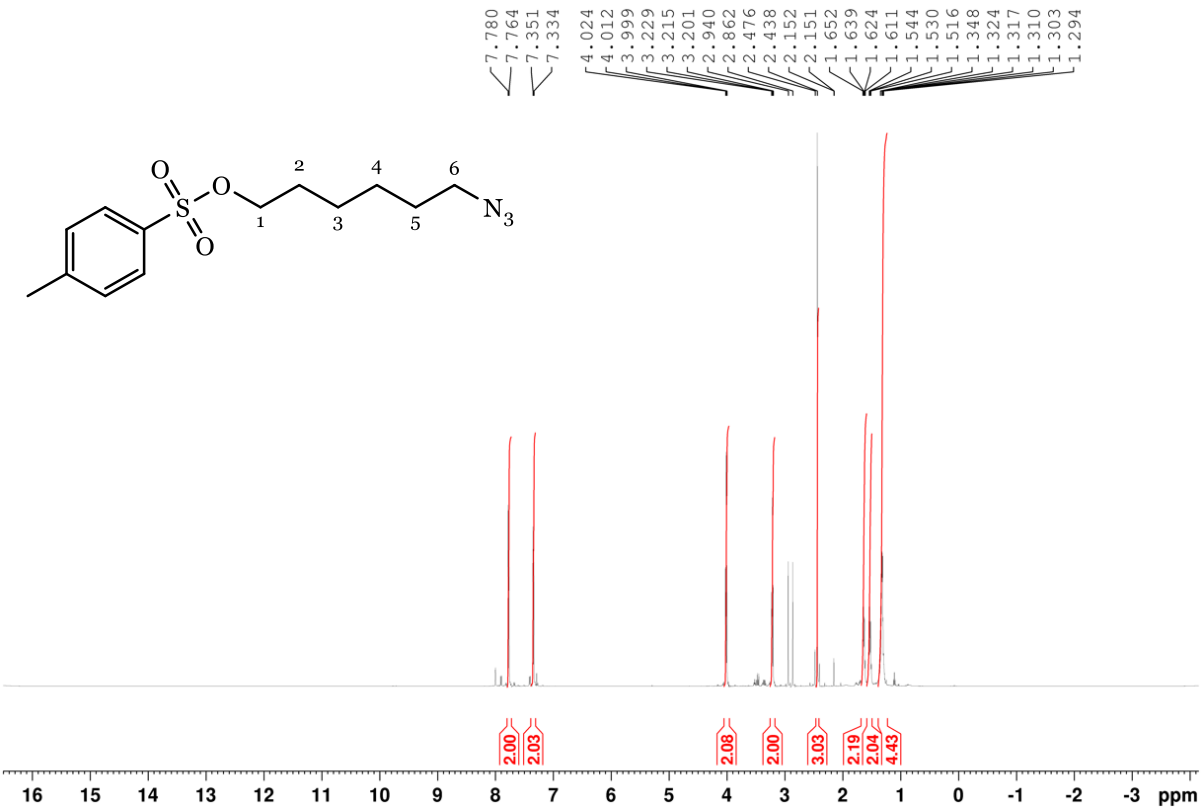


*Figure SI-62. ^{1}H NMR spectrum of 6-azido-1-toluenesufonyl hexane (compound **3c**).*

*Figure SI-63. ^{13}C NMR spectrum of 6-azido-1-toluenesufonyl hexane (compound **3c**).*

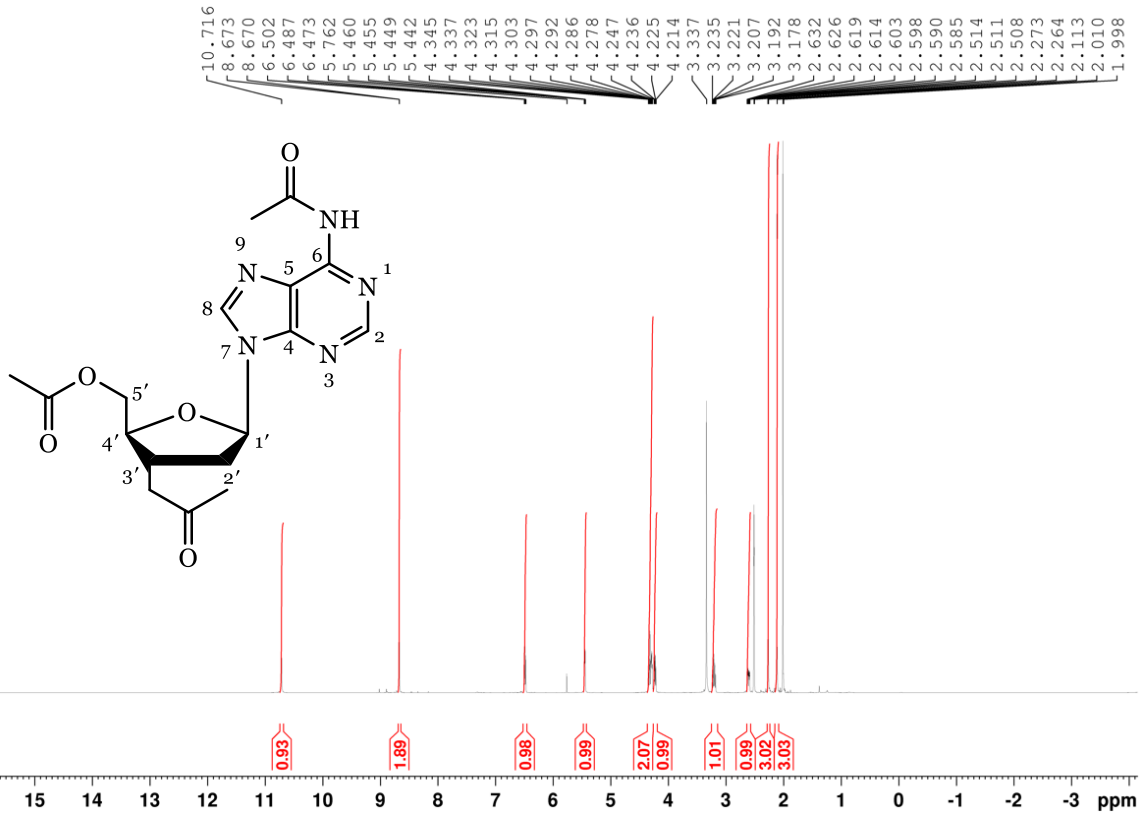


Figure SI-64. ^{1}H NMR spectrum of N^6*,5′, 3′,-triacetyl-2′-deoxyadenosine (compound **4**).*

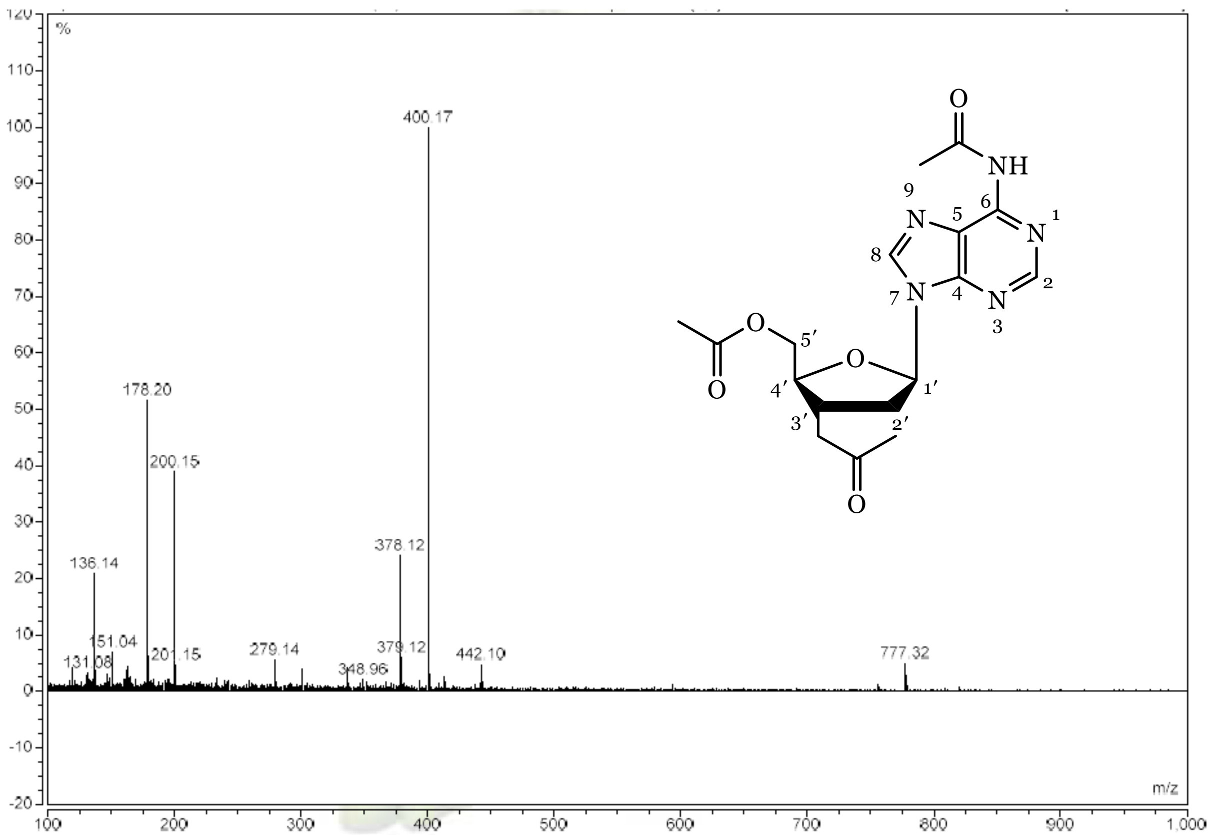


Figure SI-65. Mass spectrum of N[6]*, 5′,3′-triacetyl-2′-deoxyadenosine (compound* ***4****).*

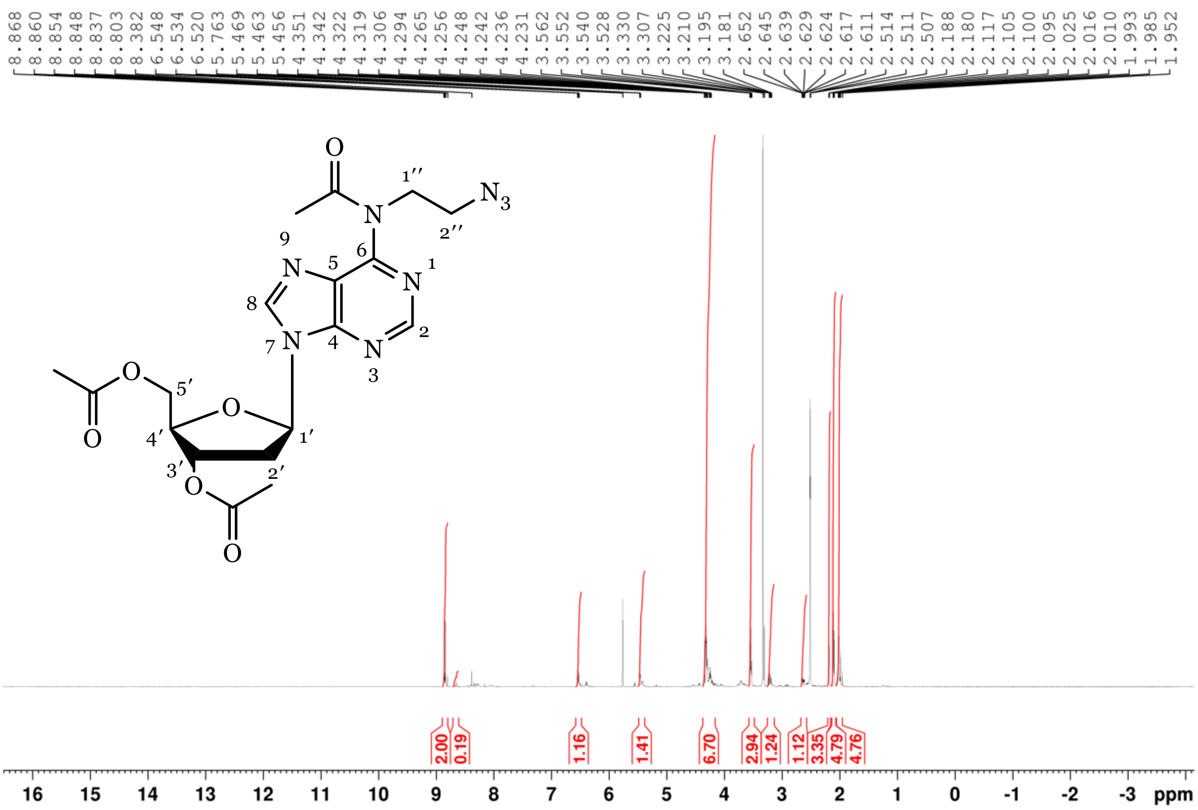


Figure SI-66. [1]H NMR spectrum of N[6]*-[2-azidoethane]-5′,3′,*N[6]*-triacetyl-2′-deoxyadenosine (compound* ***5a****).*

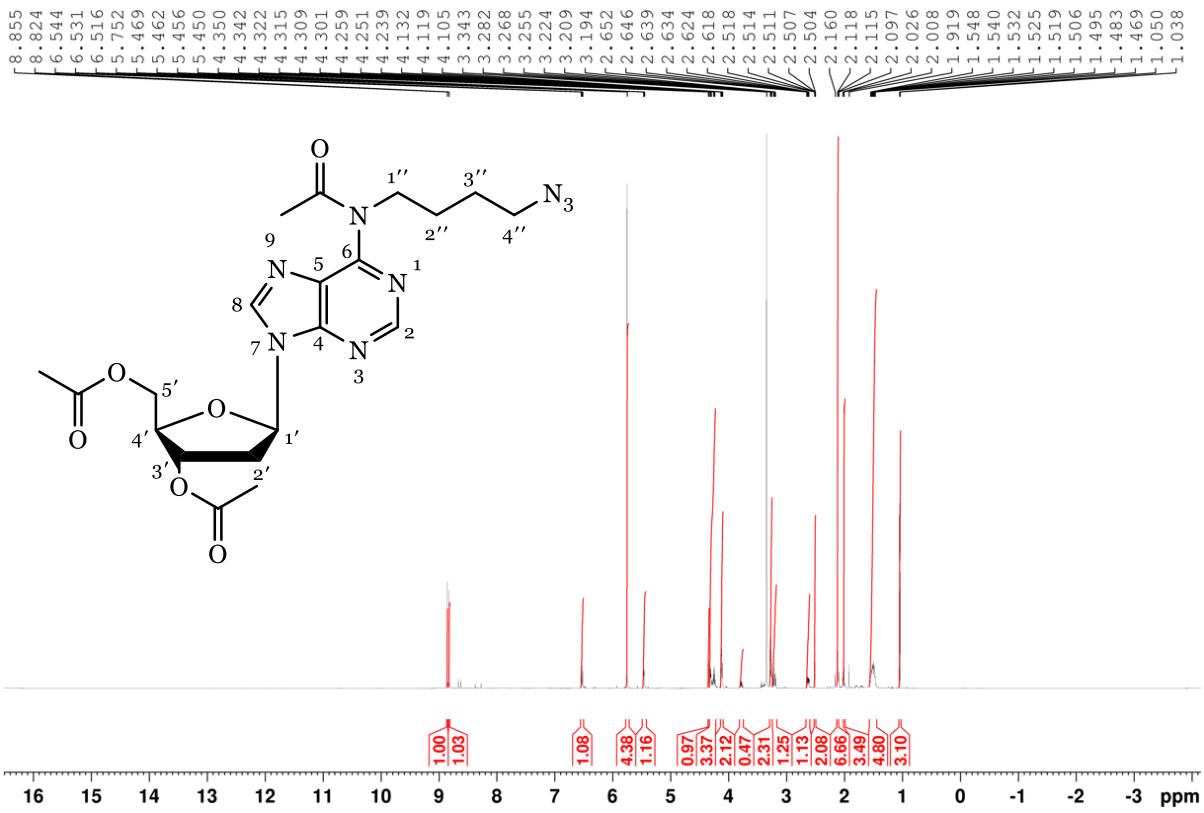


Figure SI-68. [1]H NMR spectrum of N[6]*-[4-azidobutane]-5′,3′,*N[6]*-triacetyl-2′-deoxyadenosine (compound* ***5b****).*

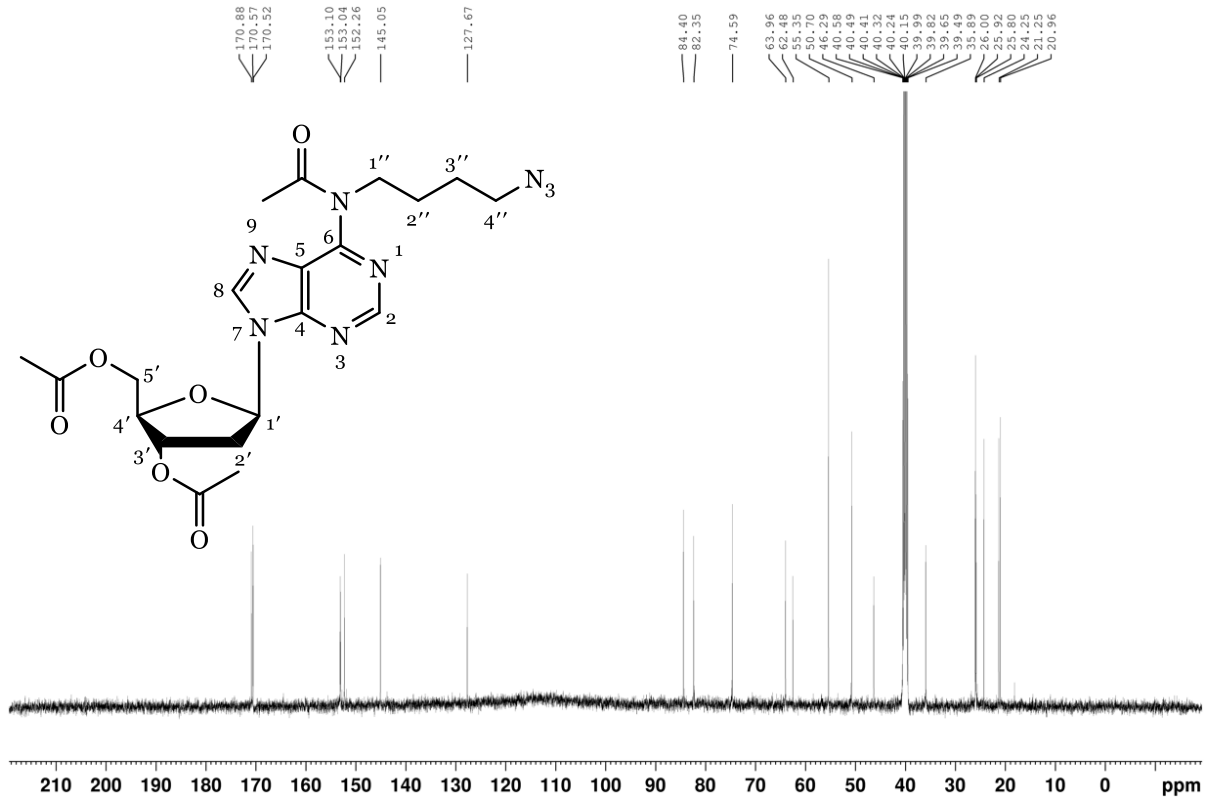


Figure SI-67. [13]C NMR spectrum of N[6]*-[4-azidobutane]-5′,3′,*N[6]*-triacetyl-2′-deoxyadenosine (compound* ***5b****).*

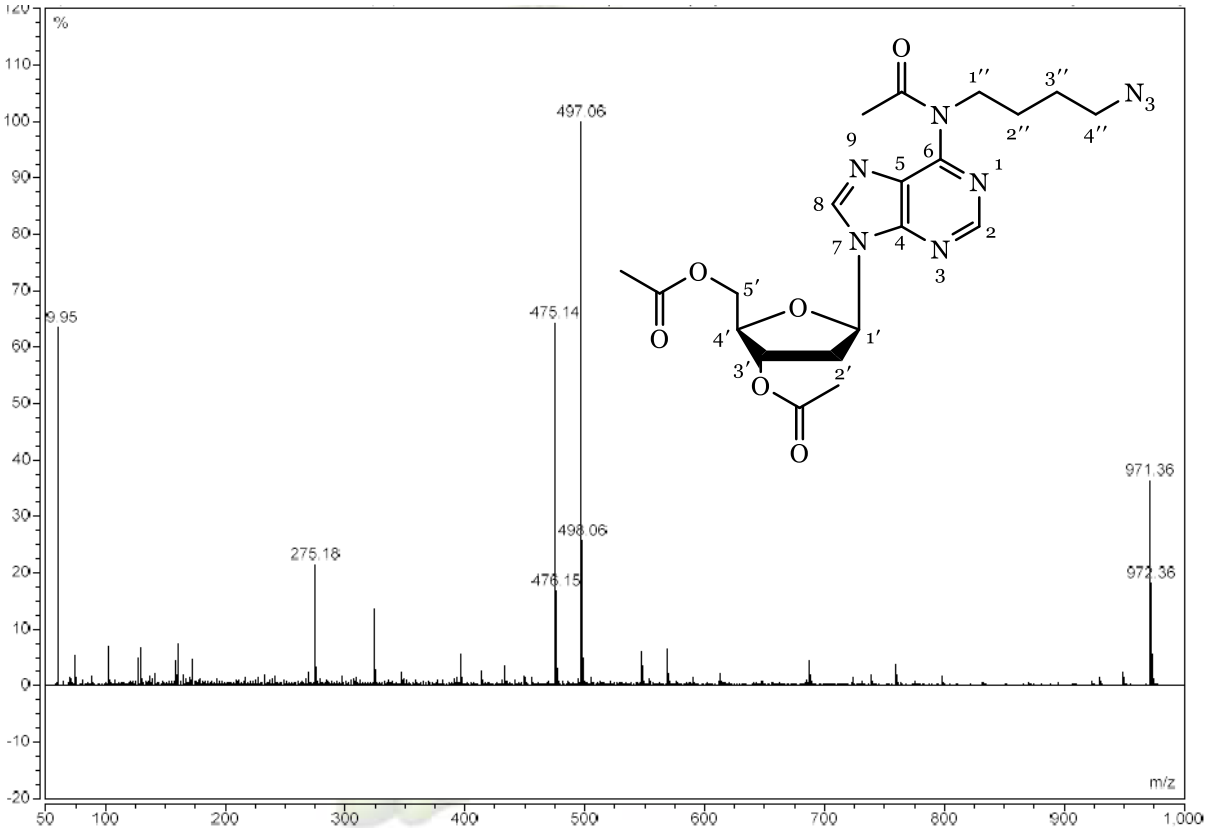


Figure SI-69. Mass spectrum of N⁶*-[4-azidobutane]-5′,3′,*N⁶*-triacetyl-2′-deoxyadenosine (compound* ***5b****).*

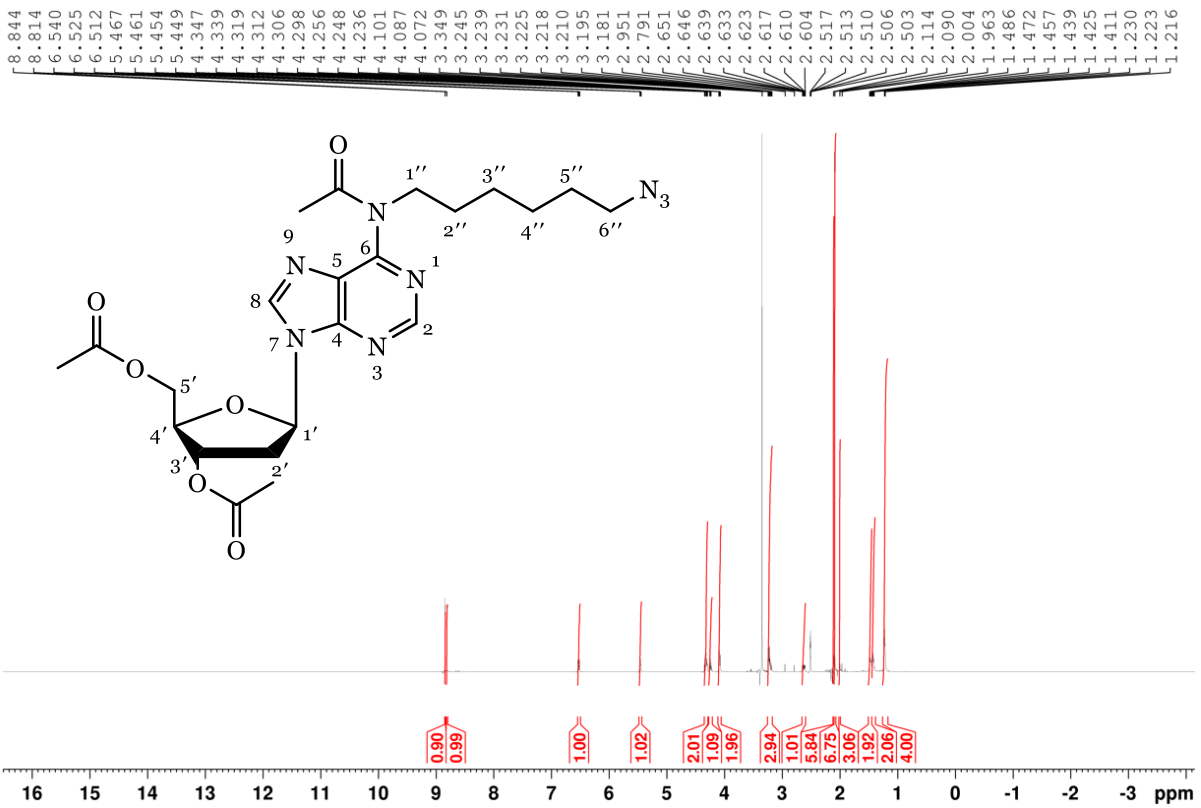


Figure SI- 70. ¹H NMR spectrum of N⁶*-[6-azidohexane]-5′,3′,*N⁶*-triacetyl-2′-deoxyadenosine (compound* ***5c****).*

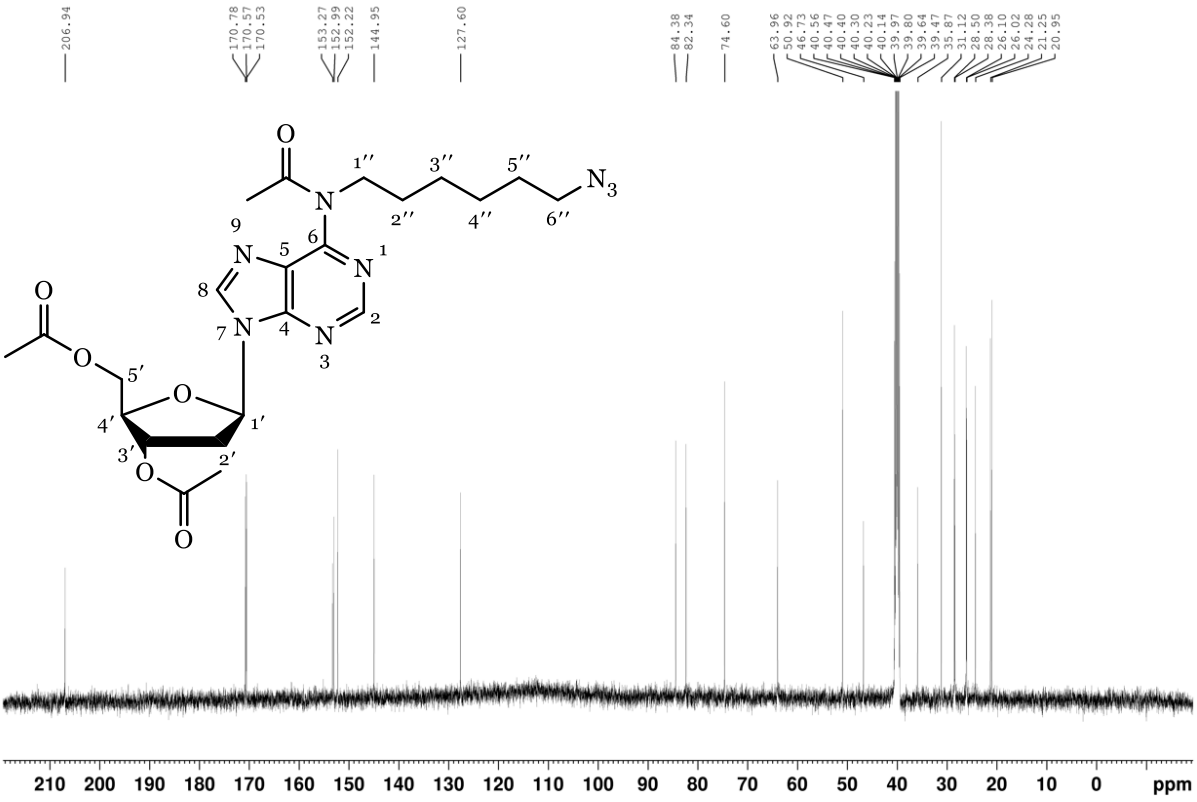


*Figure SI-71. ^{13}C NMR spectrum of N^6-[6-azidohexane]-5′,3′,N^6-triacetyl-2′-deoxyadenosine (compound **5c**).*

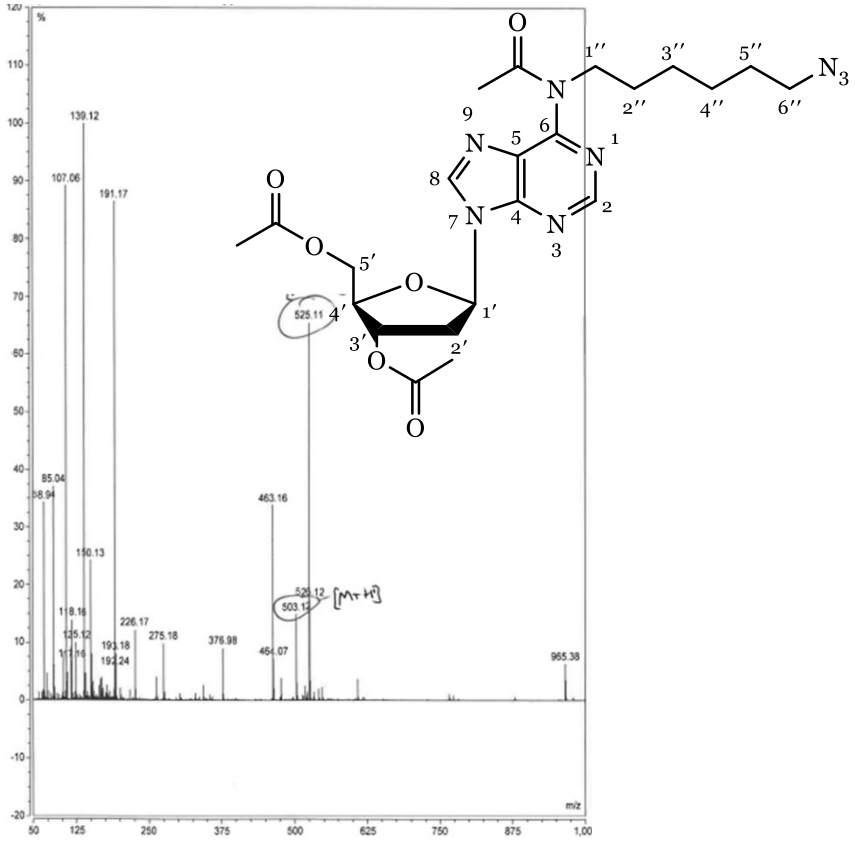


*Figure SI-72. Mass spectrum of N^6-[6-azidohexane]-5′,3′,N^6-triacetyl-2′-deoxyadenosine (compound **5c**).*

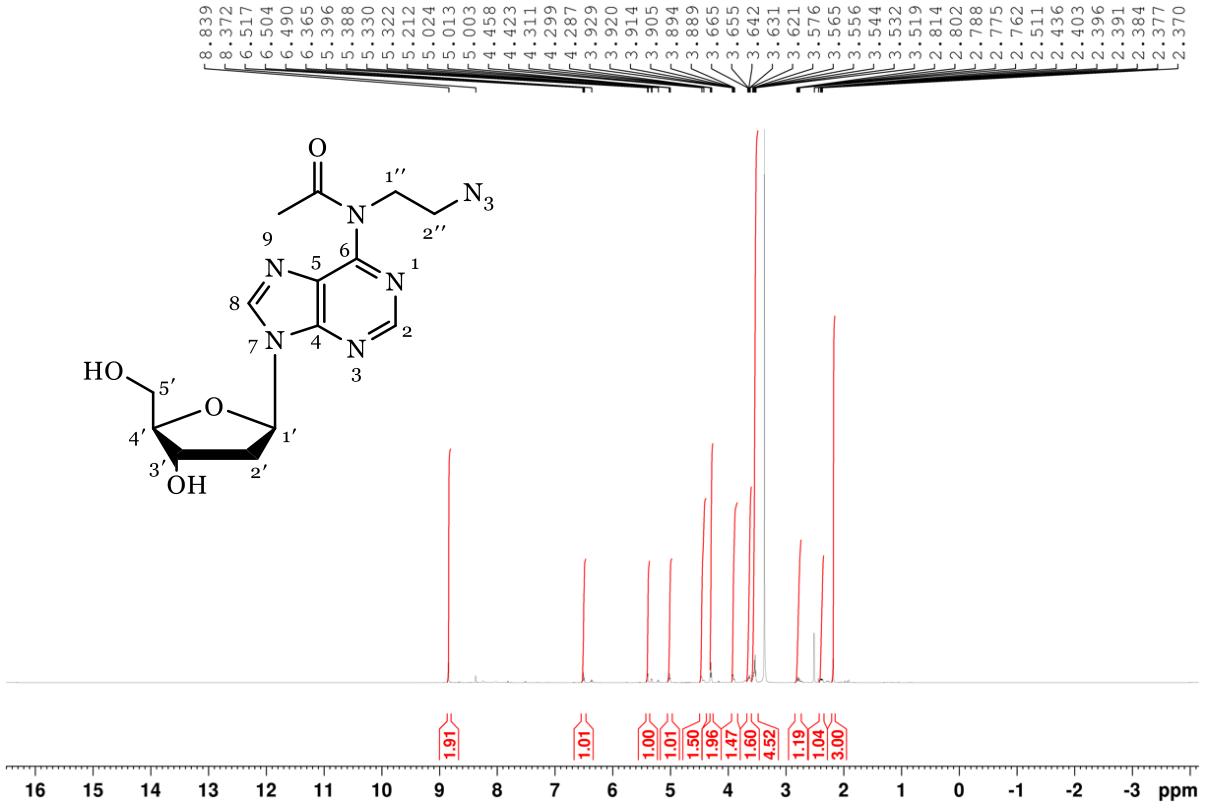


Figure SI- 73. ^{1}H NMR spectrum of N⁶-*[2-azidoethane]*-N⁶-*acetyl-2′-deoxyadenosine (compound* ***6a****).*

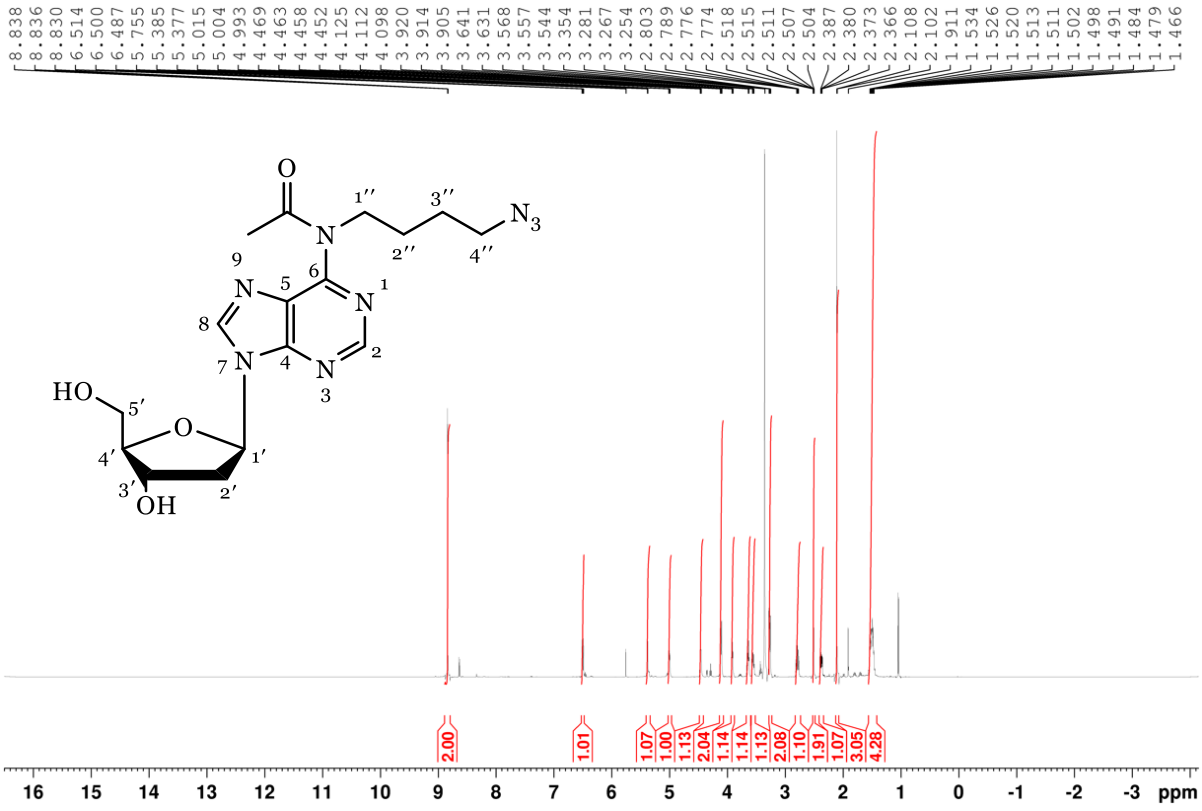


Figure SI-74. ^{1}H NMR spectrum of N⁶-*[4-azidobutane]*-N⁶-*acetyl-2′-deoxyadenosine (compound* ***6b****).*

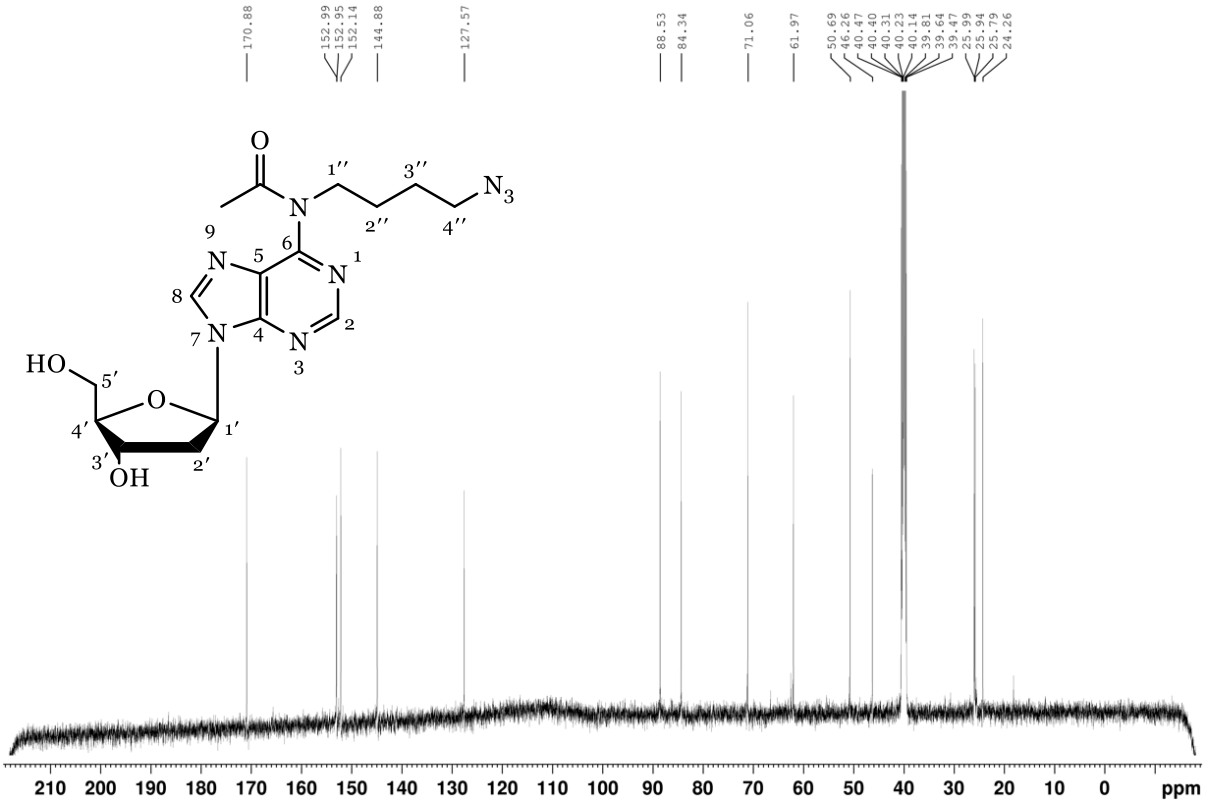


Figure SI-75. ^{13}C NMR spectrum of N⁶-*[4-azidobutane]*-N⁶-*acetyl-2′-deoxyadenosine (compound* ***6b****).*

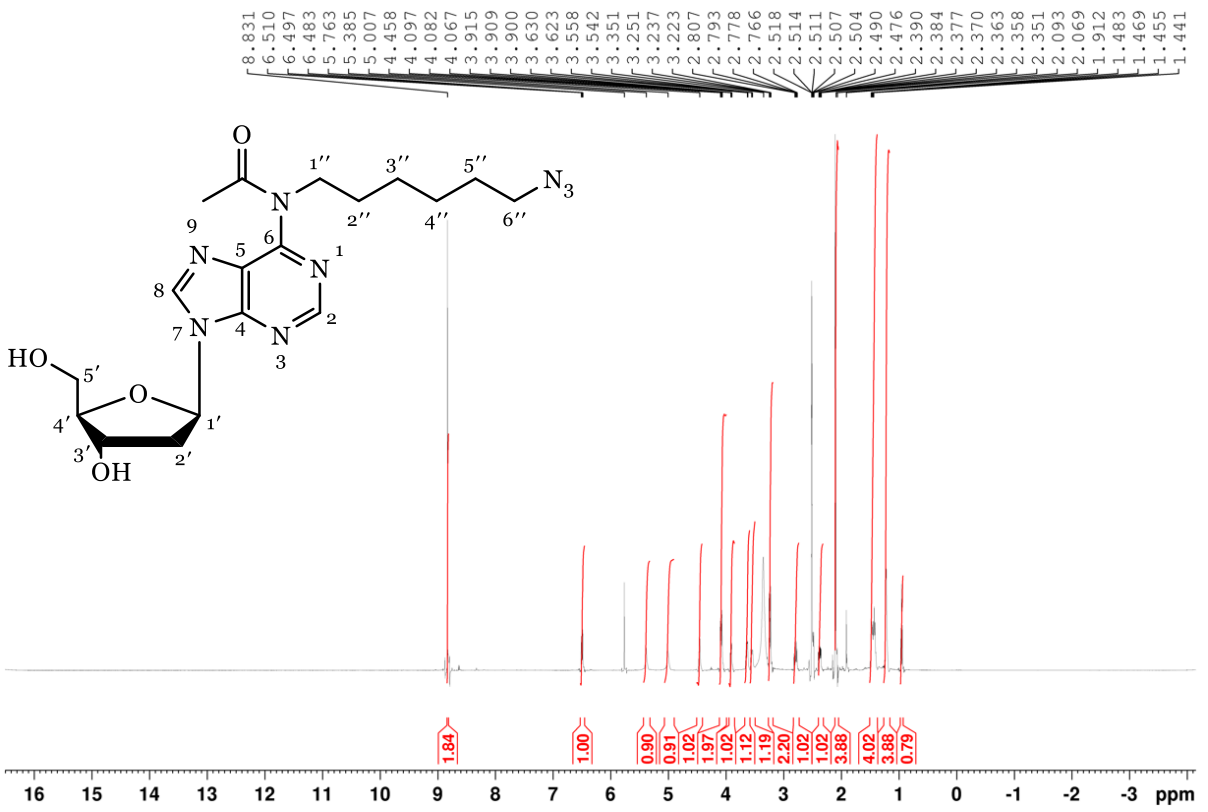


Figure SI-76. ^{1}H NMR spectrum of N⁶-*[6-azidohexane]*-N⁶-*acetyl-2′-deoxyadenosine (compound* ***6c****).*

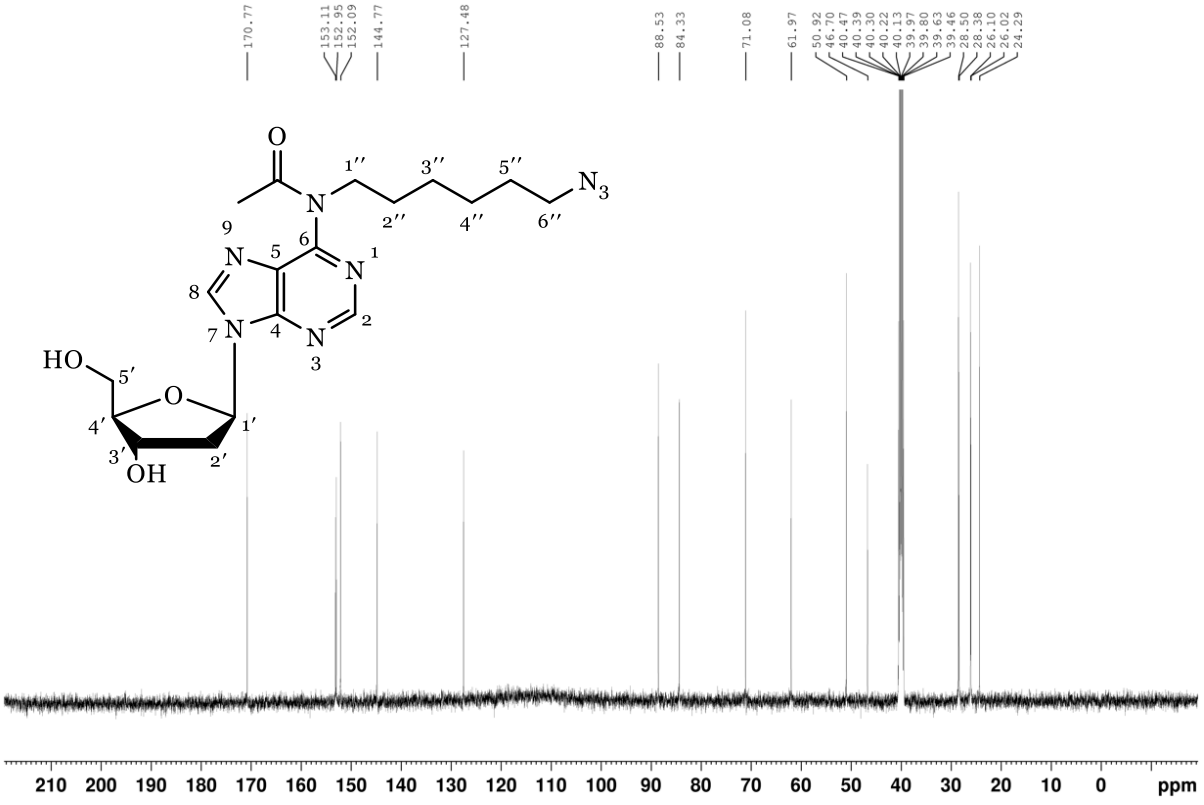


Figure SI-77. ^{13}C NMR spectrum of N^6*-[6-azidohexane]-*N^6*-acetyl-2′-deoxyadenosine (compound* ***6c****).*

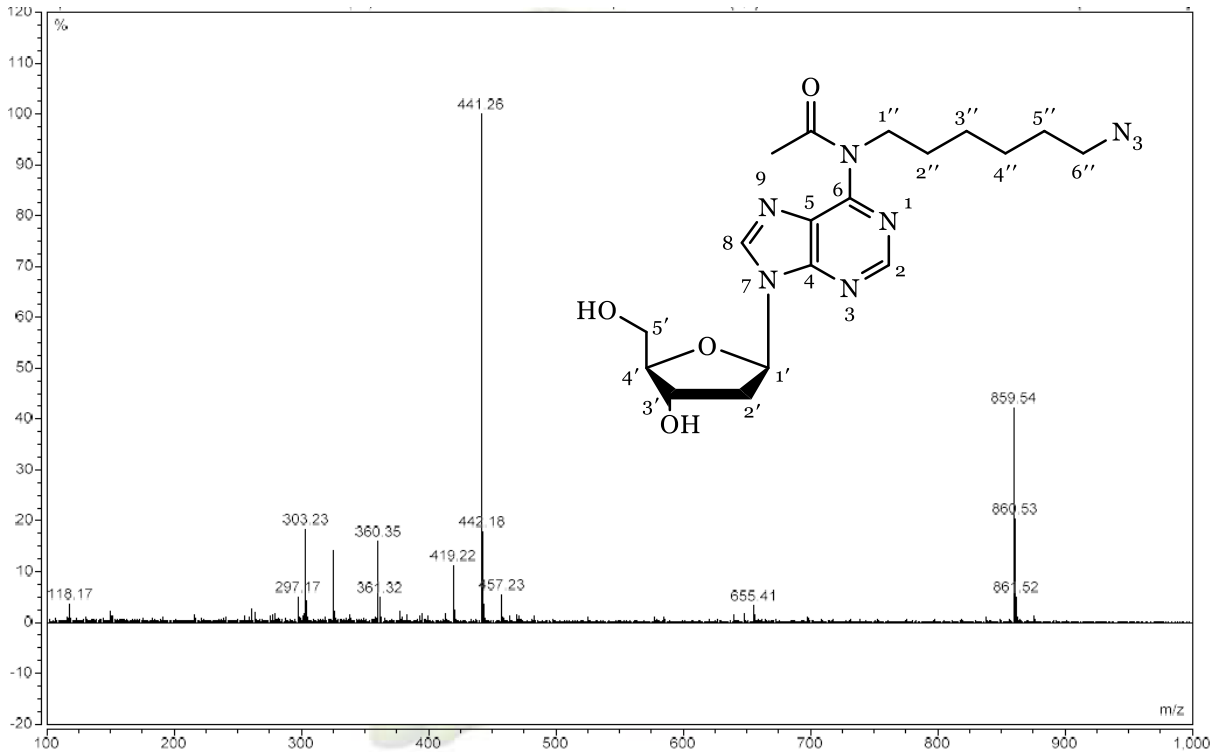


Figure SI-78. Mass spectrum of N^6*-[6-azidohexane]-*N^6*-acetyl-2′-deoxyadenosine (compound* ***6c****).*

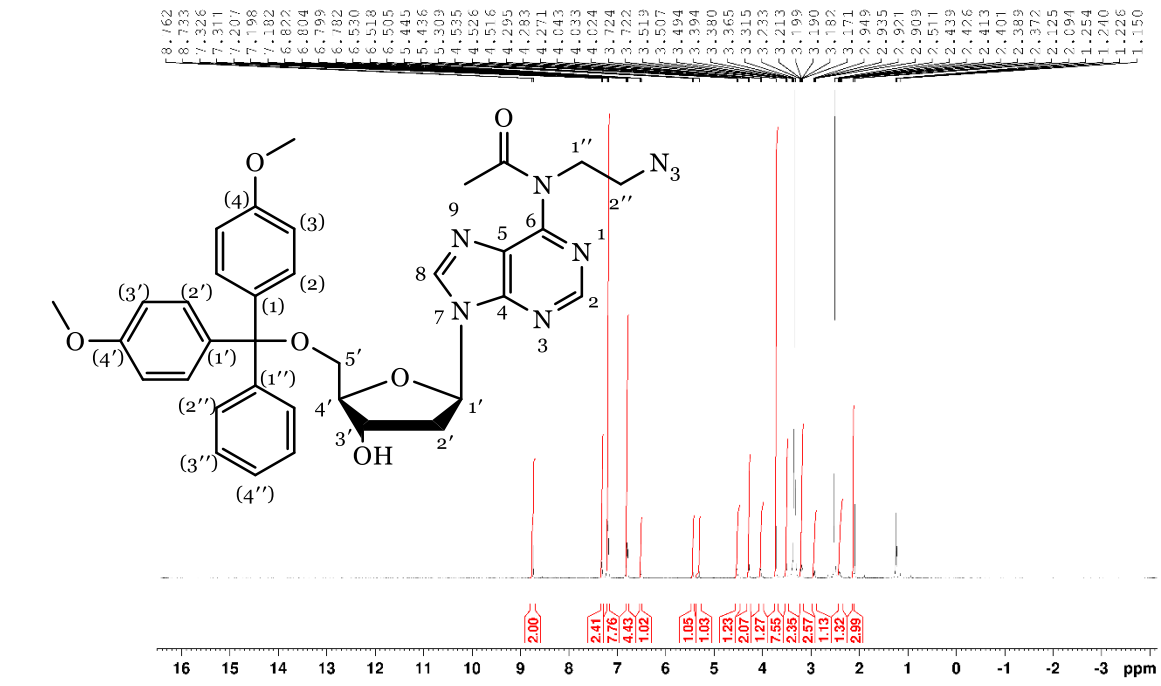


*Figure SI-79. ^{1}H NMR spectrum of 5′-DMT-N^{6}-[2-azidoethane]-N^{6}-acetyl-2′-deoxyadenosine (compound **7a**).*

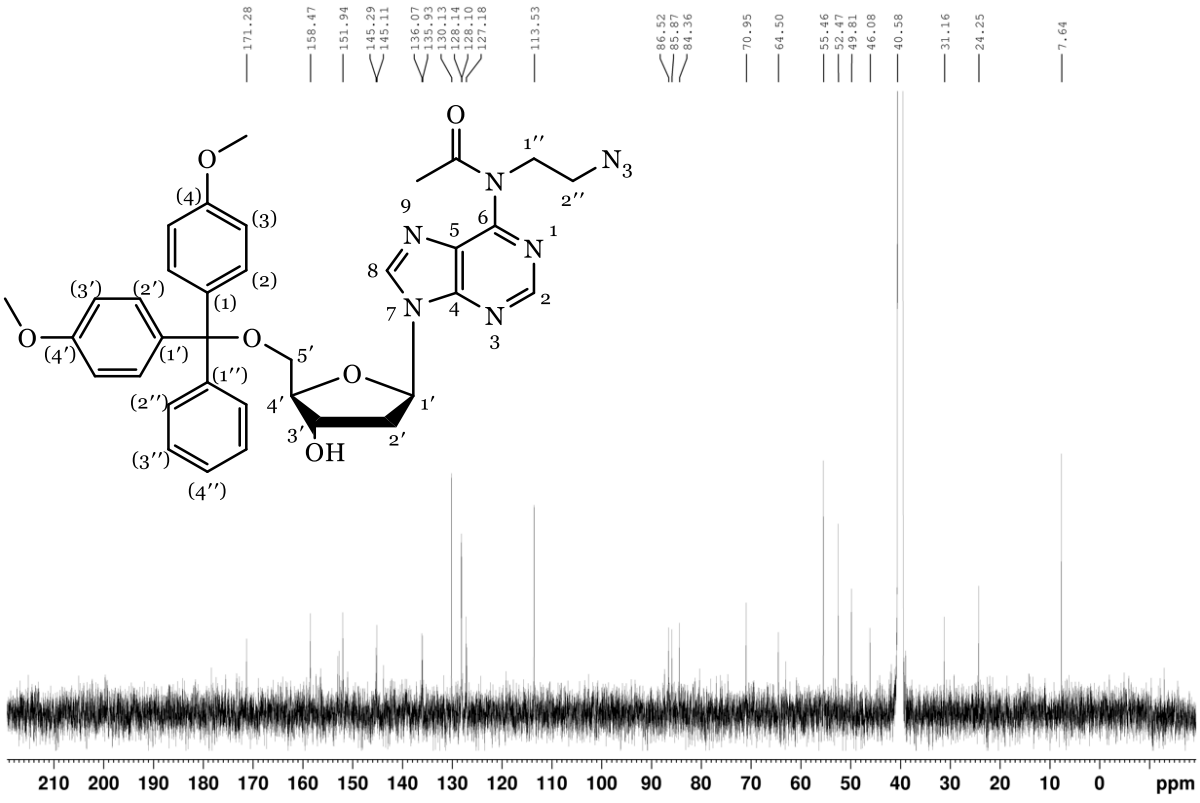


*Figure SI-80. ^{13}C NMR spectrum of 5′-DMT-N^{6}-[2-azidoethane]-N^{6}-acetyl-2′-deoxyadenosine (compound **7a**).*

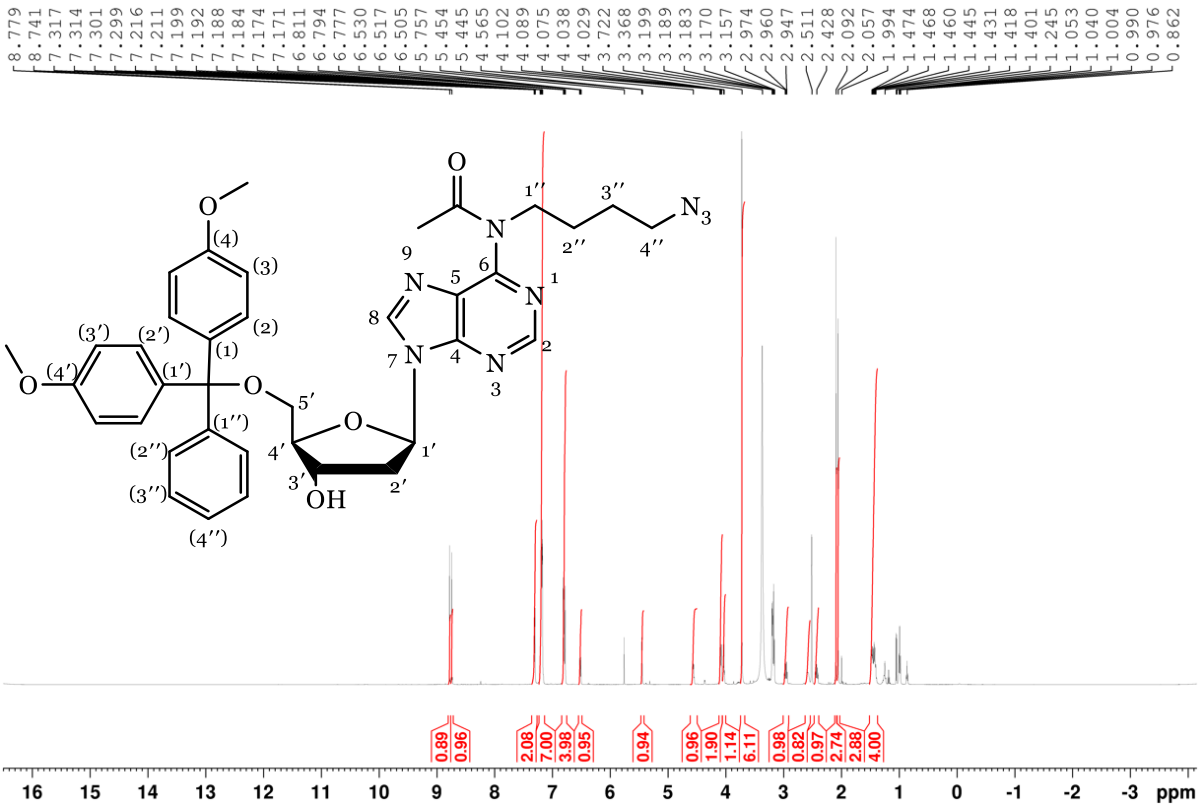


*Figure SI-81. ^{1}H NMR spectrum of 5′-DMT-N^{6}-[4-azidobutane]-N^{6}-acetyl-2′-deoxyadenosine (compound **7b**).*

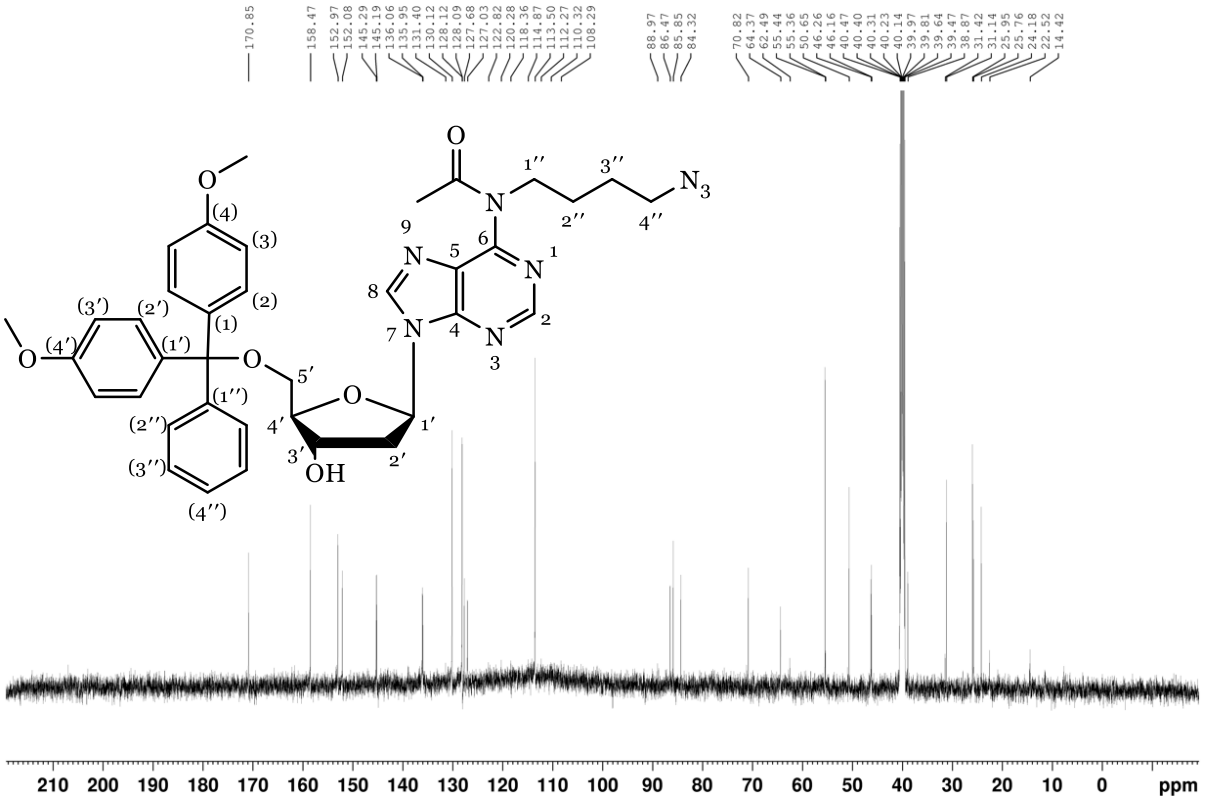


*Figure SI-82. ^{13}C NMR spectrum of 5′-DMT-N^{6}-[4-azidobutane]-N^{6}-acetyl-2′-deoxyadenosine (compound **7b**).*

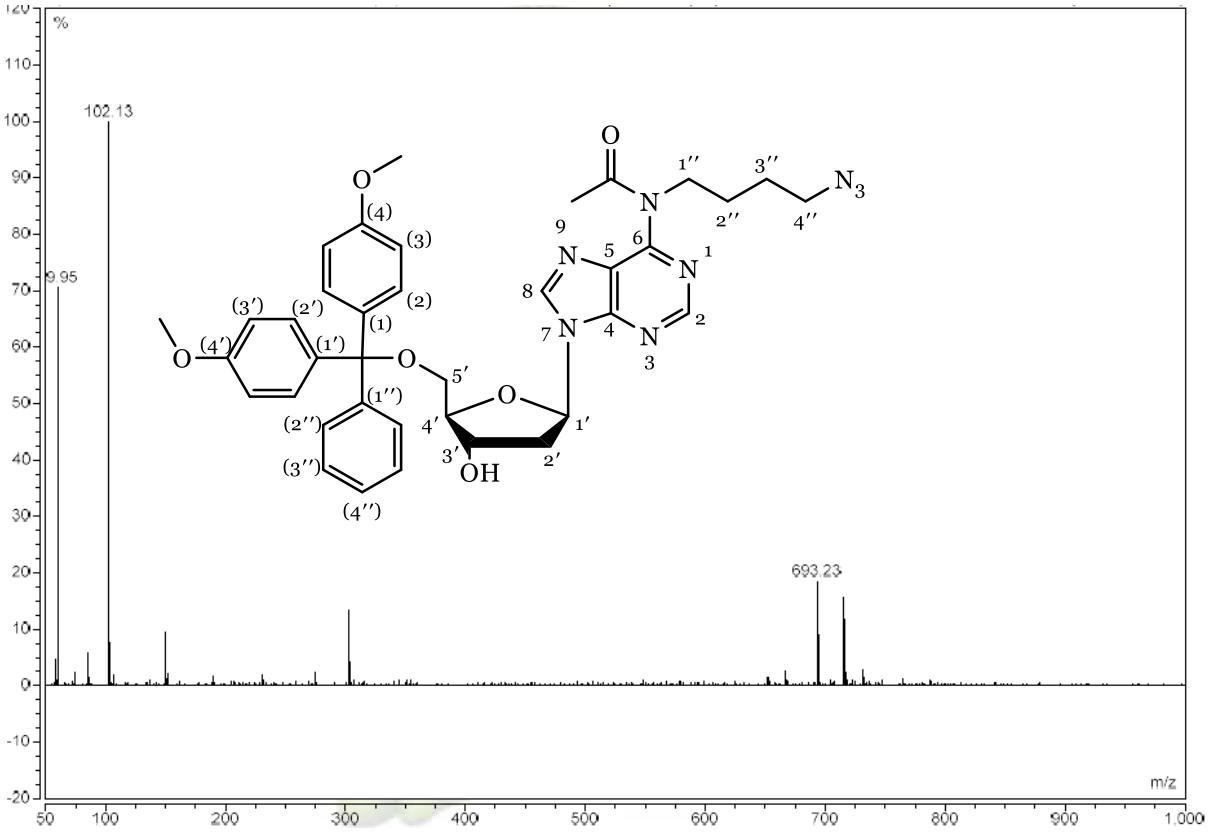


*Figure SI-83. Mass spectrum of 5′-DMT-*N^6*-[4-azidobutane]-*N^6*-acetyl-2′-deoxyadenosine (compound **7b**).*

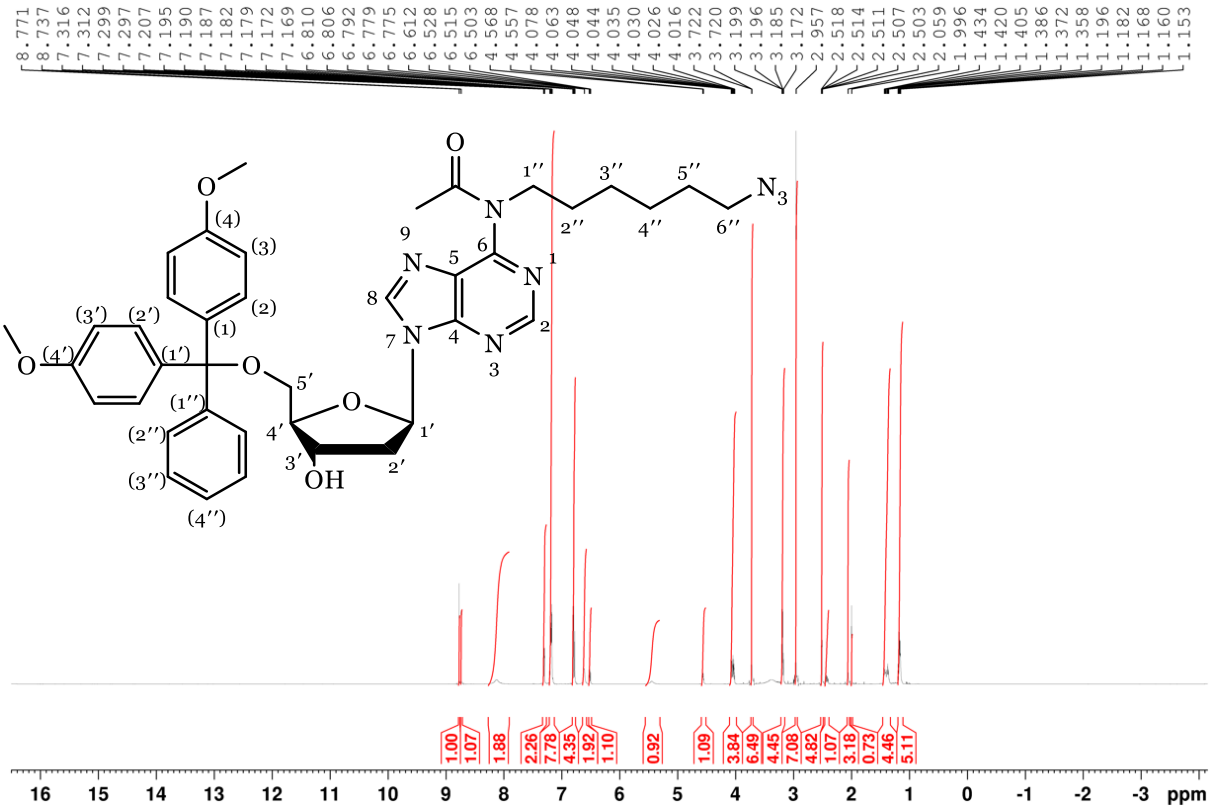


*Figure SI-84. ^{1}H NMR spectrum of 5′-DMT-*N^6*-[6-azidohexane]-*N^6*-acetyl-2′-deoxyadenosine (compound **7c**).*

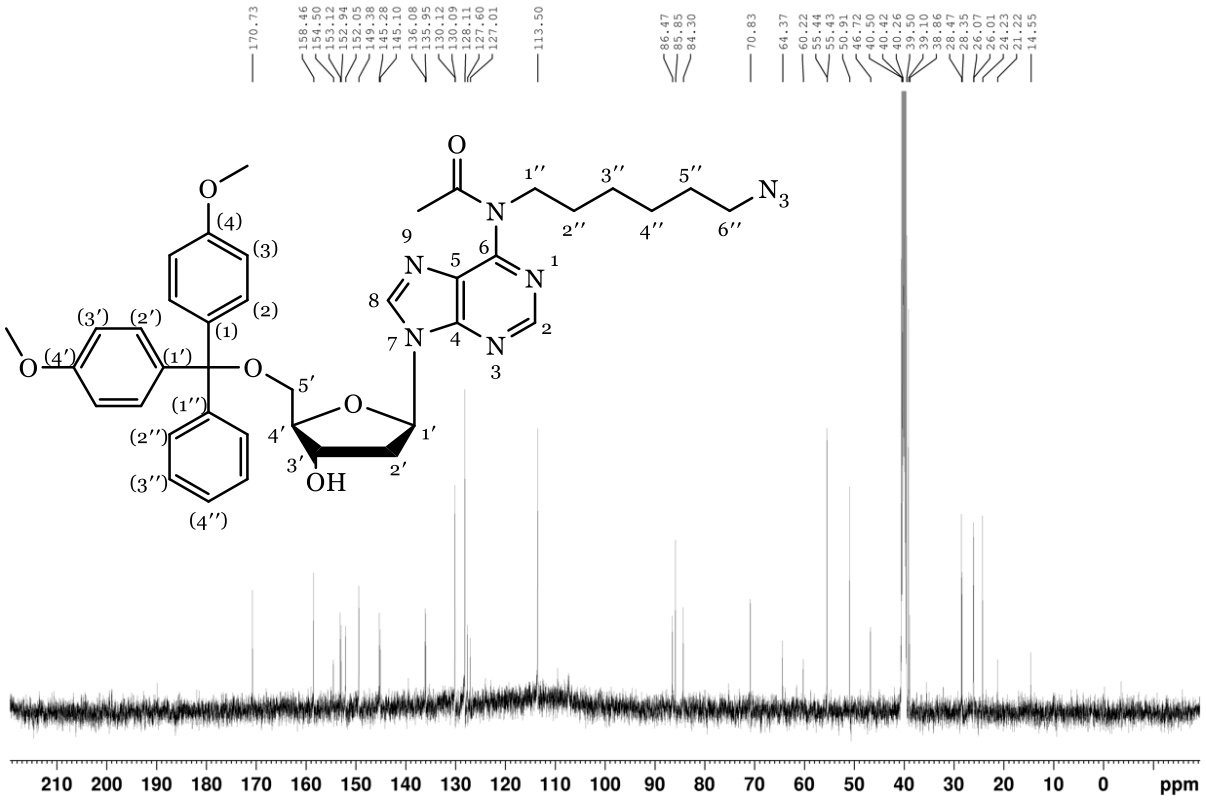


*Figure SI- 85. ^{13}C NMR spectrum of 5′-DMT-N^6-[6-azidohexane]-N^6-acetyl-2′-deoxyadenosine (compound **7c**).*

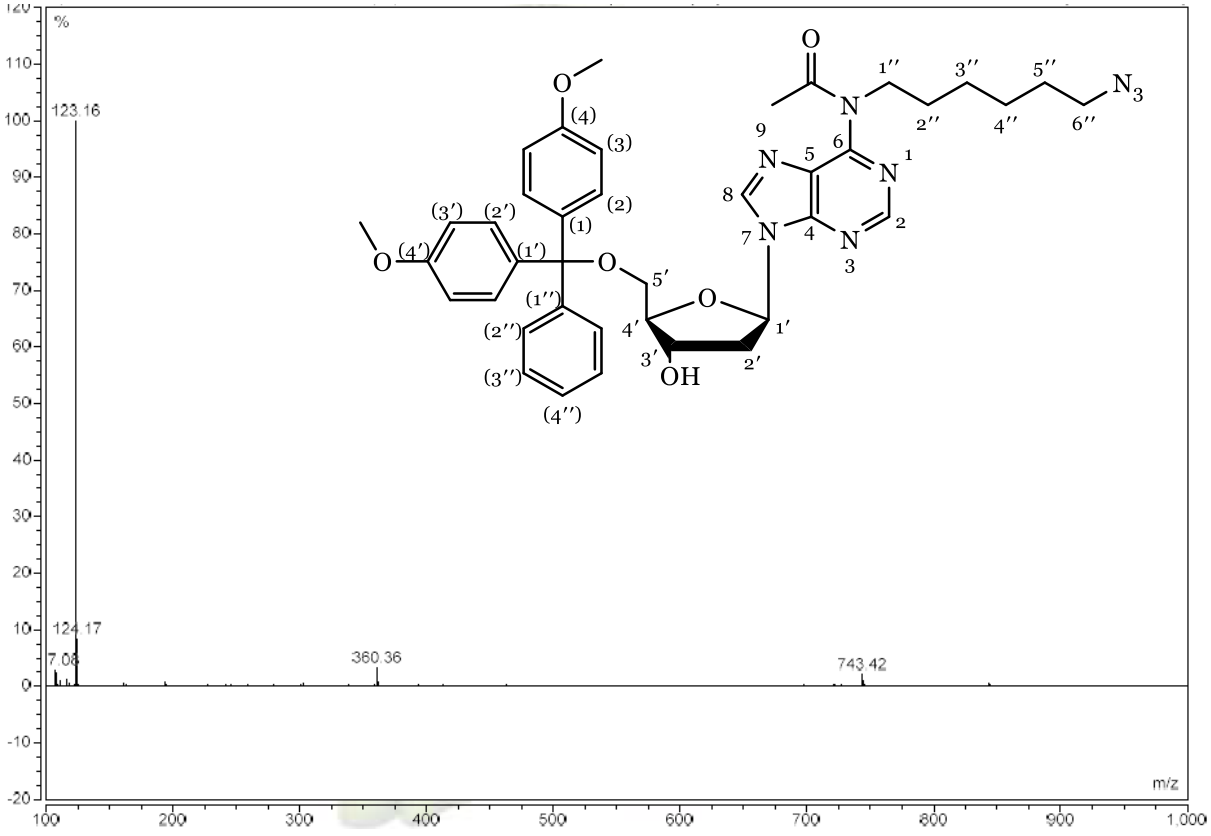


*Figure SI-86. Mass spectrum of 5′-DMT-N^6-[6-azidohexane]-N^6-acetyl-2′-deoxyadenosine (compound **7c**).*

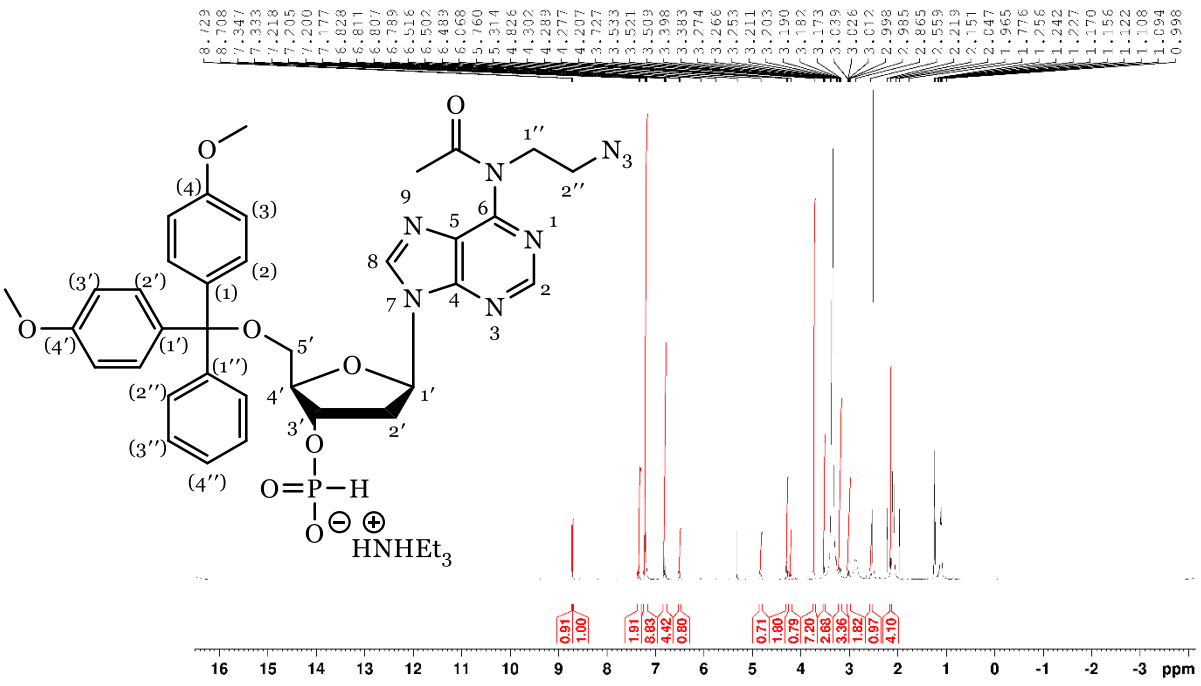


*Figure SI-87. ^{1}H NMR spectrum of 5′-DMT-N^{6}-[2-azidoethane]-N^{6}-acetyl-2′-deoxyadenosine, 3′-H-phosphonate, TEA salt (compound **8a**).*

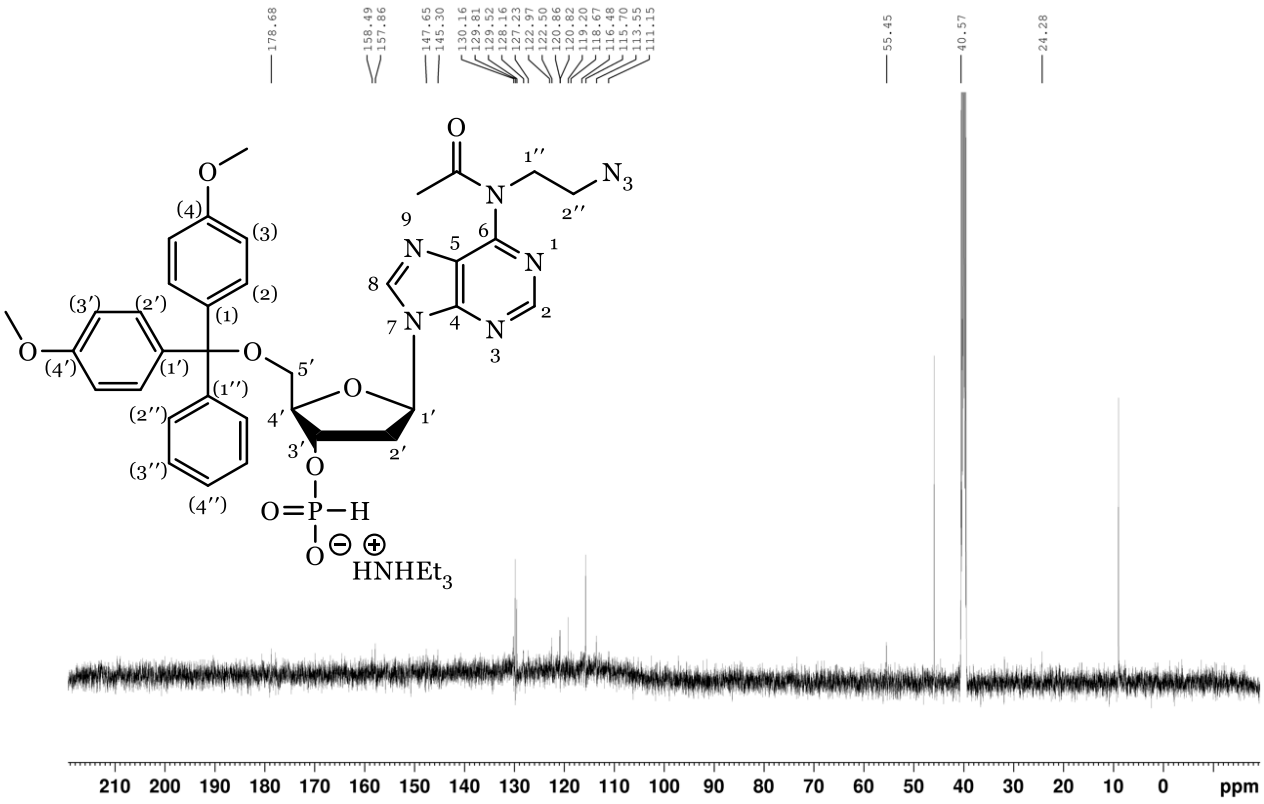


*Figure SI- 88. ^{13}C NMR spectrum of 5′-DMT-N^{6}-[2-azidoethane]-N^{6}-acetyl-2′-deoxyadenosine, 3′-H-phosphonate, TEA salt (compound **8a**).*

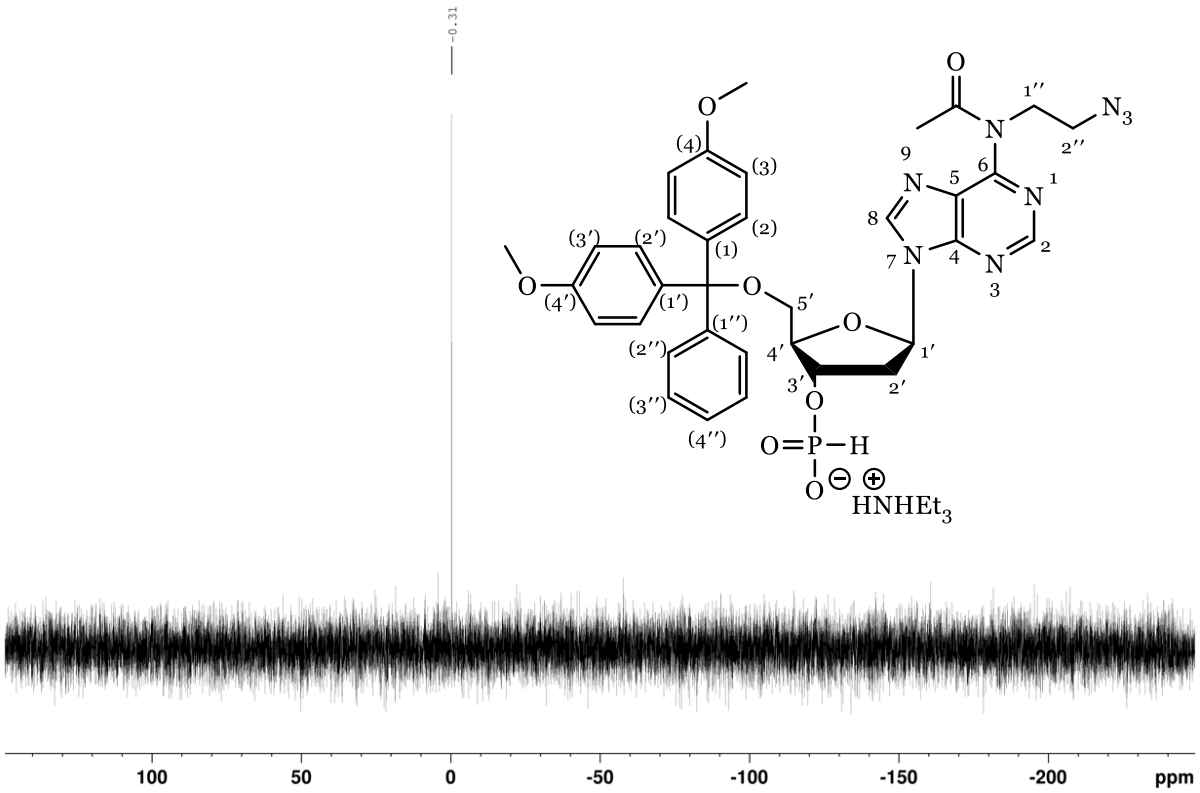


*Figure SI-89. ^{31}P NMR spectrum of 5′-DMT-N^6-[2-azidoethane]-N^6-acetyl-2′-deoxyadenosine, 3′-H-phosphonate, TEA salt (compound **8a**).*

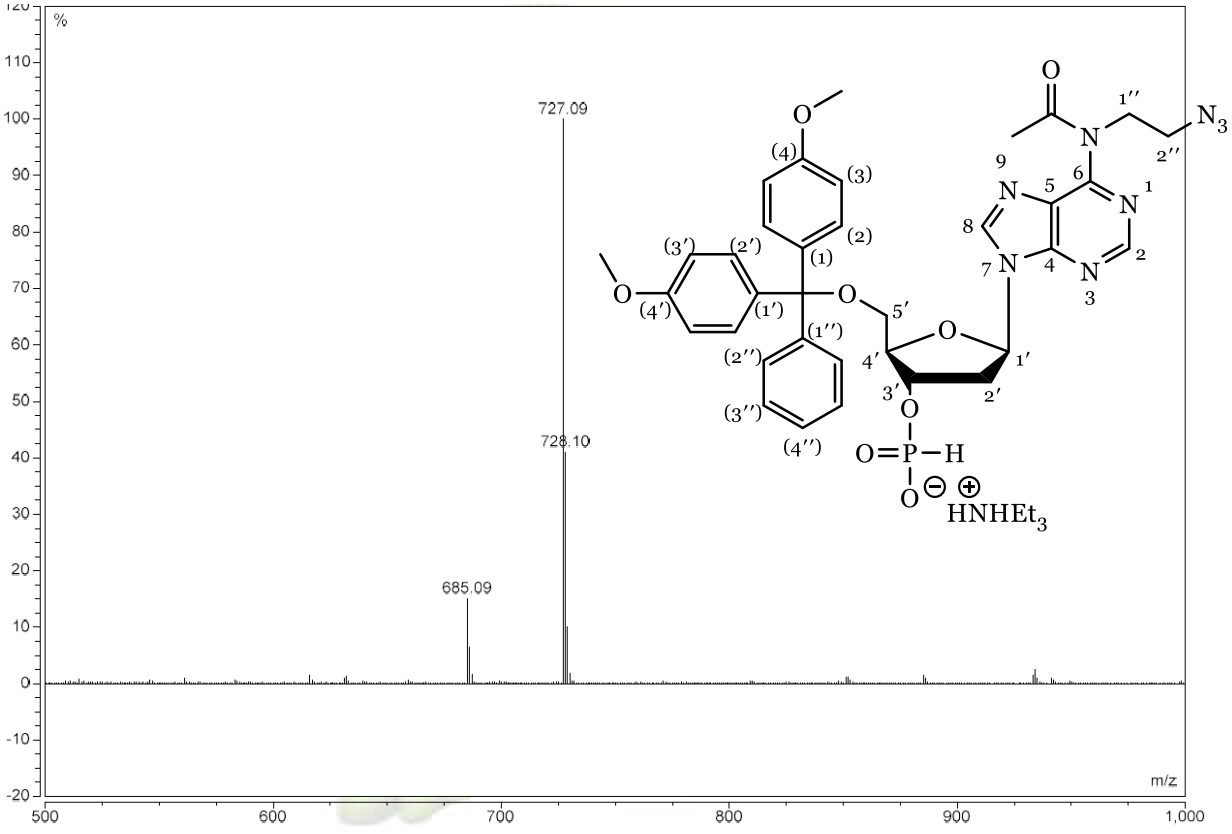


*Figure SI- 90. Mass spectrum of 5′-DMT-N^6-[2-azidoethane]-N^6-acetyl-2′-deoxyadenosine, 3′-H-phosphonate, TEA salt (compound **8a**).*

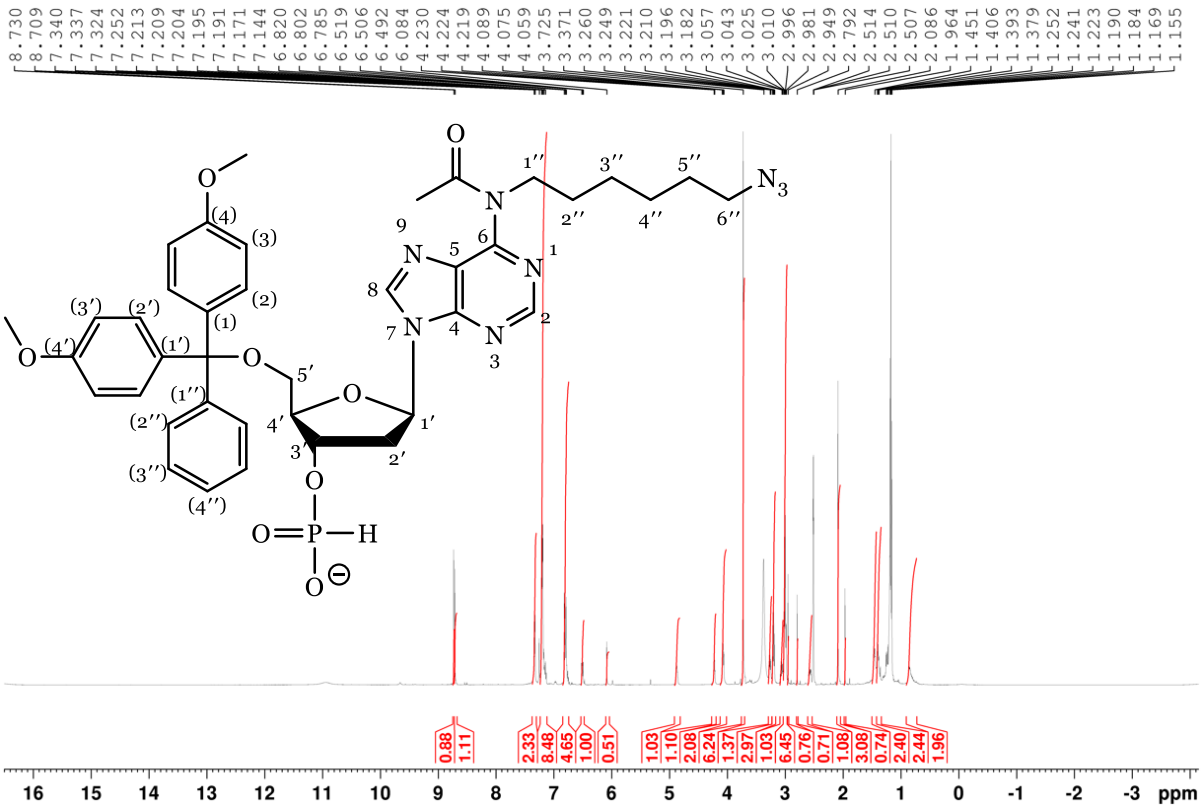


*Figure SI-91. ^{1}H NMR spectrum of 5′-DMT-N^{6}-[6-azidohexane]-N^{6}-acetyl-2′-deoxyadenosine, 3′-H-phosphonate, TEA salt (compound **8c**).*

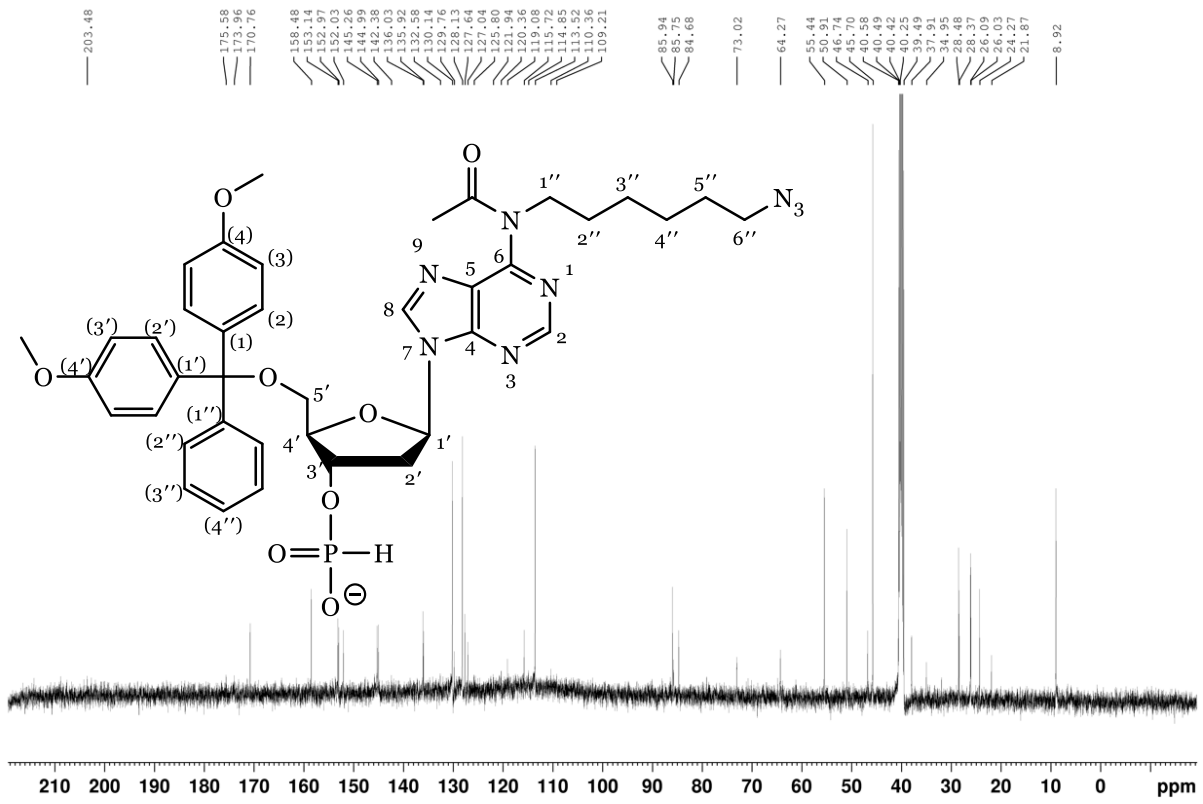


*Figure SI-92. ^{13}C NMR spectrum of 5′-DMT-N^{6}-[6-azidohexane]-N^{6}-acetyl-2′-deoxyadenosine, 3′-H-phosphonate, TEA salt (compound **8c**).*

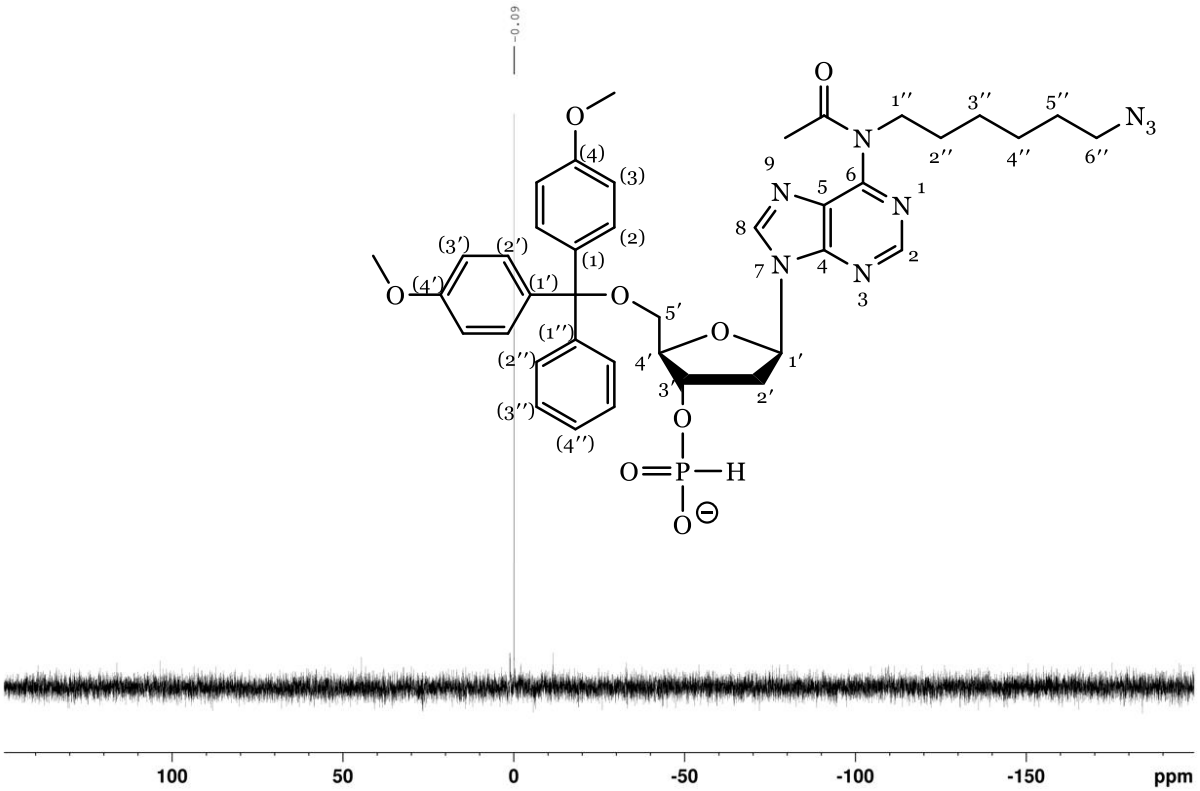


*Figure SI-93. ^{31}P NMR spectrum of 5′-DMT-N^6-[6-azidohexane]-N^6-acetyl-2′-deoxyadenosine, 3′-H-phosphonate, TEA salt (compound **8c**).*

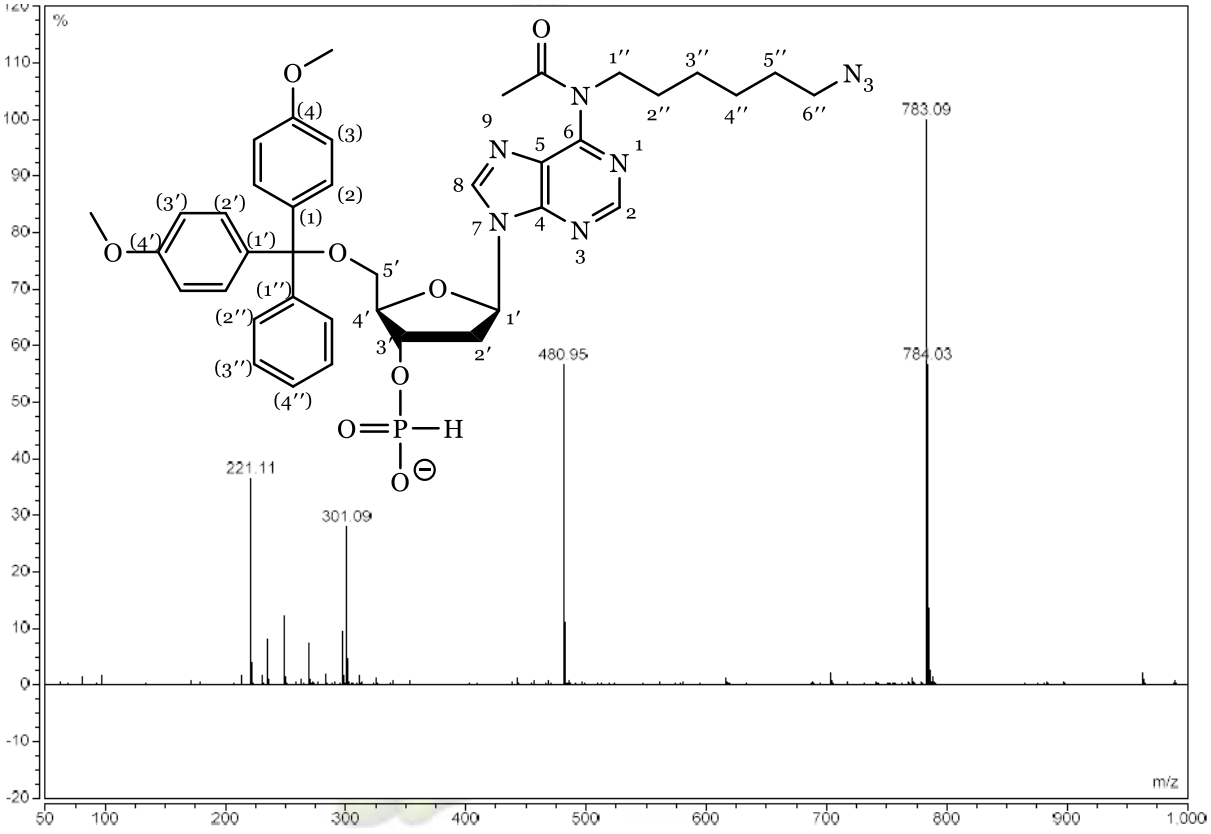


*Figure SI-94. Mass spectrum of 5′-DMT-N^6-[6-azidohexane]-N^6-acetyl-2′-deoxyadenosine, 3′-H-phosphonate, TEA salt (compound **8c**).*

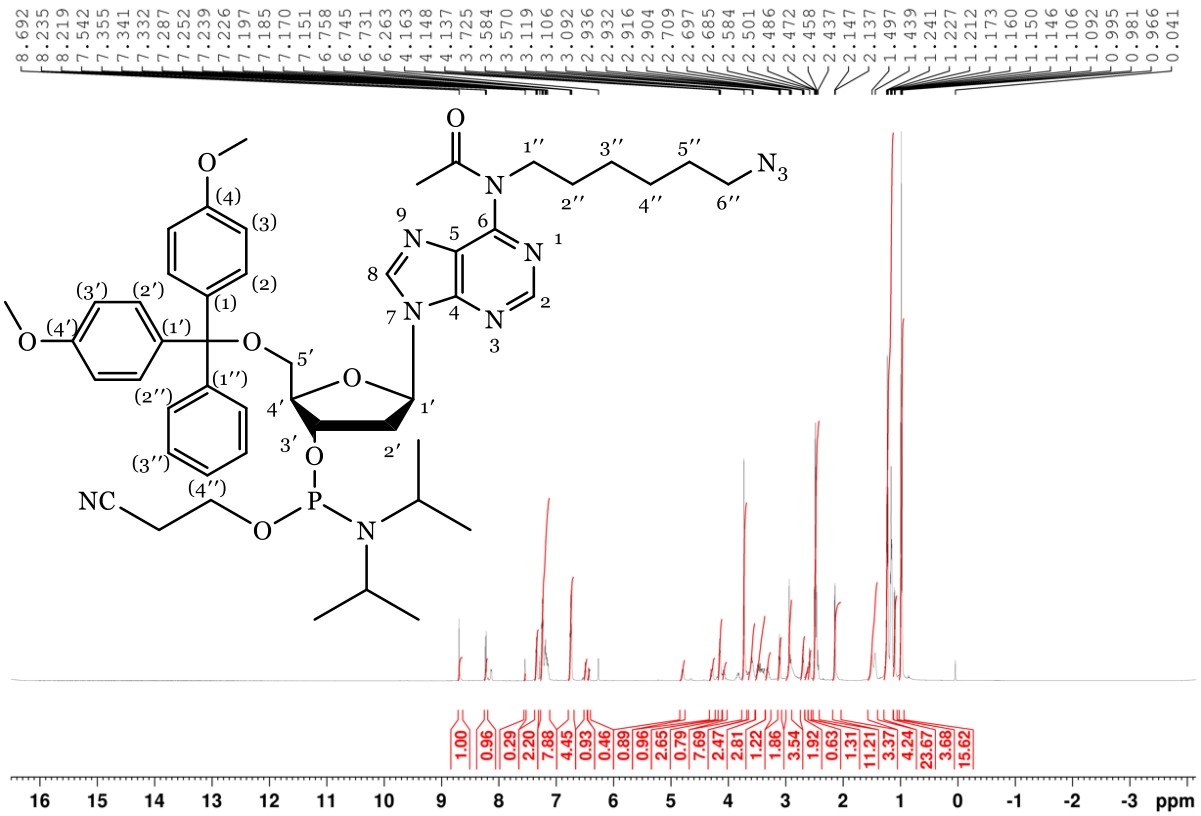


*Figure SI-96. ^{1}H NMR spectrum of 5′-DMT-N^6-[6-azidohexane]-N^6-acetyl-2′-deoxyadenosine, 3′- [2-cyanoethyl] N, N-diisopropylphosphoramidite (compound **9c**).*

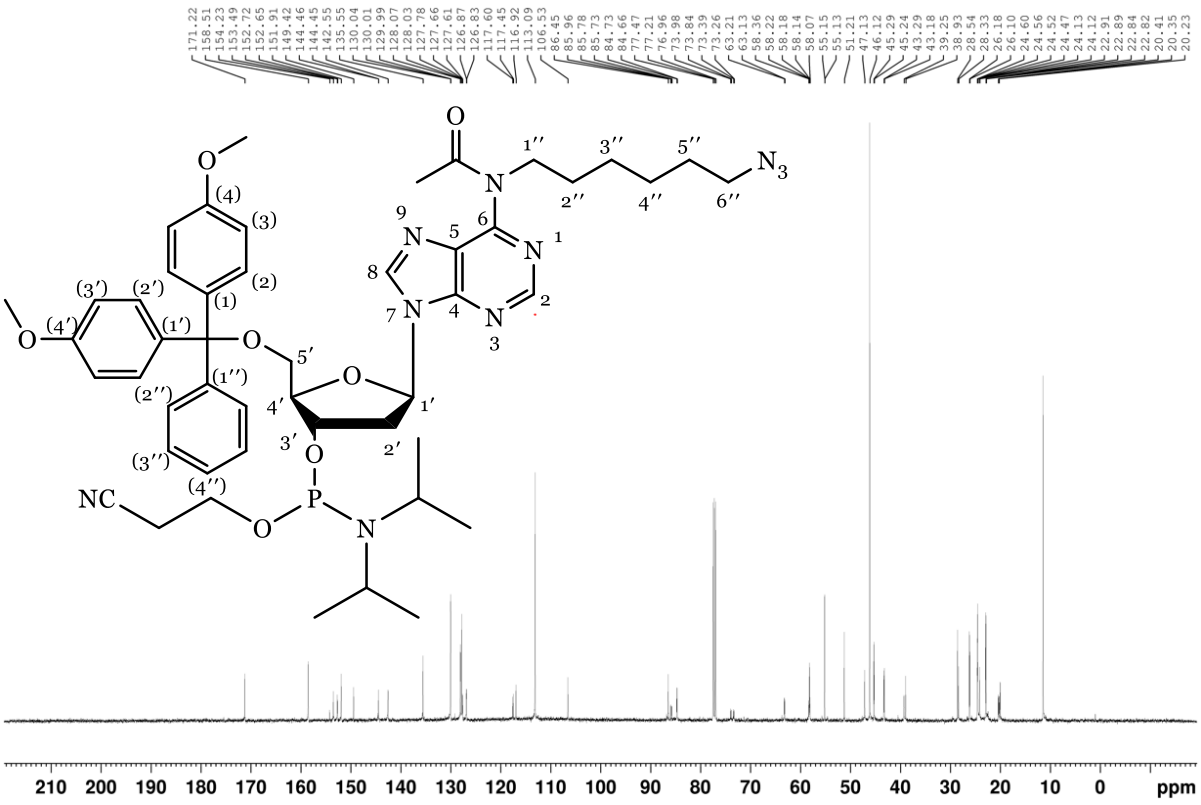


*Figure SI-95. ^{13}C NMR spectrum of 5′-DMT-N^6-[6-azidohexane]-N^6-acetyl-2′-deoxyadenosine, 3′- [2-cyanoethyl] N, N-diisopropylphosphoramidite (compound **9c**).*

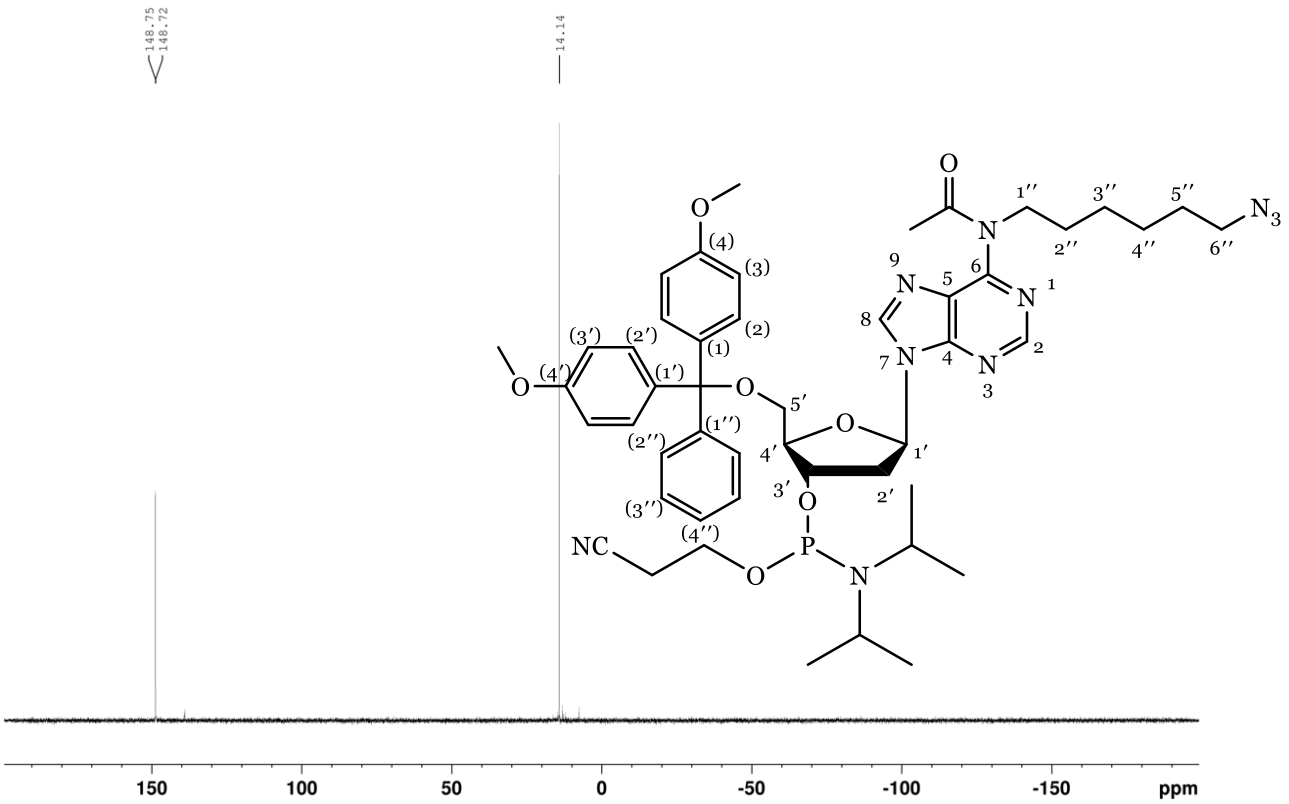


*Figure SI- 97. ^{31}P NMR spectrum of 5′-DMT-N^6-[6-azidohexane]-N^6-acetyl-2′-deoxyadenosine, 3′- [2-cyanoethyl] N, N-diisopropylphosphoramidite (compound **9c**).*

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Bruce Chilton
Name/title of Primary Supervisor:	Vyacheslav Filichev
In which chapter is the manuscript /published work:	Chapter 2

Please select one of the following three options:

○ The manuscript/published work is published or in press

- Please provide the full reference of the Research Output:

○ The manuscript is currently under review for publication – please indicate:

- The name of the journal:
- The percentage of the manuscript/published work that was contributed by the candidate:
- Describe the contribution that the candidate has made to the manuscript/published work:

◉ It is intended that the manuscript will be published, but it has not yet been submitted to a journal

Candidate's Signature:	Bruce Chilton Digitally signed by Bruce Chilton Date: 2023.09.15 11:57:43 +12'00'
Date:	15-Sep-2023
Primary Supervisor's Signature:	Vyacheslav Filichev Digitally signed by Vyacheslav Filichev Date: 2023.09.15 12:03:04 +12'00'
Date:	15-Sep-2023

This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis.

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Bruce Chilton
Name/title of Primary Supervisor:	Vyacheslav Filichev
In which chapter is the manuscript /published work:	Chapter 3

Please select one of the following three options:

○ The manuscript/published work is published or in press

- Please provide the full reference of the Research Output:

○ The manuscript is currently under review for publication – please indicate:

- The name of the journal:
- The percentage of the manuscript/published work that was contributed by the candidate:
- Describe the contribution that the candidate has made to the manuscript/published work:

◉ It is intended that the manuscript will be published, but it has not yet been submitted to a journal

Candidate's Signature:	Bruce Chilton Digitally signed by Bruce Chilton Date: 2023.09.15 11:57:43 +12'00'
Date:	15-Sep-2023
Primary Supervisor's Signature:	Vyacheslav Filichev Digitally signed by Vyacheslav Filichev Date: 2023.09.15 12:02:35 +12'00'
Date:	15-Sep-2023

This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis.

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Bruce Chilton
Name/title of Primary Supervisor:	Vyacheslav Filichev
In which chapter is the manuscript /published work:	Chapter 4

Please select one of the following three options:

○ The manuscript/published work is published or in press

- Please provide the full reference of the Research Output:

○ The manuscript is currently under review for publication – please indicate:

- The name of the journal:
- The percentage of the manuscript/published work that was contributed by the candidate:
- Describe the contribution that the candidate has made to the manuscript/published work:

◉ It is intended that the manuscript will be published, but it has not yet been submitted to a journal

Candidate's Signature:	Bruce Chilton Digitally signed by Bruce Chilton Date: 2023.09.15 11:57:43 +12'00'
Date:	15-Sep-2023
Primary Supervisor's Signature:	Vyacheslav Filichev Digitally signed by Vyacheslav Filichev Date: 2023.09.15 12:03:27 +12'00'
Date:	15-Sep-2023

This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis.

MASSEY UNIVERSITY
TE KUNENGA KI PŪREHUROA
UNIVERSITY OF NEW ZEALAND

GRADUATE RESEARCH SCHOOL

STATEMENT OF CONTRIBUTION
DOCTORATE WITH PUBLICATIONS/MANUSCRIPTS

We, the candidate and the candidate's Primary Supervisor, certify that all co-authors have consented to their work being included in the thesis and they have accepted the candidate's contribution as indicated below in the *Statement of Originality*.

Name of candidate:	Bruce Chilton
Name/title of Primary Supervisor:	Vyacheslav Filichev
In which chapter is the manuscript /published work:	Chapter 5

Please select one of the following three options:

○ The manuscript/published work is published or in press

- Please provide the full reference of the Research Output:

○ The manuscript is currently under review for publication – please indicate:

- The name of the journal:
- The percentage of the manuscript/published work that was contributed by the candidate:
- Describe the contribution that the candidate has made to the manuscript/published work:

◉ It is intended that the manuscript will be published, but it has not yet been submitted to a journal

Candidate's Signature:	Bruce Chilton Digitally signed by Bruce Chilton Date: 2023.09.15 11:57:43 +12'00'
Date:	15-Sep-2023
Primary Supervisor's Signature:	Vyacheslav Filichev Digitally signed by Vyacheslav Filichev Date: 2023.09.15 12:03:47 +12'00'
Date:	15-Sep-2023

This form should appear at the end of each thesis chapter/section/appendix submitted as a manuscript/ publication or collected as an appendix at the end of the thesis.