Development of Cross-linking Strategies for DNA G-quadruplexes

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

Master of Science

In

Chemistry

Massey University, Manawatu, New Zealand

MASSEY
UNIVERSITY

Bruce Chilton

Supervisor: Assoc. Prof. Vyacheslav Filichev

2019

Abstract:

Chemically modified DNA structures are an important development in the study of DNA properties. They allow the manipulation of biophysical properties of DNA complexes, which can give unique perspectives on the behaviours of DNA and molecules, such as enzymes, which interact with DNA. The use of chemical modification to study G-quadruplex structures has been limited, but has shown potential as a method of controlling their topology and behaviour. We plan to achieve this by functionalising positions within quadruplex forming sequences and using this to create linkages between strands. We predict this will improve the stability of the secondary structure and improve k_{on}, the association rate, of the complex. We discuss several general approaches to G-quadruplex modification, and further discuss specific strategies for carrying out these modifications.

One general method for modifying DNA is directly modifying the nucleotides that make up the sequence. G-quadruplexes are primarily composed of guanosine, modification of which is primarily possible at two positions, 2 and 8, due to the hydrogen bond arrangement within the structure. We target the 2- and 8- position of modified guanosine molecules with amine substitutions and Sonogashira coupling processes, respectively. We report the synthesis of sequences containing thiol functionalised nucleotides, and an initial assessment of their biophysical properties.

The second general method for modifying DNA is incorporating unique, non-native nucleotides with specific functionalities. This has previously been accomplished by incorporating ligands directly into sequences, which were used to form G-quadruplexes with transition metal ions such as copper(II), nickel(II) and cobalt(II). We aim to synthesise sequences containing a modified sugar and incorporate pyridine instead of a nucleobase. We hope this will provide quadruplexes which more closely mimic the structure of native complexes, improving the previously observed properties of modified G-quadruplexes formed in the presence of transition metal ions. We present the synthesis and biophysical properties of a unique quadruplex forming sequence containing a 4-pyridine ligand and pyrrolidine sugar. The thermal stability and kinetic properties of these modified structures are examined using circular dichroism spectroscopy, demonstrating the change in properties relative to native G4 DNA.

Abbreviations:

A:	Adenosine
a.m.u.:	Atomic mass units
Boc:	*tert*-Butyloxycarbonyl (*tert*-butyloxy carbamate as a protecting group)
Br-dG:	8-Bromo-2'-deoxyguanosine
Br-TBDMS-dG:	8-Br-3',5'-*tert*-butyldimethylsilyl-2'-deoxyguanosine
C:	Cytosine
CD:	Circular Dichroism
CE:	2-Cyanoethyl (protecting group in DNA synthesiser phosphoramidites)
CPG:	Controlled Pore Glass
eq:	Equivalence
DCM:	Dichloromethane
dG:	2'-Deoxyguanosine
DIAD:	Diisopropyl azodicarboxylate
DMAP:	4-Dimethylaminopyridine
DMF:	Dimethylformamide
DMSO:	Dimethyl Sulfoxide
DMT:	4,4'-Dimethoxytrityl
DNA:	Deoxyribonucleic Acid
DTT:	1,4-Dithiothreitol
EDTA:	Ethylenediaminetetraacetic acid
ε	Molar Absorptivity coefficient
EtOAc:	Ethyl Acetate
EtOH:	Ethanol
ETT:	5-Ethylthio-1H-tetrazole
ESI:	Electrospray Ionisation
^{F}dI:	2-Fluoro-2'-deoxyinosine
^{3C}dG:	2-(3-Aminopropan-1-thiol)-2'-deoxyguanosine
^{3C}dI:	2-(3-Aminopropan-1-thiol)-2'-deoxyinosine
^{4C}dG:	2-(4-Aminobutan-1-thiol)-2'-deoxyguanosine
^{4C}dI:	2-(4-Aminobutan-1-thiol)-2'-deoxyinosine
FRET:	Förster Resonance Energy Transfer
Fmoc:	9-Fluorenylmethyloxycarbonyl
G:	Guanosine
G4:	G-quadruplex
Hex:	Hexane
HPLC:	High Performance Liquid Chromatography
HMQC:	Heteronuclear Multiple Quantum Coherence (Indicates correlation between carbon atoms and protons across 1 bond)
I-dU:	5-Iodo-2'-deoxyuridine
k_{on}	Association rate
MS:	Mass Spectrometry
NBS:	N-Bromosuccinimide
NIS:	N-Iodosuccinimide
NMR:	Nuclear Magnetic Resonance
PMA:	Phosphomolybdic Acid
RNA:	Ribonucleic Acid
RP-HPLC	Reverse Phase High Performance Liquid Chromatography
R_t	Retention time
T:	Thymidine
TBA:	Thrombin Binding Aptamer

TBDMS:	*tert*-Butyldimethylsilyl
TBE Buffer:	Tris/Borate/EDTA Buffer
TBHP:	*tert*-Butyl hydrogen peroxide
TEAA:	Triethylammonium acetate
TEMPO:	(2,2,6,6-Tetramethylpiperidin-1-yl)oxyl
TFA:	Trifluoroacetic acid
TLC:	Thin Layer Chromatography
THF:	Tetrahydrofuran
TMS:	Trimethylsilane
TNB:	2-Nitro-5-thiobenzoate
SE-HPLC	Size Exclusion High Performance Liquid Chromatography
SI:	Structural Index
SF1B	Super Family 1B

Acknowledgments

I would like to thank my Supervisor, Assoc. Prof. Vyacheslav Filichev, for patiently correcting the same mistake too many times. I would also like to thank everyone who has worked in the lab with me in the last two years. In particular, Maksim Kvach, Hari Kurup, Yongdong Su, Maulik Mungalpara and David Lun for helping me to learn the techniques and develop the strategies I used in this project. Finally, thank you to Thomas Telford-Munoz for being a terrible gym-buddy, but extremely helpful when I needed someone to bounce ideas off of.

Contents

1. Introduction

The most widely recognised DNA secondary structure is the duplex structure initially proposed by Watson and Crick in 1953.[1] However, since 1953, further investigation of DNA sequences has shown a number of unique secondary structures, such as Z-DNA, G-quadruplexes, i-motifs and triplexes, which form as a result of alternate, typically less favourable, hydrogen bonding interactions. Under certain conditions these interactions may become more favourable, resulting in unique and interesting DNA behaviour. The possible sequences vary widely, allowing for the formation of numerous secondary structures.

DNA is the structure responsible for storage of genetic information, and utilising this information involves interactions with various proteins. Changes to these interactions can cause or influence the development of medical conditions including cancer, genetic diseases and viral infections. Understanding the nature of these interactions and the changes is vital to treating these conditions.

1.1 Deoxyribonucleic Acid: Structure and Behaviour

The primary structure of DNA is comprised of three building blocks: a nucleobase, a ribose sugar and a phosphate. The naturally occurring bases in DNA are Adenine (A), Thymine (T), Guanine (G) and Cytosine (C). Each nucleobase contains a unique arrangements of hydrogen bond donors and acceptors, which interact most efficiently with the inverse arrangement, resulting in canonical Watson-Crick base pairs (Figure 1.1).

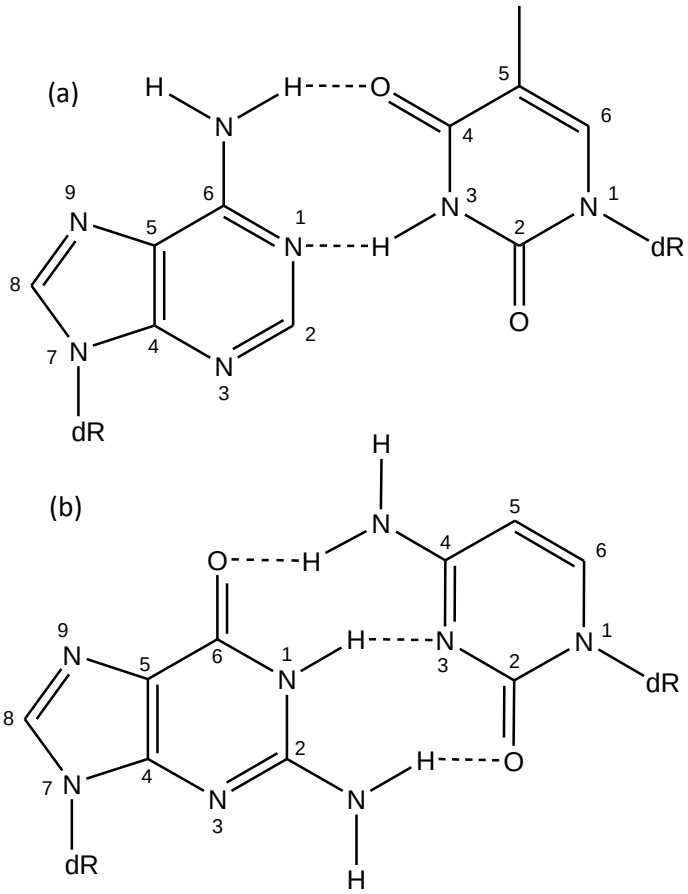


Figure 1.1 Base pairing interactions of naturally occurring DNA bases (a) Adenosine and Thymine, (b) Guanine and Cytosine. dR = 2'deoxyribose

The base is bound to the second DNA building block, a pentose sugar, by an N-Glycosidic bond to form a nucleoside. In native DNA this sugar is 2'-deoxyribose. 2'-Deoxyribose contains two hydroxyl groups, at the 3'- and 5'-position in DNA, which are conventionally used when reporting DNA sequences. According to convention, DNA sequences are reported in the direction from 5' to 3', meaning the nucleoside with a free 5'-hydroxyl is written first. 2'-Deoxyribose is distinct from the sugar of RNA, ribose, because of the absence of a 2'-hydroxyl group. The 2'-OH increases the risk of RNA degradation in acidic and basic conditions, and is an important distinction between the primary nucleic acids appearing in nature.

Each sugar is linked to phosphate groups by a phosphodiester bond, forming the sugar phosphate backbone. Phosphates are negatively charged, as seen in Figure

1.2, which influences the DNA secondary structure, by introducing a repulsive electrostatic charge between strands and nucleotides. It also plays a role in some useful experiments, such as gel electrophoresis, which separates molecules based on charge, as well as size and shape. These three components form the nucleotide, which are coupled together to form the DNA polymer.

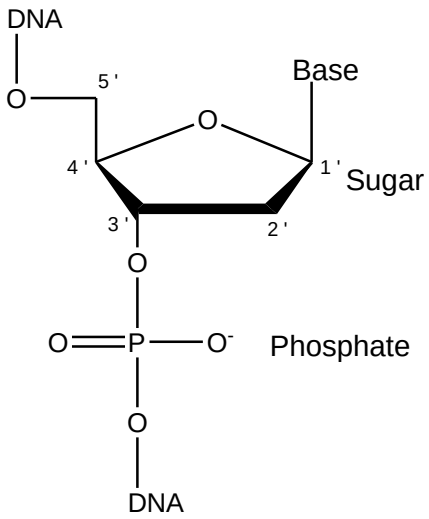


Figure 1.2 General structure of individual nucleotide, including base, sugar and phosphate.

The most common secondary structure for DNA is the double helical duplex. The stability of the duplex is derived primarily from the favourability of the hydrogen bonding interactions between base pairs. Each base pair contains several hydrogen bond donors (i.e. primary and secondary amines) and acceptors (carbonyl and imines). The high saturation obtained by the A-T and G-C base pair donors and acceptors, as shown in Figure 1.1, results in an enthalpically favourable structure, which is further influenced and stabilised by several other interactions. The duplex arrangement is favoured as it minimises hydrophobic interactions with the solvent. Hydrophobic bases are stacked towards the centre of the structure, while the negatively charged phosphate backbone interacts with polar solvents. The helix is also twisted by approximately 36° for each base pair, as this increases the distance between negative charges, while improving the strength of π-π interactions. These are a van der Waals type interaction between the π orbitals of aromatic rings. The electron rich centre and the electron poor edges overlap and attract, with the twisted structure allowing for more favourable overlap of the aromatic rings within the base pairs. Duplex formation is entropically unfavourable, as it significantly decreases the entropy of the strands, but solvent entropy favours grouping hydrophobic molecules together, which partially overcomes this. The enthalpic favourability of duplex formation due to the various hydrogen bonding, electrostatic, hydrophobic and van der Waals is sufficient to drive duplexes to form spontaneously providing that nucleobases are complementary. Temperature, pH and salt concentration changes can interfere with various driving forces of secondary structure formation, causing dissociation. While the double helical duplex is the secondary structure favoured by most DNA sequences, certain sequences can form alternate structures under the right conditions. The forces driving these structures are comparable to the forces driving duplex formation.

1.2 G-quadruplexes

G-quadruplexes (G4 DNA) form a family of alternative secondary structures for DNA. Guanosine rich sequences form a stacked arrangement of tetrads (Figure 1.3 (a)), which fold into a secondary structure comprising one to four different strands. Hoogsteen bonding, which refers to hydrogen bonds which differ from the typical bonds seen in the canonical DNA duplex, was first reported by Karst Hoogsteen[2] in 1963, 10 years after Watson and Crick first reported the DNA duplex structure. The hydrogen bonds seen in G-quadruplexes are one example of this type of hydrogen bond, although this term covers a range of non-canonical base pairing interactions.

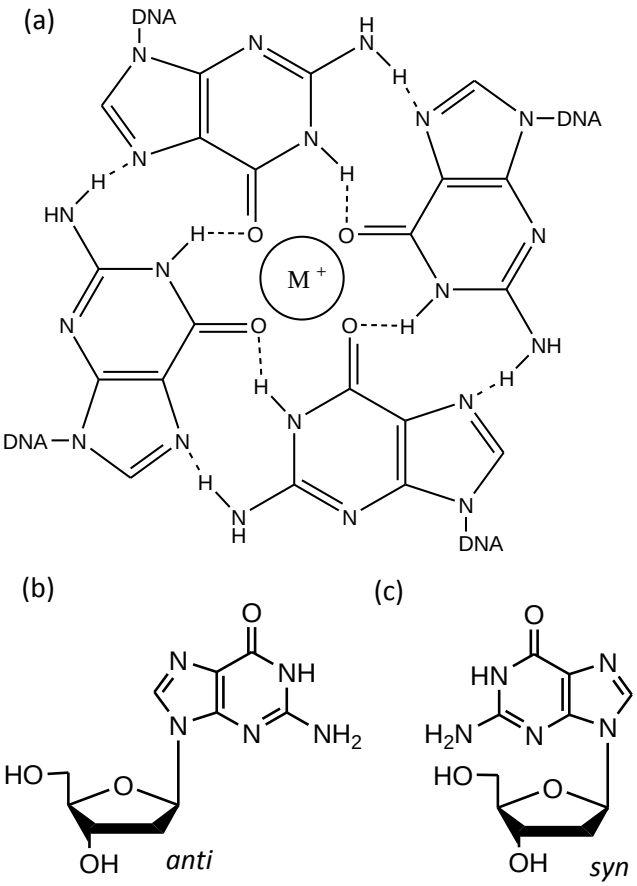


Figure 1.3 (a) Four Guanosine nucleobases arranged in a G-tetrad, indicating the Hydrogen bonding interactions formed. (b) Guanosine in anti-conformation and (c) guanosine in syn-conformation.

G-quadruplex formation is driven by similar forces to DNA duplexes. As seen in Figure 1.3 (a) the carbonyl and a tertiary amine of guanosine act as hydrogen bond acceptors to the amines. This is a distinct arrangement of hydrogen bonds to duplex base pairs, but these hydrogen bonds are also favourable, as the quadruplex structures allows for high saturation of the many amines in guanosine, with each guanosine involved in four hydrogen bonds. The negative phosphate backbones also protects the hydrophobic bases in the centre, in a similar manner to DNA duplexes. Also similar is the slight helicity of G-quadruplexes, which can be seen in the X-ray crystal structures in Figure 1.4. This helicity increases the distance between phosphate groups and strengthens π-π interactions. These

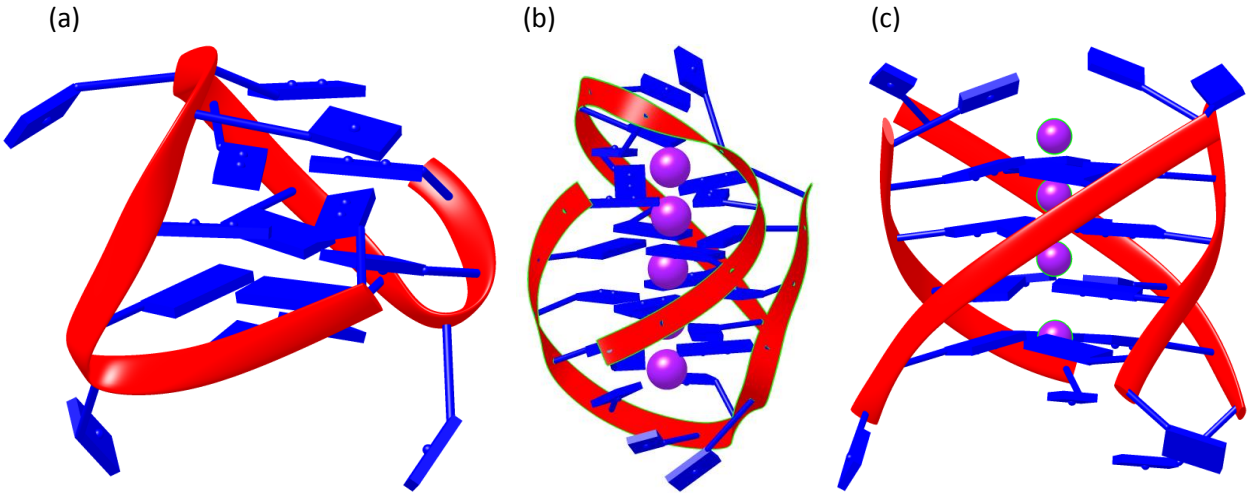


Figure 1.4 Various G4 DNA topologies: (a) unimolecular thrombin binding aptamer (GGTGGTGTGGTGG), PDB: 148D[3(a)]; (b) antiparallel bimolecular $G_4T_4G_4$, PDB: 2GWQ[3(b)]; (c) Parallel tetramolecular TG_4T, PDB: 2O4F[3(c)].

interactions are further helped by alternating syn- and anti-guanosine (Figure 1.3 (b/c) observed in many quadruplexes. Syn- and anti-guanosine are conformational isomers, a result of free rotation around the N-glycosidic bond. Syn-conformation has the base and sugar projecting in the same direction, while in anti-conformation they are opposite. The canonical Watson-Crick duplex features only the anti-conformation, but in secondary structures such as G-quadruplexes the conformation can be syn as well as anti. The conformation varies depending on the sequence and topology.

The most significant deviation from duplex DNA is the position of the carbonyl groups in guanosine. Within the tetrad they all face the centre, resulting in significant electrostatic repulsion. This is overcome by the inclusion of cations in the central cavity of the quadruplex, often K^+, Na^+ or NH_4^+ which coordinate to the guanosine O6 oxygen atoms. This means that salt is necessary to allow stable quadruplexes to form, which is not necessary for duplexes. Duplexes also form from two complementary strands, whereas G-quadruplexes can be unimolecular, bimolecular or tetramolecular. This affects the folding of G4-DNA. Tetramolecular quadruplexes, formed from sequences such as TG_4T (Figure 1.4 (c)), have significantly longer formation times than unimolecular sequences due to the increased kinetic requirement for all strands coming together to assemble. However, the driving forces for quadruplex formation are still sufficient to overcome the unfavourable entropy of secondary structure formation, with most sequences having ΔH from -15 to -25 kcal mol^{-1}, although specific conformations are found to be more favourable, with ΔH values observed up to -76 kcal mol^{-1}.[3] Sequences that form G-quadruplex can also form canonical duplexes, and some sequences have been tested under a variety of conditions to determine which secondary structure was favoured. At physiological conditions of pH, temperature and salt concentration duplex DNA was found to be favoured[3]. For example, $G_4T_4G_4$, when in the presence of its complementary sequence ($C_4A_4C_4$), formed a duplex in Na^+ buffers, but a quadruplex in K^+. Neidle and Balasubramanian (2006)[3] suggested that, while quadruplexes were marginally less stable in physiological conditions, the crowded nature of the cell made duplex formation unfavourable, resulting in the observed presence of quadruplexes in telomeric DNA. The data analysing quadruplex behaviour in genomic DNA indicates that these sequences have comparable stability to standard duplexes, suggesting that, in physiological conditions, both forms may be present.

While the G-tetrad structure described above is consistent for all quadruplex structures, the topology of these structures varies. G-quadruplex topology refers to various factors of the folding of the quadruplex, such as strand direction, loop structure and molecularity. The crystal structures[4] used in Figure 1.4 show various topologies of the different sequences. In the bimolecular $(G_4T_4G_4)_2$ (Figure 1.4 (b)) the strands are arrayed into opposite orientations to each other, indicating an antiparallel quadruplex, while in $(TG_4T)_4$ (Figure 1.4 (c)) the strands are parallel to each other. The loop structure can also vary. Figure 1.5 shows the three loop arrangements of G4-DNA. The sequences above have relatively short loops, resulting in mostly lateral and diagonal loops. Longer sequences, such as $TTG_{15}T$, displayed more flexible loop arrangements[5],

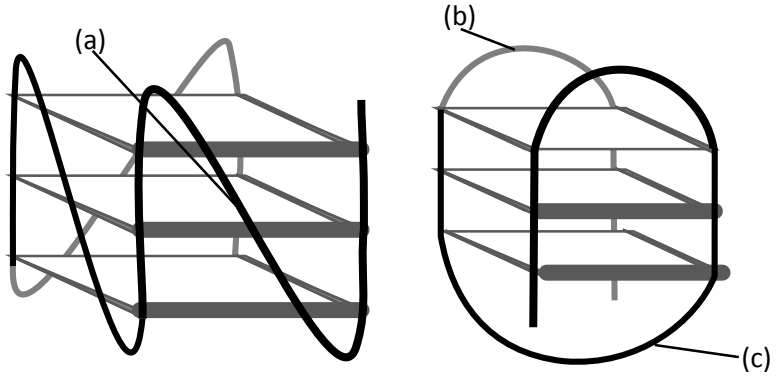


Figure 1.5 Varying topologies of G-quadruplex, showing a variety of loop arrangements. (a) propeller loop; (b) lateral loop; (c) diagonal loop.

with some loops moving diagonally from top to bottom, known as propeller loops. The increased variability of loop structures naturally allows for a greater range of topologies. However, the sequences shown in Figure 1.4 are primarily shorter, easily synthesisable sequences. One longer sequence, the human telomeric repeat ($AGGG(TTAGGG)_3$), was observed using both NMR and X-ray crystallography[4] to adopt a mixture of two distinct monomeric topologies in native conditions: parallel and anti-parallel topologies. Switching between topologies was initiated by changing pH, temperature and salt concentration, folding into the most favourable quadruplex under each condition. Manipulating this polymorphism in G-quadruplexes is interesting because different topologies can exhibit significant differences in physical and chemical properties, with different biological implications.

One enzyme which is known to interact with G-quadruplexes is Pif1 helicase, a member of the SF1B enzyme family. Helicases are a classification of enzyme which targets and unwinds DNA secondary structures, allowing for DNA to be repaired, replicated or transcribed. Therefore, understanding the mechanisms of proteins such as Pif1 is an ongoing endeavour to understand G-quadruplex behaviour in native conditions, and aid in the development of drugs for conditions related to G4-DNA. The exact mechanism for how Pif1 unwinds G4-DNA is currently unknown, but one has been proposed by Hou *et al.* (2015)[6]. They suggest that Pif1 follows a column by column approach, meaning it moves successfully down each side of the G4 structure. Helicases systematically break hydrogen bonds, destabilising the structures, and inevitably unfolding them. FRET and CD spectra of G4-DNA treated with Pif1 initially indicated the formation of a triplex intermediate, demonstrating that hydrogen bonds in one column had broken, resulting in this part of the structure unfolding. This was followed by complete disappearance of the signal, indicating complete unfolding of the secondary structure, which supported their hypothesis for this mechanism. To further understand this mechanism, and others similar to it, we develop inhibitors and structural analogues to inhibit the proteins behaviour. For example, links between columns could prevent unwinding by the helicase. Many species, including humans, use enzymes like Pif1 to unfold unwanted G4 structures. Understanding their mechanisms and the properties of the DNA they target is vital to learning how to inhibit or promote excess activity of these enzymes, which, as previously discussed, is critical to cell survival.

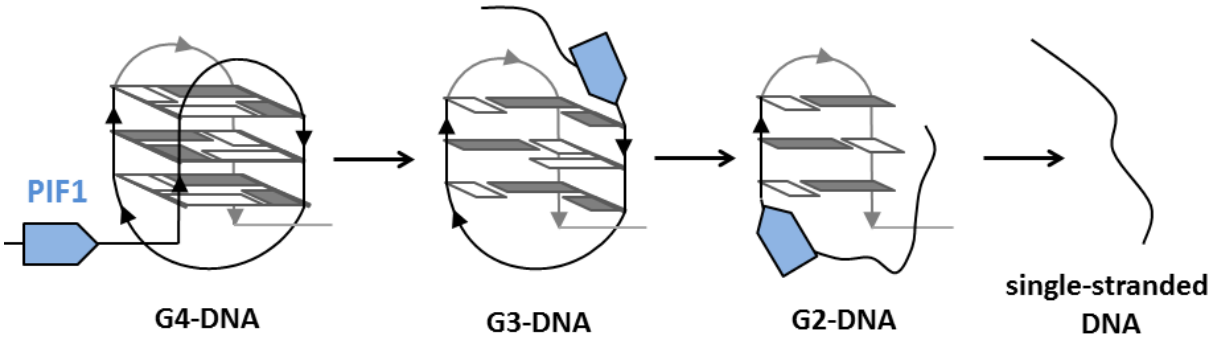


Figure 1.6 Multi-step mechanism for G4-unwinding by Pif1 proposed by Hou et al.[6] *G3-DNA is moderately stable, and was observed by FRET and CD, but G2-DNA is unstable, and FRET indicated complete dissociation to single stranded DNA.*

Genomic G-quadruplexes appear to fill several vital roles. One of their purposes is as a primary component of telomeric DNA, forming the end sequences of chromosomes. A review by Jiang *et al.* (2007)[7] gives numerous studies linking the shortening of telomeres with aging, and suggests a correlation to an increased risk of genetic defects and cancer development. Previous sources[8,9] demonstrated that overexpression of the Pif1 enzyme resulted in telomere shortening, and Pif1 inhibition resulted in telomere lengthening. Understanding the factors that lead to quadruplex stability could be critical to reducing the possibility of tumour growth. Moreover, Qi *et al.* (2006)[10] showed that introducing telomeric G-quadruplex sequences to already developed tumour cells

inhibited their growth. When Pif1 expression was decreased through gene knockdown in tumour cells an increase in the rate of apoptosis was observed.[11] High Pif1 expression decreases genome stability, increasing the potential for tumorous cells to grow, but this result suggests that Pif1 activity assists tumour cells by driving mutations that allow them to survive. By suppressing this activity, by decreasing expression or introducing inhibitors, we could reduce the viability of these cells. The addition of telomeric sequences alone was sufficient to inhibit growth, but this result only applied to the specific telomeric conformations which are targeted by enzymes. Improving the antiproliferative capabilities of G4-DNA will require control of quadruplex topology.

It was also proposed that G4 DNA was present in the promoter regions of some oncogenes.[12] An oncogene is a coding sequence in DNA whose increased expression is often linked to the development of cancer, with especially high expression often observed in tumour cells. Searching the genome database for G-rich sequences capable of quadruplex formation indicated that these occurred in the promoter regions of several oncogenes. This is not a definitive correlation, but could mean that targeting these sequences or the enzymes that interact with them could be used to inhibit their expression and delay or prevent the development of cancer cells. Telomeric and promoter regions are both vital for DNA functionality in organisms ranging from humans to bacteria and viruses. Potential G4 sequences have also been found in various viruses, including HIV-1 and Epstein-Barr Virus[13] and Zika[14], suggesting that drugs targeting G4 DNA could treat various viral and bacterial conditions. The extensive presence of G-quadruplexes in different species and regions of DNA means the potential applications for research into their behaviour is also extensive.

To accomplish this, we would like to develop G4 structures that can retain their structure under conditions which would typically cause dissociation of G-quadruplexes. This includes the presence of enzyme, but could also apply to unfavourable temperatures, salts and pH. To achieve this, modifications of G4-DNA are necessary which provide stability under these conditions.

1.3 Aim

To develop procedures for the incorporation of covalent cross-link forming chemical modifications into G-quadruplex sequences, which are resistant to the conditions used to typically dissociate G4 structure (e.g. heating, pH change, enzyme treatment) and analyse the effect these modifications have on biophysical properties.

1.4 Hypothesis

(i) Precisely placed cross-links can support the G-quadruplex structure, increasing thermal stability by maintaining the structure when hydrogen bonds are weakened.

(ii) Cross-links will retain a foundation for the secondary structure reducing polymorphism and allowing structures to form more quickly. This will increase the rate of formation, and potentially reduce the order of reaction for formation of multi stranded quadruplexes.

1.5 Methodology

The strategy we propose to increase G4-stability is covalent chemical cross-linking. This strategy requires the synthesis of DNA sequences containing functional groups which can be linked to each other. There is a broad range of possible functional groups this could apply to, from disulfide bridges (Figure 1.7 (a)), which are used by nature to stabilise proteins, to ligands coordinating a metal (Figure 1.7 (b)). This means the biggest factor in determining how this strategy is pursued is which modifications can realistically be made to G-quadruplex forming sequences. The arrangement of the G-tetrad (Figure 1.3 (a)) limits the positions in guanosine which can conceivably be modified to provide the functionality required. This is because any structural elements which are already fundamental to quadruplex integrity are not viable modification targets. This gives several potential modification targets:

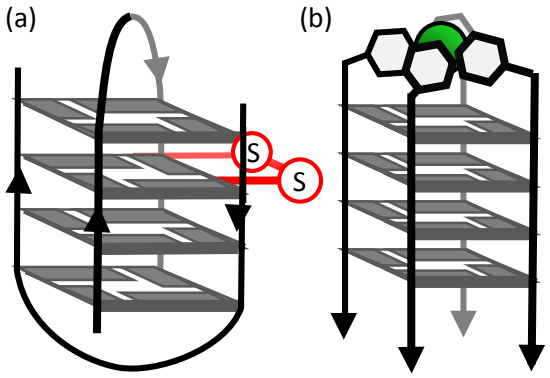


Figure 1.7 Examples of quadruplex structures for (a) covalently cross-linked sequences connecting two strands (cross-link shown in red) and (b) chelating cross-links coordinated to a metal centre.

(i) phosphate groups;
(ii) ribose sugar;
(iii) non-hydrogen bonding positions of guanosine (Figure 1.8);
(iv) alternative nucleotides that can be arranged similarly to a G-tetrad.

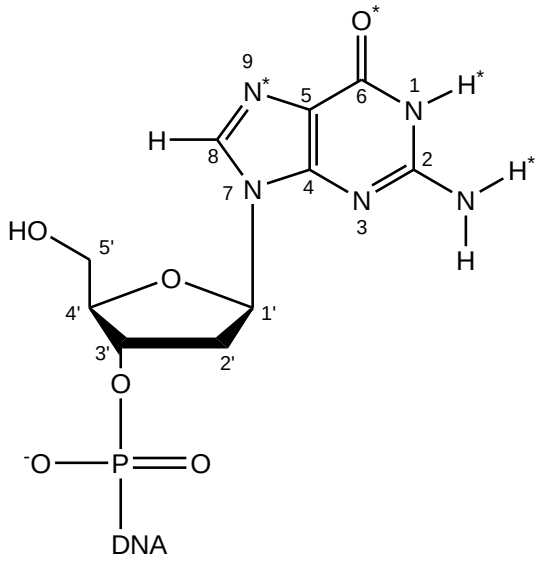


Figure 1.8 Labelled guanosine to indicate potential targets for modification. Sites labelled with (*) are involved in hydrogen bonding within G-tetrads.

Phosphate modification (i) would involve the addition of functional groups to the oxygen atoms in the phosphate backbone. This strategy is problematic because any modification here will result in diastereomers of different ratios, meaning some of the resulting material may be unusable. While this is a potentially viable strategy, there are more appealing options which do not necessarily introduce diastereoisomers.

Modification of the sugar (ii) is interesting because it should not significantly disrupt the G4-structure. However, the 3'- and 5'- positions are not available unless at the termini of the DNA sequence. The 1'- and 4'-positions are also difficult to functionalise. The 2'-position is the most commonly modified, since RNA already contains a hydroxyl-group at this position, meaning 2'-O-R modification are not too problematic. We have briefly considered some possibilities, such as 2'-O-propargyl guanosine (Figure 1.10 (a)), for this purpose, but there has been no significant exploration of this strategy.

Non-hydrogen bonding guanosine positions (iii) also imposes several limitations. Each residue takes part in four hydrogen bonds, requiring 3 nitrogens and 1 carbonyl. Several other positions are quaternary carbons in the aromatic rings, making these unappealing targets for modification. However, there are two positions, C-8 and N-2, which present possible sites for modification. Both

positions are not involved in hydrogen bonding and literature examples exist for derivatisation of these positions.

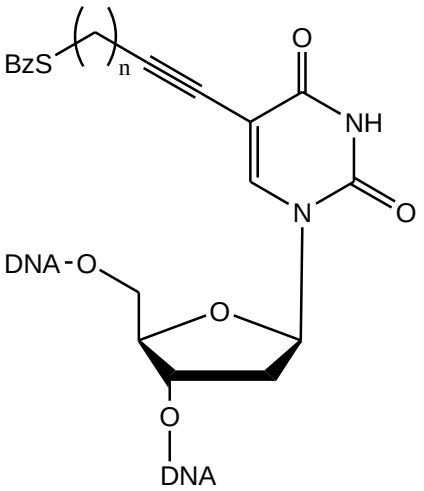


Figure 1.9 Modified nucleotide used by Prestinari and Richert (2011)[17] to introduce disulfide bridges in duplexes.

The first is the CH in the 8-position of guanosine. 8-bromo-2'-deoxyguanosine (Figure 1.10 (b)) is both a commercially available and synthesisable reagent which substitutes the 8-position proton for a halide, adding a range of options for functionalisation. The most commonly reported is Sonogashira coupling, a substitution reaction between a terminal alkyne and an aromatic halide. This reaction has been previously reported for guanosine as a strategy for introducing fluorescent labels to dG.[15,16] Furthermore, Prestinari and Richert (2011)[17] used Sonogashira coupling to modify 5-I-dU (Figure 1.10 (c)) to incorporate sulfur linkers to DNA duplexes (Figure 1.9). These linkers formed disulfide bridges in duplex DNA, resulting in increased thermal stability. We propose that a similar procedure could be used to modify guanosine, introducing disulfide linkages and giving the same increased thermal stability for quadruplexes. If successful, this strategy could extend to other functional groups which can be incorporated alongside an alkyne, providing several potential 8-position functionalities for different types of cross-links.

The second potential guanosine modification is at the N2 position. While this primary amine is involved in hydrogen bonding, only a single hydrogen atom is required for this. This means that it could be converted to a secondary amine with no loss of functionality. ^{F}dI (Figure 1.10 (d)) is a modified phosphoramidite similar to guanosine, with the 2-position substituted with fluorine. Protocol has been reported[18] for the substitution of this position with various amines, resulting in a secondary amine

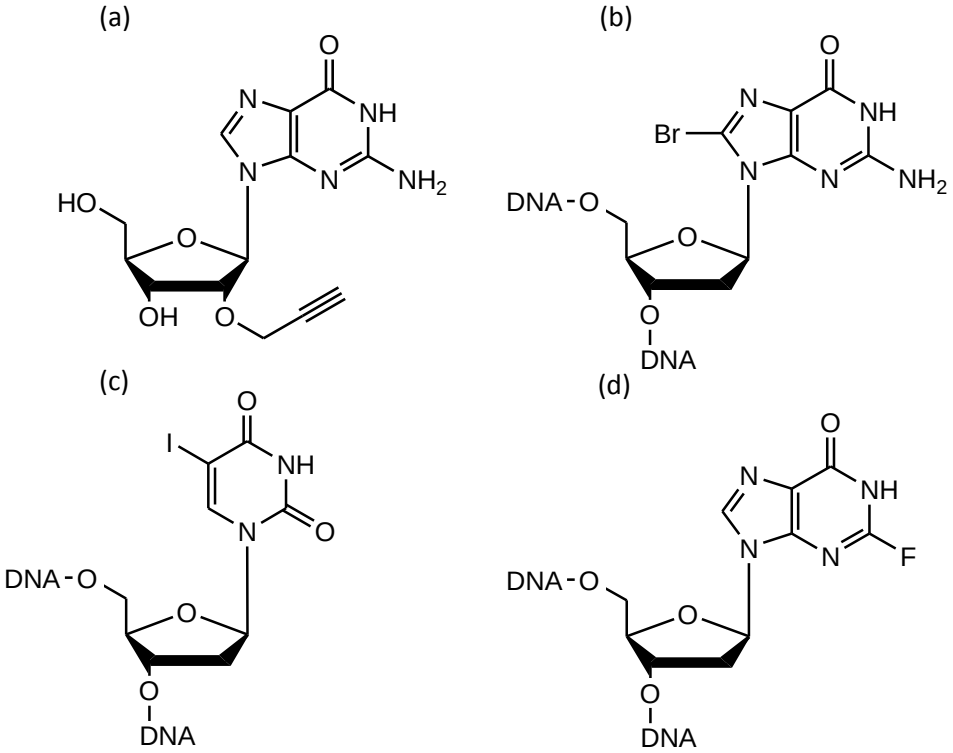


Figure 1.10 Examples of modified components for DNA synthesis used for chemical modification. (a) 2'-O-propargyl ribose; (b) 8-Bromo guanine; (c) 5-iodo uracil; (d) 2-F-Inosine;

with the desired functionality. Once again, this has primarily been used as a method of introducing fluorescent labels, but we believe it can also potentially be used to introduce functionality to form chemical cross-links. With this objective we can explore the synthesis of several amino/thiol compounds and their incorporation into DNA. As with 8-position modification, this strategy can potentially introduce a range of functional groups if our initial experiments are successful.

Finally, we consider something outside of the canonical DNA nucleotide (iv). Engelhard *et* al.[19-21] reported several modifications for the TG_4T sequence to introduce pyridine ligands in place of 5'-thymidine. These sequences formed complexes with Cu(II) and Ni(II), similar to the structures shown in Figure 1.7 (b). While they found that modified sequences were less stable than native DNA, they observed a significant increase in thermal stability upon addition of divalent cations that could coordinate to pyridine. Engelhard *et* al.[19-21] mainly suggests this strategy could be used to develop switches to carefully control G-quadruplex formation. We believe this could be expanded to a cross-linking strategy for enzyme inhibition if the stability of complexes can be sufficiently increased. We propose that the significant destabilisation of their modifications was due to the ligands deviating from the shape of a traditional nucleotide, distorting the G-tetrads and weakening hydrogen bonds. To overcome this problem, we intend to incorporate the 4-pyridine ligand into pyrrolidine, which will more closely mimic native DNA (Figure 1.11) (a). Pyrrolidine-CE phosphoramidite (Figure 1.11 (b)) is a commercially available reagent that can be converted to the secondary amine, and further functionalised using reductive amination. In this way, pyridine ligands can be introduced into DNA. We believe this will result in similar properties to the modified sequences reported by Engelhard *et al.*[19-21], but a closer approximation to a G-tetrad shape may result in better thermal stability and faster association.

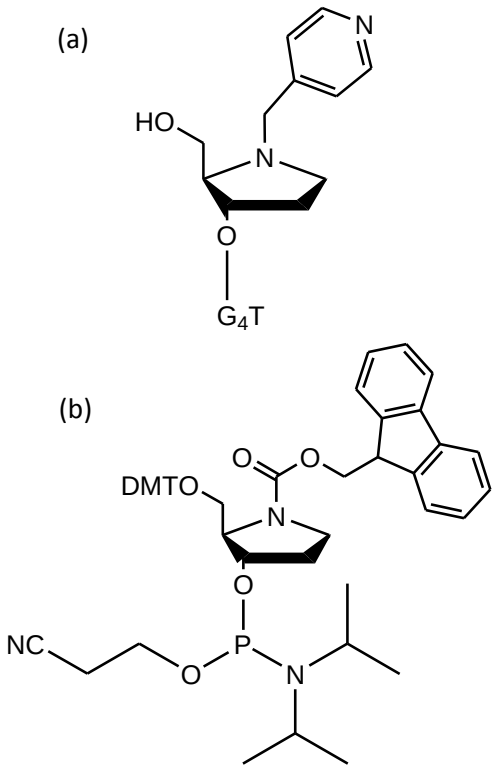


Figure 1.11 (a) Pyrrol based pyridine modification for TG_4T, (b) Fmoc- Pyrrolidine-CE phosphoramidite, a commercially available reagent.

Overall, there is a huge range of possible modifications that can be made to DNA sequences to potentially create chemical cross-links. Both the covalent and metal-ligand interactions should be stable in conditions where hydrogen bonds are broken, providing the desired additional stability. We intend to start with several basic strategies for G-quadruplex modification using these methodologies to develop a basis for more varied modifications.

2 General Methods for DNA Synthesis and Biophysical Analysis of G-quadruplexes

2.1. DNA Synthesis

DNA synthesis was performed with a Mermade 4 DNA/RNA automated synthesiser. DNA synthesis is composed of four steps, which are repeated for each nucleobase, gradually forming a specified sequence of DNA. The repetitive nature of this process makes it ideal for automation. The conditions for each step are carefully designed to maximise reaction efficiency and minimise the formation of side products. The first nucleotide is usually preloaded on a controlled pore glass (CPG), a porous silica surface, by easily cleaved moieties, and the incomplete sequences remain on the support between steps. Synthesis proceeds from 3' to 5' direction, and a universal support can be used for modification of the 3' nucleotide if necessary. This procedure allows for highly efficient coupling, and easy removal of excess reactants.

The efficiency of each reaction step is the most important factor because every additional nucleotide requires four reactions, meaning that even short sequences can require reaction pathways with dozens of steps. Lost yields over many consecutive steps results in a low final yield, so each step

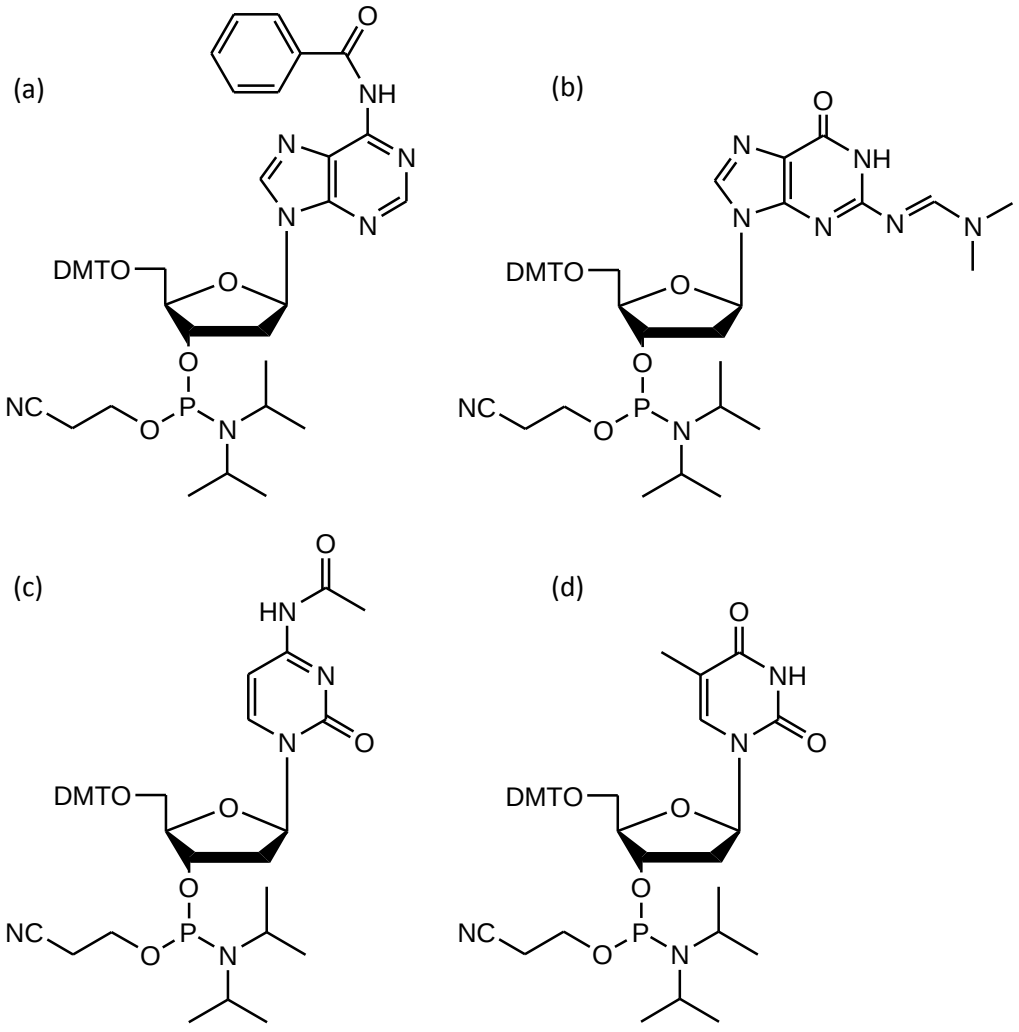


Figure 2.1 Standard phosphoramidites used for DNA synthesis: (a) 5'-DMT-3'-(2-cyanoethyl)-N,N-diisopropylamine-N6-benzoyl-2'-deoxyadenosine; (b) 5'-DMT-3'-(2-cyanoethyl)-N,N-diisopropylamine-N2-DMF-2'-deoxyguanosine; (c) 5'-DMT-3'-(2-cyanoethyl)-N,N-diisopropylamine-N4-acetal-2'-deoxycytidine; (d) 5'-DMT-3'-(2-cyanoethyl)-N,N-diisopropylamine-2'-deoxythymidine

is designed to achieve high coupling efficiency. This means that the reactants selected are highly reactive and concentrated solutions.

The nucleotides used for synthesis are typically protected by 5'-dimethoxytrityl (DMT) and 3'-β-cyanodiisopropyl phosphoramidites. Primary amines are protected by acetal, dimethylformamide dimethylacetal, or benzoyl protecting groups for cytosine, guanine and adenine respectively. These protecting groups are chosen for their stability, ease of installation and efficient removal after synthesis. DMT cleavage has the additional advantage of producing a bright orange colour, which provides an easily observable sign of success or failure at each addition.

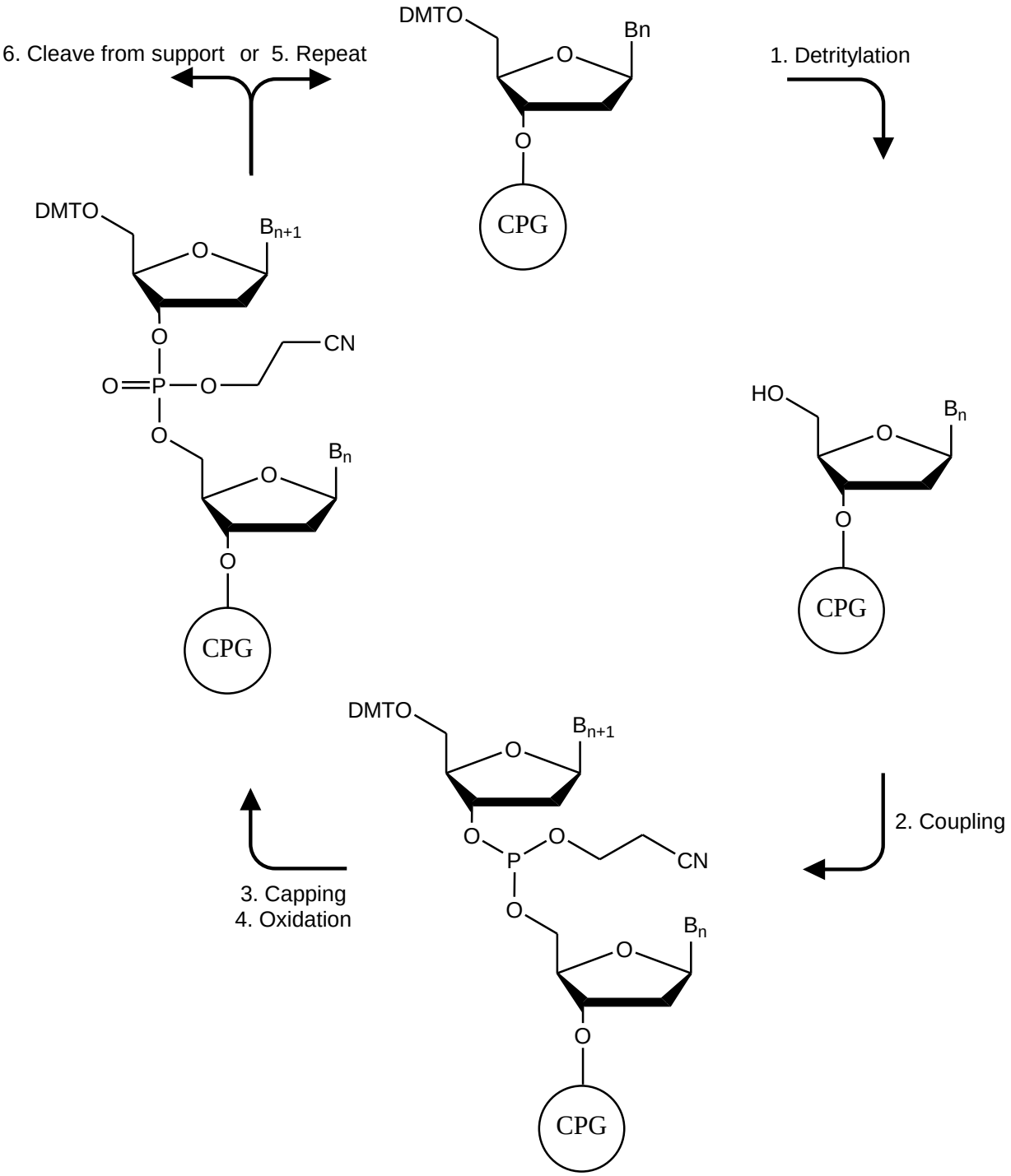


Figure 2.2 Procedure for automated DNA synthesis

Explanation of steps in DNA synthesis (Figure 2.2):

1. Detritylation: 5'-DMT is cleaved with a solution of trichloracetic acid (3% in DCM), which deprotects only the 5'-OH for the subsequent coupling step. Cleavage of DMT results in a distinct orange colour, which can indicate the success of the previous coupling.
2. Coupling: An activator, such as 1*H*-tetrazole or 4,5-dicyanoimidazole, is used to initiate a reaction of the 3'-phosphoramidite of the desired nucleotide (400 µL, 0.1 $molL^{-1}$ in freshly dried acetonitrile) with the deprotected 5'-hydroxyl of the nucleotide bound to the support. In our system 5-ethylthio-1H-tetrazole (ETT) (0.45 $molL^{-1}$ in acetonitrile) was used. This step is highly water sensitive due to the presence of the highly reactive phosphoramidite, so freshly dried acetonitrile is used as the solvent for the phosphoramidites and activator.
3. Capping: Unreacted 5'-hydroxyl groups may react in future coupling steps, resulting in undesirable sequences. To prevent this an excess of acetic anhydride (15% v/v in THF) reacts with any free 5'-hydroxyl groups, preventing further reactions at this location. This reaction is catalysed by the addition of 1-methylimidazole (16% v/v in THF).
4. Oxidation: Finally, the unstable trivalent phosphoramidite is converted to a pentavalent phosphate by treatment with a solution of I_2 in THF:pyridine:H_2O (0.02 $molL^{-1}$, Solvent: 88:10:2).
5. Repeat: Each additional nucleotide contains a new 5'-DMT group, which can be deprotected and the cycle repeats. This process can theoretically be repeated any number of times to produce oligonucleotides of any length, but in reality the efficiency of each step is less than 100%, which limits the maximum sequence length. Careful use of this procedure allows for synthesis of up to 200 nucleotide long sequences, but for optimal yields, sequences should be shorter than this.
6. Cleavage from solid support: Synthesis can be completed with DMT-on or DMT-off depending on the intended post-synthetic treatment of the support. After the synthesis is complete, DNA is cleaved from the CPG support and the protecting groups on the phosphates and nucleobases are removed by treatment with 28% aq. ammonia overnight at room temperature. This process can also be carried out by heating at 55°C for several hours, but must be left at R.T. for approximately 30 min to allow for phosphate protecting groups to be removed to prevent complete cleavage of the phosphate backbone. Aq. ammonia was evaporated by centrifugation under reduced pressure with a speed-vac (Eppendorf Concentrator Plus). In specified cases, DNA was instead precipitated using 0.3 $molL^{-1}$ $LiClO_4$ in acetone or 3 $molL^{-1}$ NaOAc in ethanol to allow for additional treatments. DNA was dissolved in water, and purification was carried out with reverse phase HPLC.

The process for addition of standard nucleotides is automated, but modified nucleotides require slight modifications to the procedure. Addition can be automated by dissolving phosphoramidites in dry acetonitrile, but the DNA synthesiser requires solution to wash and fill tubes prior to addition. This means automated synthesis requires a small excess of material which is lost. For many modified nucleotides small quantities are expensive or time consuming to produce, so the material used should be minimised. Instead, addition is carried out semi-manually. The majority of the process is automated, with the exception of the coupling step. Two separate commands are used to control the addition of the modified nucleotide, first for deblocking and a second for coupling. After deblocking occurs, synthesis is paused and solid phosphoramidite is weighed and added directly to the column. The procedure is resumed, adding activator and solvent to dissolve the solid. This gives a similar outcome to automated coupling, but minimises the loss of material and allows the coupling to be more closely controlled. This approach is used to produce sequences with modifications at specific locations.

2.2 Reverse Phase High Performance Liquid Chromatography (RP-HPLC)

Reverse phase HPLC was performed with a Thermo Scientific UltiMate 3000 UHPLC with an Alltech 250 mm x 4.6 mm, 10 µm Hypersil Gold column. This technique is used to separate and analyse DNA. This procedure uses the relative polarities of components in two phases (mobile and stationary) to separate each component based on retention time. In reverse phase HPLC a non-polar stationary phase is used. Compounds with low polarity have the strongest interaction with the non-polar stationary phase, while more polar materials have little interaction and travel through the column quickly.

DNA purification and analysis was carried out with 0.1 $molL^{-1}$ TEAA buffer and acetonitrile. In the first two minutes the column is equilibrated with 100% buffer, after which acetonitrile is gradually increased to 25% over 18 minutes. The column is washed by increasing acetonitrile to 80% over one min, washing for four minutes and returning to 100% buffer over the final min. Typically retention times for small DNA sequences are short, with 25% acetonitrile sufficient to isolate material. However, G-quadruplexes have longer retention times and were sometimes collected during the washing step. Fractions can be automatically collected using time or UV-Vis signal specifications. At specified times, or when the absorbance signal is above a specified threshold, the material is collected and the solvent can be evaporated. This is the primary technique used to purify DNA. Peaks with absorbance at 260 nm were collected, desalted using a NAP-5 size exclusion column and analysed by mass spectrometry, as described below. RP-HPLC can also be used analytically in combination with other techniques, by observing differences in retention times for each sequence.

2.3 Size Exclusion Chromatography

(i) Desalting with size exclusion chromatography:

Pre-packed NAP-5 columns were used to remove salts and collect DNA after purification on RP-HPLC. NAP-5 columns have a maximum sample volume of 0.5 mL. The column is packed with Sephadex G-25 DNA grade gel. Porous gels are used in SE-HPLC to trap small molecules and allow large molecules to pass through more quickly. This is useful for desalting as DNA passes through the gel quickly relative to the salts. RP-HPLC used TEAA buffer as a solvent, which needed to be removed before mass spectrometry and biophysical analysis. Each column was equilibrated by washing three times with the desired buffer of the final solution. Then, specific volumes of the sample, followed by the buffer were loaded onto the column. More buffer was added and a specific volume was collected. The volumes used were dependent on the initial volume of the sample (Table 2.1). When preparing for mass spectrometry the buffer used was typically water.

(ii) Analytical Size Exclusion HPLC

Size exclusion High Performance Liquid Chromatography (SE-HPLC) used a Thermo Scientific Acclaim™ SEC-300 5 µm, 300 Å, 4.6x300 mm column. This technique was used to identify the

Table 2.1 Nap-5 column protocols for different sample volumes

Sample Volume (mL)	Buffer Volume (mL)	Collection Volume (mL)
0.10	0.40	0.5
0.25	0.25	0.7
0.50	0.00	1.0

molecularity of DNA and distinguish single stranded DNA from dimeric/tetrameric quadruplexes. Molecules appear from largest to smallest, meaning quadruplexes give a shorter retention time than monomeric DNA, as shown in Figure 2.3. Comparing retention times of modified samples to known single stranded and quadruplex DNA shows a trend that represents the size of the molecules at various retention times.

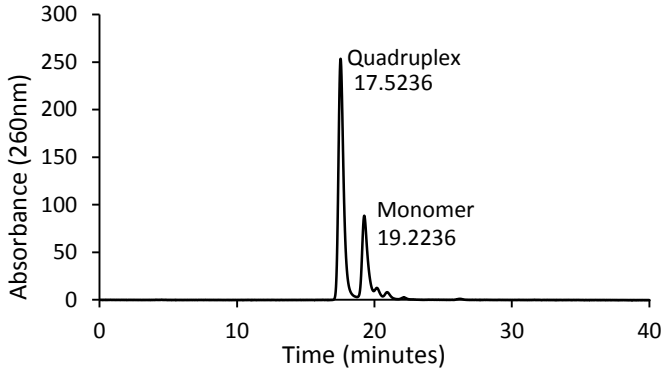


Figure 2.3 Example of size exclusion chromatogram of folded TG_4T.

Equation 2.1[22] gives the relationship between retention time and molecular weight. Structure Index (SI) is an arbitrary value. V_e is the elution volume, essentially the retention time of the peak, while V_0 is the dead volume, the retention time of an arbitrarily large molecule, in this case dextran blue (M_w = ~$2x10^6$ Da), which is used as a standard when comparing samples. The calibration curve for single stranded DNA used a range of sequences (T_5, T_{10}, T_{15}, T_{20}, T_{25}, T_{30}, T_{35}, T_{40}, T_{50}). An additional calibration was obtained for tetramolecular quadruplex forming sequences (TG_6T, TG_5T, TG_4T, G_4T), which gave a standard curve for quadruplex structures. All samples were in sodium cacodylate buffer (10 mM sodium cacodylate, 100 mM NaCl, pH 7.3), which was also used to equilibrate the column. These calibration curves (Figure 2.4) are compared with samples of modified DNA to confirm the formation of G-quadruplexes, and their molecularity.

$$SI = (V_e / V_0) \cdot \log_{10} (M_w) \qquad \text{(Equation 2.1)}^{22}$$

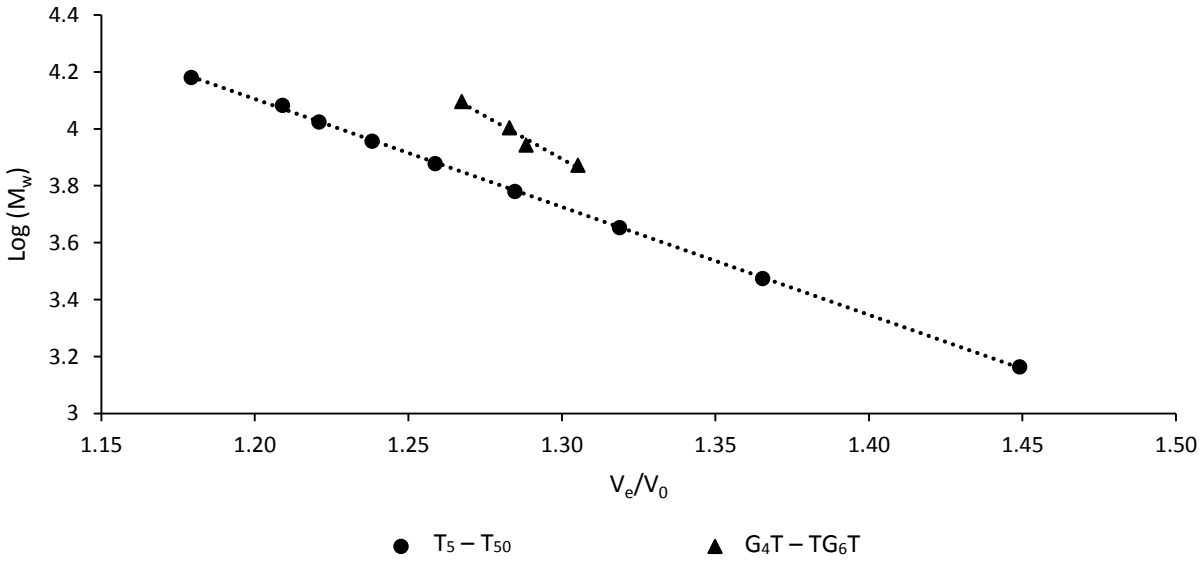


Figure 2.4 Example of SE-HPLC ladders, plotted using Equation 2.1. Conditions: pH 7.3, 10 mM Sodium Cacodylate Buffer, 100 mM NaCl.

2.4 Circular Dichroism Spectroscopy

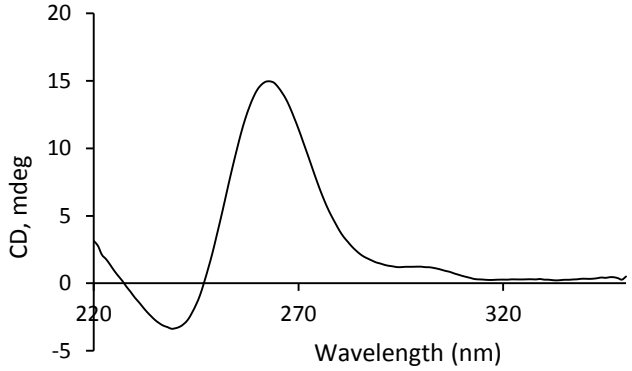


Figure 2.5 Characteristic spectra of TG_4T tetramolecular parallel quadruplex, showing peaks at approximately +265 nm and -240 nm.

Circular Dichroism (CD) was performed with a Chirascan CD spectrophotometer (150 W Xe arc) from Applied Photophysics with a Quantum Northwest TC125 temperature controller. Spectra were recorded between 220 and 350 nm, as this is the optimal range to observe the key peaks shown in most DNA secondary structures. CD is a useful technique for identifying the conformation of DNA structures, as well as the degree to which these structures are formed. The technique takes advantage of the unique interactions of chiral molecules with circularly polarised light. Similar to other light based analytical techniques, light of various wavelengths is shone through a cuvette containing the analyte. CD uses circularly polarised light and a curve is drawn to represent the difference in absorbance of left and right circularly polarised light. DNA secondary structures give a curve which is characteristic of a particular topology (Table 2.2). This is a result of the configuration of the tetrads. The nucleobases attached to the chiral sugar interact with circularly polarised light. Parallel and anti-parallel topologies have different arrangements of syn- and anti-guanosines, resulting in two distinct CD spectra.[23] This can be illustrated by the almost inverted CD spectra of B-(right handed) and Z-DNA (left handed), which contain entirely anti- and syn-configuration guanosine, respectively (Table 2.2[24]).

CD is used for several different purposes when analysing DNA. First, CD is used in combination with SE-HPLC to confirm quadruplex formation. This is shown by the appearance of characteristic peaks corresponding to the expected structure, such as in Figure 2.5. This spectrum gives no information about the sequences making up the quadruplex, as this spectrum is very similar to any other tetramolecular parallel quadruplex, but it does indicate that a parallel quadruplex structure has been formed. More specific information about the complex formed is given by other techniques such as mass spectrometry, NMR and others.

Table 2.2. Characteristic CD Peaks of DNA Secondary Structures

Topology	Characteristic peaks (nm)
Parallel G-quadruplex	+260, -240
Antiparallel G-quadruplex	+290, -260
B-DNA	+260 – 280, -245[(a)]
Z-DNA	-290, +260, -205
Other non-quadruplex structures	Various[(b)]

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. (b) Structures such as i-motifs and triplexes can also be characterised using CD, but their characteristic peaks vary significantly depending on sequence.

The second use of CD is in melting experiments to determine thermal stability of the G4 structure. Samples are heated from 5°C to 90°C over several hours, and then cooled to 5°C, giving a melting profile as shown in Figure 2.6 (a). The maximum intensity is used as a reference and signal decay at this wavelength is recorded. These values are compared to the maximum and minimum intensities using Equation 2.2.[25] α is the fraction of folded DNA, while θ_n is the intensity at 262 nm at a given temperature. θ_{max} is the maximum intensity, which was recorded prior to melting, and θ_{min} is the minimum intensity. The value used for comparison of melting profiles is $T_{½}$, which is the point at which α = 0.5 (Figure 2.6 (b)). This profile also shows the fluctuations in intensity, which introduces error when determining the $T_{½}$ value. When obtaining $T_{½}$ the values either side of the curve are used as the maxima and minima, giving an adjusted profile showing only the variation caused by melting. Similar

$$\alpha = (\theta_n - \theta_{max})/(\theta_{min} - \theta_{max}) \quad \text{(Equation 2.2)}$$

$$y = \theta_{max} + (\theta_{min} - \theta_{max}) \cdot (1 + (n - 1) \cdot k_{on} \cdot C^{n-1} \cdot t)^{1/1-n} \quad \text{(Equation 2.3)}$$

$$t_x = \frac{(1 - x)^{1-n} - 1}{C_o^{n-1} \cdot n - 1 \cdot k_{on}} \quad \text{(Equation 2.4)}$$

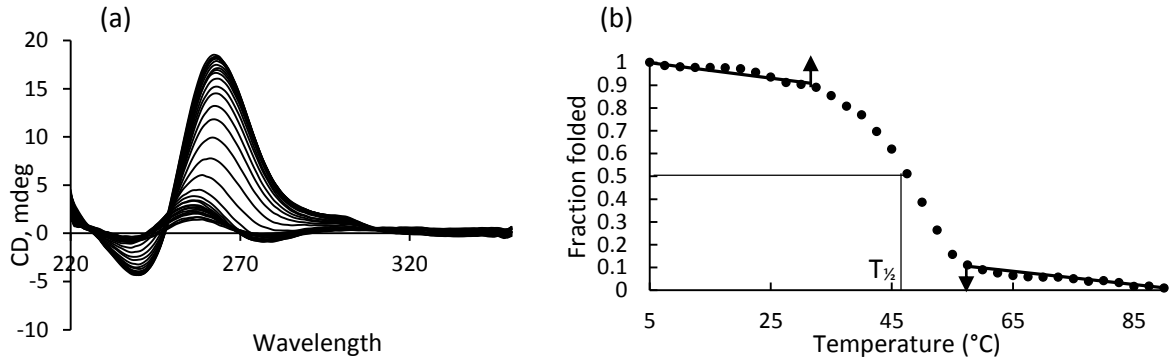


Figure 2.6 (a) Melting profile of tetramolecular parallel quadruplex; (b) melting profile showing correction using Equation 2.2.

melting profiles can also be obtained using UV-Vis spectrometry, as described below, but we found CD gave more consistent results.

The final use of CD is in isothermal kinetics experiments to examine the rate of formation of G4 structures. Samples were heated to 95°C until the spectrum indicated the structure was fully dissociated. The CD signal at 262 nm was recorded at 5°C for 10 hours, obtaining 10,000 points. These were adjusted using Equation 2.2 to show the fraction folded over time. The result was curve fitted using Origin to Equation 2.3[25], where C is concentration and t is time. This gave the rate of formation, k_{on}, and the order of the reaction, n. This equation can adjusted to give Equation 2.4[25], which gives the time required for 50% and 90% G4-formation at specific DNA concentration, giving t_{50} and t_{90} values, respectively. These values could be compared between samples with varying conditions to describe the different kinetic properties of each structure.

2.5 UV-Vis Spectroscopy

UV-Vis Spectroscopy was performed with a Cary 100 Bio UV-Vis Spectrophotometer. Light is passed through a cuvette and the absorbance is recorded. Solution concentration can be calculated using Beer's Law. DNA has an absorbance maxima at 260 nm, with ε dependent on the length and composition of the oligonucleotide, which was used to determine DNA concentration and quantity after synthesis. UV-Vis can also be used for melting experiments, with similar conditions as described

for CD spectroscopy. Samples were heated from 5°C to 90°C at a rate of 1°C min^{-1}, with several wavelengths recorded. The wavelengths were selected using a thermal difference spectrum, which uses spectra from 5°C and 90°C to show changes in absorbance, with the largest shifts being selected for thermal denaturation of each sequence.

2.6 Nuclear Magnetic Resonance Spectroscopy (NMR)

All ^{1}H and ^{13}C NMR were performed with a 500 MHz Bruker spectrometer using tetramethylsilane (TMS) as an internal standard. NMR is one of the primary techniques used to characterise the structure of small molecules. NMR observes the spin relaxation of protons and neutrons (nucleons) when they are excited by a magnetic field. Atoms will only be observable with this technique if they have a net spin greater than zero, which only occurs in the case of an odd number of nucleons. The nucleus is shielded from the magnetic field by electron density, which varies based on the presence of electron withdrawing molecules. This can be used to characterise molecules by identifying atoms based on their chemical shift as a product of proximity to electron withdrawing groups. More complex 2D NMR experiments show correlations between signals, indicating when atoms have bonds between them, or are close together in space. By combining several NMR experiments it is possible to characterise the full structure of a molecule. This technique is rarely used for large molecules because the number of overlapping signals makes it difficult to distinguish individual nuclei. When characterising these structures it is usually preferable to use techniques such as HPLC and mass spectrometry.

However, in some cases, G4-structures have been analysed with NMR,[26] for example to detect the presence of thymine or uracil by identifying specific methyl peaks. NMR can also be used to identify complex topologies by systematically labelling specific positions with ^{2}H, ^{13}C and ^{15}N, and identifying their location within the secondary structure. Comparing data from techniques such as NOESY, which identifies the distance between two atoms in space, to known distances for anti- and syn-guanosine can be used to identify the positions of each guanosine configuration, and therefore the overall structure of the complex. However, we mostly use this technique to identify small molecules, as described in Chapter 7.

2.7 Mass Spectrometry

Mass spectrometry is the primary technique for identifying the composition of molecules. It was performed with a Thermo Scientific Q Exactive Focus Hybrid Quadrupole-Orbitrap Mass Spectrometer. A variety of methods can be used to convert the analyte into ions. In this spectrometer, electrospray ionisation (ESI) is used, applying a high voltage to the analyte to produce an ionic aerosol. This is then injected into a vacuum containing a probe electrode. In the case of Orbitrap mass spectrometry ions are trapped and oscillate around the electrode, generating a current relative to the mass to charge ratio of their components. These ratios are used to identify molecules based on their mass. Since the mass of the molecules is all that matters mass spectrometry can be used to characterise large molecules such as oligonucleotides and determine molecularity of their complexes.

DNA samples are typically run with 15-20% methanol or acetonitrile to improve ionisation. Analysis of G-quadruplex structures often requires native conditions to ensure the complexes retain their structure throughout the experiment. To accomplish this, 150 mM NH_4OAc can also be added, providing NH_4^+ cations to stabilise G4-structures. However, the structure is often dissociated during ionisation, making mass spectrometry a less reliable technique for this than SE-HPLC, CD or native Gels. Many samples will also form salts with sodium and potassium when ionised, which is shown by an increase in mass corresponding to these cations.

2.8 Gel Electrophoresis

Gel Electrophoresis applies a voltage across a gel, separating the molecules passing through it based on their charge, size and shape. The phosphate groups of DNA are negatively charged, meaning this technique is very effective at separating DNA based on sequence length. This is useful for visualising the comparison between multiple DNA sequences to determine their length.

Denaturing gels were formed using 20% acrylamide solution, with 7M urea in pH 8.0 TBE buffer. N,N,N',N'-tetramethylethyldiamine (12 µL) and 10% w/v aq. ammonium persulfate (120 µL) are added to induce cross-linking. Our samples used a T_5-T_{50} oligonucleotide ladder to establish the size of the synthesised sequences. Denaturing gels contain urea, which dissociates the DNA secondary structure, meaning complexes will not be observed. This is useful for determining if two sequences have been linked together, as they will not be dissociated in urea.

Native gel can also be prepared to observe DNA complexes with a similar technique. Similar to native mass spectrometry described above, this technique uses conditions where complexes retain their structure, and is therefore useful for observing their molecularity. Gels contained using 20% acrylamide solution in pH 8.0 TBE buffer, with NaCl added (concentration can vary, but typically 100 mM). Native gels are otherwise formed using the same protocol as denaturing gels. The absence of urea, and inclusion of Na^+ allows G-quadruplexes to retain their structures.

Functionalisation of 2’-Deoxyguanosine: Preface

The G-tetrad forms through hydrogen bonding of four guanosine nucleobases, and most functional groups of the guanosine base have critical hydrogen bonding interactions for quadruplex stability (Figure 1.3). Therefore, functionalising these atoms for introduction of cross-links is not feasible without fundamentally altering the quadruplex formation. There are two exceptions: The 8-position proton and one proton of the primary amine at the 2-position, neither of which are necessary for hydrogen bonding. This makes these two sites the primary consideration for guanine modification when developing chemical cross-links. The following chapters cover these two approaches, predominantly focusing on disulfide bridge formation. The first, Sonogashira coupling, will attempt to modify the 8-position of dG through substitution of 8-bromo-2’-deoxyguanosine with various functionalised alkynes. The second, amine substitution, will attempt to modify the 2-position of dG through substitution of 2-fluoroinosine with various functionalised amines. The two approaches are similar in the goal they aim to accomplish, functionalising guanosine, but differ in the position and strategy for accomplishing this goal.

3. C-8 Modification of Guanosine

3.1 Background

Sonogashira coupling is a substitution reaction involving the substitution of a halide with a terminal alkyne, as shown in Figure 3.1. The procedure used for guanosine modification was based on a previously successfully procedure reported by Prestinari and Richert[17], used for uridine modification with sulfur functional groups (Figure 3.2 (a)). These modifications were used for a similar cross-linking process within the same strand (Figure 3.2 (b)), and resulted in improved thermal stability in the modified duplex formed with complementary DNA. We hypothesised that similar modifications would also increase thermal stability of G-quadruplexes. Furthermore we suggested that these cross-links would provide a powerful inhibitor of G4-processing enzymes (such as Pif1 helicase). Halogenated nucleosides, like those used by Prestinari and Richert[17], can be obtained on a large scale either through synthesis or from commercial sources. In some cases, they are even available as phosphoramidites. 8-Bromo-2'-deoxyguanosine is commercially available and easily synthesised, and provides a useful starting point for modification of guanosine. The protocol for Sonogashira coupling is well established, ranging from Prestinari and Richert's uridine modifications to the various sources used to adapt this method for guanosine. In these protocols large aromatic, photoactive compounds were introduced into guanosine[15,16] to use as fluorescent labels (Figure 3.2 (c)).

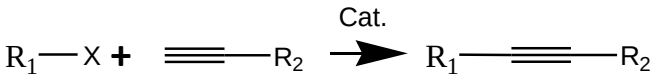


Figure 3.1 General Scheme for Sonogashira Coupling. X = F, Cl, Br, I. R_1/R_2 = desired functional groups. Cat.: Pd(0), Cu(I).

The reasons disulfide bridges were chosen for cross-linking were highlighted in Prestinari and Richert.[17] Primarily the stability and ease of formation of disulfide bonds makes them ideal for cross-linking. Disulfide bridges formed spontaneously under oxidative conditions, meaning that multiple strands containing sulfur functional groups will develop cross-links relatively quickly. Prestinari and Richert encouraged them through lowering of pH (pH 5.0) and exposing samples to air. It was also noted that their high orthogonality to other functional groups allows them to be interconverted with relatively simple reactions, and with little risk of side reactions. Once these bonds are formed they are stable under a range of conditions including heating and pH changes, but they can easily be reduced

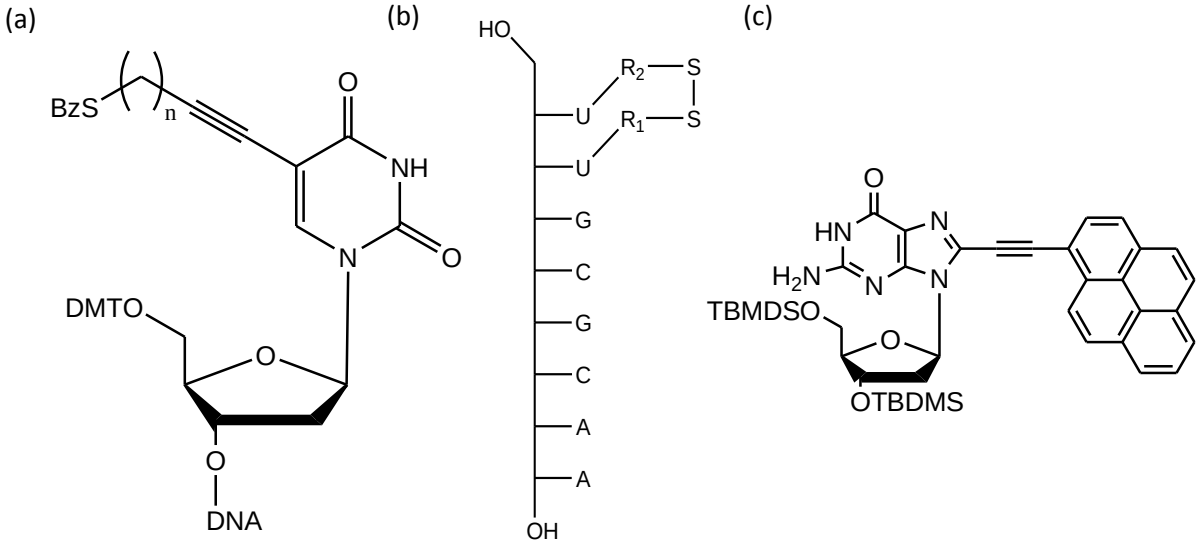


Figure 3.2 (a) Modified uridine nucleoside used by Prestinari to introduce disulfide bridges to duplex sequences. n = 2, 4. (b) Example of DNA sequence with an intrastrand disulfide bridge. R_1/R_2 are alkyne chains of varying length, as shown in (a). (c) Songashira modified guanosine nucleoside used[2] as a basis for the incorporation of our chemical cross-links into guanosine.

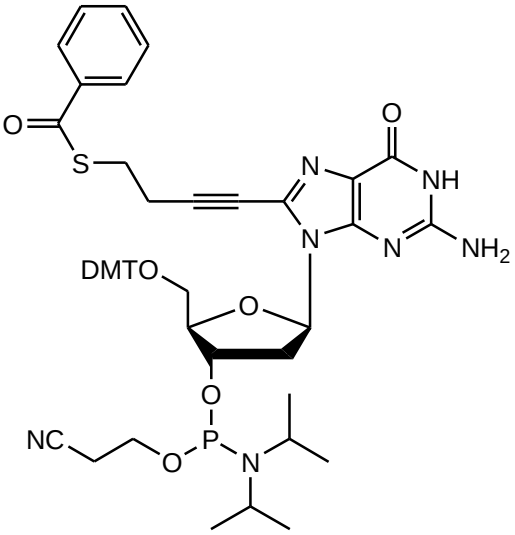


Figure 3.3 8-(S-But-3-ynyl benzothioate)-5'-O-DMT-3'-(2-cyanoethyl)-N,N-diisopropylamine-N2-DMF-2'-deoxyguanosin, the desired phosphoramidite for incorporation into G4-DNA

with reagents such as 1,4-dithiothreitol (DTT), allowing controlled formation and disruption of cross-links during experiments. Prestinari and Richert[17] observed a 5.3°C decrease in T_m when DTT was added to their modified duplexes. In addition to these benefits disulfide bonds have low steric demand, making cross-links flexible, allowing secondary structures to form more easily.

Using the nucleotide synthesised by Prestinari and Richert[17] (Figure 3.2 (a)) as the initial concept we designed the synthesis of the guanosine phosphoramidite in Figure 3.3. This material could be incorporated into a DNA sequence, and the benzoyl protecting group can be cleaved post-synthetically, resulting in free thiols required to form disulfide linkages. This was the first strategy we pursued for 8-position modification using Sonogashira coupling.

Sonogashira coupling occurs between a halide and terminal alkyne in the presence of several catalysts. The mechanism of this reaction is still somewhat debated, but the commonly proposed mechanism is demonstrated in Figure 3.4, with 8-Br-dG as an example. Et_3N deprotonates the slightly acidic terminal alkyne, which coordinates to the first catalyst, typically CuI, giving the first metal-carbon intermediate. Meanwhile, in (i) the second reactant is coordinated to the second

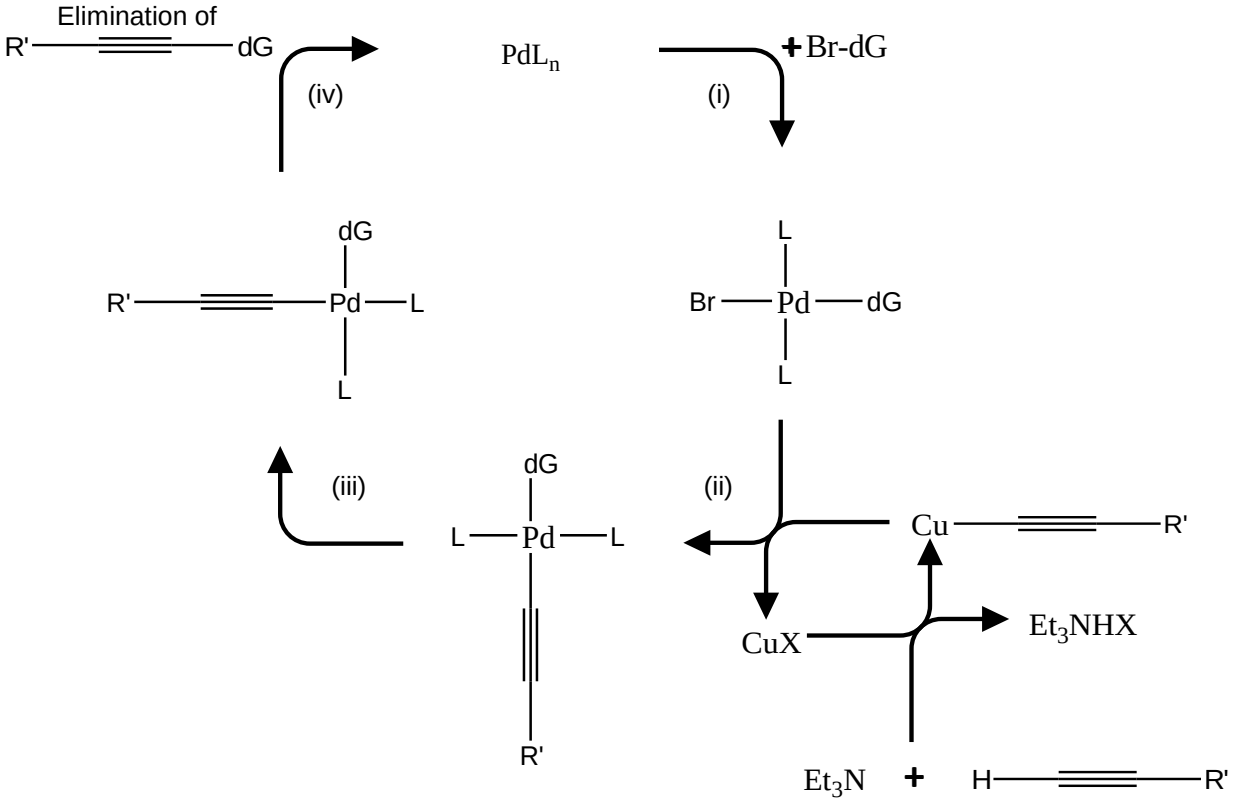


Figure 3.4 Proposed mechanism for Sonogashira coupling. Pd(0) and Cu(I) are typically provided by $Pd(PPh_3)_4$ and CuI respectively. The mechanism extends to any halide functional group, but the 8-bromo-2'-deoxyguanosinse reactant is used as an example of the reactions carried out on nucleosides. X = Halide, typically starting with CuI; L = various ligands, in our reactions typically PPh_3.

catalyst Pd(0) through oxidative addition. In almost all examples of guanosine substitution $Pd(PPh_3)_4$ was used as the source of Pd(0), but other sources of Pd, such as $Pd(PPh_3)_2Cl_2$, are also used. This addition will result in the halide and reactant coordinated opposite each other. In (ii) the two catalysts interact, exchanging ligands through a transmetallation reaction, and form the second metal-carbon intermediate, as well as reforming the copper catalyst. The two target molecules will be coordinated opposite each other so in (iii) they undergo ligand rearrangement. Finally in (iv) reductive elimination occurs, reforming the Pd catalyst and forming the product. This mechanism can theoretically be applied to any combination of terminal alkyne and halide substrates.

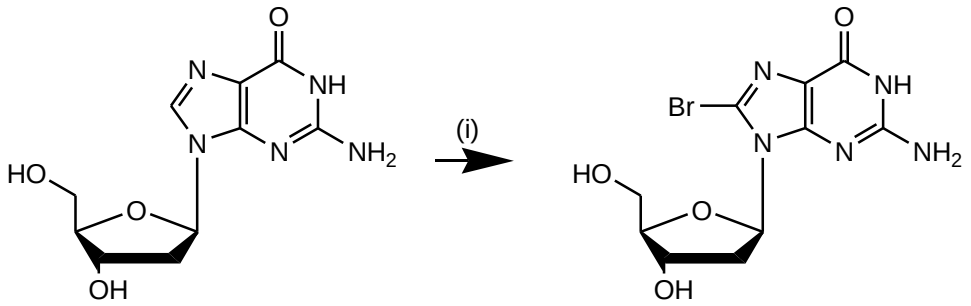


Figure 3.5 Conversion of 2'-deoxyguanosine to 8-bromo-2'-deoxyguanosine achieved with N-bromo succinimide. Reagents and conditions: (i) N-bromosuccinimide, S: 80:20 acetonitrile:water

3.2 Synthesis

8-Bromo-2'-deoxyguanosine (Br-dG) was produced through a substitution reaction of 2'-deoxyguanosine (dG) with N-bromo succinimide (NBS) (Figure 3.5). The reaction mixture was compared to the starting material using TLC (20% MeOH/DCM), and after thirty minutes a new spot was visible with no starting material present. Evaporation of the solvent resulted in a bright orange coloured solid, and after washing with acetone a pale orange powder was obtained. Unmodified deoxyguanosine could be characterised by a single peak in ^{1}H NMR at 7.93 ppm, corresponding to the proton at the 8-position. After treatment with NBS, this peak was no longer present, confirming that the reaction had gone to completion, as suggested by TLC. This reaction was initially carried out with water as the only solvent, as had been used for this reaction previously, as suggested by literature.[27] However, dG has low solubility in water, and when the reaction was tried again on a larger scale the starting material precipitated upon addition of NBS, making reactions of this scale impractical. Alternative sources[28] suggested an 80% acetonitrile/water mixture, which improved solubility significantly, allowing for large scale synthesis with 5-10 g of starting material with consistent 75-80% yields.

Later reactions would involve addition of a hydroxyl containing modification, 3-butyn-1-ol, to the 8-position, and further modification of this group, which would result in a large number of side

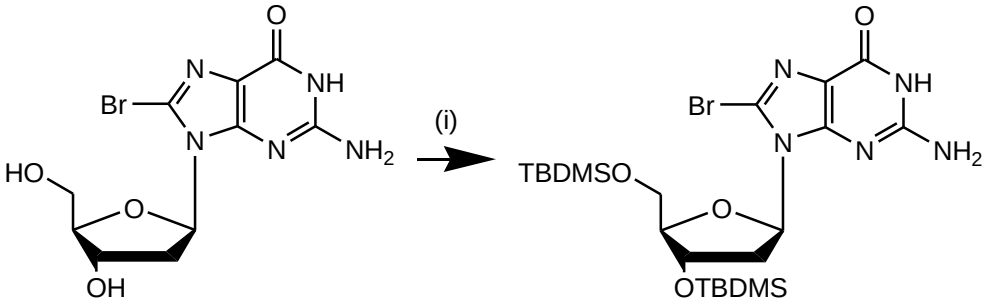


Figure 3.6 Protection of 3' and 5' Hydroxyl groups with TBDMS, to prevent side reactions after an alcohol side chain is introduced in the next step. Silyl protecting groups are reliably cleaved with F^-, making them ideal for reversible protections of hydroxyl groups. Reagents and conditions: (i) TBDMS-Cl, imidazole, S: dry DMF.

products if the 5'- and 3'-hydroxyl groups remain unprotected. *t*-Butyl dimethylsilyl chloride (TBDMS-Cl) (Figure 3.6) was chosen as the source for a protecting group because of the stability of silyl protecting groups, as well as the efficiency of both the protection and deprotection reactions. Deprotection is typically achieved through addition of F^-. Protection was carried out using the method as found in literature, using imidazole as a weak base in dry DMF.[29] The alcohol attacks the δ+ silyl chloride, and is deprotonated by the base. This procedure worked as reported, obtaining protected material with high conversion (at least 98%). This was confirmed by the presence of several characteristic peaks in both ^1H and ^{13}C NMR. In ^1H NMR the nucleoside peaks were unchanged, but two peaks at 0 ppm and 0.11 ppm with integrals of approximately six correspond to the dimethyl groups, while peaks at 0.83 ppm and 0.89 ppm with integrals of approximately nine correspond to the *t*-butyl groups. The two shifts seen in this NMR are a result of the slight difference between the 3'- and 5'-positions. In ^{13}C NMR peaks appear between 25.96-26.26 ppm and -4.99-4.24 ppm, which HMQC shows correlate to methyl and *t*-butyl protons, respectively. In addition the quaternary carbon of the *t*-butyl groups is also visible from 18.17-18.42 ppm. As will be discussed later, other procedures were eventually considered to avoid the steps which required this protection, and TBDMS protection was no longer necessary.

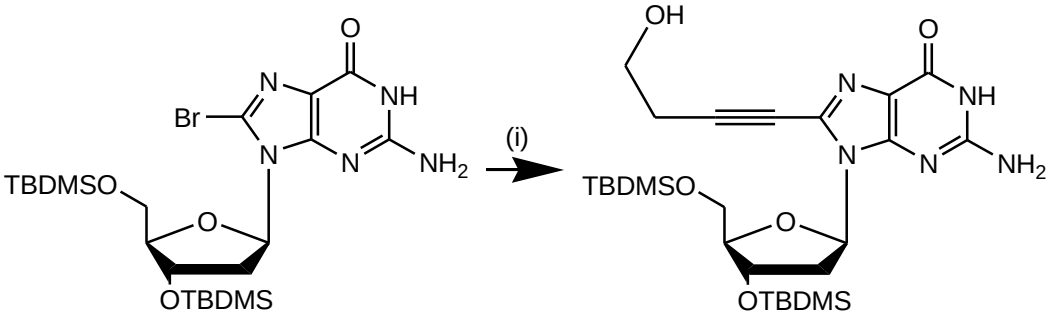


Figure 3.7 Sonogashira reaction of 8-bromo-3',5'-TBDMS-2'-dG with 3-butyn-1-ol to produce modified nucleoside. Reagents and conditions: (i) $Pd(PPh_3)_4$, CuI, Et_3N, S: dry DMF.

Once Br-dG had been protected, the first Sonogashira coupling reaction (Figure 3.7) could be attempted. This reaction was carried out in freshly distilled DMF, with 3-butyn-1-ol, $Pd(PPh_3)_4$, CuI and Et_3N, according to the protocol described in Chapter 7. This procedure was based on Prestinari and Richert's synthesis[17], which also used 3-butyn-1-ol, but quantities were adjusted based on the previous Sonogashira coupling reactions of guanosine mentioned above.[15,16]. The product was extracted with EtOAc, washed with EDTA, followed by sodium bicarbonate and finally brine. Washing with EDTA should remove the residual Pd(0) and Cu(I) catalysts from the mixture. ^1H NMR of the nucleoside showed new peaks at 2.63 ppm and 3.62 ppm, corresponding to the two CH_2 groups of 3-butyn-1-ol. A broad peak at 5.06 ppm also indicated the presence of a hydroxyl proton, which is sometimes observable in d_6-DMSO. No significant peak indicated the presence of the terminal alkyne proton, suggesting the coupling was successful. ^{13}C was also used for characterisation, with new peaks at 23.74, 60.22, 73.24 and 93.91 ppm corresponding to the 3-butyn-1-ol carbon chemical shifts. The peaks at 60.22 and 23.74 ppm correlate to the CH_2 groups of the alkyne in HMQC. The peaks at 93.91 and 73.24 ppm do not correlate to protons, and their higher chemical shift suggests they are the carbons of the alkyne.

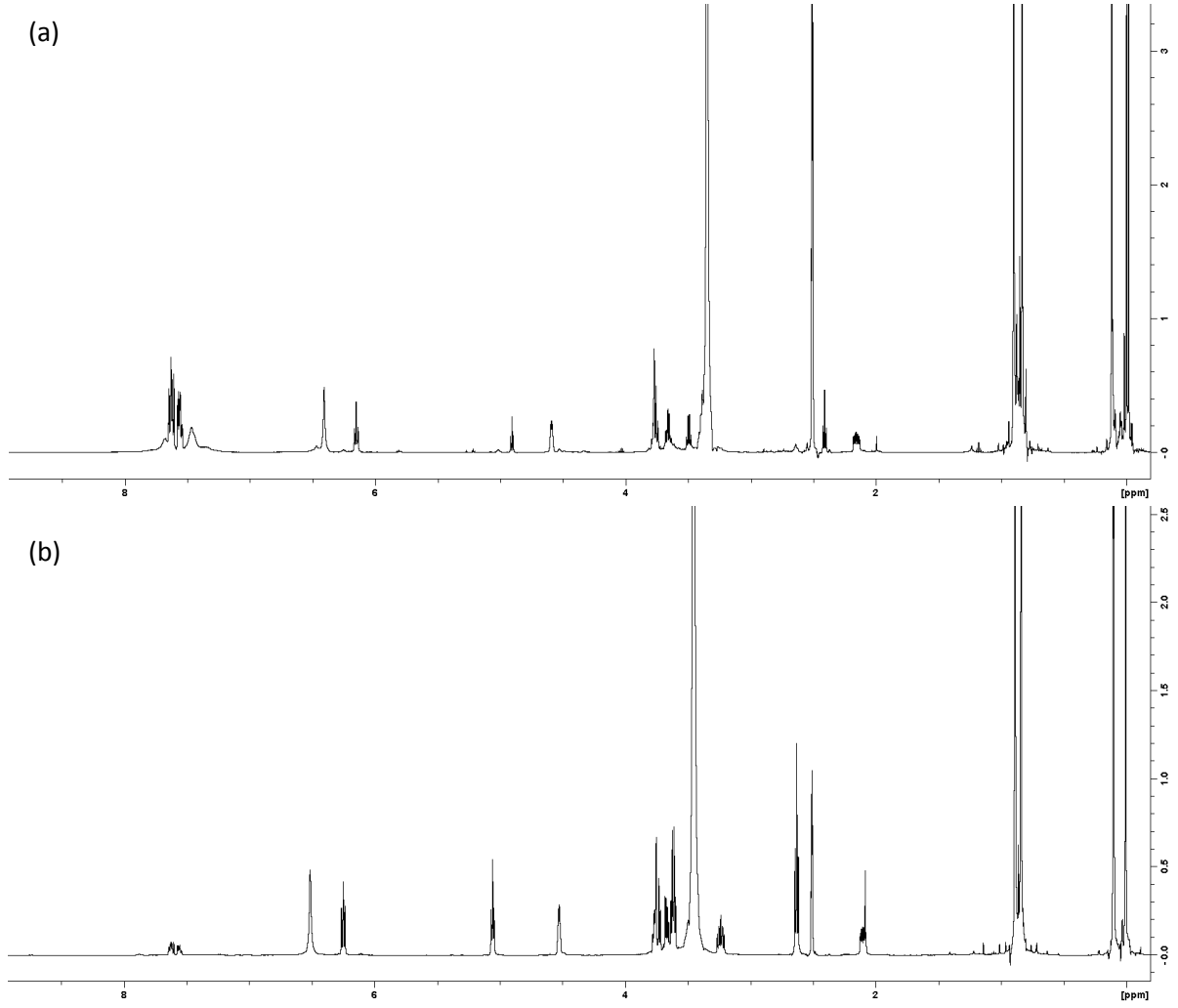


Figure 3.8 ^{1}H NMR of 8-(3-Butyn-1-ol)-3',5'-TBDMS-2'-deoxy guanosine before (a) and after (b) silica gel purification with 8:2 ethyl acetate:petroleum ether, and 99:1 acetone:toluene. Significant decreases in impurities from starting materials and side products can be seen, as well as decrease of impurity at 7.0-8.0 ppm, which was not removed after first column.

However, the reaction did result in several impurities, which created difficulty for further purification. Comparison of Figure 3.8 (a) and (b) shows a number of smaller impurities, particularly between 2.0 ppm and 5.0 ppm. However, the main side product corresponds to peaks with a chemical shift from 7.3-8.0 ppm, as seen in Figure 3.8 (a), with a pattern that suggests an aromatic material. This is possibly due to the degradation of the $Pd(PPh_3)_4$ catalyst, resulting in $O=PPh_3$. This initially suggested it could be problem with the catalyst, as $Pd(PPh_3)_4$ is sensitive to air, and degrades over time. However, the problem persisted when fresh catalyst was used in further attempts, meaning it may be an issue with the reaction conditions or the reaction with Br-dG specifically. TLC analysis of the material was carried out with varying concentrations of MeOH/DCM and EtOAc/petroleum ether, which were solvent systems used to purify similar compounds. These experiments indicated that MeOH/DCM solvents were too polar to separate any materials, but three distinct spots were observed with 8:2 EtOAc:petroleum ether. Silica gel column chromatography was performed with this solvent system, until a single spot which was believed to be the desired product was isolated. ^{1}H NMR indicated that the non-aromatic impurities had been removed but the aromatic impurity was still present in concerning quantities. Further TLC analysis found that a 99:1 acetone/toluene mixture showed two distinct spots which had not previously been separable. This solvent system was chosen because its polarity was between the two previously assessed systems. Column chromatography with this solvent isolated the desired product. According to ^{1}H NMR spectra (Figure 3.8 (b)) of this

compound, there is less than 1 mol% of the aromatic impurity remaining. The residue originally contained more than 30 mol% impurity. This material was then used in the next step.

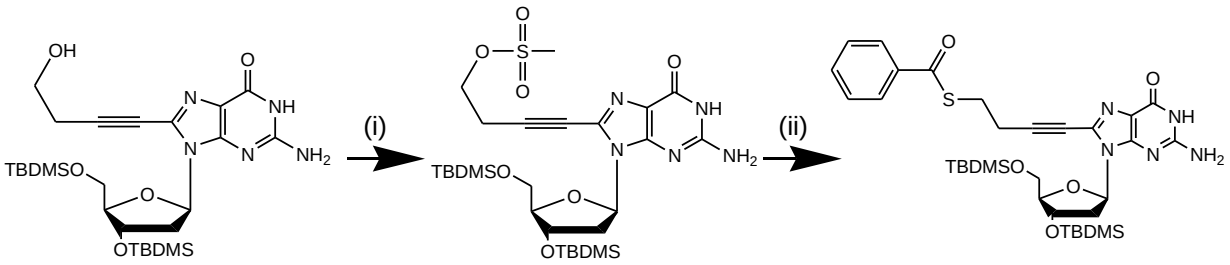


Figure 3.9 Two-step reaction carried out to introduce sulfur modification to 8-bromo-3',5'-TBDMS-2'-deoxyguanosine. Reagents and conditions: S: dry DMF, (i) methanesulfonyl chloride, Et_3N, -59° (ii) thiobenzoic acid, R.T.

The mesylation reaction was based on the method used by Prestinari and Richert, and was the key reason for the previous introduction of TBDMS protecting groups. Methanesulfonyl chloride (45 µL, 1.5 eq) was added to the nucleoside (0.22 g), dissolved in 12 mL of dry DMF and 0.136 mL of Et_3N (2.5 eq) at -59°C (Figure 3.9 (i)). Under basic conditions the methanesulfonyl chloride activates the alcohol, creating a good leaving group for the next step. After one hour thiobenzoic acid (0.081 g, 1.5 eq) was added and the solution was allowed to warm to room temperature (Figure 3.9 (ii)). After 2.5 hours, when TLC analysis (45% Petroleum Ether/EtOAc) seemed to indicate no starting material was remaining, the product was extracted with 35 mL of DCM and 25 mL of water. The organic phase was evaporated to give a black residue which was analysed by NMR. Unfortunately, this reaction appeared to result in complete loss of the nucleoside, because no nucleoside peaks were observed in NMR. Mesyl sulfonate is a good leaving group, and should have allowed substitution using the thiobenzoic group. This synthetic route was attempted several times, but its unreliability and inefficiency meant a new strategy was developed as a more effective strategy.

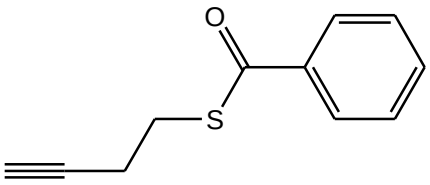


Figure 3.11 Newly designed sulfur containing modification, S-But-3-ynyl benzothioate, to be used in place of 3-Butyn-1-ol

The previous reaction encountered problems because of an ineffective mesylation reaction and decreased yields due to impurities that resulted from Sonogashira coupling. The molecule shown in Figure 3.11 was proposed to overcome these problems. The material would be synthesised first, and coupled with 8-Br-dG. This synthesis of S-but-ynyl benzothioate had been reported[30] previously with thioacetic acid, but thiobenzoic acid was more readily available, and benzoyl protecting groups had been used for nucleoside modification previously.[31] The synthesis could begin with 3-butyn-1-ol, the starting material used previously, and a similar mesylation reaction with Et_3N and methanesulfonyl chloride was used. Following this, cesium carbonate (Cs_2CO_3) was used with thiobenzoic acid to give the product. This reaction is similar to the procedure used for nucleoside modification, but we hoped it would eliminate the problems that arose when the reaction took place on the nucleoside. In addition, the previous linear synthetic route had resulted in low yields, whereas this procedure uses a convergent synthetic route, hopefully addressing problems with low yield of the previous Sonogashira coupling. Finally, because the protocol did not carry out mesylation in the

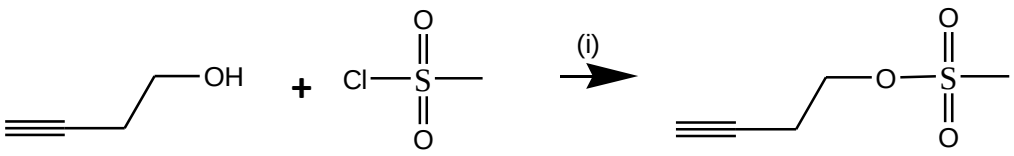


Figure 3.10 Mesylation reaction of 3-butyn-1-ol. Reagents and conditions: (i) methanesulfonyl chloride, Et_3N, S: DCM, 0°C.

presence of the 3'- and 5'-hydroxyl groups protection of these groups may not be necessary, further reducing the required number of steps.

3-Butyn-1-ol and Et_3N were dissolved in DCM and cooled to 0°C. Approximately 1 equivalent of methanesulfonyl chloride was added slowly, and this reaction mixture was stirred for two hours. After two hours TLC analysis (30% EtOAc/hexane) showed no starting material remaining. This TLC was visualised with Phosphomolybdic Acid (PMA) because neither reactants nor products are UV active. Water was added to quench the reaction, and the organic layer was washed with brine. Evaporation of DCM gave a 71% yield of pure product, which was confirmed by NMR. ^{1}H NMR showed 3-butyn-1-ol peaks, with the addition of a 3 proton peaks at 3.08 ppm, corresponding to the methyl group of the sulfonate. This peak correlates to a ^{13}C peak at 37.5 ppm. This product was used in the following step without further purification.

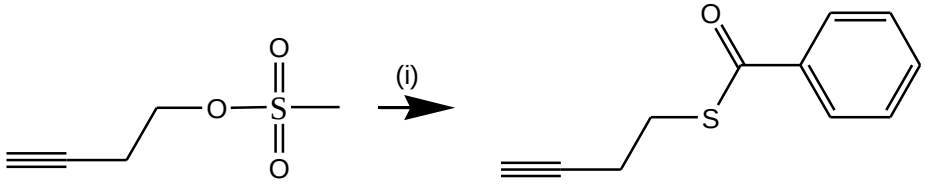


Figure 3.12 Substitution of methane sulfonate with thiobenzoic acid. Reagents and conditions (i) Thiobenzoic acid, Cs_2CO_3, S: DMF, 0°C.

The product was dissolved with 1.2 equivalences of Cs_2CO_3 in DMF and cooled to 0°C. Approximately 1.2 equivalences of thiobenzoic acid was dissolved in DMF and added slowly to the mixture while stirring. Cs_2CO_3 acts as a base, and the thiobenzoic acid substitutes the methane sulfonyl group. TLC analysis (45% EtOAc/Hex) was performed regularly until no starting material was observed after approximately two hours. The product is UV active, but PMA was used once again to visualise the starting material. The reaction mixture was diluted with EtOAc and washed with brine before evaporating to obtain an orange oil. Formation of the product was indicated by disappearance of the methyl group at 3.08 ppm and appearance of three aromatic peaks at 7.4 – 8.0 ppm. These peaks correlate to ^{13}C peaks a 133.4, 128.6 and 128.2 ppm, with peaks at 192.3 and 136.7 ppm corresponding to the quaternary thioester and aromatic carbons respectively. NMR indicated that the material required no further purification.

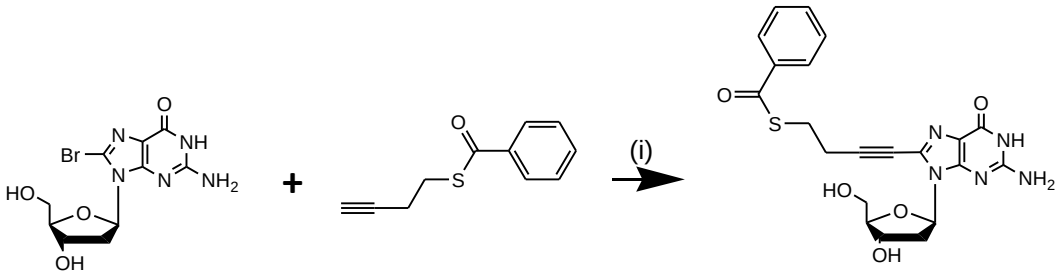


Figure 3.13 Modification of 8-bromo-dG with S-but-3-ynyl benzothioate. Carried out using a microwave synthesiser. Reagents and conditions: (i) $Pd(PPh_3)_4$, CuI, bases (Et_3N, Dowex 1X-8, Amberlite IRA-400 or Amberlite IRA-67), S: dry DMF.

Table 3.1 Solvents tested for Sonogashira coupling of 8-Br-2'-dG with S-but-3-ynyl benzothioate

	Solvents (8-Br-2'-dG)							
	Et_3N	Piperidine	Morpholine	Dioxane	THF	Acetonitrile	DMF	DCM
(a)	No	Yes	Yes	Yes	Yes	No	Yes	No

(a) Formation of product is indicated by the disappearance of the starting material and appearance of a new material as observed by TLC analysis (5% MeOH/DCM). Yes: Starting material disappeared, No: Starting material present.

This material was reacted with 8-Br-dG using the same protocol described for 3-butyn-1-ol modification in Chapter 7. However, the reaction was first attempted in a small scale (0.05 g of nucleoside) in a range of solvents (Table 3.1) to determine if the choice of DMF as a solvent had caused the previous impurity formation. This was carried out parallel to several experiments discussed below. Two distinct products were observed by TLC analysis (5% MeOH/DCM). One product with $r_t \approx 0.8$ appeared in DMF, piperidine and THF, while a second with $r_f \approx 0.5$ appeared in dioxane and morpholine. However, the starting material had low solubility in some solvents and was recollected after reaction. Table 3.1 shows the solvents in which the reaction appeared to proceed according the TLC analysis. ^{1}H NMR of the products showed that the first product ($r_f \approx 0.8$) could be the desired material, while the second product ($r_f \approx 0.5$) contained no nucleoside peaks. We chose to continue using DMF, as the solubility of the starting material was comparable for all solvents, but the purity of the product, as observed by TLC analysis of the reaction mixture, was optimal in DMF.

Petricci *et al*. (2003)[32] established a protocol for synthesis of modified pyrimidines using Sonogashira coupling under microwave irradiation. More recently, Zhang *et al.* (2018)[33] extended this procedure to their modification of a purine, 8-bromo-2'-deoxyadenosine. The conditions for microwave synthesis gave good results for nucleoside modification, and we suggested this could improve the coupling of guanosine as well. Garg *et al*. (2005)[34] reported that when IRA-67 resin was used in Sonogashira coupling in place of Et_3N it resulted in increased yield, particularly in the case of Br-dG based reactions. They note that Br-dG reactions are particularly slow and unreliable without the resin. This observation supported our previous observations of the coupling. Petricci *et al.*[32] also reported the use of IRA-67 in their modification, while Zhang *et al.*[33] used both Et_3N and Amberlite IRA-400 together.

We decided to test a range of conditions to determine if Sonogashira coupling of Br-dG could be effectively carried out with a microwave synthesiser using resins. In addition to Et_3N, Amberlite IRA-400, Dowex 1X-8 and Amberlite IRA-67 were tested as bases. The same catalysts and solvents were used as previously described for Sonogashira coupling. Microwave reactions were initially carried out at 50°C for 2.5 hours at 6.8 atm. Initially, monitoring with TLC indicated that reactions were incomplete when carried out at 50°C for 2.5 hours, but further addition of reagents (CuI, $Pd(PPh_3)_4$, base) and increasing the temperature to 65°C resulted in complete disappearance of starting material. In future reactions the time was increased to 6 hours. The product was extracted with EtOAc as described previously and a small quantity of insoluble material was collected in addition to the residue after EtOAc evaporation. All of these materials were analysed with NMR.

NMR of the residue showed no nucleoside peaks, suggesting the product could be degraded. However, NMR of the insoluble material contained peaks which could correspond to modified nucleoside, although a large number of impurities also appeared to be present. Mass spectrometry indicated a peak at 456.13 a.m.u. (49%) which corresponded to the expected product, but several smaller molecules were also present. This made it difficult to identify the product without a larger quantity of material for purification. This reaction was increased in scale to allow for better characterisation and further synthesis. However, large scale reactions produced no product. The previous result suggested that this reaction would occur and produce the desired material, so further work was done to examine conditions to encourage the reaction to occur. First, reactions were run at 40°C, 50°C and 60°C to observe the effect of temperature on the reaction, but no product was observed at any temperature. This reaction was tried several times with variations made to quantity of resins and catalysts, but no product was ever observed in these experiments. This initially appeared to be a promising method, but further attempts made it clear that this method suffered from the same problems with unreliability and impurities as previous Sonogashira couplings. The results of

microwave synthesis suggest that it may be a possibility for future modifications, but other reactions had more success, and were given priority.

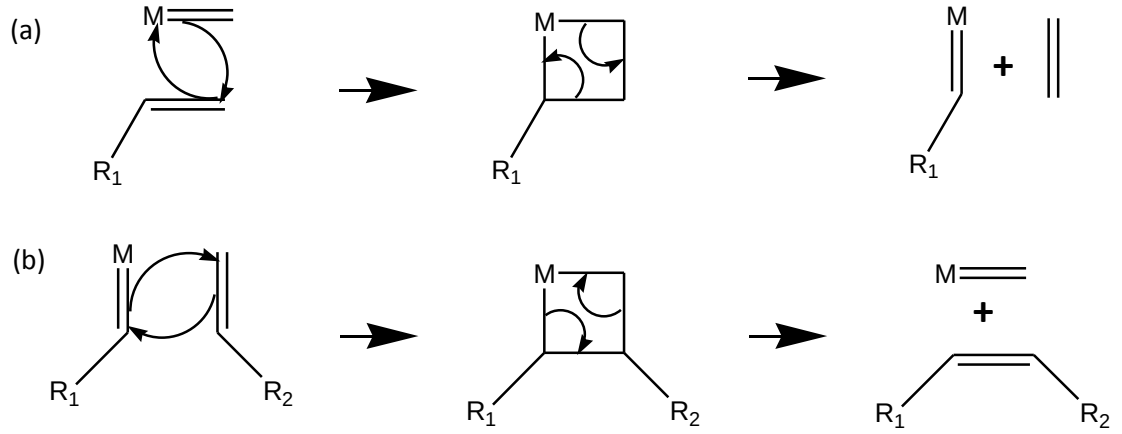


Figure 3.14 Mechanism for cross metathesis of multiple alkenes.

One alternative strategy we considered to disulfide formation is cross metathesis using Grubb's catalysts. Grubb's catalysts are a series of ruthenium based molecules used to couple two alkenes. In many cases these reactions are known as ring-closing metathesis because they are used in the final step of a ring formation. The two alkenes react together to form a single alkene, along with the elimination of ethene (Figure 3.14). The common feature in these catalysts is the Ru=C bond. The first step of the mechanism for these reactions involves the formation of a ruthenium complex through a reaction of an alkene with this bond, as seen in Figure 3.14 (a). This complex rearranges, eliminating ethene, and essentially reforms the M=C bond. This new molecule behaves with a similar mechanism to the original catalyst, forming a new complex with the second alkene. A second rearrangement results in the removal of the metal, reforming the catalyst, and the formation of a new alkene, as seen in Figure 3.14. In both steps elimination results in the dissociation of an alkene, first ethene, and then the product. This means that the product, due to the loss of ethene, is always a shorter chain than the starting material. This is an important consideration when developing cross-links using this method, because cross-links will need to be long enough to allow the formation of secondary structures after this chain shortening occurs.

To use cross metathesis for cross-link formation we required a molecule containing both an alkyne and an alkene functional group. The alkyne allows for the functionality to be introduced through Sonogashira coupling, while the alkene allows cross-links to be formed with Grubb's catalyst. Alkenes are stable in a wide range of conditions, although they may lack some of the flexibility that was appealing with disulfide bridges. However, one key advantage is their stability in DTT, which avoided the concerns for enzyme assays later in testing because DTT is usually added to the buffer to stabilise enzymes. The first proposed molecule was 1-Hexen-5-yne (Figure 3.15 (a)), which was a simple molecule containing both groups. However, this molecule had two disadvantages over the molecule eventually chosen for this experiment. First, commercial availability was low, and as such the material was prohibitively expensive, as well as being difficult to synthesise from readily available chemicals. Second, it had limited flexibility, especially once the chain was shortened during cross metathesis. Instead the molecule chosen was 3-(2-propyn-1-

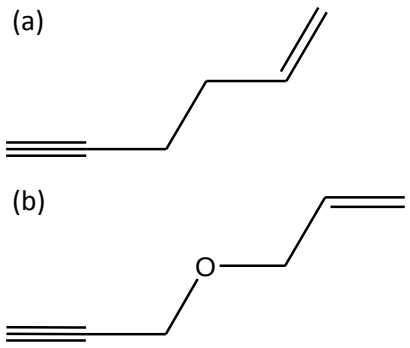


Figure 3.15 (a) 1-Hexen-5-yne, the first molecule considered for alkene cross-links. (b) 3-(2-Propyn-1-yloxy)-1-propene, the modification chosen for experimentation with alkene cross-link formation.

yloxy)-1-propene (Figure 3.15 (b)). This synthesis for this molecule was described by Guermont (1953)[35], using allyl bromide and propargyl alcohol, both of which were available, and the ether in the middle gives added flexibility which is highly desirable in the eventual formation of crosslinks.

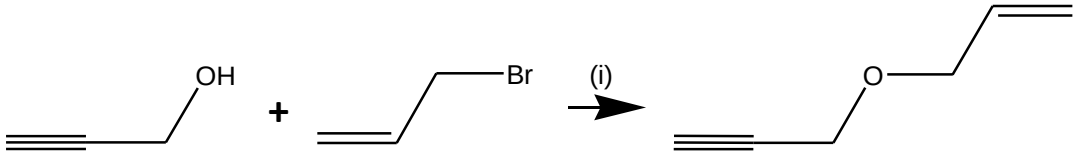


Figure 3.16 Reaction of allyl bromide and propargyl alcohol gives 3-(2-propyn-1-yloxy)-1-propene. Reagents and conditions: (i) 4M KOH (aq.), 70°C.

The two materials (Figure 3.16) were reacted by heating gradually from 0°C to 70°C while aq. 4M KOH was added dropwise. After four hours, the resulting material was distilled at 105-106°C to produce pure compound. Both starting materials have widely differing boiling points, but the NMR of the product is very similar to the spectra of the individual starting materials. NMR spectra indicate that the material collected between these temperatures contains peaks from both, with the correct integration, suggesting this is a single product, rather than a mixture. The original source did not report NMR, but ^{1}H NMR was compared to NMR reported by a later source[36] to identify key peaks. The reaction only had a yield of approximately 40%; however, the reactions at this stage required only small quantities of material, so this reaction was sufficient for all the following steps.

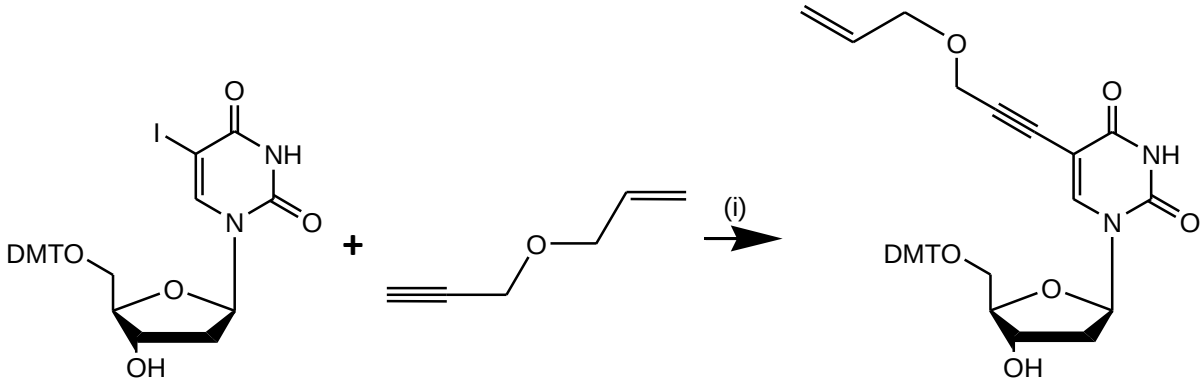


Figure 3.17 Reaction scheme for Sonogashira coupling of 5-iodo-5'-DMT-2'-deoxyuridine and 3-(2-propyl-1-yloxy)-1-propene. Reagents and conditions: (i) $Pd(PPd_3)_4$, CuI, Et_3N, S: dry DMF.

This modification was initially tested with the commercially available and cheaper 5'-DMT-5-iodo-2'-deoxyuridine (Figure 3.17) to determine if the alkyne synthesised material was reactive in Sonogashira conditions. The reaction was carried out based on the same procedures used previously.[15,16,31] The product showed a number of impurities in ^{1}H NMR (Figure 3.18). Peaks between 6.5 – 8.0 ppm seemed to indicate an unexpected aromatic system, and various smaller peaks exist in the remainder of the spectra. The anomeric proton is visible at approximately 5.8 ppm, suggesting the nucleoside is present.

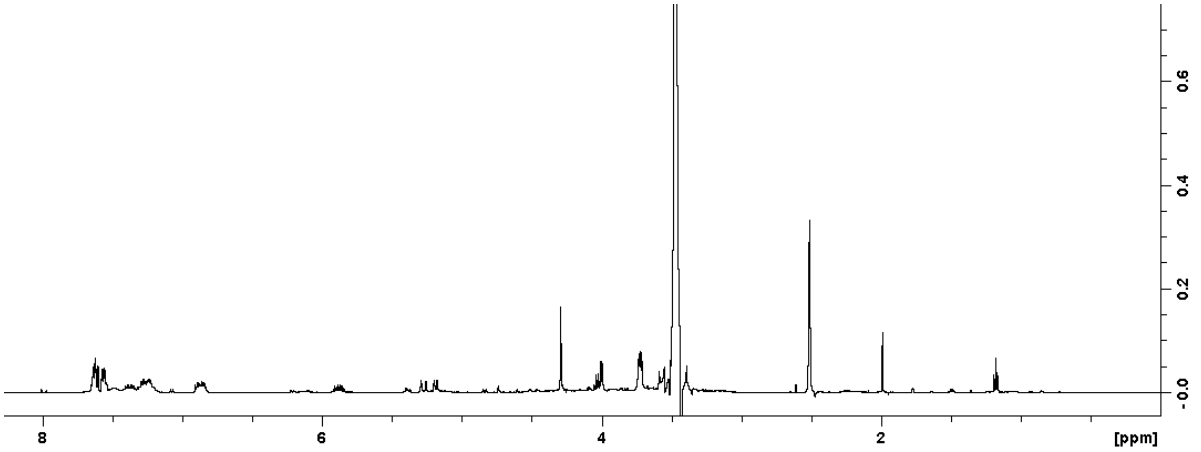


Figure 3.18 ^{1}H NMR of product of Sonogashira coupling with 5-iodo-5'-DMT-2'-deoxyuridine and 3-(2-propyl-1-yloxy)-1-propene. Significant impurities are visible at 6.8-8.0 ppm, which are attributed to either PPh_3O (a product of catalyst degradation) or a cyclisation of the modification with an unprotected hydroxyl of uridine.

The removal of these impurities was attempted using column chromatography. TLC analysis was used to evaluate potential solvent systems. Silica is normally acidic, so TLC plates and the column were pre equilibrated with Et_3N to avoid cleavage of DMT. Initially a silica gel column using 45:54:1 acetone:DCM:Et_3N elution system was shown to separate several materials. However, the impurities were still observed after purification, and a further column was carried out with 5:94:1 EtOH:DCM:Et_3N. This material showed only one major spot on TLC after purification, but NMR still showed the presence of the impurities, as shown in Figure 3.18. Further TLC analysis, varying the polarity of eluents, did not show any more separation. This appears to be the same degradation of the $Pd(PPh_3)_4$ catalyst to give $O=PPh_3$, which was observed for previous Sonogashira coupling.

Table 3.2 Solvents tested for Sonagashira coupling of 5-I-5'-DMT-2'-dU with 3-(2-propyn-1-yloxy)-1-propene

Solvents (5-I-5'-DMT-2'-dU)								
	Et_3N	Piperidine	Morpholine	Dioxane	THF	Acetonitrile	DIPEA	DCM
(a)	No	Yes	Yes	No	No	Yes	No	No

(a) *Formation of product is indicated by the disappearance of the starting material and appearance of a new material as observed by TLC analysis (5% MeOH/DCM). Yes: Starting material disappeared, No: starting material remained.*

It was unclear what the actual cause of this product is, so the reaction was carried out in a variety of solvents (Figure 3.19) to determine if different conditions could prevent formation of this material. These reactions were carried out parallel to the previously described experiments with the sulfonate. TLC analysis (5% MeOH/DCM, 1% Et_3N) indicated that the reaction occurred in piperidine, morpholine and acetonitrile, but all systems resulted in similar products to the previous synthesis with DMF. The products resulting from these coupling reactions suggested that Sonogashira coupling did occur, but excessive side reactions made it difficult to obtain the desired product.

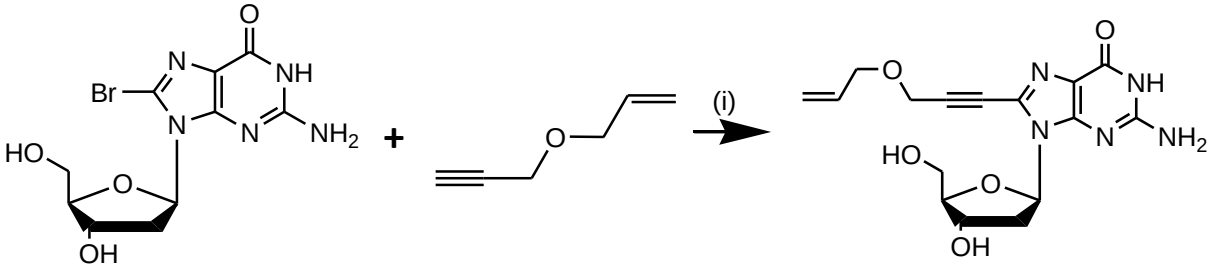


Figure 3.19 Reaction scheme for Sonogashira coupling of 8-bromo-2'-deoxyuridine and 3-(2-propyl-1-yloxy)-1-propene. Reagents and conditions: (i) $Pd(PPd_3)_4$, CuI, Et_3N, S: dry DMF.

The same reaction protocol was used to couple 3-(2-propyn-1-yloxy)-1-propene and 8-Br-dG in parallel with the previously discussed microwave irradiated synthesis with S-but-3-ynyl benzothioate. This reaction used DMF as a solvent, as none of the other solvents appeared to be preferable. TLC analysis (5% MeOH/DCM) suggested the reaction proceeded to completion, but after several attempts no nucleoside peaks were observed in ^{1}H NMR. While S-but-3-ynyl benzothioate reactions gave a product which was difficult to replicate, this synthesis gave no indication of success. This was a discouraging result for the introduction of alkene functionality to guanosine using Sonogashira coupling, but we will explore the possibility of introducing functionality through alternative modifications in Chapter 6.

3.3 Conclusion

Sonogashira coupling seemed like a convenient method of modifying guanosine due to the availability of Br-dG as a starting material and the previously reported protocols. Several of the modifications appeared to be successful, and some mass spectra indicate the formation of unique, modified nucleosides. However, while this initially seemed promising, repeated attempts could not produce modified nucleosides. There is still potential for this method to be used, but other methods presented in the following chapters gave more reliable results.

4. N-2 Modification of Guanosine

4.1 Background

The alternative synthetic approach to the previously described modification at the 8-position is modification at the 2-position, which we considered based on examples from previous literature. [18] 5'-DMT-2'-deoxy-O6-(trimethylsilyl ethyl)-2-fluoro-Inosine phosphoramidite, the starting material shown in Figure 4.1, is a convertible nucleoside for guanosine having a fluoride at the 2-position. This fluoride can be substituted post-synthetically with a secondary amine, as shown in Figure 4.1, and form the N-2 modified guanosine. The Hoogsteen bonds in a G-tetrad only involve one of the protons of the primary amine, meaning this nitrogen can be functionalised in this way without disrupting the tetrad. Previous applications of this modification had used inosine to create labelled sequences or transition state analogues, but not for cross-linking. We intend to investigate new applications of this modification for G4 analysis.

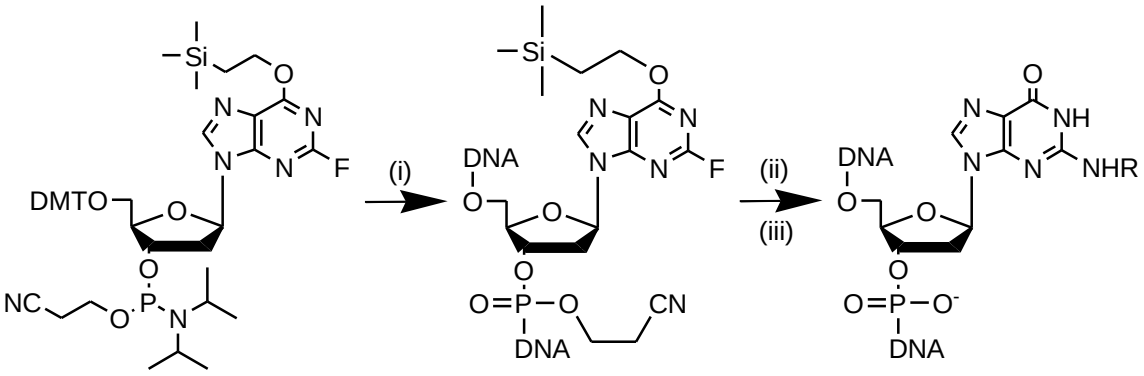


Figure 4.1 Reaction scheme for (i) incorporation of 2-fluoro-2'-deoxyinosine into DNA, (ii) substitution of F by R-NH_2, and (iii) cleavage from solid support and deprotection. Deprotection uses 28% aq. NH_4OH to remove cyanoethyl protecting groups and cleave from the solid support, followed by 0.1% acetic acid to remove the silyl protecting group. The resulting sequence will contain the desired functionality on this base, but should otherwise behave the same as native guanosine.

The mechanism for this substitution is not shown in the literature; however, my suggested mechanism, similar to the mechanism of the Chichibabin reaction, is shown Figure 4.2. The amine acts as the nucleophile, attacking the carbon, which is δ+ due to the electron withdrawing halide. The pyrimidine of guanosine behaves similarly to pyridine, with the aromaticity of this system increasing the stability of the intermediate. F^- acts as a leaving group, and subsequently removes the proton from the amine, completing the oxidation of the amine and producing the substituted guanosine. In an alternative S_n2 mechanism, F^- could act as leaving group without the formation of the intermediate. However, F^- is typically considered a poor leaving group and the relative stability of the intermediate helps to encourage the process. Regardless, fluoride elimination would be the rate determining step, and because of this the reaction is slow.

The choice of modification is different from the Sonogashira modifications for two reasons. First, at this position the modification needs to introduce a proton for hydrogen bonding, meaning the modification is most likely an amine. Secondly, the 8-position and 2-position of guanosine face different directions when folded into the quadruplex, which is shown by the diagram of $(G_4T_4G_4)_2$ in Figure 4.4 (a). This means that different nucleotides will be better situated for 2- or 8-position modifications. To determine the optimal position for these modifications before attempting to

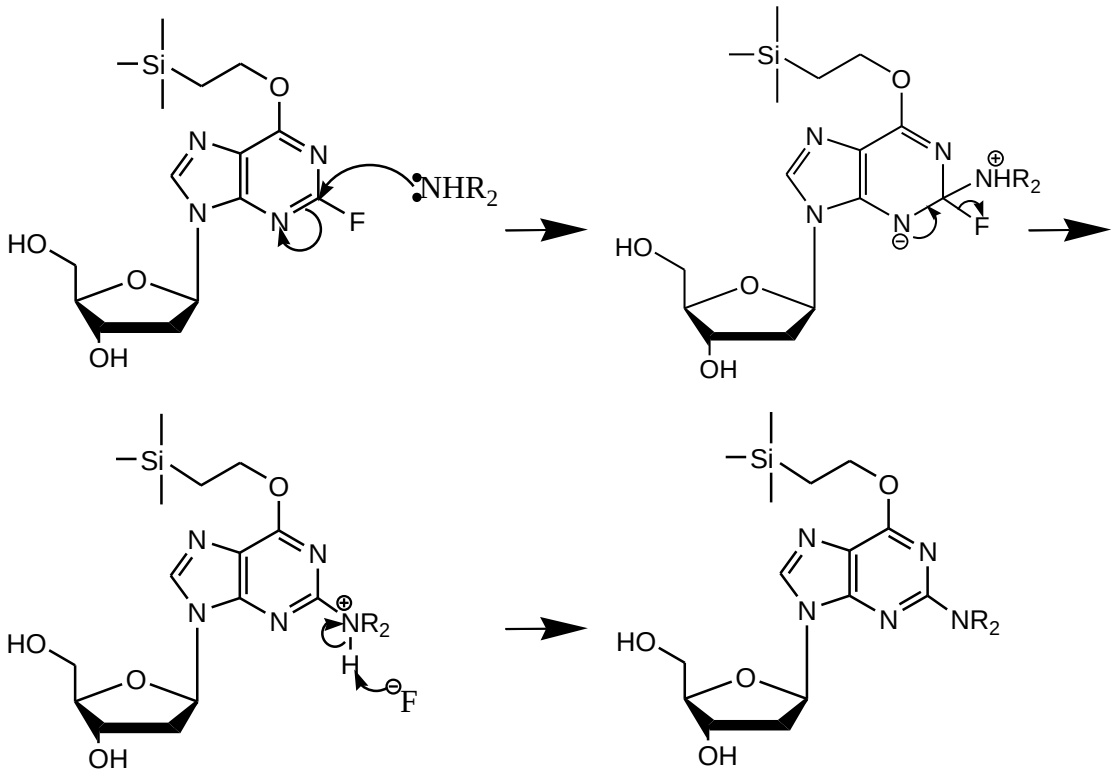
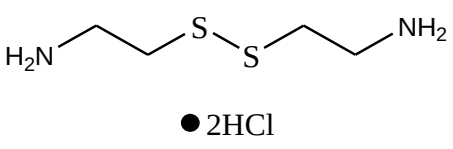


Figure 4.2 Proposed mechanism for amine substitution of 2-fluoro-2'-deoxyinosine. The mechanism is similar to the Chichibabin reaction because the aromatic rings of purine can behave similarly to pyridine.

synthesise sequence the crystal structure of $(G_4T_4G_4)_2$ (pdb = 2WGQ)[4(b)] was used to measure distances between the modifiable N2-position nitrogen atoms of each guanosine, as shown in Figure 4.4 (b). Comparing these distances and the size of the desired modification demonstrates that cross-links can be introduced without significantly altering the G4-structure. This is a bimolecular, anti-parallel G4-DNA formed by the same sequence in the presence of monovalent cations (Na^+, K^+ and NH_4^+). This provides the possibility to introduce a single modification in a sequence, which can then form a cross-link if the modified bases will be close to each other upon G4-folding. Figure 4.4 (a) shows both the configuration and modifiable positions of each guanosine, and this can give a general idea of the viability of each position for modification. This meant several positions were disqualified because the position of the NH_2 groups were incompatible due to the orientation of guanosine. Distances between the possible positions on opposite strands are shown in Figure 4.4 (c). The positions chosen for this purpose were 2-2 and 11-11. We wanted to use the same position on each strand to allow a single modified strand to form cross-link with the identical molecule. The 2- and 11-positions have optimal orientations and distances for cross-link formation with their counterpart. Two further modification 4-4 and 10-10 would also be synthesised. 4-4 modifications should be too far apart for short linkers, while 10-10 have the incorrect orientation. We intend to use these sequences for two purposes. They will either act as negative controls, indicating that cross-link formation prevents G-quadruplex folding, or they will form alternate topologies as a result of the cross-link. These four modifications would give an idea of the how successful the synthesis and subsequent substitution of 2-fluoro inosine modified sequences was, as well as allowing for analysis of the resulting complexes.

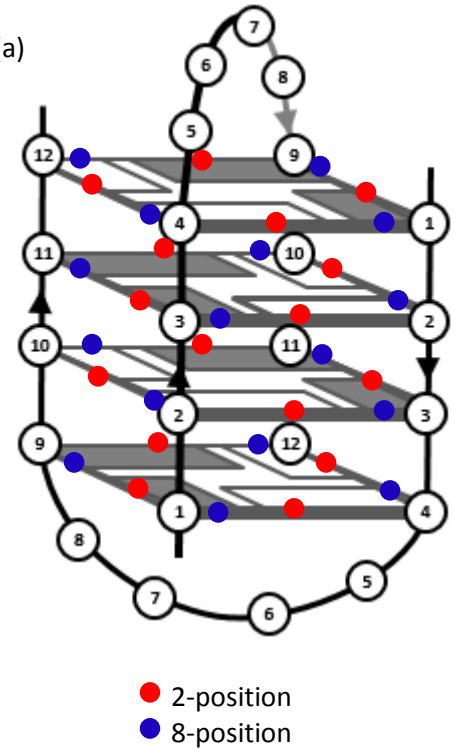


● 2HCl

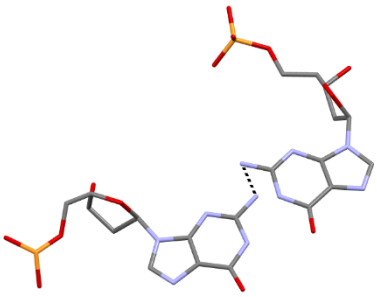
Figure 4.3 2,2'-dithiobis-ethanamine (Cystamine dihydrochloride), used as the starting point of 2-fluoro Inosine modification.

We decided to introduce sulfur functionality initially for the same reasons discussed previously for Songashira coupling: ease of formation, stability and flexibility. This required a molecule with both amine and

(a)



(b)



(c)

Position on strand one					Position on strand two
G-2	G-4	G-10	G-11		
		3.4905	7.6705	G-1'	
4.983	3.199			G-2'	
		5.796		G-3'	
	10.731			G-4'	
		7.711	3.3605	G-9'	
			6.208	G-11'	

Figure 4.4 (a) Structure of bimolecular $G_4T_4G_4$ quadruplex indicating the locations of modifiable sites within the structure. The tetrad colours indicate syn- and anti-configuration. Grey: anti, White: syn. (b) X-ray crystal structure (2WGQ) showing two guanosine residues with correct orientation to interact, and indicating the distance between positions where cross-links would be incorporated. (c) Distances (Å) between sites with favourable orientations for N2-modification, based on the X-ray crystal structure, 2WGQ.

thiol functional groups. Cystamine (2,2'-dithiobis- ethanamine) dihydrochloride was chosen as the starting material (Figure 4.3). The measured distance between nitrogen atoms of cystamine 7.117 Å, (using the X-ray crystal structure of Cystamine tartrate[37]). This is slightly larger than the distance measured within the quadruplex, allowing for some flexibility. Cystamine dihydrochloride is readily available, so it could be used as a test of the coupling reaction.

4.2 Synthesis

Before reaction with ^{F}dI-modified oligonucleotides cystamine dihydrochloride was converted to the free amine by treating the salt with sodium hydroxide in methanol. This conversion was a typical acid/base type reaction, which produced NaCl, and H_2O as side products. NaCl was removed by evaporating methanol until it precipitated. Ethanol was added, and the precipitate was removed by filtration. Ethanol was then evaporated, assisting the removal of H_2O through formation of an azeotrope with a more favourable boiling point. ^{1}H NMR of cystamine dihydrochloride contained an NH peak at 8.37 ppm, which was shifted to 3.09 ppm after conversion. Cystamine was stored as a hydrochloride salt due to the relative instability of its free amine form. ^{1}H NMR carried out two days after the first synthesis indicated that the material had already degraded significantly, and it was discarded. Therefore, this conversion was always carried out immediately prior to the coupling reaction.

Modified $G_4T_4G_4$ sequences were synthesised with the functionalised phosphoramidite incorporated through semi-manual addition at the positions indicated in Table 4.1. After DNA synthesis 0.5 molL^{-1} cystamine solution in DMSO was added to the solid support, and the mixture was

Table 4.1 Modified sequences $G_4T_4G_4$ synthesised with 2-fluoro-2'-deoxyinosine (^{F}dI)

Modification	Sequence
2-position	G^{F}**dI**G$_2$T$_4$G$_4$
4-position	G$_3$^{F}**dI**T$_4$G$_4$
10-position	G$_4$T$_4$G^{F}**dI**G$_2$
11-position	G$_4$T$_4$G$_2$^{F}**dI**G

continuously shaken to disturb the solid material from the bottom of the vial. The reaction mixture was removed by filtration, and the solid support was washed with DMSO and methanol. DNA was cleaved from the solid support overnight with 1 mL of 28% aq. NH_4OH. This was evaporated, and the

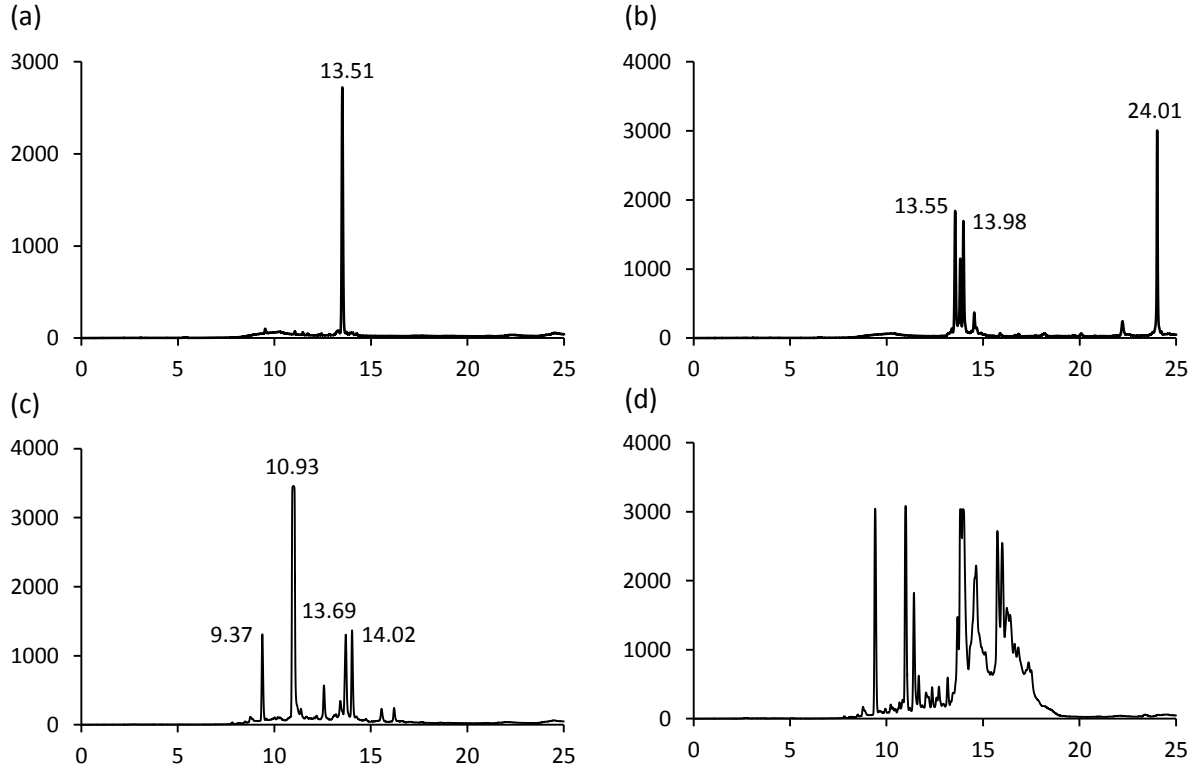


Figure 4.5 HPLC of various $G_4T_4G_4$ based sequences: (a) unmodified, pure $G_4T_4G_4$; (b) $G_4T_4G_2{}^{F}\underline{\mathbf{dI}}G$; (c) $G_4T_4G^{F}\underline{\mathbf{dI}}G_2$; (d) $G_3{}^{F}\underline{\mathbf{dI}}T_4G_4$. 100 mmolL^{-1} TEAA buffer, pH 7.0 in Acetonitrile. 0 – 25%, 2 - 20 mins, 25 – 80%, 20 – 25 mins.

residue was dissolved in 1 mL of 0.1% acetic acid solution. After two hours the solution was neutralised with 3M NaOAc and precipitated from ethanol. The pellet was dissolved in Milli-Q water. Insoluble materials were removed by centrifugation, and the solution was purified by reverse phase-HPLC.

Figure 4.5 (a) shows the RP-HPLC chromatogram for the control sequence of unmodified $G_4T_4G_4$, with a single peak at R_t = 13.51 min. Modified sequences might vary slightly, but are expected to have similar retention times to the unmodified sequence. Similar peaks with retention times of approximately 13.5 min were observed in both (b) and (c). These materials were isolated by UV-Vis controlled fraction collection under the same HPLC conditions, and were then desalted by size exclusion with a NAP-5 column. However, the isolated materials gave mass spectra indicating a mixture of products, primarily 2-fluoro-2'-deoxyinosine modified sequences without the cystamine modification. However, the chromatogram in Figure 4.5 (b) also contains a peak with R_t = 24.01 min. This peak was also isolated and gave the correct mass in ESI-MS (3923 gmol^{-1}) for the cystamine modified material. We suggested this material was the quadruplex due to the significant difference in retention time, but the secondary structure was disrupted in mass spectrometry and we observed the isolated product. Unfortunately the G-10 modification forms mostly a truncated sequence with R_t = 10.93 min Figure 4.5 (c)), which gave a mass of 3339.55 gmol^{-1}, indicating two missing nucleotides. This means the failure likely occurred during DNA synthesis, and the synthesis could be attempted again. A number of problems can arise during synthesis, from unexplained coupling failures. The modification of the G-4 (Figure 4.5 (d)) and G-2 produces numerous products which could not be isolated. Two sequences (G-10 and G-11 modification) are promising, and both are near the end of the sequence. DNA synthesis occurs from 3' to 5', meaning the semi-manual addition

occurs early in the protocol. The two less promising sequences (G-2 and G-4) both attempt semi-manual coupling near the end of DNA synthesis. None of these results means the modifications are impossible, but the modified phosphoramidite is an expensive material to use, so the successful G-11 modification was used to give an initial idea of the effect these modifications could have on biophysical properties. This also gave evidence that the amine substitution with cystamine was possible, and we decided to attempt to synthesise the same modification with longer linkers.

Longer amine chains with sulfur functional groups are not commercially available. The structures shown in Figure 4.6 were suggested as alternatives to cystamine which could be synthesised from 3-amino-1-propanol and 4-amino-1-butanol, both of which are available. Mitsunobu coupling can be used to convert the hydroxyl to a sulfur group, and the only other steps required protection and deprotection to prevent substitution of the amine. The synthesis was reported by Aufort *et al.* (2011)[38] with thioacetic acid, as well as several additional steps to convert the amine to secondary amines and an imine. We decided to adapt their method for the more available thiobenzoic acid and a primary amine salt.

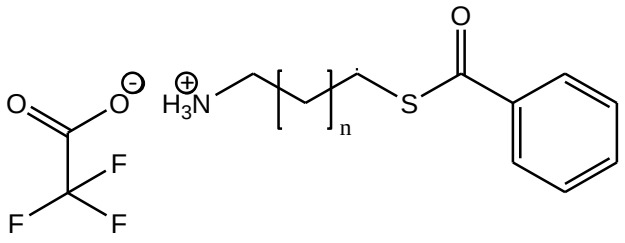


Figure 4.6 Amine salts used for initial synthesis of extended chain modifications of $G_4T_4G_4$. n = 1,2

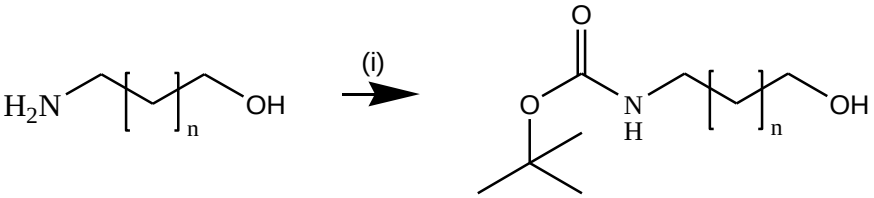


*Figure 4.7 Protection of amine with Boc_2O. Reagents and conditions: (i) di-*tert*-butyl dicarbonate (Boc_2O), DMAP, S: methanol. n = 1, 2.*

t-Butyl carbonate (Boc) was used to protect both amines (Figure 4.7), preventing the substitution of these groups in the subsequent Mitsunobu coupling. Boc protection was carried out in methanol with a slight (1.1 eq) excess of di-*tert*-butyl dicarbonate, with 5 mol% of 4-dimethylaminopyridine (DMAP) used as a catalyst. Both reactions were monitored with TLC, visualised with ninhydrin as the materials are not UV active. Ninhydrin reacts with amines, producing Ruheman's purple when heated. TLC analysis (20% MeOH in DCM) showed a small quantity of starting material remaining and additional Boc was added (2 x 0.1 equivalences) until starting material disappeared. In all future reactions a slightly larger quantity of Boc was used, as reported in Chapter 7. The products were extracted with ethyl acetate and washed with citric acid solution to remove any remaining reactants. The amine starting material becomes protonated, going to the aqueous phase, which helps to isolate the protected material. The organic solvent was evaporated *in vacuo* and both products were analysed by NMR. ^{1}H NMR spectra of the three carbon chain showed new peaks characteristic of the *t*-Butyl group at 1.43 ppm. HMQC NMR showed correlation between this peak and a ^{13}C peak with chemical shift of 28.3 ppm. In addition several new peaks appeared at 79.54 ppm and 157.13 ppm, corresponding to the quarternary carbons of the *t*-butyl and carbamide, respectively. The same pattern of peaks was observed for the four carbon material. These materials were used without any further purification.

The next reaction, Mitsonobu coupling (Figure 4.8 (a)), is sensitive to air, requiring careful formation of a phosphine intermediate under argon, but it offers the most effective protocol for substituting the alcohol with the sulfur modification. The mechanism of this reaction is shown in Figure

4.8 (b), which occurs in two steps. First, triphenyl phosphine (PPh_3) reacts with the oxidising agent diisopropyl azodicarboxylate (DIAD) to produce the positively charged phosphine intermediate. This is carried out in freshly distilled THF, dried over sodium and benzophenone, under an argon atmosphere at 0°C to prevent oxidation of the sensitive phosphine material. Over thirty minutes the

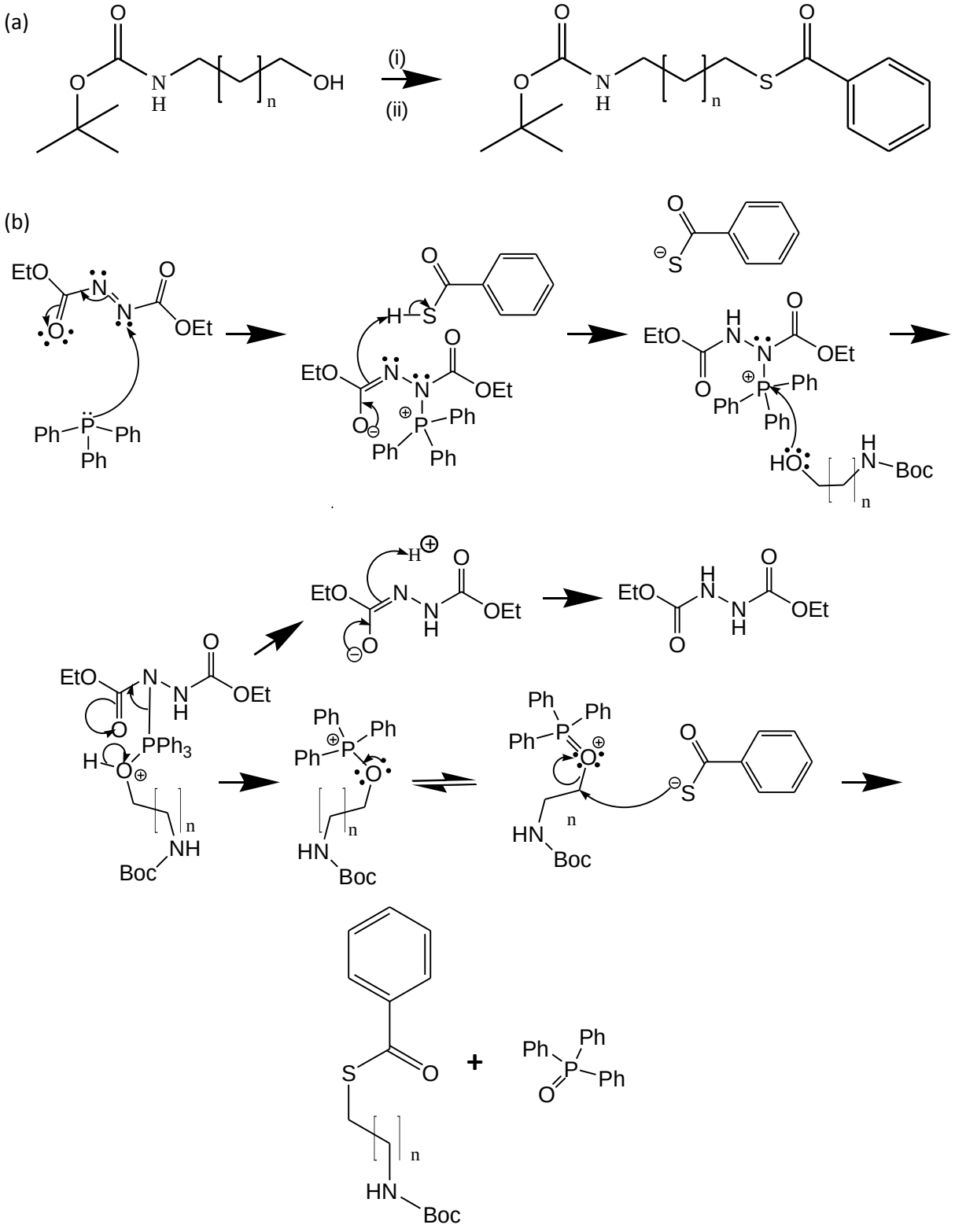


Figure 4.8 Mitsunobu coupling: (a) Scheme for Mitsunobu coupling of previously protected material with thiobenzoic acid. Reagent and conditions: S: dry THF(i) PPh_3, DIAD, 0°C, 30 mins, (ii) thiobenzoic acid, reactant, R.T. n = 1, 2. (b) Mechanism for this reaction including DIAD and PPh_3 reagents. n = 2, 3.

solution becomes viscous, as some intermediate precipitates. One of the Boc protected alcohols prepared in the previous step was added, followed by gradual addition of thiobenzoic acid. This resulted in disappearance of the precipitate and a rapid colour change from white to orange or green for the three and four carbon chains, respectively. The DIAD component of the phosphine intermediate deprotonates thiobenzoic acid. A nucleophile, such as the lone pairs of an alcohol or amine, attacks the electrophilic phosphorous. The Boc protecting group on the amine prevents nucleophilic attack by the amine, avoiding the formation of a mixture of products at this step. DIAD is a good leaving group, resulting in a new phosphine intermediate with the nucleophile. The hydroxyl group is deprotonated by DIAD and the $Ph_3P{=}O^+$-R intermediate is formed. The previously deprotonated acid can now attack the δ+ carbon due to the extremely electron withdrawing O^+. Triphenyl phosphine oxide is a good leaving group, allowing an S_n2 substitution to occur. The solvent was evaporated *in vacuo* and the residue was dissolved in toluene for purification by silica gel column chromatography. However, this resulted in the precipitation of triphenyl phosphine oxide, which was removed by filtration. The material was then purified by silica gel column chromatography with a 0-20% ethyl acetate/toluene gradient, which was determined through experimentation with various solvent systems by TLC analysis. These materials could be identified on ^{1}H NMR by the appearance of multiple aromatic peaks from 7.4 – 8.0 ppm, corresponding to ^{13}C peaks from 125 – 140 ppm. In addition, the quaternary thioester peak is visible at approximately 190 ppm in ^{13}C NMR, and the peaks closest to the new thiobenzoic group exhibit slight shifts as a result of a change in polarity between the alcohol and thioester functional groups.

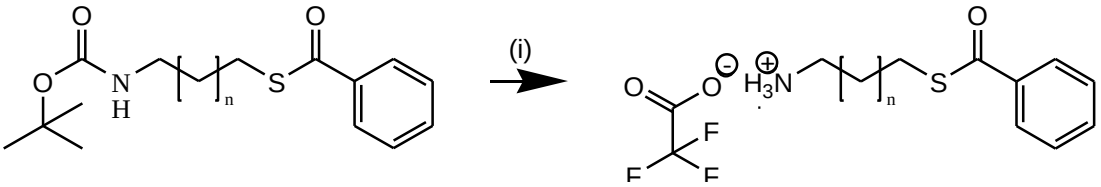


Figure 4.9 Cleavage of Boc protecting group and formation of TFA salt of benzoyl protected thio-amines. Reagents and conditions: 1:1 DCM:TFA. n =1,2.

The material was then treated with trifluoroacetic acid (TFA) to cleave the Boc protecting group (Figure 4.9). This was done by stirring in 50% TFA/DCM for two hours, after which the solvent was evaporated *in vacuo*. Several major changes indicated this procedure was successful. The NMR sample used D_2O as a solvent, as the salt is not soluble in $CDCl_3$. The peaks used to identify the Boc protecting group earlier were no longer present. Instead, a broad peak appeared at approximately 5.0 ppm, which did not correlate to any ^{13}C peaks. This suggests the presence of $^+NH_3$. In ^{13}C NMR, in addition to the disappearance of the characteristic Boc peaks, TFA^- has several identifying peaks. Fluoride (^{19}F) has a nuclear spin of ½, meaning it can be observed in NMR. The ^{13}C peak of CF_3 is split into a quartet, as is the adjacent COO^-, for the same reason we observe peak splitting in ^{1}H NMR from adjacent protons. Observing this splitting is one clear indicator of TFA's presence, and CF_3 appears as a quartet from 110 – 120 ppm, while COO^- appears as a quartet at approximately 160 ppm. In mass

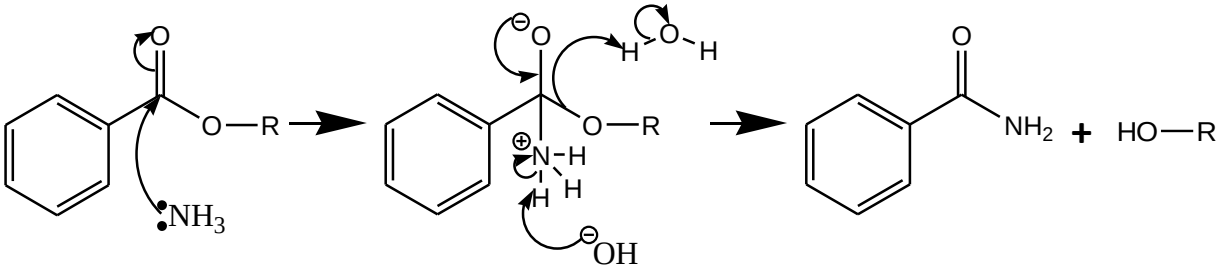


Figure 4.10 Mechanism for cleavage of Benzoate protecting group in aq. NH_3, resulting in benzamide and alcohol. Thioesters behave similarly in the presence of bases, theoretically resulting in free thiols.

spectrometry both TFA^- and the amine salt are visible, in the negative (approximately 112 a.m.u.) and positive mode (n=1: 195 a.m.u., n=2: 209 a.m.u.) respectively.

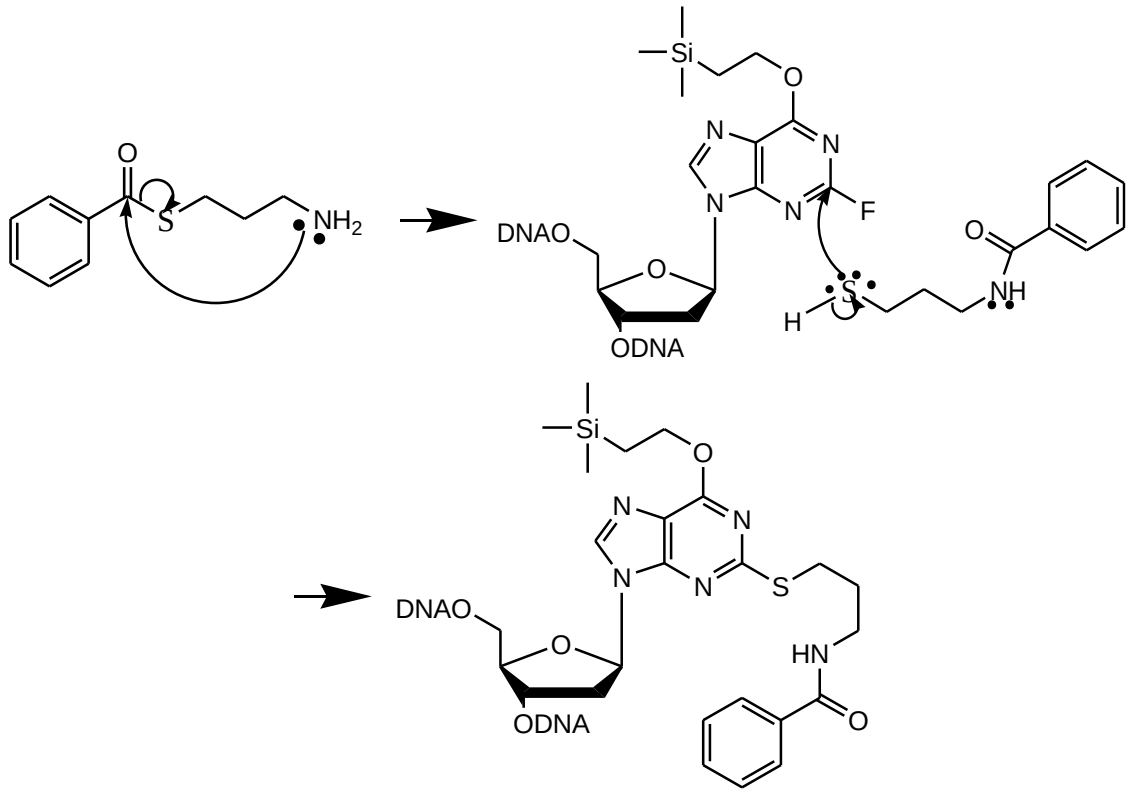


Figure 4.11 Proposed alternative mechanism for inosine substitution, similar to Figure 4.2 and Figure 4.10, except the lone pairs of sulfur attack the δ+ carbon. It is unclear if the benzoyl migration occurs earlier in the reaction. The mass of the oligonucleotide will be the same as the desired product, but the benzoyl amide is not as easily cleaved as thiobenzoyl.

The purified material was reacted with $G_4T_4G_2{}^{F}\underline{\mathbf{dI}}G$ sequences, using the same protocol established for cystamine. The exception is the addition of 10% Et_3N to the $0.5 molL^{-1}$ solution of reactant in DMSO before the reactant was added to the solid support. The solution was stirred briefly, then immediately added to the solid support. This was done to neutralise TFA, producing the free amine immediately before synthesis. Cystamine is stored as a dihydrochloride salt because its free amine is only stable for short periods of time. It is unknown if this is also true of these compounds, but we stored them as TFA salts until immediately before the reaction. Mass spectrometry was used to analyse the sequences containing both three carbon (3963.63 $gmol^{-1}$) and four carbon (3977.68 $gmol^{-1}$) linkers. Benzoyl protecting groups, such as the one on our thiol, are typically cleaved using basic conditions, as shown in Figure 4.10. This is typically achieved with KOH, but most bases should be acceptable, including NH_4OH. The mechanism above includes NH_3 as excess aqueous ammonia is used to cleave DNA from the solid support, and under these conditions the nitrogen lone pair of NH_3 should be able to attack this group. We expected that the benzoyl deprotection would occur at this step, but mass spectrometry indicated that benzoyl was still present. The nucleophile should attack the δ+ carbonyl to produce the free thiol, but we did not see this occurring. One possible explanation is migration of the benzoyl protecting group to the amine, substituting the thiol at the 2-position of guanosine instead, as shown in Figure 4.11. These steps may not happen concurrently, but the benzoyl group migration could happen spontaneously upon addition of Et_3N if the intramolecular reaction is favourable in basic conditions. The benzamide functional group would not be as easily removed by aqueous ammonia.

We attempted to circumvent these issues by using a different synthetic route, in which the Bz-protecting group would be removed and replaced by a preformed disulfide bridge (Figure 4.12) prior to DNA synthesis. Disulfides bridges would be cleaved as originally intended, using dithiothreitol (DTT), and after DTT is removed the bridges should reform between strands. The first two steps of the synthesis remained the same, but the benzoyl protecting group was then cleaved with aq. NH_4OH (Figure 4.13), with a similar mechanism as in Figure 4.10.

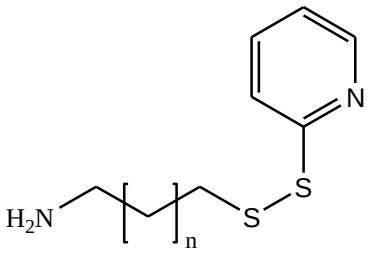


Figure 4.12 Proposed sulfur modifications. Bz-protecting group is removed and replaced with 2-mercapto pyridine which is easily removed with DTT. n = 1,2

Initially, aqueous ammonia was added to the synthesised material and stirred overnight (Figure 4.13). Solubility in aqueous conditions was limited, so in future reactions methanolic ammonia was used instead. This solvent was evaporated and the resulting oil was dissolved in DCM. We suggested that it may be possible to evaporate aqueous ammonia and form the disulfide bridge (Figure 4.15 (a)) before purifying this material. 2-Mercapto pyridine and *t*-butyl hydrogen peroxide (TBHP) were added, followed by N-iodosuccinimide (NIS) after approximately thirty seconds. Upon addition of NIS the solution heated rapidly and changed to a dark red colour. The mechanism for oxidation of thiols with NIS has not been reported, but it appears that some intermediate was formed initially using TBHP as an oxidant, and this intermediate reacted with NIS, to produce the desired disulfide bond. Similar mechanism with alcohols are described, but those are also unconfirmed.

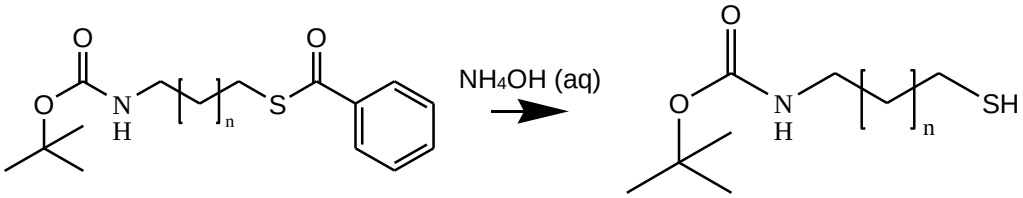


Figure 4.13 Removal of benzoyl protecting group using aq. NH_4OH, n = 1, 2.

We expected this would produce only the molecule seen in Figure 4.12, as TLC analysis (10% EtOAc:toluene) had only indicated the presence of a single material. However, we discovered the presence of a second material (Figure 4.14). The initial ^1H NMR spectra (Figure 4.15 (b)) showed two peaks at 2.83 ppm and 2.70 ppm with integrals of 1.1 and 0.9, respectively, rather than the single CH_2 peak we expected. ^{13}C NMR also showed twice as many peaks as expected, with pairs of peaks with slightly different chemical shifts. Mass spectrometry (Figure 4.15 (c)) was used to identify the products, indicating two molecules, with masses of 301.16 a.m.u. and 380.2 a.m.u., corresponding to the 2-mercapto pyridine coupled material (Figure 4.12) and a dimer of the starting material (Figure 4.14), respectively. In some cases, materials in the positive mode have an additional 23 a.m.u. due to formation of Na^+ adducts, which are seen in peaks at 323.22 a.m.u. and 403.19 a.m.u.. The peak visible at 245.11 a.m.u. was an impurity in

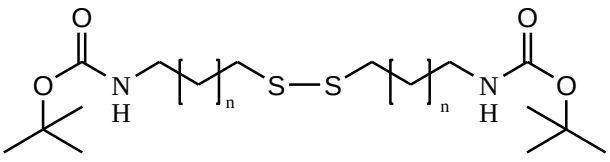


Figure 4.14 Dimeric products of N-iodo succinimide oxidation of previously synthesised thiols. n = 1, 2.

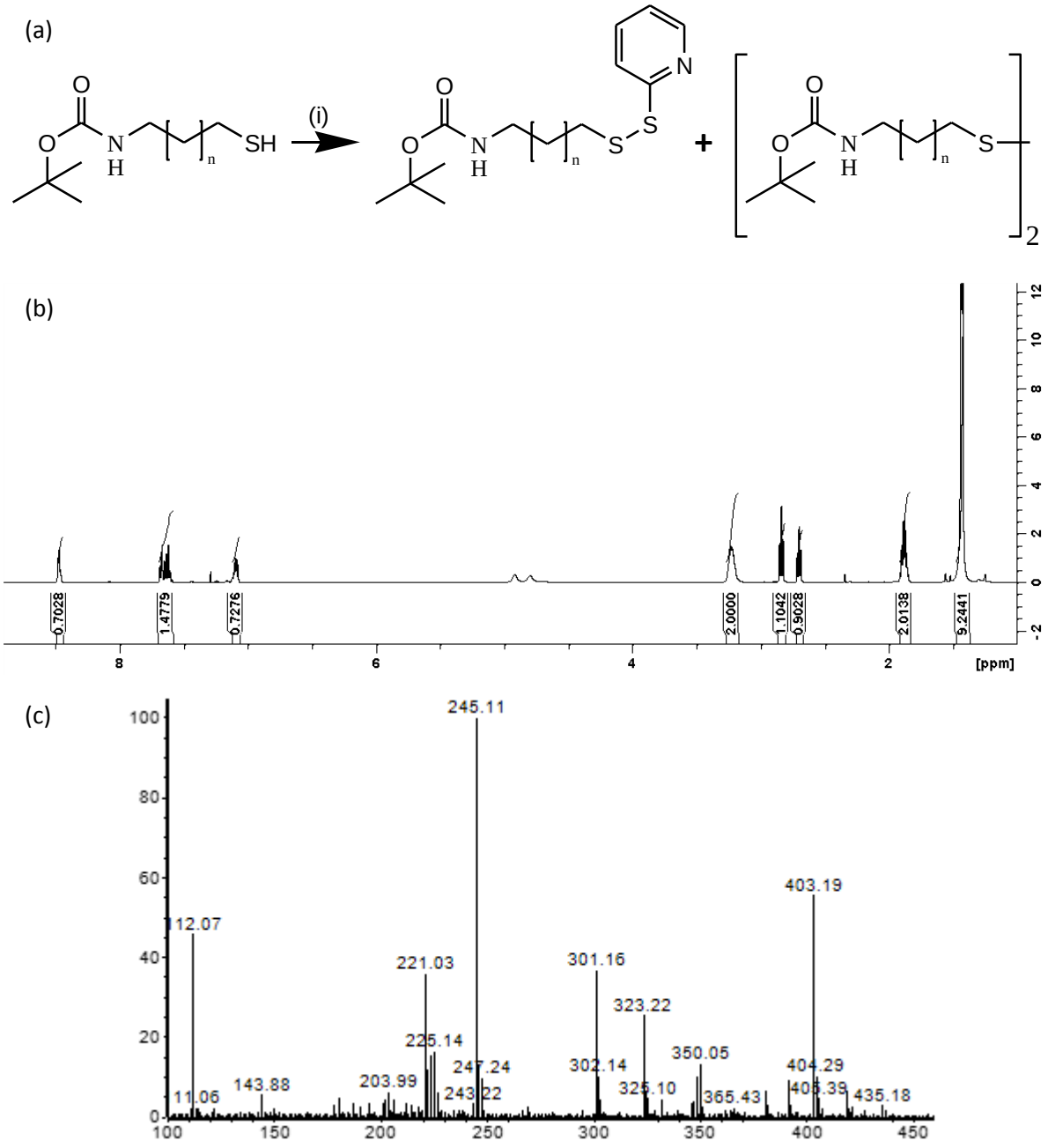


Figure 4.15 (a) Reaction of deprotected thiol with 2-mercapto pyridine to give a disulfide material to be incorporated into G4-DNA. Reagents and conditions: (i) TBHP, NIS, 2-mecapto pyridine, S: DCM. (b) ^{1}H NMR spectra of product from (a). Most of the spectra suggests a single material, except two peaks at approximately 2.9 ppm, which should only be one peak. (c) ESI-MS of product of (a) indicating the presence of two materials, with mass of 300.16 a.m.u. and 380.19 a.m.u.

the spectrometer, which consistently appeared while running blanks. In addition, a number of peaks corresponding to fragments and starting materials are visible in this spectra. ^{1}H NMR of each material was almost identical, except for the peak closest to the disulfide, and the presence of pyridine in one material. TLC analysis failed to indicate separation of these materials, suggesting their polarities were too similar to isolate by silica gel column chromatography.

We wanted to know when the dimer formed, to determine if it was possible to prevent this reaction, so the product of ammonia deprotection (Figure 4.13) was examined separately. TLC analysis (10% EtOAc in toluene) showed that during NH_4OH treatment, two products formed quickly. ^{1}H NMR of the crude material indicated similar peaks to those observed in previous experiments. This suggested that two products were formed by the deprotection: both the monomer and dimer. The

TLC indicated that these molecules had significant differences in polarity, meaning it was possible to distinguish the two materials by silica gel column chromatography. A 0-20% ethyl acetate in toluene gradient was used as an eluent and both the monomer and dimer were isolated. ^{1}H NMR spectra of these materials are almost identical, with the only notable difference being a slightly higher shift for the CH_2 protons adjacent to the disulfide bridge. This shows that the dimer formed during deprotection, but could be removed by column chromatography. It was unclear if it had also formed during disulfide formation with 2-mercapto pyridine.

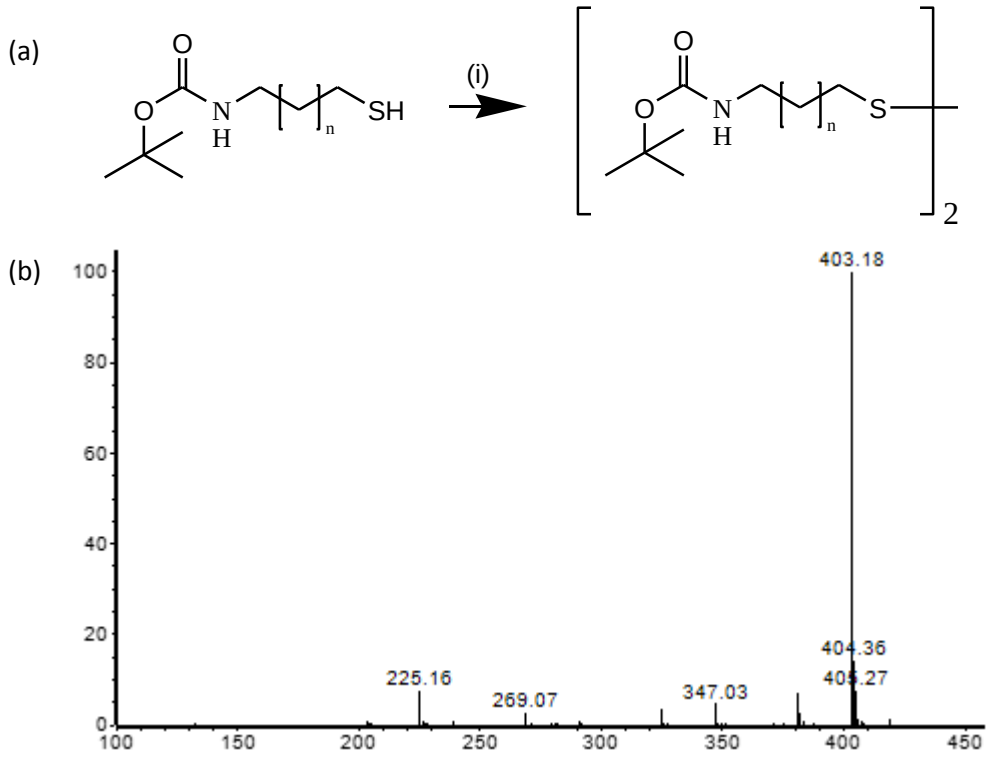


Figure 4.16 (a) Scheme for reaction of previously synthesised thiol t form dimer. Reagents and conditions: (i) TBHP, NIS, S: DCM. (b) Mass spectra of product of (a), purified by silica gel column chromatography (0-20% ethyl acetate/toluene). Expected Mass [M + Na^{+}] = 403.2 gmol^{-1}

To test this, a number of small scale experiments were carried out with the now pure monomer to investigate the dimer formation. 2-Mercapto pyridine, TBHP and NIS were added with the same conditions as before, but ^{1}H NMR spectra showed that both products formed in a similar ratio. This outcome made it clear that both steps of the reaction had resulted in some dimer formation.

We felt that the presence of oxygen could be enough to encourage disulfide formation so deprotection was attempted in an inert environment using argon. Under these conditions no reaction occurred. Upon exposure to air the reaction proceeded as previously observed, forming both products. We next decided to test if it was possible to fully oxidise the material to form only the dimer. TBHP was added to the monomer, and TLC analysis (10% EtOAc in toluene) showed two materials, the monomer and the dimer. When NIS was added the solution heated and changed colour, becoming red. TLC analysis (10% EtOAc in toluene) showed complete conversion to the dimer (Figure 4.16 (a)). This suggested that in the absence of 2-mercapto pyridine the dimer would still be formed. This was confirmed using both NMR and with mass spectrometry which both showed only a single material. Mass spectrometry showed a single peak at 403.18 a.m.u. (Figure 4.16 (b)), very similar to the peak seen in Figure 4.15 (c). ^{1}H, ^{13}C, COSY and HMQC NMR spectra for this material is similarly identical to the peaks seen in previous reactions, with only a single CH_2 peak at 2.72 ppm in ^{1}H NMR. Conversely,

when the equivalence of 2-mercapto pyridine was doubled, no effect on the ratio of products was observed. This suggested that the dimer formation was highly favourable, as it occurred spontaneously following deprotection, and was unaffected by a large excess of 2-mercaptopyridine in NIS oxidation. These experiments implied that the synthesis of the pure 2-merapto pyridine product shown in Figure 4.12 would not be possible under these conditions due to the more favourable formation of the dimer.

However, the dimer could fill the same role as the 2-mercapto pyridine material. It included a disulfide bridge rather than a benzoyl protecting group and contained the amine required for substitution. The main reason it was not been considered originally was that the molar quantity was halved by dimerization, meaning more starting material was consumed by this process. After observing the difficulties with separating the two products however, we decided to proceed with modification using this material instead.

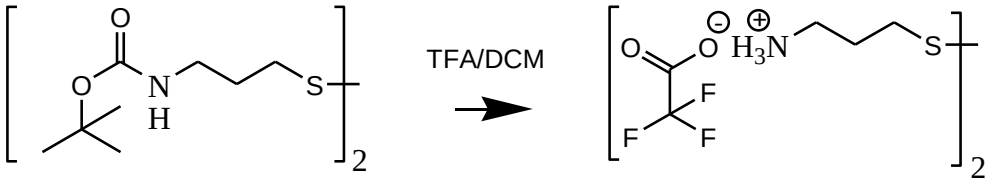


Figure 4.17 Scheme for Boc deprotection of new disulfide linkers.

To remove the Boc protecting group, the dimer was treated with a 50% trifluoroacetic acid (TFA)/DCM mixture, and the solvent was removed *in vacuo*. The changes observed in NMR were also similar: disappearance of characteristic Boc peaks, a broad $^{+}NH_3$ peak, and the characteristic TFA peaks in ^{13}C NMR.

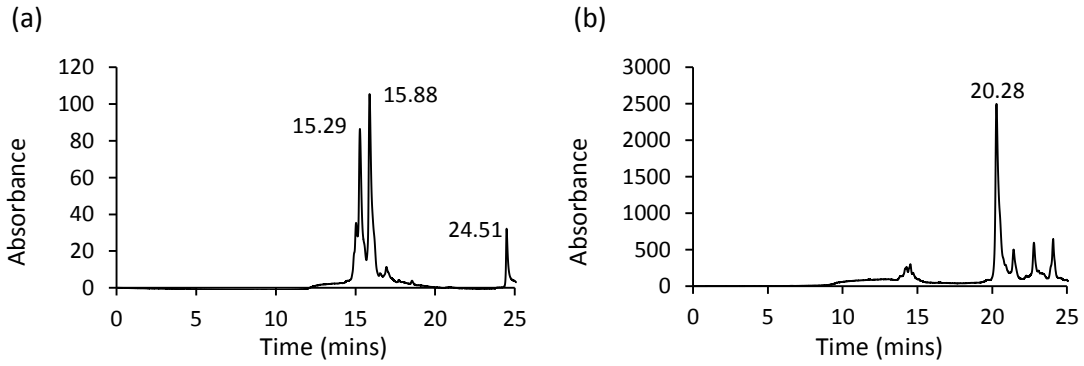


*Figure 4.18 RP-HPLC Chromatograms of (a) $G_4T_4G_2$**^{3C}dI**G and (b) $G_4T_4G_2$**^{4C}dI**G. 100 mmolL^{-1} TEAA buffer, pH 7.0 in acetonitrile: 0 – 25%, 2 - 20 mins, 25 – 80%, 20 – 25 mins.*

The TFA salts of both three and four carbon linkers were reacted with $G_4T_4G_2$**^{F}dI**G using the same protocol described for benzoyl protected TFA salts above. As before, the solution was shaken for several days to continuously disturb the solid support. After NH_4OH deprotection and removal of silyl protecting groups, RP-HPLC was used to purify the resulting materials (Figure 4.18). The material collected from the three carbon linker (Figure 4.18 (a)) with r_t = 15.88 min gave the correct mass, 3949.66 a.m.u., while the other peaks contained unmodified sequences (r_t = 15.29) and a mixture of sequences (r_t =24.51). This mixture included some modified material, but was mostly composed of several truncated materials. The four carbon linker had one major product (Figure 4.18 (b)), which we speculate is a quadruplex based on the r_t of 20.28 min. Mass spectrometry indicated this material ws the correct sequence but with no modification, with a mass of 3789.6 a.m.u.. This means the

substitution reaction did not take place, and we suggest that the aq. NH_4OH conditions used to cleave the sequence from the solid support was sufficient to aminate ^{F}dI, giving the mass of unmodified $G_4T_4G_4$. This reaction may require further experimentation to obtain a useful sequence.

4.3 Results and Discussion

Unmodified and the two successfully modified sequences were dissolved in sodium phosphate buffer (0.1 $molL^{-1}$ sodium phosphate buffer, pH 7.0, 1 $mmolL^{-1}$ EDTA, 0.1 $molL^{-1}$ NaCl). All samples were then heated at 90°C for 5 mins, before cooling to room temperature, and annealing at 4°C overnight. The samples were analysed with circular dichroism from 220 – 350 nm. Modified sequences exhibit the same key peaks as unmodified $(G_4T_4G_4)_2$ as shown in Figure 4.19, which suggests the structures form anti-parallel quadruplexes, as expected. The CD spectra for $G_4T_4G_4$ is the same as reported in literature.[24] Melting profiles were obtained by heating the solution from 5°C to 90°C in 5°C steps and recording a CD spectra at each temperature. Samples were also cooled at the same rate and hysteresis was observed. Melting curves were obtained and corrected using the method described in Chapter 2. The $T_{½}$ value, as listed in Figure 4.20 (d) and indicated in melting curves in Figure 4.20 (a-c), is the temperature at which half of the structure is unfolded, which is used as a point of comparison because hysteresis occurring means these are non-equilibrium denaturing/annealing profiles.

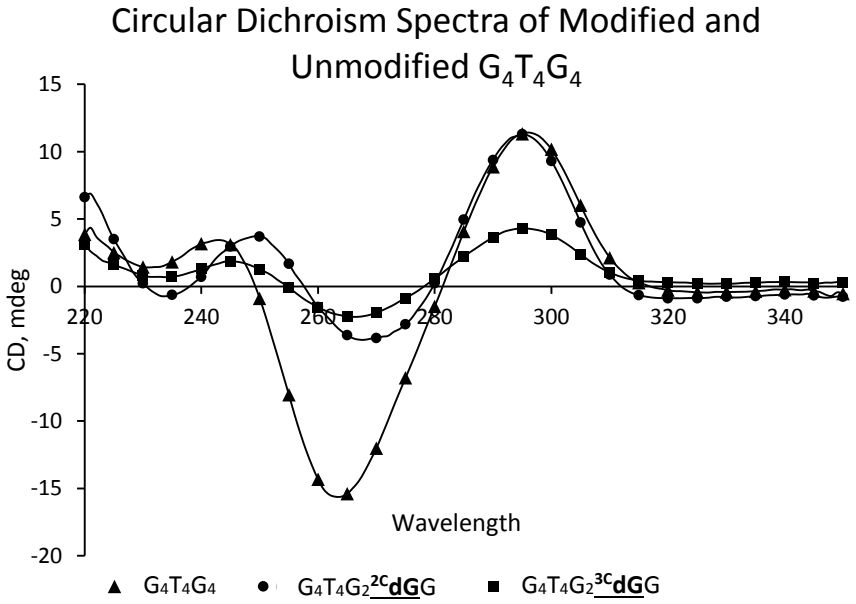
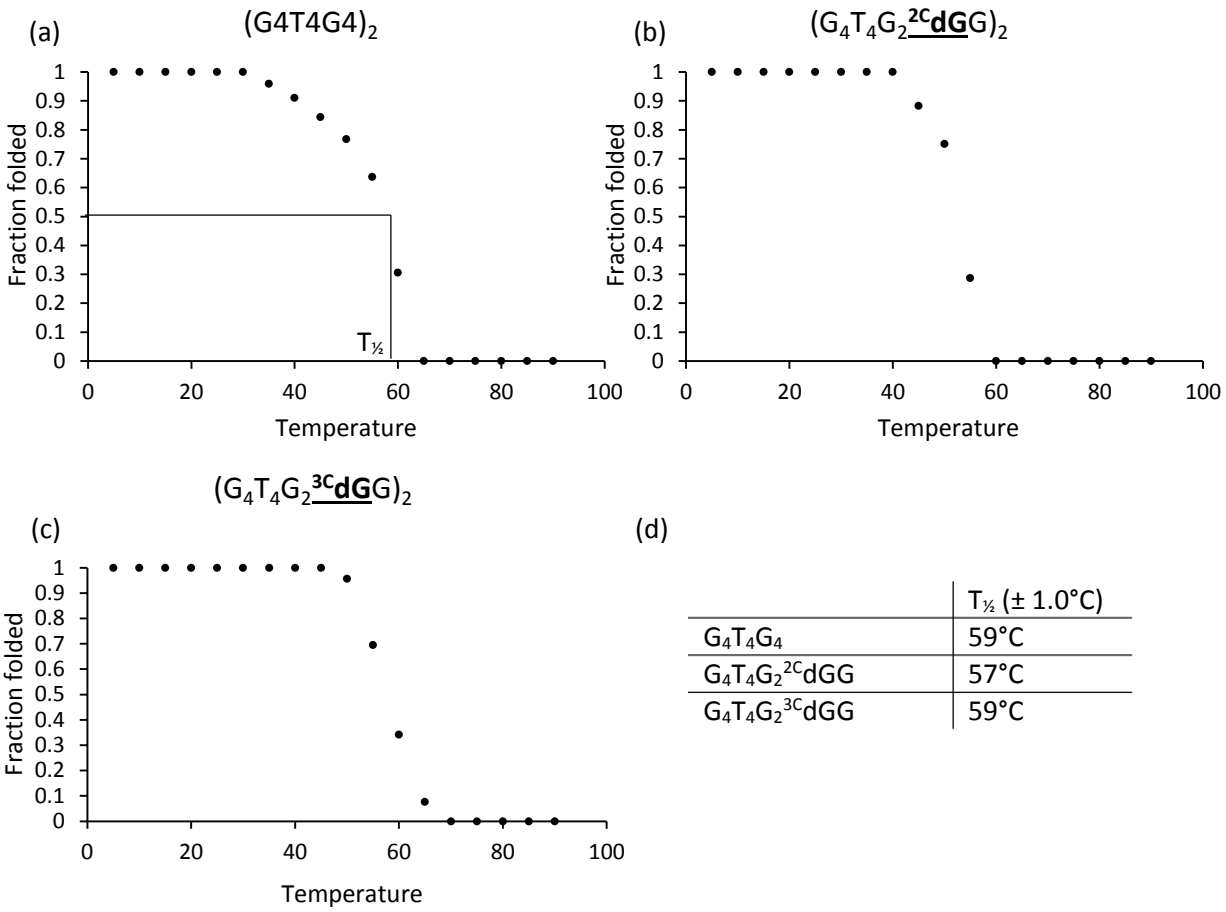


Figure 4.19 CD profiles of unmodified and modified $G_4T_4G_4$ sequences, with characteristic peaks at approximately 295 nm and 260 nm. 0.1 $molL^{-1}$ sodium phosphate buffer, 1 $mmolL^{-1}$ EDTA, 0.1 $molL^{-1}$ NaCl, pH 7.0.

No significant difference in $T_{½}$ was observed between the modified and unmodified sequences. This suggested thiol linkages might not be formed, or did not influence the thermal stability. However, the structures did not show significant destabilisation, meaning the introduction of our modification also did not have a negative impact on the melting or annealing of our G4-DNA. The approach showed promise as the modification reaction was viable, and the resulting structures were still able to form stable quadruplexes. We needed to determine if cross-links could $G_4T_4G_2{}^{3C}\mathbf{G}G$



(d)

	$T_{½}$ (± 1.0°C)
$G_4T_4G_4$	59°C
$G_4T_4G_2{}^{2C}dGG$	57°C
$G_4T_4G_2{}^{3C}dGG$	59°C

*Figure 4.20 Melting curves for (a) $G_4T_4G_4$ and modified $G_4T_4G_2$**R dG**G at 262 nm, with R = (b) Cystamine, (c) 3-aminopropan-1-thiol. (d) Table showing $T_{½}$ values (± 1.0°C) for each complex. 40 µmolL^{-1} DNA, 0.1 molL^{-1} sodium phosphate buffer, pH 7.0, 1 mmolL^{-1} EDTA, 0.1 molL^{-1} NaCl.*

be formed in either of these complexes, so we attempted several oxidising techniques described in literature with the 3-aminopropan-1-thiol modified sequence.

Some literature[39-42] reported that disulfide bond formation in synthetic protein structures had been encouraged with oxidising agents. The first instance of this we found was the synthesis of human insulin, where the correct formation of the secondary structure was controlled by successive deprotection of thiol groups and linking them by treatment with I_2[39]. Since then, numerous sources discuss different oxidation techniques for controlled disulfide bridge formation. It seemed reasonable that these methods could be translated to DNA cross-links.

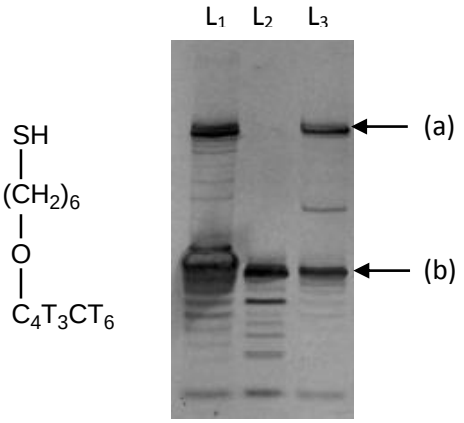


Figure 4.21 Denaturing gel image obtained with this strand, showing the anticipated outcome of thiol formation. L_1 = Untreated strand. L_2 = Strand treated with DTT. L_3 = Strand treated with DTT and left overnight in H_2O, exposed to air. (a) Dimer; (b) monomer.

Previous experiments (Figure 4.21) had shown, using denaturing gel electrophoresis, that sequences modified at the 5'-end formed disulfide bridges. L_1 shows the untreated material with two major bands indicating the monomeric and dimeric products. Addition of DTT resulted in complete disappearance of the dimer, as shown in L_2. Finally, when DTT is removed the dimer forms again,

as shown in L_3. This experiment intended to show a complex formation with a protein, which can be seen as a faint central band in L_3, but the major product was the dimer of two strands. While this is not a quadruplex structure, it demonstrates what we expect to observe with this type of modification. One possible difference between these experiments is the position of the modification. While our modifications are internal, the modifications shown in Figure 4.21 are only at the ends of sequences. This effect was observed with no oxidant present, and the solutions were not degassed.

Our $G_4T_4G_4$ and $G_4T_4G_2$**^{3C}dG**G sequences were run on denaturing gels containing 7M urea with a T_5 to T_{50} ladder to see if the same result could be observed, indicating formation of the disulfide bridges. We expected to observe two bands, corresponding to approximately T_{12} and T_{24} on the ladder, corresponding to the monomeric DNA and the complex, respectively. It should be noted that the gel does not contain any reducing agents, such as DTT, so disulfide bridges will be preserved.

With no additional oxidant, samples were dissolved in phosphate buffer and some were heated and annealed as previously described. The formation of thiol linkages requires interaction between modifications on different strands, so we felt that we might see cross-linking if the quadruplex structure was formed first. Initially the introduction of phosphate buffer caused the DNA

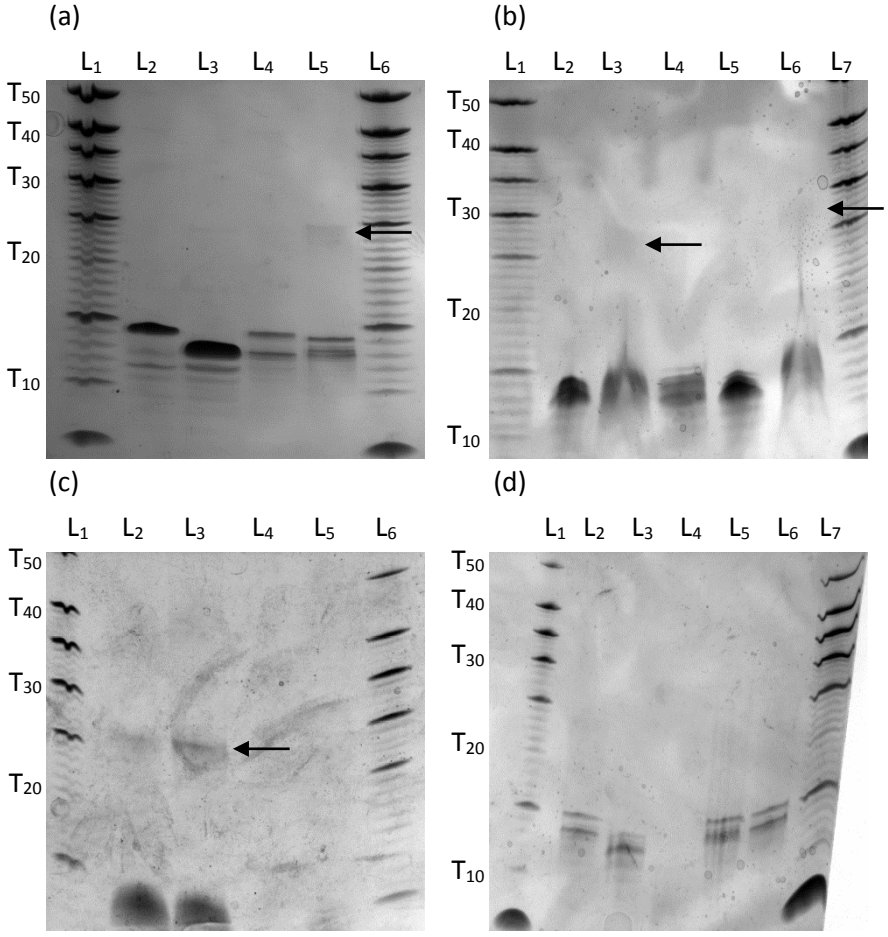


*Figure 4.22 Denaturing gel of modified and unmodified $G_4T_4G_4$. Ladder: T_5, T_{10}, T_{15}, T_{20}, T_{25}, T_{30}, T_{35}, T_{40}, T_{50} sequences. Gel: 20% Acrylamide, 7 molL^{-1} Urea. (a) L_1, L_6 = $T_5 – T_{50}$ Ladder, L_2 = $G_4T_4G_4$ unheated, L_3 = $G_4T_4G_4$ heated L_4 = $G_4T_4G_2$**^{3C}dG**G unheated, L_5 =$G_4T_4G_2$**^{3C}dG**G heated (b) L_1, L_7 = $T_5 – T_{50}$ Ladder, $L_2 – L_6$ = $G_4T_4G_2$**^{3C}dG**G treated with: L_2 = I_2, L_3 = NIS, L_4 = H_2O_2, L_5 = I_2 + H_2O_2, L_6 = NIS + H_2O_2. (c) L_1, L_6 = $T_5 – T_{50}$ Ladder, $L_2 – L_5$ = $G_4T_4G_2$**^{3C}dG**G treated with NIS: L_2 = 0.05M, L_3 = 0.1M, L_4 = 0.2M, L_5 = 0.5M. (d) L_1, L_7 = $T_5 – T_{50}$ Ladder, $L_2 – L_6$ = $G_4T_4G_2$**^{3C}dG**G conditions:: L_2 = pH 7.5, L_3 = pH 7.5 + I_2, L_4 = pH 7.5 + NIS, L_5 = pH 6.0, potassium ferricyanide, L_6: 1% DMSO.*

bands to be distorted, so in future the samples were all precipitated with $LiClO_4$/acetone before treatment and loading to remove the salts. All of these samples, both modified and unmodified, were seen to be approximately equivalent to a 12-mer (Figure 4.22 (a)). A band, possibly representing the longer sequence, is faintly visible in the Lane 5 as indicated, which is approximately equivalent to T_{24}. These bands were more clearly visible on the stained gel, but not a clear in the picture. However, this is still a minor product. This seemed to suggest the dimer could be forming when the secondary structure was folded, but the majority of the material is monomeric. This result led to further experiments to encourage the formation of this product by treating samples with commonly used oxidising agents.[39-42]

Hydrogen peroxide (H_2O_2), iodine (I_2) and N-iodo succinimide (NIS) were the first methods tested. Reactions were run over night before being treated with Urea and loaded onto a denaturing gel. Samples treated with H_2O_2 and I_2 showed only monomeric products, but sample treated with NIS showed significant distortion of bands for short sequences, as well as similar bands coinciding with approximately T_{24} sequences (Figure 4.22 (b)). It seemed like NIS could be encouraging disulfide formation so we varied the concentration in an attempt to find optimal conditions for disulfide formation. Figure 4.22 (c) shows lower concentrations (0.05 $molL^{-1}$ and 0.1 $molL^{-1}$) resulted in some conversion, but a significant amount of DNA remained monomeric. However, 0.2 and 0.5 $molL^{-1}$ did not show any converted or monomeric product. These samples were run several times, but no DNA was ever observed in these conditions, suggesting the conditions were too harsh and sequences could be degraded. If the longer sequences observed were disulfide containing dimers, it appeared that disulfide formation could be encouraged through treatment with NIS. We decided to try new conditions suggested by other literature.[41] Figure 4.22 (d) shows the results of these alternatives, none of which showed any dimerization. In lane 2-4 higher pH was used by dissolving DNA in pH 7.5 phosphate buffer, and both I_2 and NIS were added, like in Figure 4.22 (b), but this yielded no observable conversion. 0.01 $molL^{-1}$ potassium ferricyanide in pH 6.0 phosphate buffer showed no conversion, nor did addition of 1% DMSO, as seen in lane 5 and 6, respectively. So far very few methods had indicated the formation of disulfide bridges, and no methods with the expected reliability and high conversion.

The lack of success of these procedures suggests two possible issues:

(i) The modifications are incorrect despite the sequences showing the expected mass of the correct material.
(ii) Disulfide bridge formation is not possible under any of the conditions tested so far, and further conditions need to be found.

The same reaction mixtures used for denaturing gels were purified with reverse phase HPLC to see if distinct peaks were observable for monomeric and dimeric sequences, potentially allowing for further mass spectrometry analysis to confirm disulfide formation. However, the resulting HPLC was unhelpful because peaks were not clearly differentiable using this technique. This shows that if dimeric materials were formed it was only present in a small quantity.

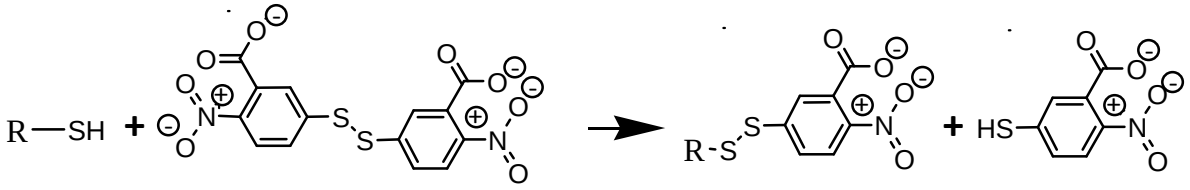


Figure 4.23 Reaction of Ellman's Reagent (DTNB) with free thiol to form disulfide with thiol and TNB, which is deprotonated in basic conditions to TNB^{2-}, resulting in measurable absorbance at 412 nm (pH 7.0).

One useful procedure to test for the presence of free thiols referenced in the literature was the Ellman test. The Ellman reagent, the disulfide starting material shown in Figure 4.23. The reaction of the reagent with free thiols results in the formation of 2-nitro-5-thiobenzoate (TNB) and a disulfide compound of TNB and the thiol (Figure 4.23). TNB is deprotonated to the TNB^{2-} anion in neutral and basic conditions, which has a reported absorbance maxima at 412 nm, and molar absorptivity coefficient of 13,700 $Lmol^{-1}cm^{-1}$. Monitoring the increase in absorbance at this wavelength is a useful method of determining the concentration of any thiols present. The test is carried out in pH 8.0 phosphate buffer with 100mM NaCl and 1mM EDTA.

A standard curve was set with L-cysteine hydrochloride monohydrate from 25-250 µM, testing for a DNA concentration of approximately 100µM. This test aimed to determine if the modified DNA

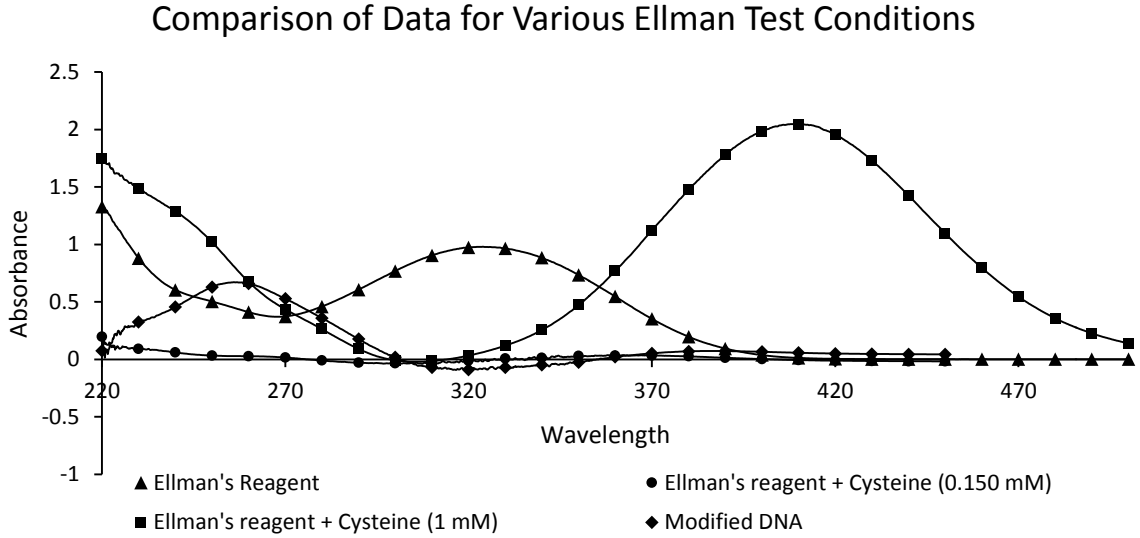


Figure 4.24 Comparison of peaks showing disappearance of Ellman's reagent and appearance of TNB^{2-}. These peaks are not visible in 0.15 mM Cysteine solution, which is comparable to the concentration of DNA seen here. Modified DNA sample did not change upon addition of Ellman's Reagent.

contained thiols capable of reaction with Ellman's reagent, or if the reason for a lack of success previously was due to the low reactivity of the linkers. The expected outcome was a gradual increase in absorbance at 412 nm seen when L-Cysteine hydrochloride reacts with Ellman's reagent. However, the difference within the range was insignificant compared to the background, while also having a large margin of error. The suggested procedure for Ellman's test gave a working range of 0.1 – 2.0 mM. With our DNA having an expected concentration of approximately 0.1 mM, any standard curve will be in the low end of this range.

To ensure this was the issue several further reactions were carried out to verify the functionality of these reagents. Addition of Ellman's reagent to a 1 mM sample of L-Cysteine, as shown in Figure 4.24, showed a significant shift in the peak from approximately 320 nm (■) to 412 nm (▲), as expected. Further addition of Ellman's reagent showed no change in absorbance, and the concentration given by this data was correct, showing the test and reagents were reliable. Figure 4.24

also shows UV spectra of 0.1 mM L-Cysteine and modified DNA at the same concentration. λ_{max} of DNA samples (◆) did not change upon addition of Ellman's reagent. The increase in absorbance at 412 nm for these samples when Ellman's reagent is added is far too small to be measurable, and is indecipherable from the background. This test could have been an extremely useful way of detecting free thiols. However, the low quantity of DNA produced by this synthetic approach makes it undetectable with a reagent with a low molar absorptivity coefficient. In future, if we have higher concentration samples this techniques could be used to verify the presence of free thiols on DNA.

4.4 Conclusion

So far this strategy has demonstrated and reported the synthesis of several small molecules which could be used to modify 2-fluoro inosine. We then also demonstrated the synthesis of unique sequences containing these modifications. Finally, we attempted to form chemical cross-links between the modified sequences and analyse the resulting change in biophysical properties these modifications induced. However, the result of this modification was discouraging. While our chemical modifications appear to have been incorporated as expected, we could not produce evidence of the formation of cross-links, and no significant change in $T_{½}$ was observed. Modifications of this type did not have a significant negative impact on biophysical properties, but it is also clear that more work is necessary to obtain useful modified structures.

5. Biophysical Evaluation of Tetramolecular G-quadruplexes Modified with (2R,3S)-N-(4'-pyridinylmethyl)-3-hydroxy-2-pyrrolidinemethanol

5.1 Background

TG_4T is a DNA sequence often used for analysis of modified G-quadruplexes. The native sequence forms a tetrameric quadruplex with well documented properties. The sequence is short, allowing for reliable synthesis with high yields. However, it does have two main disadvantages as a model. First, tetrameric quadruplexes form slowly, so experiments can take longer than formation of monomeric or dimeric sequences complexes. Secondly, most native DNA is monomeric, meaning that results obtained with modified TG_4T sequences usually need to be demonstrated with other biologically relevant sequences.

Several modification for tetrameric quadruplexes have been investigated and reviewed in the past.[43] From the perspective of cross-links the modification was primarily the addition of linkers to the

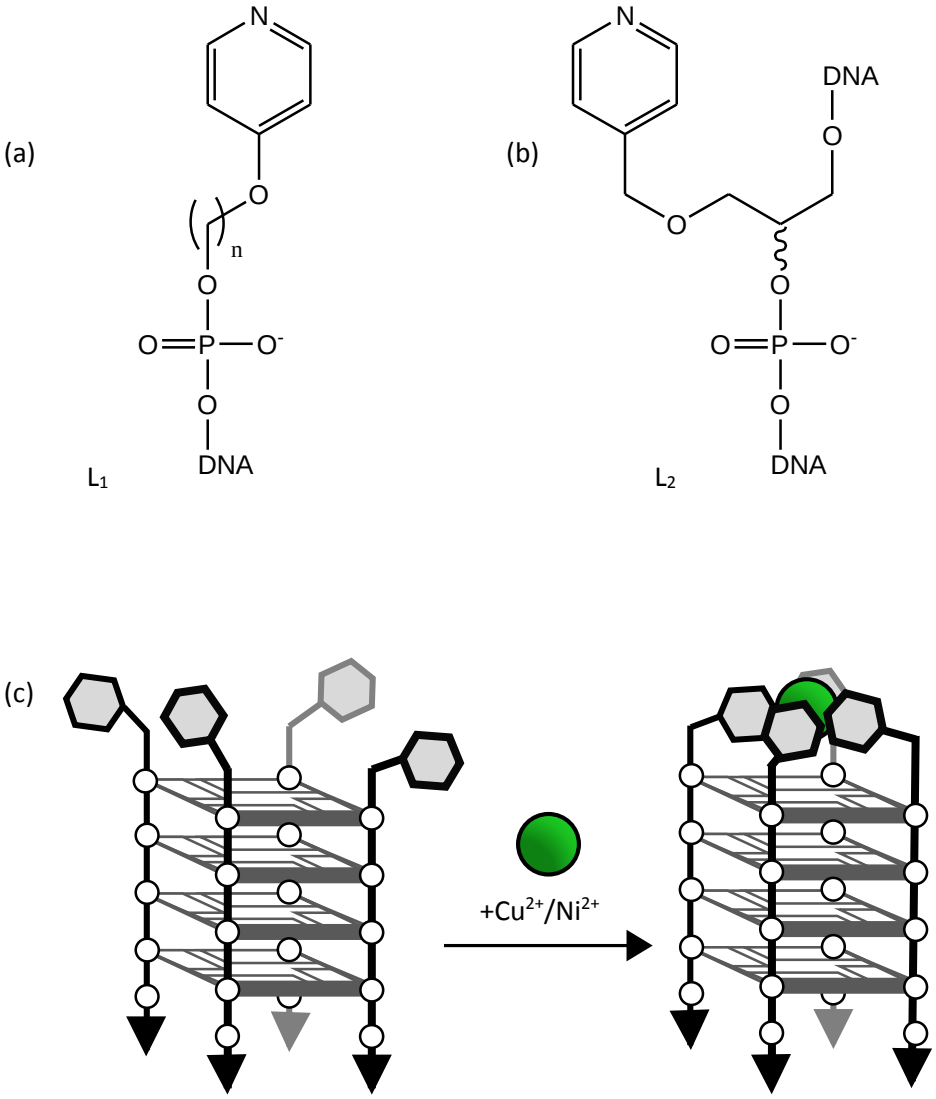


Figure 5.1 a) Variable linker (L_1) used in tetrameric XG4 sequences by Engelhard *et al.* (2018).[18] n = 1-4. (b) Linker (L_2) used in Engelhard *et al.* (2017)[17] for internal incorporation in monomeric G4 sequences. (c) Example of *XG_4T sequences modified with L_1 folding into tetrameric parallel quadruplexes (left) and pyridine ligands coordinating with cations (right) to give DNA-metal*

end of sequences with ligands such as pyridine, as shown in Figure 5.1, which form complexes with Cu^{2+} or Ni^{2+}. One example[19] synthesised XG_4 sequences with L_1 modifications at the 5' end, as shown in Figure 5.1 (a). The linkers ranged from 1 to 4 carbon atoms. Modified complexes had lower thermal stability than native species. However, Engelhard *et al.* (2018)[19] demonstrated a relationship between the linker length and thermal stability, with longer linkers giving higher $T_{½}$ values relative to shorter linkers. This is attributed to longer chains giving increased flexibility, which gives stronger π-π stacking interactions with neighbouring G-tetrads. Engelhard *et al.*[19] further observed that upon copper(II) addition the $T_{½}$ increased. The longest linker (n =4) gave an increase of only +3.9°C, but the shortest linker (n =1) gave an increase of +25.1°. It was proposed that while high flexibility improved the effect of π-π stacking of G-tetrads, the lower entropy and steric strain of short ligands increased stability when coordinating to the metal centre. Engelhard *et al.* (2013)[20] also demonstrated that the process was reversible by addition of ethylenediaminetetraacetic acid (edta), which removes copper(II) from chelation with pyridine.

Previous modifications had placed ligands within the loops of G4 structures, and the smallest change in T_m was less than 1°C, with copper(II) addition making the structure more stable than native DNA. Next, L_2 linkers (Figure 5.1 (b)) were used to replace a G-tetrad, which resulted in significantly destabilised quadruplexes (e.g. T_m = 63.9°C for unmodified G4-DNA to T_m = 18.1°C for modified G4-DNA). Introduction of copper(II) improved these values (18.1°C to 33.0°C), although they were still substantially less stable than native assemblies. This is potentially very useful, as it indicated the ability to control the formation of these structure with Cu^{2+} addition. The ttel24 sequence forms a hybrid structure under native conditions, meaning multiple G4 topologies are present. With the same modification one G-tetrad $(TTGGGL_2TT\ L_2GGG)_2$ the thermal stability was again decreased. However, upon addition of Cu^{2+} characteristic CD peaks of various conformations decreased, while peaks indicating a particular antiparallel G-quadruplex conformation with three lateral loops increased. This demonstrated that ligand coordination could be used to initiate topological change if metal coordination favoured one topology over another.

Another study[44] carried out similar experiments also indicating that topological control could be achieved with metal chelation in G4 structures. They used the common bimolecular quadruplex forming sequence, $(G_4T_4G_4)_2$, which was discussed in previous chapters. All four thymidines in the loop were replaced with a linker containing a 2,2'-bipyridine ligand. It was observed that addition of salts, such as $MgCl_2$, $CaCl_2$, $ZnCl_2$, $CoCl_2$ and $NiCl_2$ resulted in reduced intensity of the CD signal characteristic of antiparallel quadruplexes. In the case of some metals, such as nickel(II) and cobalt(II), the CD spectra suggested formation of a parallel quadruplex. This result was supported by native gel electrophoresis, wherein addition of metal cations blurred the band corresponding to an antiparallel quadruplex, and in some cases showed almost complete conversion to the parallel conformation. It is suggested that the high affinity of ligands for metal atoms means artificial base pairs are favoured over native base pairs, which in the case of 2,2'-bipyridine modified $G_4T_4G_4$ was sufficient to encourage the formation of an atypical quadruplex topology.[45]

Based on the success of these modifications, we developed a strategy to incorporate a pyridine moiety, similar to Engelhard *et al.* (2018).[19] We hypothesised that this modification, shown in Figure 5.2 (a), could lead to quadruplex structures with increased stability. Our modification more closely mimics the typical DNA structure due to the inclusion of a sugar. The previous modifications had shown significant decreases in thermal stability when structurally essential nucleobases were replaced with modified nucleobases. We hoped that by more closely matching the shape and size of unmodified nucleosides our secondary structures would not be distorted, giving more stable analogues of native structures. The previous studies also examined only melting behaviour of modified sequences. While thermal stability is an important property of G4 structures, kinetic properties determine if structures form rapidly upon addition of divalent cations to act as a switch for G-quadruplex folding. Examination of kinetic properties gives the rate of formation, which we expect will be faster than for native G4-DNA.

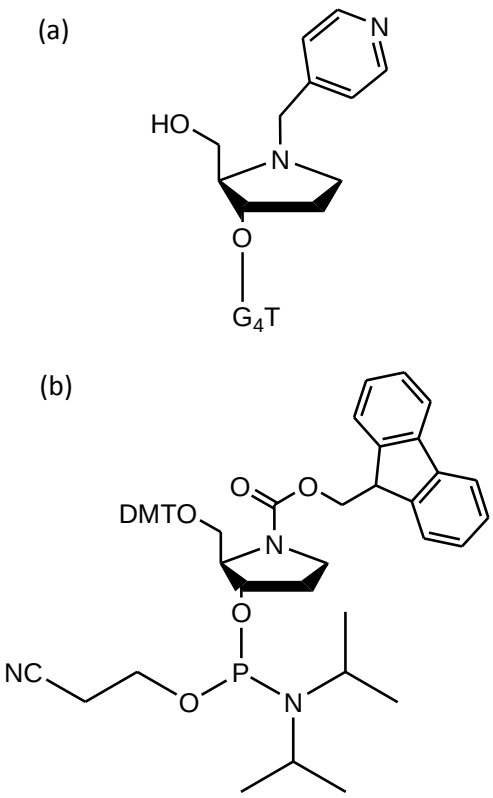


Figure 5.2 (a) Target modified sequence with 4-pyridine containing ligand; (b) Pyrrolidine-CE Phosphoramidite used in modified DNA sequences for ligand introduction

Pyrrolidine-CE Phosphoramidite (Figure 5.2 (b)) was chosen as the starting material for introduction of chelating ligands because it is commercially available and is easily modified with various aldehydes.[46] Previous uses of this protocol used it as a transition state analogue, meaning this would be the first application for cross-linking. After synthesising sequences with the modification shown in Figure 5.2 (b), the Fmoc-protecting group can be removed and modified sequences can be obtained by reductive amination. The use of different ligands will depend on the potential shown by substitution with 4-pyridine carbaldehyde, which was an easily obtainable reagent to test this type of modification. The first modification attempted was XG_4T, which incorporates the phosphoramidite at the 5'-end in TG_4T instead of thymidine. TG_4T is used as a starting point to determine if the modification is efficient and offer the expected chemical properties before attempting more complex procedures. The 5'-position was chosen for simplicity, as DNA synthesis typically uses preloaded CPG, making 3'-modifications more difficult. We also did not want to attempt to replace guanosine at this stage.

5.2 Synthesis

XG_4T (X = pyrolidine phosphoramidite) was synthesised on a 5 µmol scale, using semi-manual addition to couple the Fmoc-pyrollidine phosphoramidite, as described in Chapter 2. The DNA was treated by gradually passing several solutions through the column while oligonucleotides were still bound to the solid support. First, the cyanoethyl protecting group was cleaved with 1 mL of 10% diethylamine/acetonitrile solution over 10 min. These protecting groups are normally removed when DNA is cleaved from the solid support, but was removed here instead to prevent side reactions in later steps. Next, the fluorenylmethyloxycarbonyl (Fmoc) protecting group was cleaved by a 20% piperidine/DMF solution (Figure 5.3). Piperidine acts as a nucleophile, removing an acidic proton. This carbanion intermediate is stable because the resulting system is highly resonant. The conjugated system now acts as a leaving group, forming a carboxylate anion. The leaving group has further reactions with piperidine to form a final side product, which is UV active. The presence of this compound indicates the success of Fmoc cleavage, but it has no further reactions with the DNA. In the final step carbon dioxide is the leaving group and the nitrogen becomes protonated, resulting in the deprotected amine. 20% Piperidine/DMF was passed through the solid support over approximately ten minutes. The support was washed with DMF to remove any residual reactant, but no further purification was possible before DNA.

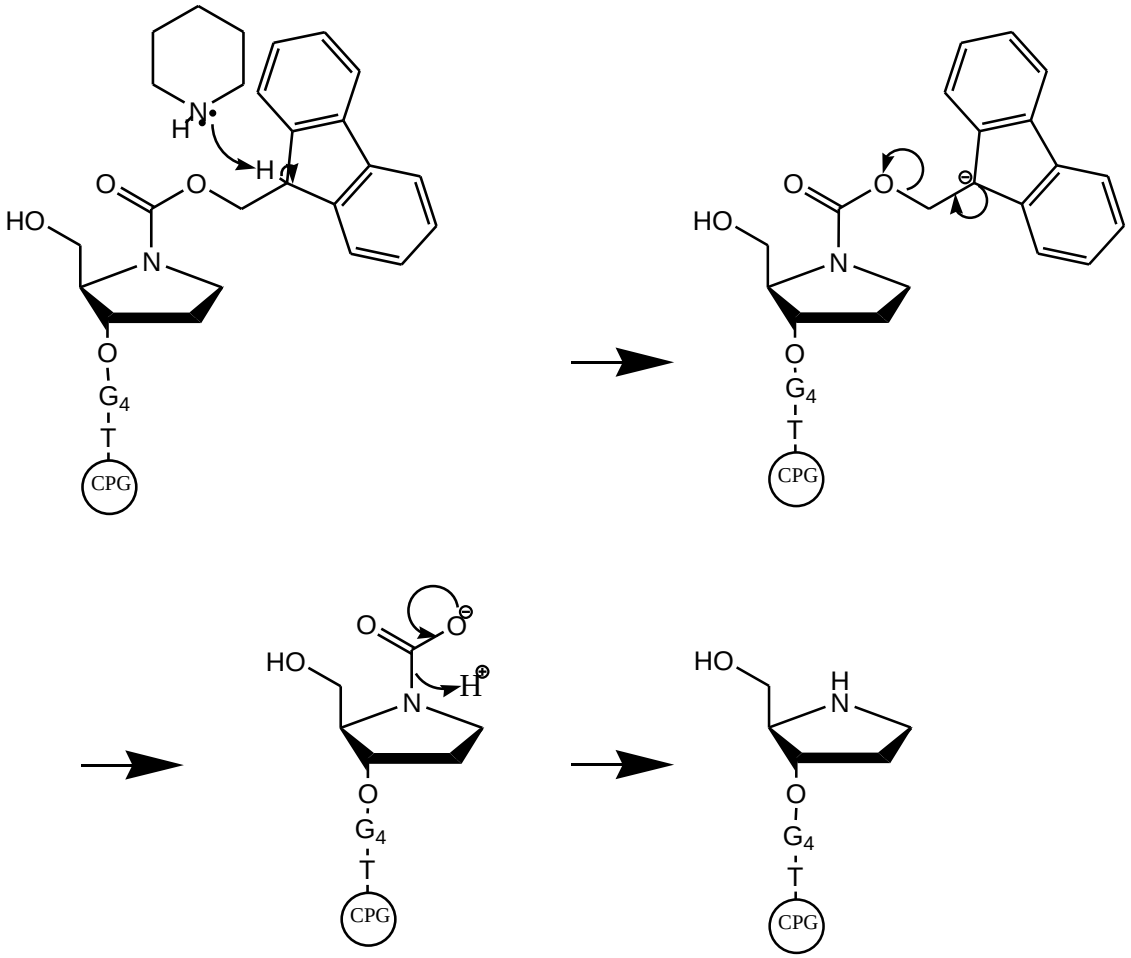


Figure 5.3 Reaction Mechanism of Fmoc deprotection using 20% Piperidine/DMF solution

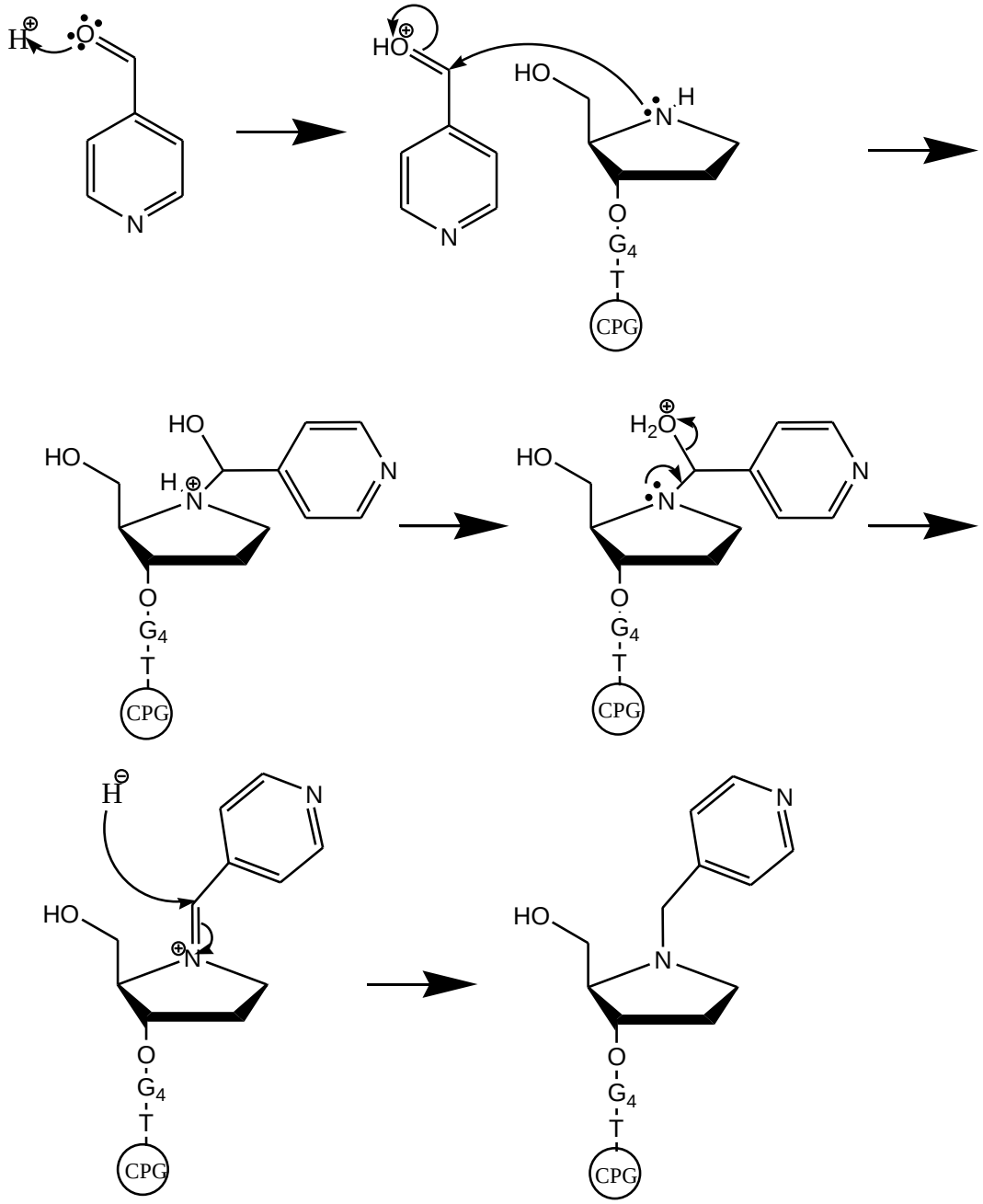


Figure 5.4 Mechanism for imine formation with aldehyde and reductive amination of imine to target modified DNA sequence. H^- is provided by Sodium Cyanoborohydride.

With the Fmoc group removed, the DNA can be functionalised. The solid support was removed from the column and suspended in 100 µL of pH 6.0, 0.05 $molL^{-1}$ sodium acetate/acetic acid buffer. A solution of 7.5µL Pyridine-4-carbaldehyde in 300µL DMF was added. This reaction mixture was shaken for 10 mins at room temperature. A 100 µL solution of 0.8 $molL^{-1}$ sodium cyanoborohydride in DMF is added as a reducing agent (Figure 5.4). The slight acidic conditions were necessary to protonate the carbonyl group of the aldehyde. The lone pair of nitrogen is a nucleophile, attacking the δ+ carbon, and forming the alcohol containing intermediate, with the modification substituted at this position. Next, a rearrangement occurs resulting in H_2O, a good leaving group. The final step uses the lone pair of nitrogen to form the imine, and H_2O is lost. At this stage the $NaCHBH_3$ provides H^-, which attacks the carbon of the imine, reducing it, in this case to a tertiary amine. This reaction mixture was shaken for 1 hr, then filtered. The solid support was washed with DMF and methanol, and DNA was cleaved with 28% aq. NH_4OH. Ammonia and water are evaporated *in vacuo* using speed-vac (Eppendorf

Concentrator Plus) and the residue was dissolved in 500 µL of Milli-Q H_2O. Insoluble materials were removed using a centrifuge and the resulting solution was purified by RP-HPLC (Figure 5.5).

The RP-HPLC chromatogram shown in Figure 5.5 indicates four products in the mixture, all absorbing at 260 nm. Each fraction was collected and desalted using a NAP-5 size exclusion column. Mass spectrometry indicated that the first peak (R_t = 10.33 minutes) contained G_4T, a truncated sequence resulting from poor coupling of the modified phosphoramidite. The phosphoramidite used

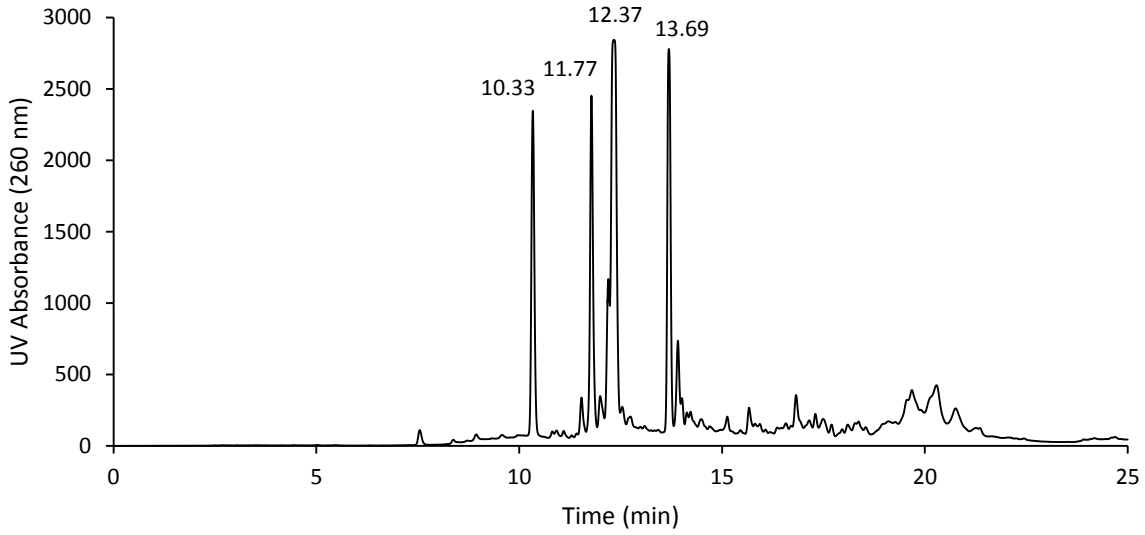


Figure 5.5 Analytical RP-HPLC of XG_4T. Purification and mass spectrometry was used to further characterise each peak.

for this synthesis was subsequently replaced, and this product was significantly less prominent in future reactions. In addition, all semi-manual coupling times were increased to ten minutes. The second peak (R_t = 11.77 minutes) contains the modification, but the mass indicated that Fmoc was removed, but the amine was not substituted. This indicated a potential failure at two steps: Fmoc-deprotection or functionalization of the amine. Treatment with aqueous ammonia also introduces a base capable of deprotecting the amine, which could mean that the Fmoc was not cleaved efficiently while on the solid support, and was instead cleaved afterwards. To test this, Fmoc concentration was tested in the piperidine/DMF solution. Fmoc absorbs UV light at 301 nm, meaning that UV analysis of this solution will indicate whether cleavage has occurred. In addition to this, the treatment time with piperidine/DMF solution was increased, which also meant a larger quantity of solution to ensure it was constantly passing though the column. Future attempts seemed to indicate that Fmoc cleavage had fully occurred, meaning that the issue may have been during the reductive amination. However, Fmoc cleavage should be monitored, and it may be necessary to adjust concentrations and coupling times for the substitution step of the synthesis in future. The third peak (R_t = 12.37 minutes) was the expected product, with the 4-pyridine modification correctly incorporated, giving the mass of 1828.35 a.m.u.. The final peak (R_t = 13.69 minutes) contained a mixture of these materials, but primarily the materials from the truncated and unreacted species.

The correctly modified XG_4T sequences were folded into G-quadruplexes, at a concentration of 100 µM, by adding buffer and heating each sample to 90°C for approximately 5 mins before allowing to gradually cool to room temperature and annealing at 4°C for several days. Samples were then diluted to the desired concentration for each experiment. The first buffer used was 150 mM NH_4OAc, a common buffer used to encourage quadruplex formation for analysis of complex formation with mass spectrometry. As discussed below, this buffer was incompatible with some biophysical tests, and pH 7.3 sodium cacodylate buffer was used instead. Metal solutions were prepared with $CuSO_4.5H_2O$ and $NiSO_4.6H_2O$, which were dissolved in Milli-Q water at 1 $mmolL^{-1}$ and added to solutions prior to heating and annealing. It was expected that solutions with no metal present would be partially stabilised by π-π interactions, as suggested in Engelhard *et al.* (2018)[19] , but the addition of cations should result in the change in structure shown in Figure 5.1 (c), giving the anticipated change in physical properties, such as thermal stabiity.

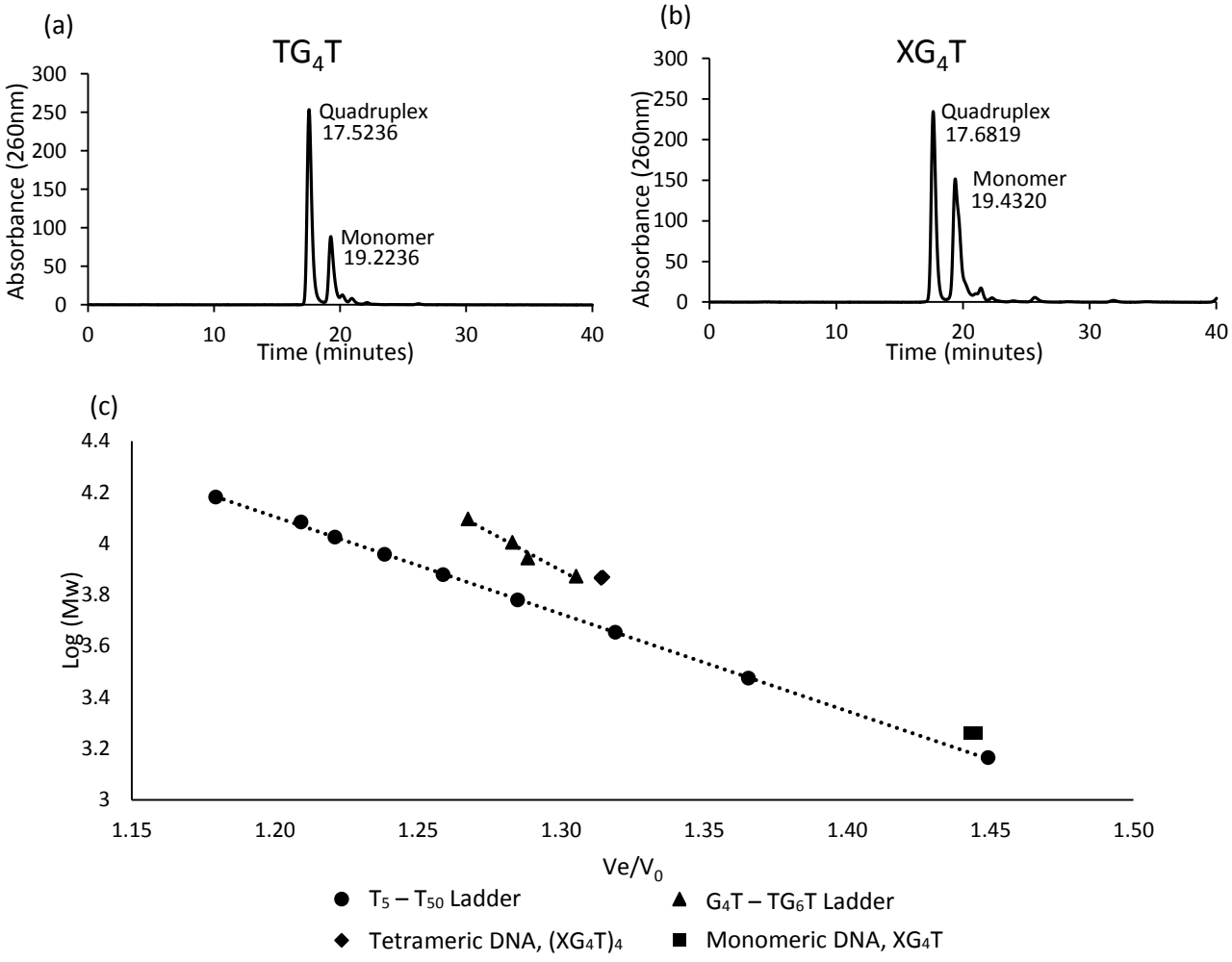


Figure 5.6 (a/b) Chromatograms of modified and unmodified TG_4T sequences. Distinct retention times are observed for both quadruplex and monomeric DNA, which are similar for both sequences. (c) Size Exclusion data indicates molecularity of XG_4T sequences. 250µM XG_4T, 10 mM sodium cacodylate Buffer, pH 7.3, 100 mM NaCl.

5.3 Results and Discussion

To prove that these quadruplexes form as expected, two experiments were performed to demonstrate that the modified sequences form G-quadruplexes with the same topology as unmodified DNA. Size exclusion chromatography was used to observe the trends of monomeric and tetrameric DNA sequences of varying size. This was carried out using pH 7.3, 10 mM sodium cacodylate buffer with 100 mM NaCl. A number of sequences containing only T_5 – T_{50} were used as a ladder for

single stranded DNA, while G_4T, TG_4T, TG_5T and TG_6T were used as a ladder for tetrameric structures. Dextran blue is a large molecule that is used as a standard to establish a dead volume of the size-exclusion column (it has a very short retention time). The column was pre-equilibrated with sodium cacodylate buffer, which was also used as the eluent throughout the experiment. Figure 5.6 (a/b) shows the chromatograms of the unmodified and modified sequences respectively, each with two peaks, suggesting the presence of both quadruplex and single stranded DNA. The relationship of retention time and molecular weight for these materials is compared to the ladder using Equation 5.1, shown below. This equation gives an arbitrary value, the structure index (SI), which is used to compare various sequences. V_e (elution volume) and V_0 (dead volume) are the retention times for the DNA samples and the dextran blue standard respectively. For size exclusion HPLC the ratio of these values is directly related to the $\log_{10}$ of the molecular weight of tetramolecular complexes, which results in the standard curves seen in Figure 5.6 (c). The relationships between G-quadruplex $(XG_4T)_4$ peaks and the G_4T – TG_6T ladder and monomeric XG_4T and the T_5 – T_{50} ladder supports this assignment of the peaks Figure 5.6 (a/b). This indicates that our modified sequences fold into tetrameric G-quadruplexes.

Circular dichroism was used to further demonstrate quadruplex formation and topology, by identifying characteristic structural peaks. As shown in Table 2.1, parallel quadruplexes have characteristic peaks with positive ellipticity at approximately 265 nm and negative ellipticity at 240 nm. Figure 5.7 shows these peaks for both TG_4T and XG_4T sequences, including in the presence of both Cu^{2+} and Ni^{2+}. Both of these experiments suggest that our modified sequences form similar secondary structures to unmodified DNA, including in the presence of cations.

Following this, melting experiments were carried out to determine the thermal stability of modified complexes. It was expected that the introduction of copper(II) or nickel(II) would result in an increase in melting temperature, as had been observed previously.[19] The first experiments were carried out in 150 mM NH_4OAc buffer, with 200 µM DNA and 50 µM copper(II) or nickel(II). These concentrations would be reduced in future because lower concentrations produced comparable results while requiring less DNA stock solution. Controls included both TG_4T and XG_4T without metals to give experimental melting data for an unmodified species and a modified but uncoordinated species. Samples were heated from 5°C to 90°C and then returned to 5°C in 2.5°C steps every 150 seconds, with CD spectra recorded at each temperature from 220 nm to 350 nm (Figure 5.9 (a)). The data was adjusted as described in Chapter 2. The result of these adjustments can be seen in the curves in Figure 5.10. This adjustment emphasises the melting curve, providing a more accurate $T_{½}$ value. $T_{½}$ is the temperature at which half of the structure is folded as indicated in (Figure 5.9 (b)). It is used to compare the thermal stability of each complex in non-equilibrium meltings. We know this is not an equilibrium because hysteresis was observed when the sample is

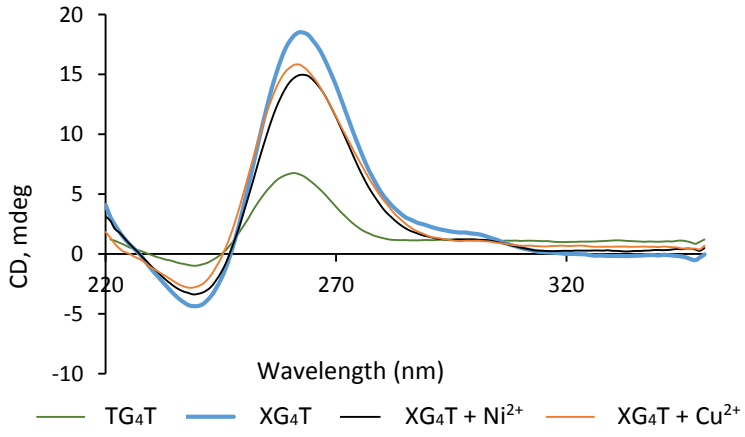


Figure 5.7 Circular dichroism spectra for modified and unmodified TG_4T. Peaks with positive ellipticity at 262 nm and negative ellipticity at 240 nm are characteristic of parallel G-quadruplex structures. 5 µM XG_4T, 5 µM $CuSO_4$/$NiSO_4$, 10 mM sodium cacodylate Buffer, pH 7.3, 100 mM NaCl.

cooled. Under these conditions our modification slightly destabilised the quadruplex, but no change was observed when cations were added (Table 5.1).

$$[Cu(H_2O)_6]^{2+} + nNH_3\ (aq) \rightleftharpoons Cu(NH_3)_n(H_2O)_{6-n}]^{2+} + n(H_2O)\ n = 1\text{-}5,\ \log(Q_c) = 1.48\text{-}4.06^{47}$$

$$[Ni(H_2O)_6]^{2+} + nNH_3\ (aq) \rightleftharpoons Ni(NH_3)_n(H_2O)_{6-n}]^{2+} + n(H_2O)\ n = 1\text{-}5,\ \text{Log}\ (Q_c) = 1.12\text{-}2.82^{48}$$

Figure 5.8 Equation for complexation of copper(II) and nickel(II) with ammonia. Copper complexation at 30°C, using NH_4NO_3 and $Cu(NO_3)_2$. Log Q_c decreases for each successive exchange, ranging from 4.06 to 1.48.[47] Nickel complexation at 25°C, using $Ni(ClO_4)_2$ and NH_4ClO_4. Log (Q_c) also decreases for subsequent substitutions, ranging from 2.82 to 1.12.[48]

However, the first experiments had been conducted in 150 mM NH_4OAc buffer. We believed the results of the melting experiment was due to interference from the buffer, as ammonium or acetate ions in the buffer could form complexes with the cations used. The equations in Figure 5.8 give examples for the complexation of both Cu^{2+} and Ni^{2+} with ammonia in aqueous conditions.[47,48] The presence of NH_4OAc will result in an equilibrium with NH_3 or ^-OAc, and the subsequent ligand exchange. The exchange is favoured according to the reported equilibrium constants, and the excess of buffer (150 mM) relative to cation concentration (40 µM) further drives the equilibrium to the right, meaning cation availability for coordination with DNA will be extremely low. To correct this a different buffer with non-coordinating ions was required. 10 mM sodium cacodylate/cacodylic Acid, with 100 mM NaCl was prepared at pH 7.3 and used as the buffer for the new melting experiments. Cacodylate buffer was also used for the previously described size exclusion chromatography to ensure consistent behaviour of the G-quadruplexes.

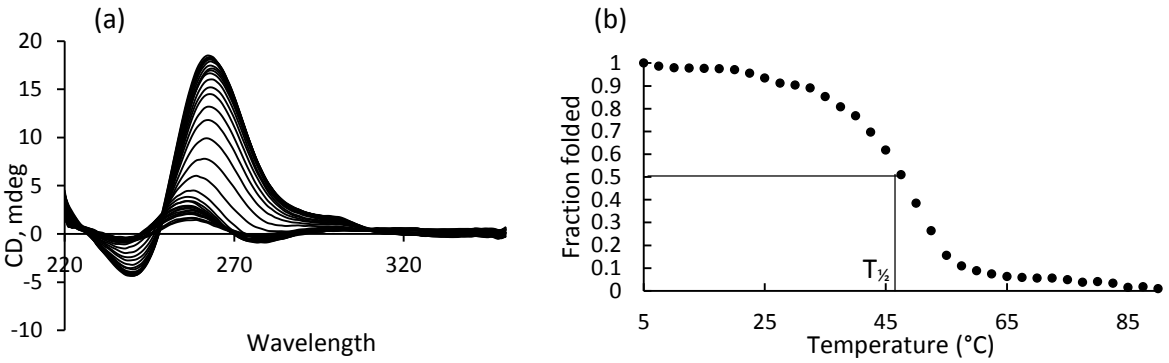


Figure 5.9 (a) Melting profile of XG_4T without metal present, from 5°C to 90°C. Decreases in characteristic signals is used to observe the degradation of the secondary structure. (b) Uncorrected melting curve of modified sequence at 262 nm. Intercept with y axis at 0.5 indicates the temperature at which half of the secondary structure is unfolded. This is the value shown in Table 5.1. 40 µM XG_4T, 10 mM sodium cacodylate Buffer, pH 7.3, 100 mM NaCl.

In cacodylate buffer the melting point for $(TG_4T)_4$ was lower than in NH_4OAc, but the difference between modified and unmodified sequences was more significant (Table 5.1). In addition to this, a further significant decrease in thermal stability was observed in the presence of both copper(II) and nickel(II), with nickel(II) having a smaller effect. The most likely explanation for the decrease in thermal stability is that coordination to metal centres introduces additional strain, twisting the strands and disrupting hydrogen bonds. Because our method incorporates the modification directly onto the ribose sugar, the modification is bulkier than previous examples, potentially resulting in unforeseen steric effects. The bond lengths or angles for Cu---N or Ni---N chelation with our pyridine modifications may also be unfavourable with nitrogen in the 4-position. Both Miyoshi *et al.* (2007)[44] and Engelhard *et al.* (2018)[19] indicated that metal ligand interactions were sufficiently strong to enforce topological

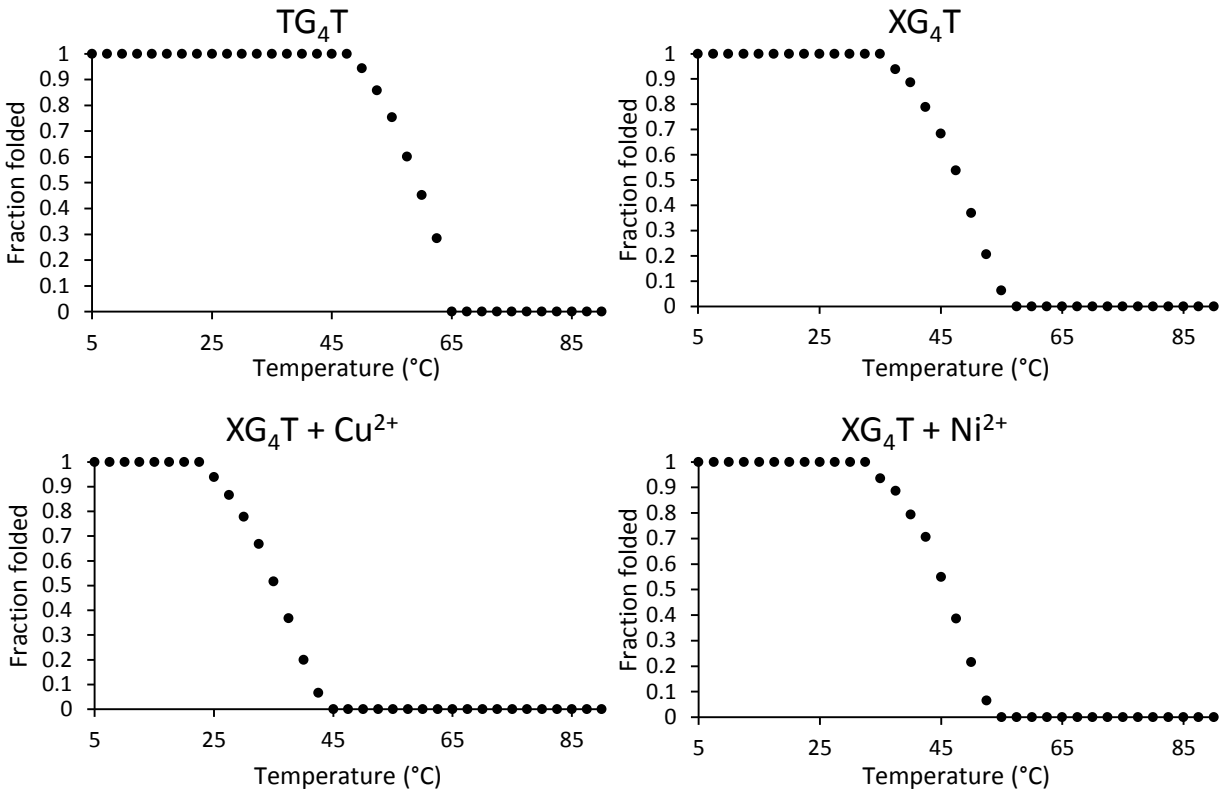


Figure 5.10 Corrected melting curves of XG_4T sequences with different cations from 5°C to 90°C. 40 µM XG_4T, 40 µM Cu^{2+}/Ni^{2+}, 10 mM Sodium Cacodylate Buffer, pH 7.3, 100 mM NaCl.

Table 5.1 $T_{½}$ Values of TG_4T and XG_4T complexes.

Sequence	$T_{½}$ (NH_4OAc Buffer)	$T_{½}$ (Sodium Cacodylate Buffer)
TG4T	76°C	58°C
XG4T	73°C	49°C
XG4T + Cu^{2+}	73°C	38°C
XG4T + Ni^{2+}	73°C	43°C

Comparison of $T_{½}$ (± 1.5°C)values for melting of XG_4T sequences with and without modification, and after addition of Cu^{2+} and Ni^{2+} to observe effect of ligand coordination. 40 µM XG_4T, 40 µM Cu^{2+}/Ni^{2+}, 150 mM NH_4OAc, or 10 mM Sodium Cacodylate Buffer, pH 7.3, 100 mM NaCl.

changes. It is possible we see a similar effect here. The optimal formation of metal-ligand bonds is favoured over the weaker hydrogen bonds, distorting the structure and resulting in lower $T_{½}$ values. This may be reflected in the kinetic properties, as hydrogen bonds may form more quickly with the stronger metal-ligand interactions holding the G4 template together, even under unfavourable conditions. We also address this in Chapter 6 with several suggested alterations to improve coordination. Importantly, 4-pyridine carbaldehyde had been successfully incorporated into quadruplex forming sequences, resulting in unique physical properties. The decrease in melting temperature was comparable to decreases observed by previous literature[19] and, while we did not see the same increase upon Cu(II) and Ni(II) addition, the successful incorporation of pyridine using pyrrolidine CE-phosphoramidite is promising for further experimentation with this strategy.

Previous literature[19-21,44] had not reported the effect that introducing ligands had on the kinetic properties of G- quadruplexes or procedures for analysing these properties. We had predicted that coordination to the metal centre would retain a template of the quadruplex structure, resulting

in faster formation times. To test this we heated samples at 95°C for 20 minutes to completely denature any secondary structure. The samples were then immediately cooled to 5°C in a CD spectrophotometer. The maximum intensity (θ_5) is recorded at 5 °C before the sample is melted and the minimum intensity (θ_{90}) is recorded after denaturing the sample. These are obtained by scanning between 220 nm and 350 nm. The maximum intensity in θ_5 is at 262 nm. Kinetic behaviour was then observed by measuring the intensity at 262 nm for 8 hours, with a total of 10,000 points. Similar adjustment was made to melting experiments, as described in Chapter 2, using the minimum and maximum intensity to calculate the fraction folded (α), at t = 0, α = 0; at t = ∞, α = 1. This was plotted over time to give the curves shown in Figure 5.11.

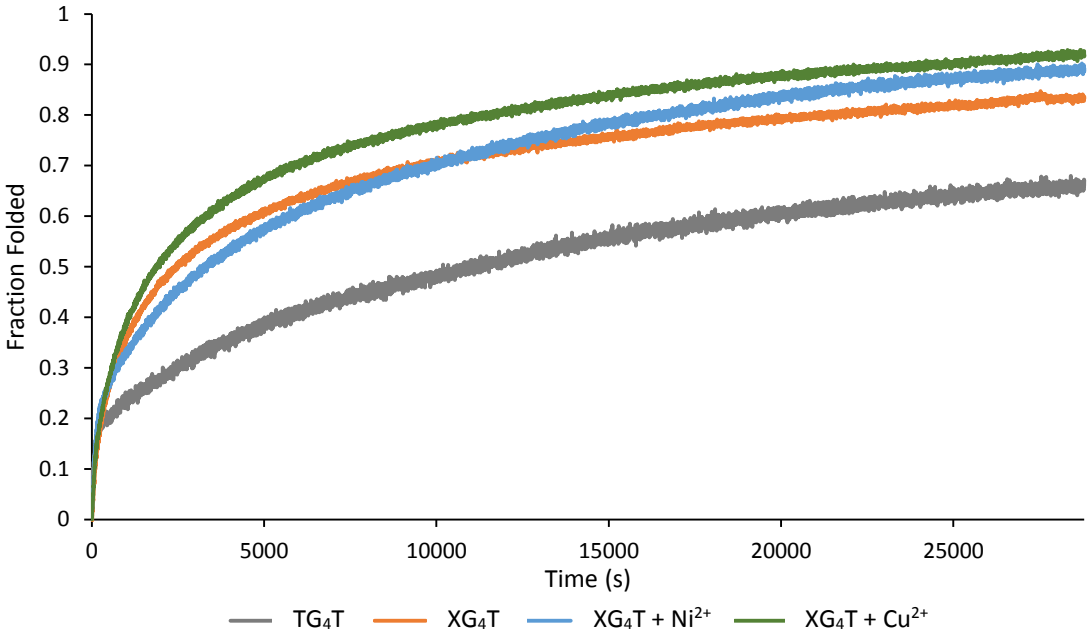


Figure 5.11 Circular Dichroism intensity at 262 nm for different XG_4T quadruplex formations at 5°C. Samples were heated to 90°C for 20 minutes, before cooling over eight hours to observe the change in CD signal. 40 μM XG_4T, 40 μM Cu^{2+}/Ni^{2+}, 10 mM sodium cacodylate buffer, pH 7.3, 100 mM NaCl.

Several interesting variables can be obtained from this data. Origin is used to fit the adjusted curves to Equation 5.2. θ_5, θ_{90} and t are obtained experimentally, as described. C is the strand concentration, which was verified by UV-Vis spectroscopy at 260 nm. This programme provides two variables: n, the strand cooperativity, and k_{on}, the rate constant for quadruplex formation (Figure 5.11). The quadruplex requires all four strands to interact before it can fold into the secondary structure. The cooperativity is therefore comparable to the order of reaction. This gives a cooperativity of four for TG_4T, which is slightly reduced for some modified sequences. We suggest this is because coordination to Cu^{2+} or Ni^{2+} gives a foundation for quadruplex formation, reducing the entropic restrictions for strands interacting. The effect is not large, but we believe it could be more signifcant in modified complexes with more favourable properties. The associtation rate constant, k_{on}, gives a rate for the formation of the G4 secondary structure. This rate is dependent on concentration and n,

$$y = \theta_5 + (\theta_{90} - \theta_5) \cdot (1 + (n-1) \cdot k_{on} \cdot C^{n-1} \cdot t)^{1/1-n} \qquad \text{(Equation 5.2)}^{25}$$

$$t_x = \frac{(1-x)^{1-n} - 1}{C_o^{n-1} \cdot (n-1) \cdot k_{on}} \qquad \text{(Equation 5.3)}^{25}$$

Table 5.2 Kinetic Properties of TG_4T and XG_4T Complexes

Sample	k_{on} ($M^{-3}s^{-1}$)	n	t_{90}
TG_4T	$2.94x10^7$	4	49,200 ± 15,600 hours (2048 days)
XG_4T	$2.94x10^8$	3.58	30.9 ± 10 hours (1.3 days)
$XG_4T + Cu^{2+}$	$3.89x10^8$	3.85	606 ± 190 hours (25 days)
$XG_4T + Ni^{2+}$	$1.91x10^9$	4	757 ± 240 hours (32 days)

k_{on} (± 30%) and n (± 10%) values for multiple XG_4T sequences. Curves from Figure 5.11 were fitted to Equation 5.2, with t_{90} values from Equation 5.3. 40 µM XG_4T, 40 µM Cu^{2+}/Ni^{2+}, 10 mM sodium cacodylate buffer, pH 7.3, 100 mM NaCl.

meaning it is a poor comparison for these samples, as n varies between samples. To account for differences between samples Equation 5.3 is used to calculate the formation time for a certain fraction (x) for a given strand concentration and n. The resulting t_{90} (Table 5.2) gives a time in hours required for the structure to be 90% folded. This is a more useful value for comparison because it accounts for variance in cooperativity and concentration. In this case we see significantly faster formation of modified structures, as we had predicted. However, much like the thermal stability experiments, this effect was reduced upon addition of cations. While these structures still form more quickly than unmodified TG_4T, they are not ideal and better ligands may need to be found to improve these results as well.

5.4 Conclusion

We have demonstrated a procedure for post-synthetic introduction of pyridine ligands into pyrrolidine-modified DNS sequences. Melting experiments with our modification did not give the same change in $T_{½}$ seen with similar modifications in the past. While previous literature had observed similar decreases in thermal stability for modified sequences, they saw an increase upon addition of cations. We proposed that the type of modification needs to be carefully selected to obtain better results. However, we did demonstrate a procedure for analysing kinetic properties of modified G-quadruplexes, and showed the improved kinetic properties of our modifications compared to TG_4T. These results give a useful starting point for future exploration of modifications based on pyrolidine-CE phosphoramidite and similar modified phosphoramidites.

6. Future Directions

Chemical cross-links are an area of oligonucleotide research which has only begun to be explored. As discussed previously, several applications have been investigated, with promising results for improving secondary structure stability and controlling topologies. This section will discuss the significance of the collected data, as well as potential future directions to improve where these modifications did not give satisfactory results.

One major change in future is the choice of the sequences used for quadruplex formation. Our modifications were incorporated only into dimeric and tetrameric quadruplexes. While these structures provide a useful model for G-quadruplex behaviour, the majority of native G-quadruplexes are monomeric. To study these modifications in sequences that more closely resemble native DNA we need to synthesise monomeric quadruplex strands, which comes with additional challenges and benefits. The benefit of monomeric sequences is that they do not have entropic requirements for assembly, potentially assisting in the formation of cross-links. For example, the disulfide bridge formation using N-2 functionalized dG was problematic in the formation of dimeric $(G_4T_4G_4)_2$ complexes. We suggested that one cause of this could be that thiols on different strands cannot interact before the structure folds, making cross-linking unlikely to occur at μM DNA concentrations. Monomeric structures avoid this problem because the linkages are already in the same sequence.

However, this comes at the cost of simplicity during synthesis. Monomeric sequences require more than one modification, which we did not attempt in any of our previous modifications. We note several times, both for 2-position inosine modification and XG_4T ligand substitution that we observe low efficiency for phosphoramidite coupling or functionalization of nucleotides. This problem is amplified for monomeric sequences where multiple modifications are required, and it will be necessary to improve coupling efficiency to obtain useful products. The Thrombin Binding Aptamer (TBA, GGTTGGTGTGGTTGG) is a useful model because it is a natively occurring sequence, while being short enough to be easily synthesised. If multiple modifications can be incorporated into monomeric G-quadruplex sequences it could bring us one step closer to successfully developing cross-linked G-quadruplexes.

C-8 Modification of Guanosine

Modification the 8-position of guanosine was attempted using Sonogashira coupling of Br-dG, but it was the least successful approach we considered. Several procedures showed promise: Sonogashira coupling with 3-butyn-1-ol, the synthetic routes of S-But-3-ynyl benzothioate and 3-(2-propyn-1-yloxy)-1-propene and microwave irradiation in general. This strategy, as discussed, could offer a useful route, if it could be optimised. Future investigation of these methods could continue in several directions, if desired.

(i) Improve conditions for Sonogashira coupling: In several instances these reactions could produce the desired products. However, the reactions were unreliable and only worked in small quantities. It may be possible to screen conditions by further adjusting solvents, reaction temperature, quantities or type of catalysts and reagents used etc. I believe this is the least promising route, as many of these changes have been attempted already with little or no effect, and it is a time consuming process with little anticipated success.

(ii) Post-synthetic Sonogashira Coupling: 8-Br-dG can also be converted to a phosphoramidite for DNA synthesis. This material is also commercially available. It has been previously demonstrated that Sonogashira coupling can be used for modification of oligonucleotides on the solid support.[49] However, previous experimentation with this method had primarily resulted in Glaser homocoupling between terminal alkynes[50] rather than the target modification on Br-dG in the DNA. The difficulties with purification could still be an issue based on previous literature and the pre-synthetic reactions, but post-synthetic modification could still be worth exploring.

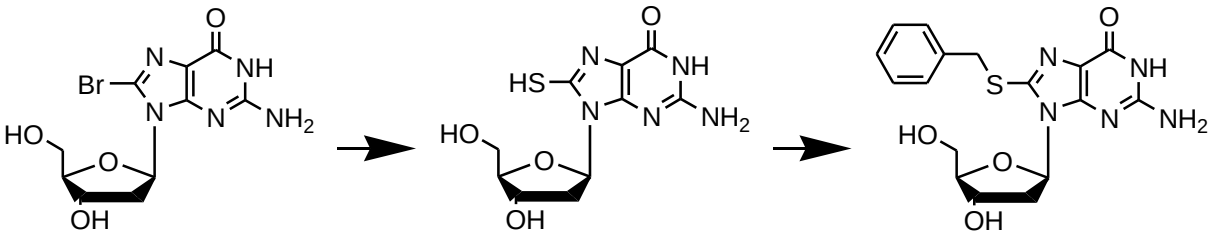


Figure 6.1 Modification of 8-Br-dG described by Hwu (2019)[49] using a two-step conversion from 8-Br-dG. Various functional groups could theoretically be incorporated through substitution of the thiol, in this case used for a benzyl group.

(iii) Achieving similar modifications with different protocols: It is possible that Sonogashira coupling is not an effective method for modifying 8-Br-dG However, there are numerous methods for nucleophilic halide substitution, which could be attempted on Br-dG. These could include substitutions with functional groups such as amines, carboxylic acids or nitriles. For 8-Br-dG in particular, substitution with methanol has been previously demonstrated,[51] and other known halide substitutions could be explored with this starting material. Hwu *et al.* (2019)[52] report substitution with a thiol, which was further reacted to incorporate a carbon chain at the 8-position (Figure 6.1). Similar strategies could be used to circumvent Sonogashira coupling for cross-link formations. This is the route I believe has the best chance of successfully producing C-8 modifications of guanosine which can be incorporated into quadruplexes.

N-2 Modification of Guanosine

Modification of 2-fluoro inosine with various thiol functionalized amines showed successful modification with mass spectrometry analysis. However, no significant change to biophysical properties was observed, and gel electrophoresis of modified G4-structures suggested that disulfide bridges had not been formed. Previous literature of this modification did not use it to create cross-links, and modifications of this type had mostly been used for fluorescent labelling. We demonstrated that this modification had potential for cross-link formation, and we have several suggested directions for continuing this strategy.

(i) Further investigation of disulfide formation: Several methods were used to try to encourage disulfide formation, which were unsuccessful. However, some methods suggested partial formation of the disulfide bridge, and other methods were not tested. These modifications could still be successful, and more exploration of disulfide forming conditions could be sufficient to form cross-links.

(ii) Pre-synthetic synthesis: This modification was only carried out using 2-F-dI phosphoramidites and post-synthetic modification. Modified inosine could be used as a starting point for similar modifications, which are incorporated directly into sequences. As previously discussed, we want to expand our cross-linking strategies to monomeric sequences, and direct incorporation like this is a potential strategy for improving the efficiency of modification to allow for multiple modifications with acceptable yields.

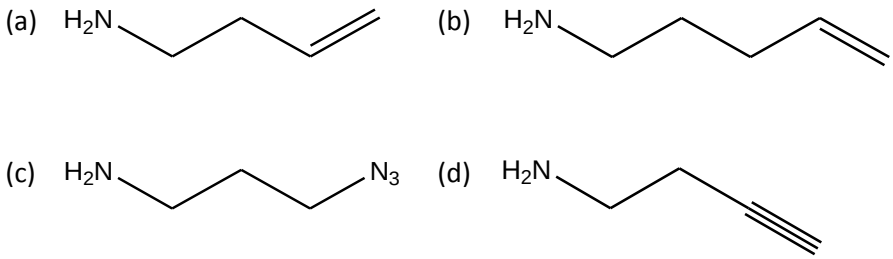


Figure 6.2 Potential modifications for 2-position cross-link formation in G-quadruplex sequences. (a) 3-buten-1- amine; (b) 4-penten-1-amine; (c) 3-azido-1-propanamine; (d) 1-amino-3-butyne.

(iii) Alternative Modifications: Several alternative modifications have been considered for cross-link formation. 3-Buten-1-amine and 4-penten-1-amine, Figure 6.2 (a) and (b) respectively, are both commercially available reagents and could be introduced into N-2 position of dG and after cross metathesis create a covalent link. 3-Azidopropanamine (Figure 6.2 (c)) could introduce azide functionality, with 1-amino-3-butyne (Figure 6.2 (d)) providing an alkyne. These two modifications give the possibility for a copper catalyzed click reaction (Figure 6.3). While the synthetic route explored previously did not give satisfying results, they did provide evidence for the synthesis of functionalised guanosine using this protocol. This protocol could now be extended to other functionalities to see if they are more effective than disulfide bridges.

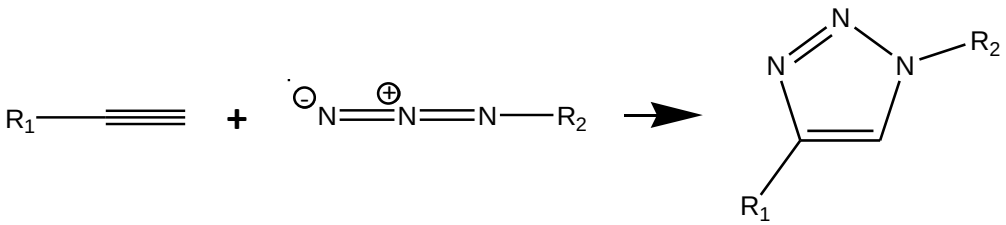


Figure 6.3 General equation for the synthesis of a 1,4-triazole linker using copper(I) catalyzed azide-alkyne cycloaddition.

Incorporation of Ligands into G4-structures

The incorporation of pyridine ligands into TG_4T sequences is the most successful cross-linking structure we discussed. Contradicting one of our original hypotheses, the inclusion of the sugar with pyridine modifications did not improve the stability of TG_4T. In fact, the addition of cations appeared to have a slight destabilising effect on the complex. However, the decrease in thermal stability was comparable to previous literature,[19-21] and the exploration of kinetic properties was promising. Association properties of sequences with our modification did not improve significantly on divalent cation addition, but the formation of all modified structures was substantially faster than unmodified TG_4T. This result had not been explored in the previous literature, and it partially proved our hypothesis. This limited success gives several routes forward to improve on the results.

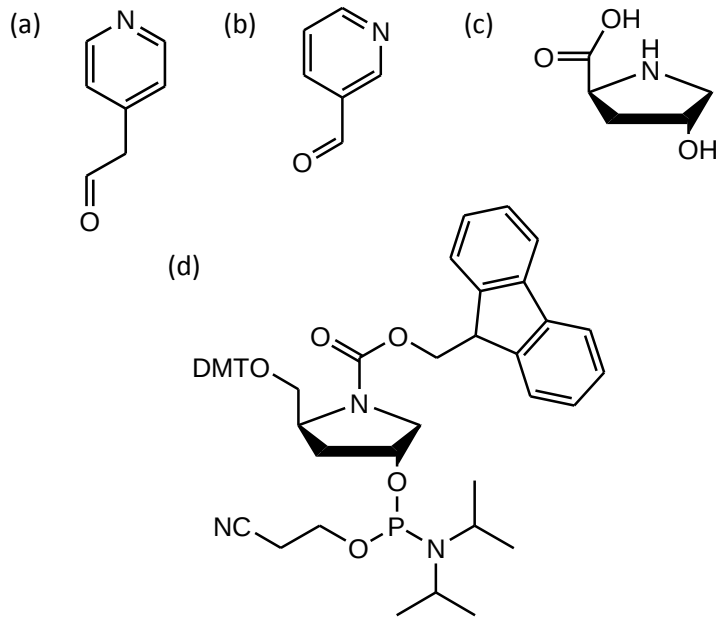


Figure 6.4 Potential molecules for further modification of XG4T. (a) pyridine-4- acetaldehyde; (b) pyridine-3-carbaldehyde; (c) L-4-hydroxyprolinol; (d) Target phosphoramidite, with trivalent phosphine at the 2'-position.

(i) Chain Elongation: The chain length had been shown to impact the change in thermal stability for metal coordinated complexes. While Engelhard *et al.* (2018)[19] observed shorter chains being more effective, it is possible the use of 4-pyridinecarbaldehyde modification is too short, increasing the strain on the quadruplex. This would explain the decrease in thermal stability observed upon addition of cations. We could develop longer linkers which behave similarly to the 4-pyridine modification (Figure 6.4 (a)), but increase flexibility to allow for more stable coordination to divalent cations. Kenta and Yasahiro (2010)[53] demonstrate a protocol for the addition of methanol to 4-methyl pyridine to produce 2-(pyridin-4-yl) ethanol (Figure 6.5 (i)). Chatterjee *et al.* (2015)[54] uses TEMPO oxidation to convert the alcohol to an aldehyde (Figure 6.5 (ii)). They also use the same protocol for several longer linkers. The desired product, 2-(pyridin-4-yl) acetaldehyde (Figure 6.4 (a)), should behave similarly to the carbaldehyde used previously, resulting in a two carbon linker. Neither of these compounds is commercially available, but this relatively short synthesis could be used to extend the linker length. Modification with these sequences should use the same protocol as described previously.

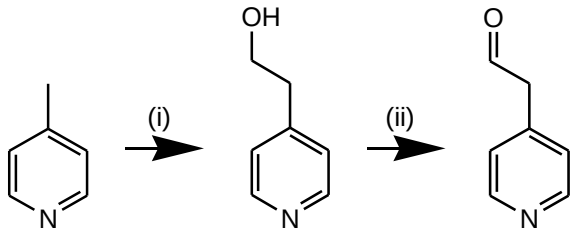


Figure 6.5 Synthetic rote for synthesis of 4-pyridine acetaldehyde (i) Addition of Methanol, Et3N; (ii) Tempo oxidation of alcohol, TEMPO, NaClO4, Bu4NHSO4, S: DCM.

(ii) Different ligands: In addition to the flexibility of the linker potentially influencing complex stability, the position of the coordination site, and therefore the angle of the interaction. 3-Pyridine carbaldehyde (Figure 6.4 (b)) could be incorporated using the same method as previously described, possible giving a better coordination angle with less strain on the complex.

(iii) Different sugars: Native DNA has a 3'-hydroxyl group bound to the phosphate, but the N-glycosidic bond is attached at the 1'-position. Our base analogue is attached to an amine which replaces the oxygen in the ribose ring. A similar phosphoramidite could possibly be synthesised from L-4-hydroxyprolinol (Figure 6.4 (c)) with the hydroxyl group at the 2'-position.[55] This would require reduction of the acid, followed by Fmoc and DMT protection and conversion to the phosphoramidite (Figure 6.4 (d)). The modification in oligonucleotides would position the base more closely to its position in native DNA.

7. Experimental Procedures

7.1 General Methods

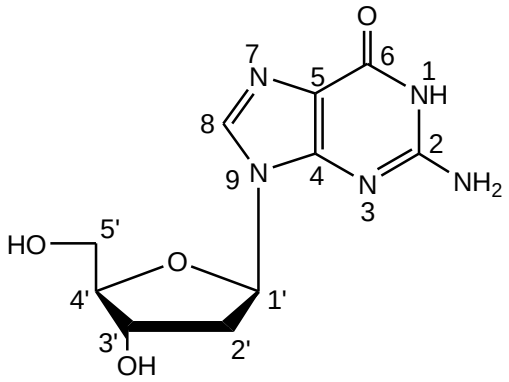
Solvents and reagents were obtained from a range of sources, certified ACS grade. Unmodified DNA sequences were obtained from IDT (USA). Modified phosphoramidites were obtained from ChemGenes or Glenresearch. Materials were used as received without further purification. Solvents were dried using standard techniques for distillation. 'Brine' refers to a saturated solution of NaCl and water.

All reaction were performed in oven-dried glassware. Sensitive reactions were performed under an inert atmosphere of argon, using standard 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).

Analytical thin layer chromatography (TLC) was performed with MERCK precoated silica gel 60- F254 (0.5-mm) alumina plates. Spots were visualised using UV, or various chemicals. Ninhydrin solution (1.5 g, 0.08 M) in 100 mL of butanol and 3 mL acetic acid was used to identify amines by dripping it over TLC plates and heating them, resulting in formation of Ruheman's purple. Silica/I_2 powder (100 mL of silica powder, 1 g of I_2 crystals) was used to visualise unsaturated materials by dipping TLC plates into the powder several times until yellow/brown spots appear. PMA solution (10 wt% PMA in ethanol) was used to visualise various materials using a similar procedure to ninhydrin.

^{1}H and ^{13}C NMR were recorded using Bruker 400 and 500 MHz spectrometers using tetramethylsilane (TMS) as an internal standard. Chemical shifts are reported in parts per million (ppm) downfield of TMS. 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 in methanol using Electrospray ionisation MS. Masses are reported in atomic mass units (a.m.u.). RPHPLC of synthesised DNA sequences was performed using TEAA buffer/acetonitrile, using the method described in Chapter 2.

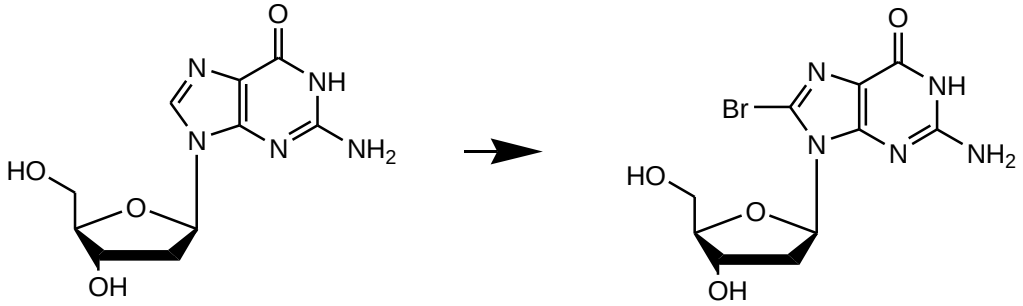
Numbering



All spectra of guanosine derivatives use the numbering system shown above. This is the standard numbering system for purine nucleosides.

7.2 Synthesis of C-8 Modifications

8-Bromo-2'-deoxyguanosine (1)

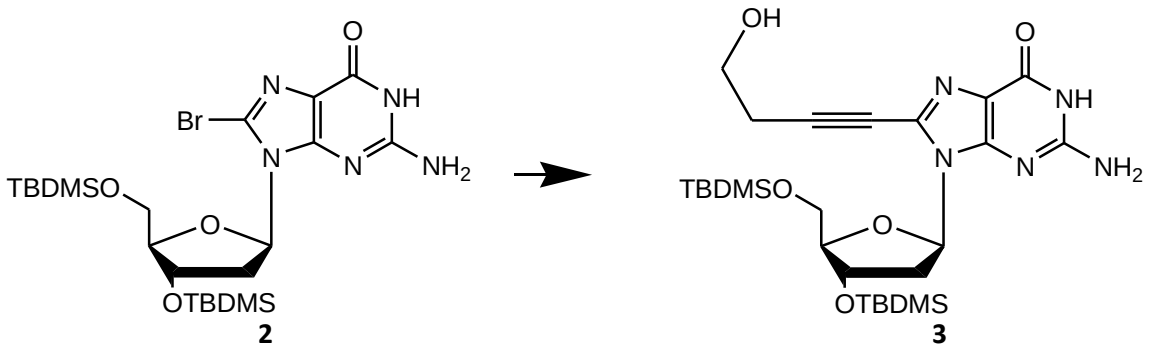


2'-Deoxyguanosine (1 g, 3.75mmol) was dissolved in an 8:2 mixture of acetonitrile: water (50 mL). The reaction mixture was cooled to 0°C and N-bromosuccinimide (1 g, 5.625 mmol) was added. The reaction mixture was stirred at room temperature for one hour, then evaporated to dryness *in vacuo*. The residue was suspended in acetone (20 mL) and stirred at room temperature for two hours. The product was collected (0.98 g, 76%) as a pale orange powder by vacuum filtration. ^{1}H NMR (500 MHz, d_6-DMSO) δ = 10.81 (1H, s, NH), 6.50 (2H, s, NH_2), 6.17 (1H, dd, J = 7.85 Hz, 6.71 Hz, 1'), 5.26 (1H, s, OH), 4.86 (1H, s, OH), 4.41 (1H, s, 4'), 3.81 (1H, sex, J = 2.96 Hz, 3'), 3.65-3.48 (2H, m, J = 5.40 Hz, 5'), 3.17 (1H, quint, J = 6.46 Hz, 2a'), 2.11 (1H, br.s., 2b'). ^{13}C (500 MHz, d_6-DMSO) δ = 155.92 (6), 153.81 (4), 152.48 (2), 121.02 (8), 117.98 (5), 88.38 (5'), 85.55 (1'), 71.51 (4'), 62.53 (3'), 36.93 (2'). NMR results were in agreement with Silerme *et al*. (2014)[26].

8-Bromo-3',5'-TBDMS-2'-deoxyguanosine (2)

Compound **1** (1.9 g, 5.49 mmol) was dissolved in 40 mL of dry DMF. Imidazole (1.9 g, 27.4 mmol) and *t*-butyldimethylsilyl chloride (2.43 g, 16.5 mmol) were added, and the mixture was stirred overnight at room temperature. 50 mL of saturated $NaHCO_3$ was added to stop the reaction, followed by extraction with 80 mL ethyl acetate. The organic layer was washed twice with H_2O, dried over $MgSO_4$, filtered and evaporated to dryness *in vacuo*. The product (2.88 g, 78%) was obtained as an orange solid, which was used in the next step without further purification. ^{1}H NMR (500 MHz, d_6-DMSO) δ = 10.84 (1H, s, NH), 6.42 (2H, s, NH_2), 6.15 (1H, t, J = 7.11 Hz, 1'), 4.59 (1H, quint, J = 3.00 Hz, 4'), 3.80 – 3.74 (2H, m, 5'), 3.69 – 3.64 (1H, m, 3'), 2.51 (1H, quint, J = 1.87 Hz, 2a'), 2.19 – 2.12 (1H, m, 2b'), 0.911-0.824 (18H, m, Si – C($C\underline{H}_3$)$_3$), 0.126-⁻0.039 (12H, m, Si – ($C\underline{H}_3$)$_2$). ^{13}C NMR (500 MHz, d_6-DMSO) δ = 155.99 (6), 153.84 (4), 152.59 (2), 121.08 (8), 117.92 (5), 87.61 (5'), 85.19 (1'), 72.86 (4'), 63.31 (3'), 36.52 (2'), 25.96 – 26.26 (m, Si – C($\underline{C}H_3$)$_3$), 18.3 (d, J = 30.9 Hz, Si – $\underline{C}$(CH_3)$_3$), -4.99 – 4.24 (4x Si-($\underline{C}H_3$)$_2$). Results in agreement with Münzel *et al.* (2011).[27]

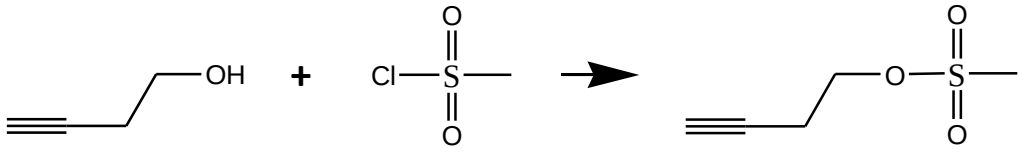
8-(4-Hydroxy-1-butyn-1-yl)-3',5'-TBDMS-2'-deoxyguanosine (3)



2 **3**

Compound **2** (0.5 g, 0.87 mmol) was dissolved in 40 mL of dry DMF with 3-butyn-1-ol (0.17 mL, 2.2 mmol). Argon was bubbled through this solution for 30 mins. Copper iodide (0.03 g, 0.07 mmol), palladium tetrakistriphenylphosphine (0.18 g, 0.08 mmol) and triethylamine (0.5 mL, 1.78 mmol) were added and the mixture was stirred under argon at 55 °C. The reaction was monitored by TLC (8:2 EtOAc:petroleum ether) until disappearance of the starting material was observed. The solution was evaporated to dryness *in vacuo*. The residue was dissolved in 20 mL of ethyl acetate. This mixture was washed with 20 mL of 5% EDTA solution, 30 mL of saturated $NaHCO_3$ and 30 mL of brine. The organic layer was dried over $MgSO_4$, filtered and evaporated to dryness *in vacuo*. The final product was then purified with silica gel column chromatography (99:1, acetone:toluene) to obtain a brown solid (0.22 g, 45%). This product had 4.6 mol% O=PPh_3. ^{1}H NMR (500 MHz, d_6-DMSO) δ = 10.73 (1H, s, NH), 6.51 (2H, s, NH_2), 6.25 (1H, t, J = 7.25 Hz, 1'), 5.06 (1H, t, J = 5.46 Hz, C≡CCH$_2$CH$_2$O<u>H</u>), 4.53 (1H, quintet, J = 2.88 Hz, 3'), 3.77 (2H, m, 5') 3.67 (1H, q, J = 5.16 Hz, 4'), 3.62 (2H, q, J = 6.27 Hz, C≡CCH$_2$C<u>H</u>$_2$OH) 3.27 – 3.21 (1H, m, 2_a'), 2.63 (2H, t, J = 6.78 Hz, C≡CC<u>H</u>$_2$CH$_2$OH), 2.13 – 2.10 (1H, m, 2_b'), 0.91-0.82 (18H, m, Si – C(C<u>H</u>$_3$)$_3$), 0.12-0.00 (12H, m, Si – (C<u>H</u>$_3$)$_2$). ^{13}C NMR (500 MHz, d_6-DMSO) δ = 170.80 (8), 156.52 (6), 154.19 (4), 151.25 (2), 117.22 (5), 93.91 (<u>C</u>≡CCH$_2$CH$_2$OH), 87.65 (5'), 83.75 (1'), 73.24 (C≡<u>C</u>CH$_2$CH$_2$OH), 72.14 (4'), 63.70 (3'), 60.22 (C≡CCH$_2$<u>C</u>H$_2$OH), 59.64 (2'), 26.14 – 26.24 (Si – C(<u>C</u>H$_3$)$_3$), 23.74 (C≡C<u>C</u>H$_2$CH$_2$OH), 14.55 (Si – <u>C</u>(CH$_3$)$_3$), -4.94 – -4.24 (Si-(<u>C</u>H$_3$)$_2$). Method developed based on method from Okamoto *et al.* (2005)[23] and Tainake *et al.* (2007).[24]

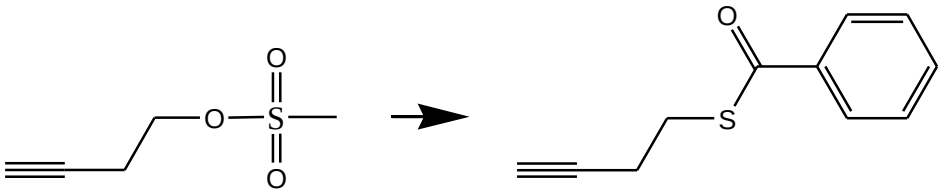
3-Butynyl methanesulfonate (4)



4

3-Butyn-1-ol (5.4 mL, 2.45 mmol) and Et_3N (14.9 mL, 0.48 mmol) were stirred in 80 mL of DCM under argon. The solution was cooled to 0° C and methanesulfonyl chloride (5.5 mL, 2.34 mmol) was added dropwise. The solution was stirred at room temperature for two hours, when TLC (30% EtOAc/Hex, visualised in PMA solution) indicated no starting material remained. The solution was then quenched with 50 mL of water. The organic layer was washed with water, followed by brine, dried with Na_2SO_4, filtered and DCM was removed *in vacuo*. The resulting orange liquid (6.19 mL, 70.2% yield) was used without further purification in the next step. ^{1}H NMR (500 MHz, $CDCl_3$) δ = 4.32 (2H, t, *J* = 6.63 Hz, CH$_2$C<u>H</u>$_2$OS), 3.08 (3H, s, SO$_2$C<u>H</u>$_3$), 2.67 (2H, td, *J* = 6.67 Hz, 2.83 Hz, C<u>H</u>$_2$CH$_2$OS), 2.09 (1H, t, *J* = 2.74 Hz, C≡C<u>H</u>). ^{13}C (500 MHz, $CDCl_3$) δ = 78.88 (<u>C</u>≡CH), 71.04 (C≡<u>C</u>H), 67.39 (CH$_2$<u>C</u>H$_2$OS), 37.50 (SO$_2$<u>C</u>H$_3$), 19.63 (<u>C</u>H$_2$CH$_2$OS). Procedure used from Iannazzo (2016)[28]; NMR results were in agreement with spectra from Jackson *et al.* (1988).[56]

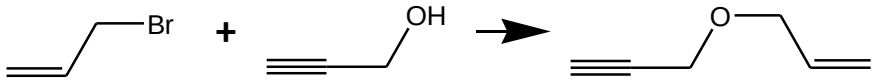
***S*-3-Butynyl benzene carbothioate (5)**



4 **5**

Compound **4** (0.3 g, 2.02 mmol) and Cs_2CO_3 (0.79 g, 2.43 mmol) were dissolved in 5 mL of dry DMF. This solution was stirred at 0° C, followed by dropwise addition of thiobenzoic acid (0.416g, 2.43 mmol) in 5 mL DMF. Stirring was continued at room temperature, until TLC (45% EtOAc/Hex) indicated that no starting material remained. The solution was diluted with ethyl acetate, washed with brine, dried with Na_2SO_4, filtered and solvent was removed *in vacuo*, resulting in an orange oil (0.61g, 47.4% yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ = 7.99 (2H, dd, *o*-C<u>H</u>), 7.62-7.58 (1H, m, *p*-C<u>H</u>), 7.50-7.46 (2H, m, *m*-C<u>H</u>), 3.27 (2H, t, J = 7.08 Hz, $CH_2C\underline{H}_2S$), 2.61 (2H, td, J = 7.27 Hz, 2.87 Hz, $C\underline{H}_2CH_2S$), 2.08 (1H, t, J = 2.50 Hz, C≡C<u>H</u>). ^{13}C NMR (500 MHz, $CDCl_3$) δ = 191.26 (S<u>C</u>=O), 136.74 (quaternary <u>C</u>) 133.47 (*p*-<u>C</u>H), 128.58 (*m*-<u>C</u>H), 128.17 (*o*-<u>C</u>H), 82.50 (<u>C</u>≡CH), 69.79 (C≡<u>C</u>H), 27.91 ($CH_2\underline{C}H_2S$), 19.50 ($\underline{C}H_2CH_2S$). Procedure used from Iannazzo *et al.* (2016) [30], NMR results are in agreement with spectra from Held *et al.* (2003).[31]

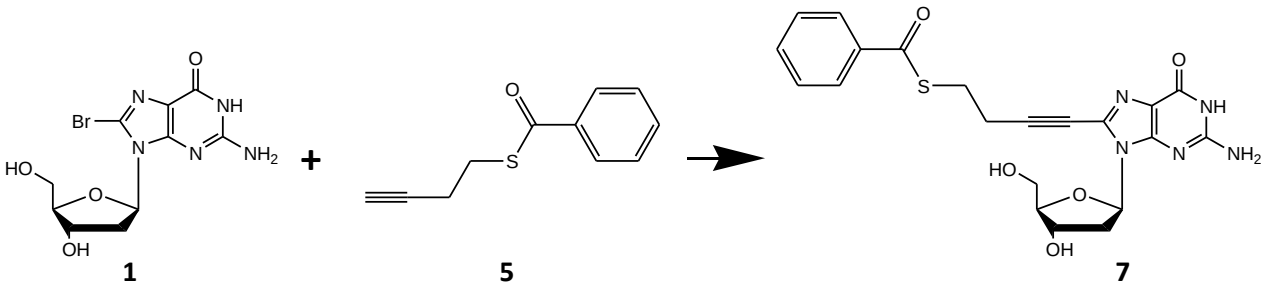
3-(Prop-2-ynyloxy)prop-1-ene (6)



6

Allyl bromide (17.3 mL, 0.2 mol) and propargyl alcohol (14.4 mL, 0.25 moL) were mixed in a round bottom flask at 0 °C. The solution was gradually heated while adding 4M aq. KOH (62.5 mL, 0.25 mol) dropwise, then stirred at 70 °C temperature for four hours. The organic phase was washed with water, then dried with Na_2SO_4 and filtered. The product was distilled (b.p. 105-106 °C, atm. pressure) to obtain 7.6 g of pure product (37% yield). ^{1}H NMR (500 MHz, $CDCl_3$) δ = 5.87 (1H, q, H_2C=C<u>H</u>), 5.29 (1H, d, $\underline{H}_2C$=CH), 5.19 (1H, d, $\underline{H}_2C$=CH), 4.12 (2H, d, HC≡C$C\underline{H}_2$-O), 4.04 (2H, dt, O-$C\underline{H}_2CH$=CH_2), 2.43 (1H, t, <u>H</u>C≡C). ^{13}C NMR (500 MHz, $CDCl_3$) δ = 133.9 (H_2C=<u>C</u>H), 117.83 ($H_2\underline{C}$=CH), 117.74 (HC≡<u>C</u>), 74.40 (H<u>C</u>≡C), 70.45 (O-$\underline{C}H_2CH$=CH_2), 56.98 (HC≡C$\underline{C}H_2$-O). Method translated from French and used from Guermont (1953)[35], NMR spectra are in agreement with spectra reported by Daniel *et al.* (1996).[36]

8-(*S*-3-Butynylbenzenecarbothioate)-2'-deoxyguanosine (7)



1 **5** **7**

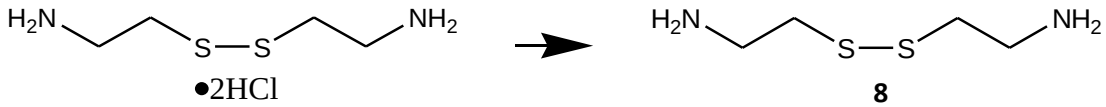
Compound **1** (0.2 g, 1 eq) was dissolved in 3 mL of dry DMF with $Pd(PPh_3)_4$ (0.032 g, 0.029 mmol, 0.05 eq), CuI (0.011g, 0.058mmol, 0.01 eq) and various bases (Et_3N, IRA-400, IRA-67, Dowex 1X-8, 2 eq).This solution was degassed under a vacuum and bubbled with argon. Compound **5** (0.86mmol, 1.5 eq) was added and the mixtures were reacted under argon in a microwave synthesiser at 50 °C for four hours. TLC (10% MeOH/DCM) indicated starting material remained so additional $Pd(PPh_3)_4$, CuI and base (half of original amount) were added, and the mixture was reacted for a further four hours under the same conditions. The solvent was removed *in vacuo*, and the residue was dissolved in EtOAc (10 mL). The organic layer was washed with 5% EDTA solution (20 mL), followed by sat. $NaHCO_3$ (20 mL) and brine (20 mL). The product (0.05g, 19.2% yield) was collected as a brown solid which was insoluble in water and EtOAc. This quantity was insufficient for full characterisation.

ESI-MS (MeOH, positive) $[M + H^+]$ = calculated ($C_{21}H_{22}N_5O_5S$) : 456.49; experimental: 456.133. Peaks: m/z (z = 1) = 456.133 (49%, Compound **7**), 347.078 (14%, Compound **1**), 340.086 (100%, Compound **7** - ribose), 263.098 (12%, Compound **7** – ribose/Bn), 229.967 (25%, Compound **1** – ribose).

This reaction was run in parallel with several similar reactions. This was the only reaction which resulted in an analysable product. For more details, see Chapter 3.

7.3 Synthesis of N-2 Modifications

2,2'-Dithiobis-ethanamine (cystamine) (8)

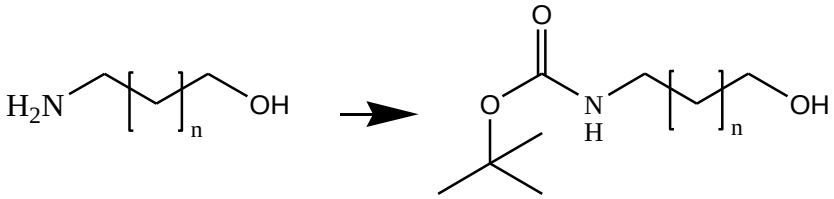


8

Cystamine dihydrochloride (1 g, 4.4 mmol) was dissolved in 20 mL of methanol and NaOH (0.363 g, 9.1 mmol) was added. The solution was stirred at room temperature for two hours. Methanol was evaporated *in vacuo* until a precipitate was observed. NaCl was removed by filtration and approximately 20 mL of ethanol was added. Ethanol and water were evaporated *in vacuo*, giving a pale yellow oil (0.56 g, 84% yield). Method used from Kim *et al.* (2004)[57]; NMR in agreement with Potta *et al.* (2009).[58]

^{1}H NMR (500 MHz, d_6-DMSO) δ = 3.06 (4H, br.s., [$\underline{H}_2NCH_2CH_2S]_2$), 2.78 (4H, t, J = 6.60 Hz, [$H_2NCH_2C\underline{H}_2S]_2$), 2.71 (4H, t, J = 6.74 Hz, [$H_2NC\underline{H}_2CH_2S]_2$).

***tert*-Butyl *N*-(3-hydroxypropyl)carbamate (10)** and ***tert*-Butyl *N*-(4-hydroxybutyl)carbamate (18)**



9: n = 1
17: n = 2

10: n = 1
18: n = 2

Compound **9** (2.3 mL, 30 mmol) or **17** (2.8 mL, 30 mmol) were dissolved in 30 mL of methanol. Di-*tert*-butyl dicarbonate (8.51 g, 39 mmol, 1.3 eq) was added at 0°C, followed by DMAP (0.18 g, 1.5 mmol, 5 mol%) and the solution was stirred overnight at room temperature. The reaction was analysed with TLC (20% methanol in DCM, visualised with ninhydrin solution). When TLC analysis indicated no starting material the solvent was removed *in vacuo*, and the residue was dissolved in ethyl acetate, washed with 10% w/v citric acid solution, followed by brine, dried over Na_2SO_4, filtered and EtOAc was removed *in vacuo*. The resulting oil was used without further purification in the next step. NMR spectra are in agreement with Aufort *et al.* (2011).[38]

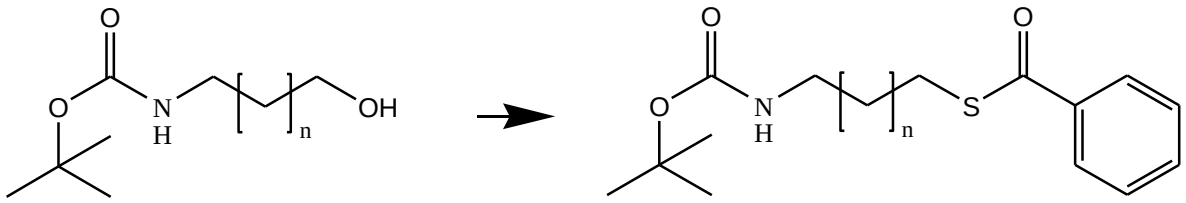
Compound **10**: Yield: 4.80 g (91%). ^{1}H NMR (500 MHz, $CDCl_3$) δ = 4.98 (1H, s, NH), 3.64 (2H, t, J = 5.22 Hz, N-C$\underline{H}_2$), 3.26 (2H, q, J = 6.04 Hz, HO-C$\underline{H}_2$), 1.66 (2H, quintet, J = 5.86 Hz, $CH_2C\underline{H}_2CH_2$), 1.43 (9H, s, *t*-Butyl). ^{13}C NMR (500 MHz, $CDCl_3$) δ = 157.13 (N-$\underline{C}$OO), 79.54 ($\underline{C}(CH_3)_3$), 59.23 (N-$\underline{C}H_2$), 36.99 (OH-$\underline{C}H_2$), 32.73 ($CH_2\underline{C}H_2CH_2$), 28.35 ($C(\underline{C}H_3)_3$).

ESI-MS (MeOH, positive) [M + Na^+] = calculated ($C_8H_{17}NO_3$.Na): 198.12; experimental: 198.41 (Na^+ adduct).

Compound **18**: Yield: 5.39 g (95%). ^{1}H NMR (500 MHz, $CDCl_3$) δ = 4.38 (1H, s, NH), 3.54 (2H, t, J = 6.01 Hz, N-C$\underline{H}_2$), 3.03 (2H, m, HO-C$\underline{H}_2$), 1.48 (4H, m, C$\underline{H}_2$C$\underline{H}_2$), 1.35 (9H, s, *t*-Butyl). ^{13}C NMR (500 MHz, $CDCl_3$) δ = 156.32 (HN-$\underline{C}$OO) 61.8 (N-$\underline{C}H_2$), 60.4 ($\underline{C}(CH_3)_3$) 40.2 (HO-$\underline{C}H_2$), 29.5, 26.4 ($\underline{C}H_2\underline{C}H_2$), 28.3 (*t*-Butyl).

ESI-MS (MeOH, positive) [M + Na^+] = calculated ($C_9H_{19}NO_3$.Na): 212.14; experimental: 212.29.

***tert*-Butyl-*N*-(3-benzothioatepropyl)carbamate (11)** and ***tert*-butyl-*N*-(4-benzothioatebutyl)carbamate (19)**



10: n = 1 **11: n = 1**
18: n = 2 **19: n = 2**

Triphenyl phosphine (4.04 g, 15.4 mmol, 1.4 eq) and diisopropyl azodicarboxylate (3.03 mL, 15.4 mmol, 1.4 eq) were both stirred in 50 mL of dry THF for 30 minutes at 0°C under argon. Then compound **10** (2 g, 11 mmol,1 eq) or **18** (2 g, 11 mmol) was added, followed by dropwise addition of thiobenzoic acid (2.12 g, 15.4 mmol, 1.4 eq). The solution was allowed to warm to room temperature and stirred overnight in argon. The reaction was monitored with TLC (15% EtOAc in toluene, visualised with UV then silica/iodine powder) until no starting material was observed. Solvent was removed *in vacuo*. The residue was dissolved in EtOAc, washed with 10% w/v citric acid, followed by sat. aq. $NaHCO_3$ then brine, dried with Na_2SO_4 and filtered. EtOAc was removed *in vacuo*. The resulting oil was dissolved in toluene, and PPh_3 precipitated and was removed by filtration. The product was then purified by silica gel column chromatography (0-20% EtOAC/Toluene). The method was based on Aufort *et al.* (2011)[38] and NMR spectra were compared to their substitution with thioacetic acid, because experimental spectra for benzoyl material were not available.

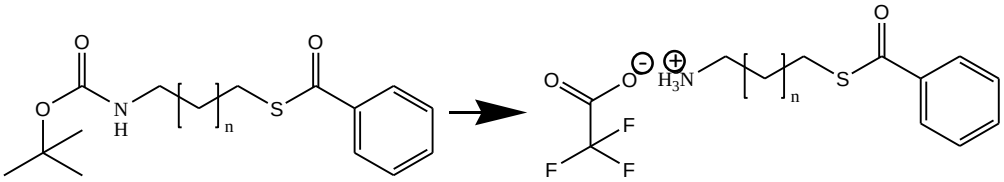
Compound **11**: Yield: 2.04 g (63%). ^{1}H NMR (500 MHz, $CDCl_3$) δ = 7.98 (2H, d, J = 7.59 Hz, *o*-C<u>H</u>), 7.59 (1H, t, J = 7.43 Hz, *p*-C<u>H</u>), 7.46 (2H, t, J = 7.68 Hz, *m*-C<u>H</u>), 3.24 (2H, q, J = 6.28 Hz, N-C<u>H_2</u>), 3.13 (2H, t, J = 7.00 Hz, S-C<u>H_2</u>), 1.88 (2H, quintet, J = 6.75 Hz, $CH_2C\underline{H_2}CH_2$), 1.47 (9H, s, *t*-butyl). ^{13}C NMR (500 MHz, $CDCl_3$) δ = 192.2 (S-<u>C</u>=O), 156.0 (N-<u>C</u>OO), 137.0 (aromatic/quaternary), 133.4 (*p*-<u>C</u>H), 128.6 (*m*-<u>C</u>H), 127.2 (*o*-<u>C</u>H), 79.5 (<u>C</u>$(CH_3)_3$), 39.1 (N-<u>C</u>H_2), 30.1 ($CH_2\underline{C}H_2CH_2$), 28.4 (*t*-butyl), 26.1 (S-<u>C</u>H_2).

ESI-MS (MeOH, positive) [M + Na^+] = calculated ($C_{15}H_{21}NO_3S.Na$): 318.13; experimental: 318.80. Peaks: m/z (z =1) = 318.80 (56%, Compound **11** + Na^+), 195.82 (100%, Compound **11** – Boc).

Compound **19**: Yield: 1.98 g (58%). ^{1}H NMR (500 MHz, $CDCl_3$) δ = 7.98 (2H, d, J= 7.04 Hz, *o*-C<u>H</u>), 7.55 (1H t, J = 7.45 Hz, *p*-C<u>H</u>) 7.43 (2H, t, J = 7.71 Hz, *m*-C<u>H</u>), 3.17 (2H, t, J = 6.78 Hz, N-C<u>H_2</u>), 3.09 (2H, t, J = 7.15 Hz, S-C<u>H_2</u>), 1.72 (2H, quintet, J = 7.82 Hz, S-$CH_2C\underline{H_2}$), 1.62 (2H, quintet, J = 7.05 Hz, N-$CH_2C\underline{H_2}$), 1.48 (9H, s, *t*-butyl). ^{13}C NMR (500 MHz, $CDCl_3$) δ = 191.8 (S-<u>C</u>=O), 169.6 (N-<u>C</u>OO), 127.8 (aromatic/quaternary), 133.3 (*p*-<u>C</u>H), 128.3 (*m*-<u>C</u>H), 127.2 (*o*-<u>C</u>H), 79.0 (<u>C</u>$(CH_3)_3$), 40.1 (N-<u>C</u>H_2), 29.3 (N-$CH_2\underline{C}H_2$), 28.6 (S-<u>C</u>H_2), 28.5 (*t*-butyl), 26.9 (S-$CH_2\underline{C}H_2$), 21.5 (S-<u>C</u>H_2).

ESI-MS (MeOH, positive) [M + Na^+] = calculated ($C_{16}H_{23}NO_3S.Na$): 332.14; experimental: 332.23. Peaks: m/z (z = 1) = 332.23 (48%, Compound **19** + Na^+), 210.22 (100%, Compound **19** - Boc).

3-Aminopropyl-1-benzothioate trifluroacetate (12) and **4-aminobutyl-1-benzothioate trifluroacetate (20)**



11: n = 1
19: n = 2

12: n = 1
20: n = 2

Compound **11** (1 g, 3.4 mmol) or Compound **19** (1g, 3.2 mmol) was dissolved in DCM (5 mL) followed by addition of trifluoroacetic acid (5 mL). Reaction mixture was stirred at room temperature for two hours and monitored with TLC (15% EtOAc in toluene) to ensure no starting material remained. The solvent was evaporated *in vacuo*, giving a brown oil.

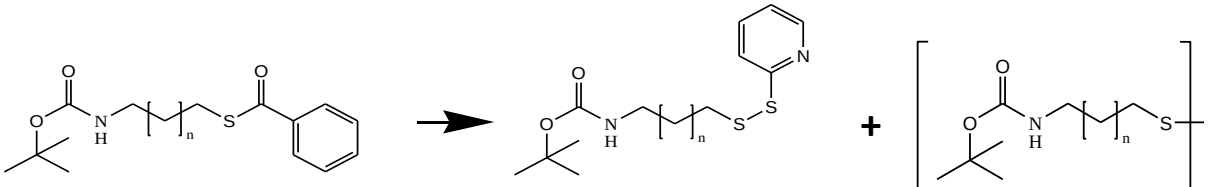
Compound **12**: Yield: 0.91 g (98%). ^{1}H NMR (500 MHz, D_2O) δ = 8.16 (2.5H, br.s., NH_3^+), 7.93 (2H, d, *o*-CH, J = 7.72 Hz), 7.59 (1H, t, *p*-CH J = 7.52 Hz), 7.45 (2H, t, *m*-CH, J = 7.62 Hz), 3.17 (2H, t, S-CH_3, J = 6.52 Hz), 3.11 (2H, br.s., H_3N^+-CH_2), 2.11 (2H, quintet, $CH_2CH_2CH_2$, J = 6.82 Hz). ^{13}C NMR (500 MHz, D_2O) δ = 193.29 (S-C=O), 164.1 (q, COO^-), 136.45 (aromatic/quaternary), 133.86 (*p*-CH), 128.72 (*m*-CH), 127.34 (*o*-CH), 120-113 (q, CF_3), 38.20 (H_3N^+-CH_2), 27.67 ($CH_2CH_2CH_2$), 25.14 (S-CH_2).

ESI-MS (MeOH, positive) [M + H^+] = calculated ($C_{10}H_{14}NOS$): 195.07; experimental: 195.22; ESI-MS (MeOH, negative) [M^-] = calculated ($C_2F_3O_2^-$): 112.99; experimental: 113.23.

Compound **20**: Yield: 0.95 g (91%). ^{1}H NMR (500 MHz, D_2O) δ = 8.06 (3H, br.s., $^+NH_3$), 7.93 (2H, d, J = 7.56 Hz, *o*-CH), 7.57 (1H, t, J = 7.98 Hz, *p*-CH), 7.44 (2H, t, J = 7.98 Hz, *m*-CH), 3.08 – 2.98 (4H, m, CH_2CH_2,CH_2CH_2), 1.90 – 1.80 (2H, m, S-CH_2), 1.80 – 1.70 (2H, m, H_3N^+-CH_2). ^{13}C NMR (500 MHz, D_2O) δ = 192.4 (S-C=O), 162.4 (q, COO^-), 136.9 (aromatic/quaternary), 133.5 (*p*-CH), 130 – 125 (q, CF_3), 128.6 (*m*-CH), 127.2 (*o*-CH), 39.4 (H_3N^+-CH_2), 28.0 (S-CH_2), 26.49, 26.44 (CH_2CH_2,CH_2CH_2).

ESI-MS (MeOH, positive) [M + H^+] = calculated ($C_{11}H_{16}NOS$): 209.31; experimental: 209.17; ESI-MS (MeOH, negative) [M^-] = calculated ($C_2F_3O_2^-$): 112.99; experimental: 113.11.

***tert*-Butyl *N*-[3-(2-mercaptopyridine)propyl]carbamate (13)** and ***tert*-Butyl *N*-[4-(2-mercaptopyridine)butyl]carbamate (21)**



11: n = 1
19: n = 2

13: n = 1
21: n = 2

14: n = 1
22: n = 2

Compound **11** (0.2 g, 0.67 mmol) was dissolved in methanol, and excess 28% aq. NH_4OH (10 mL) was added. The reaction was monitored by TLC (15% EtOAc in toluene, visualised by UV and silica/iodine powder) until no starting material was observed. Solvents were evaporated *in vacuo* and the residue was dissolved in DCM. 2-Mercaptopyridine (0.07 g, 1 eq) and *tert*-butyl hydrogenperoxide (72 µL, 1.1 eq) were added under argon. Finally, N-iodosuccinimide (0.015 g, 0.1 eq) was added and the solution was stirred at room temperature until TLC (same conditions as previously used) showed minimal starting material remaining. The reaction was quenched by addition of 10 mL of water, and the mixture was extracted three times with 20 mL of EtOAc. EtOAc was evaporated *in vacuo* to give a yellow oil. This material showed a single spot in TLC analysis (15% EtOAc in toluene, visualised with

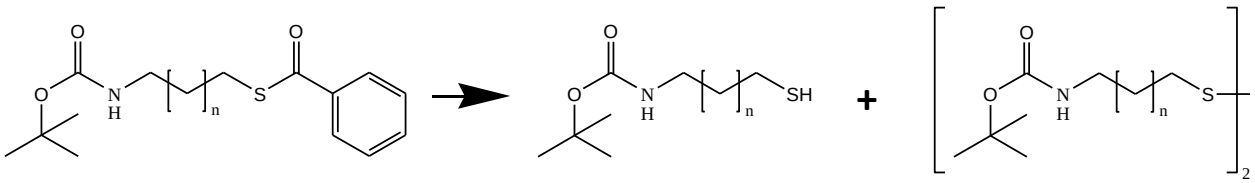
silica/iodine powder), but NMR and MS analysis showed two products which could not be isolated. Method developed based on Yuan *et al*. (2015).[59]

Compound **13**/**14**: Yield: Not recorded. ^{1}H NMR (500 MHz, $CDCl_3$) δ = 8.49 (0.78H, J = 4.63, d, H-6 $_{aromatic}$), 7.67 (0.7H, m, H-3 $_{aromatic}$), 7.62 (0.7H, m, *m*-C<u>H</u> pyrdine) 7.12 (0.8H, m, H-4$_{aromatic}$), 3.14 (2H, m, H-5 $_{aromatic}$), 2.83 (1.2H, t, J = 7.2 Hz, S-C<u>H</u>$_2$ (**14**)), 2.70 (0.8H, t, J = 7.18 Hz, S-C<u>H</u>$_2$ (**13**)), 1.73 (2H, m, $CH_2C\underline{H}_2CH_2$), 1.60 (1.6H, m, $CH_2C\underline{H}_2CH_2$), 2x 1.45 (9H, s, $C(C\underline{H}_3)_3$). ^{13}C NMR (500 MHz, $CDCl_3$) δ = 160.03 (N-<u>C</u>OO), 158.92 (C-2 $_{aromatic}$), 149.66 (C-6 $_{aromatic}$), 137.41 (C-5 $_{aromatic}$), 121.12 (C-4 $_{aromatic}$), 119.87 (C-3 $_{aromatic}$), 119.68 (3), 79.23 ($\underline{C}(CH_3)_3$), 39.25 (N-<u>C</u>H$_2$), 39.09 (N-<u>C</u>H$_2$), 36.13 (S-<u>C</u>H$_2$), 36.00 (S-<u>C</u>H$_2$), 29.59 ($CH_2\underline{C}H_2CH_2$), 29.20 ($CH_2\underline{C}H_2CH_2$), 28.41 ($C(\underline{C}H_3)_3$).

ESI-MS (MeOH, positive) [M + H$^+$] = calculated ($C_{13}H_{21}N_2O_2S_2$): 301.10; experimental: 301.18, 403.14. Peaks: m/z (z = 1) = 301.18 (38%, Compound **13**), 323.16 (28%, Compound **13** + Na$^+$), 381.21 (7%, Compound **14**), 403.14 (81%, Compound **14** + Na$^+$).

For a more detailed analysis of these spectra see Chapter 4. It was determined that after NH_3 was used to remove the protecting group the free thiols formed a disulfide bond spontaneously, resulting in two products, Compounds **13** and **14** with slightly different spectra. Purification of these materials is covered below, but was not possible with this method.

***tert*-Butyl *N*-(3-thiopropyl)carbamate (14)** and ***tert*-butyl *N*-(4-thiobutyl)carbamate (22)**



11: n = 1 **14: n = 1** **15: n = 1**
19: n = 2 **22: n = 2** **23: n = 2**

Compound **11** (1 g, 3.4 mmol) or **19** (0.1 g, 0.3 mmol) was dissolved in 10 mL of methanol and 10 mL of MeOH saturated with NH_3 was added. The reaction was monitored by TLC (15% EtOAc in toluene, visualisation by UV followed by silica/iodine powder) until no starting material was observed. Solvents were removed *in vacuo*. The product were purified by silica gel column chromatography (0-20% EtOAc in tluene) to obtain two pure roducts: a minor product (**14/22**, 40%) and major product (**15/23**, 60%).

Compound **14**: Yield: 0.14 g (22%). ^{1}H NMR (500 MHz, $CDCl_3$) δ = 3.23 (2H, q, J = 5.48 Hz, N-C<u>H</u>$_2$), 2.55 (2H, q, J = 7.45 Hz, S-C<u>H</u>$_2$), 1.79 (2H, quintet, J = 6.86 Hz, $CH_2C\underline{H}_2CH_2$), 1.43 (9H, s, $C(C\underline{H}_3)_3$). ^{13}C NMR (500 MHz, $CDCl_3$) δ = 156.1 (N-<u>C</u>OO), 79.1 ($\underline{C}(CH_3)_3$), 39.2 (N-<u>C</u>H$_2$), 34.1 ($CH_2\underline{C}H_2CH_2$), 28.4 ($C(\underline{C}H_3)_3$), 21.9 (S-<u>C</u>H$_2$).

ESI-MS (MeOH, positive) [M + Na$^+$] = calculated ($C_8H_{17}NO_2S$.Na): 214.1; experimental: 214.10, 403.14. Peaks: m/z (z = 1) = 214.10 (40%, Compound **14** + Na$^+$), 403.14 (100%, Compound **15** + Na$^+$).

Compound **15**: Yield: 0.36 g (56%). ^{1}H NMR (500 MHz, $CDCl_3$) δ = 3.24 (4H, q, J = 6.0 Hz, N-C<u>H</u>$_2$), 2.72 (4H, t, J = 7.41 Hz, S-C<u>H</u>$_2$), 1.90 (4H, quintet, J = 7.07 Hz, $CH_2C\underline{H}_2CH_2$), 1.46 (18H, s, *t*-Butyl). M/z = 403.18 (100%, Na). ^{13}C NMR (500 MHz, $CDCl_3$) δ = 156.0 (N-<u>C</u>OO), 79.2 ($C(CH_3)_3$), 39.1 (N-<u>C</u>H$_2$), 36.1 (S-<u>C</u>H$_2$), 29.5 ($CH_2\underline{C}H_2CH_2$), 28.4 ($C(\underline{C}H_3)_3$).

ESI-MS (MeOH, positive) [M + Na$^+$] = calculated ($C_{16}H_{32}N_2O_4S_2$.Na): 403.18; experimental: 403.18.

Compound **22**: Yield: 0.016 g (24%). ^{1}H NMR (500 MHz, $CDCl_3$) δ = 2.98 (2H, t, J = 6.91 Hz, N-C<u>H</u>$_2$), 2.41 (2H, q, J = 7.47 Hz, S-C<u>H</u>$_2$), 1.65 – 1.52 (4H, m, $CH_2C\underline{H}_2C\underline{H}_2CH_2$), 1.31 (9H, s, *t*-butyl). ^{13}C NMR (500 MHz,

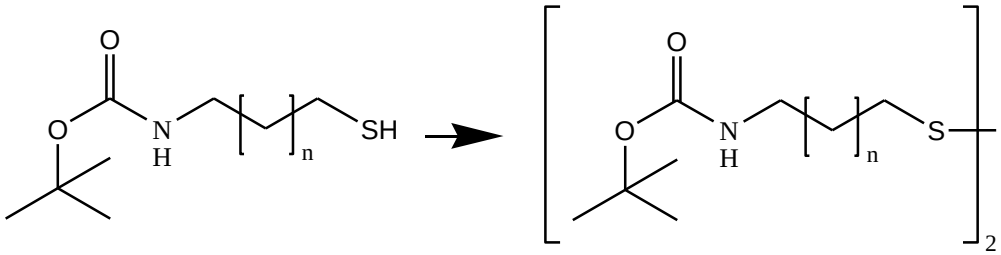
$CDCl_3$) δ = 155.9 (N-C=O), 79.3 (C(CH_3)$_3$), 40.1 (S-CH_2), 31.1 ($CH_2CH_2CH_2CH_2$), 28.4 (*t*-butyl), 24.3 (N-CH_2).

ESI-MS (MeOH, positive) [M + Na^+] = calculated ($C_{16}H_{32}N_2O_4S_2$.Na): 228.12; experimental: 228.20, 431.26. Peaks: m/z (z = 1) =431.26 (6%, Compound **23** + Na^+), 332.12 (22%, Compound **19** + Na^+), 228.20 (100%, Compound **22** + Na^+).

Compound **23**: Yield: 0.042 g (64%). ^{1}H NMR (500 MHz, $CDCl_3$) δ = 3.16 (4H, q, J = 6.74 Hz, S-CH_2), 2.70 (4H, t, J = 7.21 Hz, N-CH_2), 1.73 (4H, quintet, J = 7.44 Hz, S-CH_2CH_2), 1.60 (2H, quintet, J = 7.44 Hz, N-CH_2CH_2), 1.46 (18H, s, *t*-butyl). ^{13}C NMR (500 MHz, $CDCl_3$) δ = 155.9 (N-C=O), 79.2 (C(CH_3)$_3$), 38.4 (S-CH_2), 28.9 ($CH_2CH_2CH_2$), 28.4 (C(CH_3)$_3$), 26.1 (N-CH_2).

ESI-MS (MeOH, positive) [M + Na^+] = calculated ($C_{18}H_{36}N_2O_4S_2$.Na): 431.22; experimental: 431.3.

Bis[N, N'-(*tert*-butyl aminocarbamate)propyl]disulfide (15) and **Bis[N, N'-(*tert*-butyl aminocarbamate)butyl]disulfide (23)**



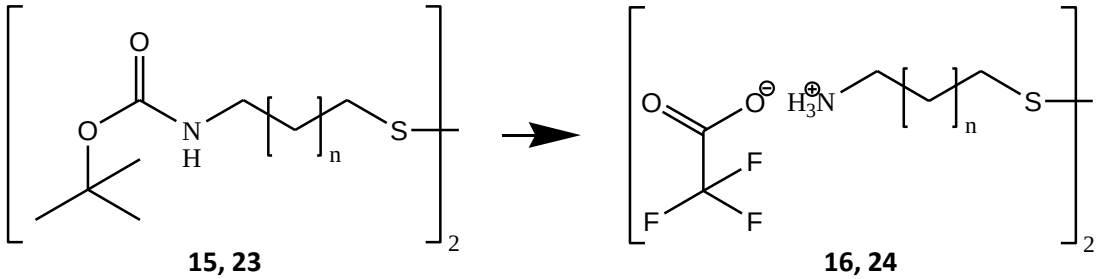
14: n = 1 **15: n = 1**
22: n = 2 **23: n = 2**

Compound **14** (0.1 g, 0.52 mmol) or **22** (0.1 g, 0.48 mmol) was dissolved in 1 mL of DCM, and *t*-Butyl hydrogen peroxide (50 µL, 1 eq) was added. Monitoring with TLC (10% EtOAc/toluene) indicated that the reaction did not proceed to completion under these conditions. N-Iodosuccinimide (0.011 g, 0.1 eq) was added and the reaction mixture became hot and changed to a dark red colour. Further monitoring with TLC showed no starting material remaining, and the reaction was stopped by addition of H_2O. The reaction mixture was extracted with 3x 5 mL of EtOAc. The combined organic layers was dried with Na_2SO_4, filtered and concentrated *in vacuo*, and the residue was purified by silica gel column chromatography (0-20% EtOAc/Toluene). This method gave a single dimeric product. For full assignment see **15** and **23** above.

Compound **15**: Yield: 0.076 g (77%)

Compound **23**: Yield: 0.072 g (73%)

3,3'-diaminopropyl-1,1'-Bisdisulfide ditrifluroacetate (16) and **4,4'-diaminobutyl-1,1'-Bisdisulfide ditrifluroacetate (24)**



15, 23 **16, 24**

Compound **15** (1 g, 2.6 mmol) or **23** (1 g, 2.4 mmol) was dissolved in DCM (5 mL) followed by addition of trifluoroacetic acid (5 mL). Reaction mixture was stirred at room temperature for two hours and monitored with TLC (15% EtOAc in toluene) to ensure no starting material remained. The solvent was removed *in vacuo* to obtain a brown oil.

Compound **16:** Yield: 1.05 g (97%). ^{1}H NMR (500 MHz, D_2O) δ = 2.90 (4H, t, J = 7.54 Hz, H_3N^+-C<u>H</u>$_2$), 2.59 (4H, t, J = 7.04 Hz, S-S-C<u>H</u>$_2$), 1.88 (4H, quintet, J = 7.50 Hz, $CH_2C\underline{H}_2CH_2$). ^{13}C NMR (500 MHz, D_2O) δ = 162.5 (q, <u>C</u>OO^-), 119-112 (q, CF_3), 37.99 (H_3N^+-<u>C</u>H_2), 33.63 (S-S-CH_2), 25.89 ($CH_2\underline{C}H_2CH_2$).

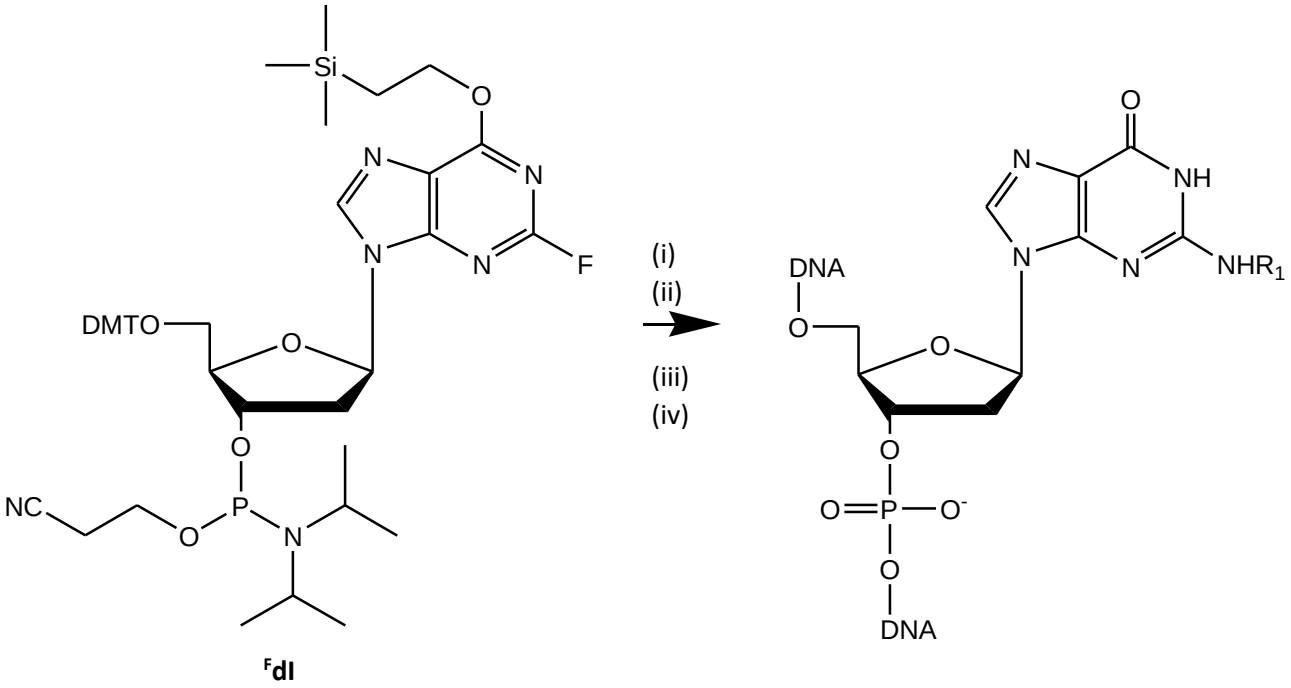
ESI-MS (MeOH, positive) [M + H^+] = calculated ($C_6H_{18}N_2S_2^{2+}$): 182.09; experimental: 181.25.

Compound **24:** Yield: 0.98 g (92%). ^{1}H NMR (500 MHz, D_2O) δ = 2.76 (4H, t, J = 6.54 Hz, H_3N^+-C<u>H</u>$_2$), 2.50 (4H, t, J = 6.59 Hz, S-S-C<u>H</u>$_2$), 1.51 (8H, quintet, J = 3.02 Hz, $CH_2(C\underline{H}_2)_2CH_2$). ^{13}C NMR (500 MHz, D_2O) δ = 162.2 (q, COO^-), 116.1 (q, CF_3), 38.9 (H_3N^+-<u>C</u>H_2), 36.8 (S-S-<u>C</u>H_2), 25.3, 25.0 ($CH_2(\underline{C}H_2)_2CH_2$).

ESI-MS (MeOH, positive) [M + H^+] = calculated ($C_8H_{22}N_2S_2^{2+}$): 210.12; experimental: 209.22.

7.4 Post Synthetic Modification of DNA Using Convertible Nucleoside Phosphoramidites

General Protocol for modification of ^{F}dI incorporated into DNA sequence



^{F}dI

(i) DNA synthesis; (ii) amine substitution; (iii) treatment with acetic acid to cleave silyl protecting group; (iv) cleavage from solid support and deprotection with aq. NH_4OH.

Phosphoramidite was obtained from Chemgenes and used for several post-synthetic chemical modifications of DNA. These reactions all used a similar protocol, previously reported by DeCorte *et al.* (1996)[18], with R_1 being:

- cystamine (**8**)
- 3-benzothioate-1-amine trifluoroacetate (**12**)
- 4-benzothioate-1-amine trifluoroacetate (**20**)
- 3,3'-dithiobis-1-propanamine (**16**)
- 4,4'-dithiobis-1-butanamine (**24**)

Cystamine and 3-azido-propanamine were both obtained from commercial sources. 3,3'-dithiobis-1-propanamine and 4,4'-dithiobis-1-butanamine were synthesised using the method described above.

Modified $G_4T_4G_4$ (5 µmol) was synthesised using semi-manual addition of the ^{F}dI phosphoramidite at selected positions (Table 7.1), as described in Chapter 2. 66.7 mg (75 µmol/g) of preloaded dG CPG was loaded into the column and synthesis was carried out as described in Chapter 2. Synthesis was paused after deblocking prior to addition of modification. ^{F}dI phosphoramidite (34.9 mg, final concentration: 0.1M) was weighed onto the column. The activator (400 µL, 0.45 $molL^{-1}$ ETT in acetonitrile) was added and the vacuum was held for thirty seconds to allow for phosphoramidite to be dissolved using a pipette. Coupling time for modified phosphoramidite was increased from 90 s to 10 mins to ensure the reaction went to completion. Synthesis was then completed as previously described for unmodified phosphoramidites.

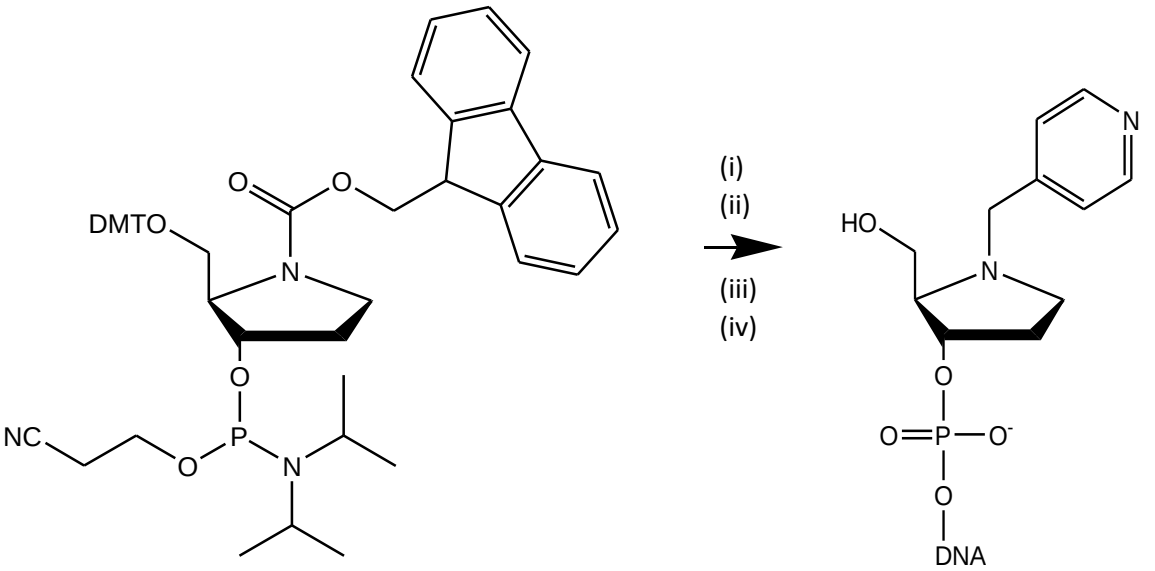
Each amine was dissolved in dry DMSO (1 mL, 0.5 molL^{-1}). The TFA salts of compound **12, 16, 20** and **24** were converted to free amines by adding 10% Et_3N to this solution. These solutions were added to the solid support (66.7 mg, 5 µmol) containing the modified DNA and shaken at room temperature for three days. The reaction mixture was filtered and the solid support was washed with dry DMSO. DNA was cleaved from the solid support with 1 mL of 28% aq. NH_4OH. DNA overnight, then precipitated with $LiClO_4$ (1 mL, 0.3 molL^{-1}) in 9 mL of acetone. The pellet was treated with aq. acetic acid (1 mL, 0.1%) for two hours to cleave the O6-Silyl protecting group. The solution was neutralised with sodium acetate (1 mL, 3 molL^{-1}) and DNA precipitated from 9 mL of ethanol. Finally, oligonucleotides were purified with RP-HPLC, desalted with NAP-5 size exclusion columns and analysed by ESI-MS (Table 7.1).

Table 7.1. Calculated and Experimental Results for ʃdI Modified DNA sequences

Sequence	Modification	RP-HPLC R_t (min)	Expected M_w (Da)	Experimental M_w (Da)
$G_4T_4GXG_2$	X = **8**	---	3923	3339.55
$G_3XT_4G_4$	X = **8**	---	3923	2511.57
$G_4T_4G_2XG$	X = **8**	24.012	3923	3923.48
$GXG_2T_4G_4$	X = **8**	---	3923	3128.51
$G_4T_4G_2XG$	X = **12**	24.226	3860	3963.63
$G_4T_4G_2XG$	X = **20**	24.368	3874	3977.68
$G_4T_4G_2XG$	X = **16**	15.878	3949	3949.66
$G_4T_4G_2XG$	X = **24**	---	3977	3789.61

RP-HPLC Conditions: 100 mmolL-1 TEAA buffer, pH 7.0 in Acetonitrile. 0 – 25%, 2 - 20 mins, 25 – 80%, 20 – 25 mins., Mass Spectrometry: 20 µmolL^{-1} DNA, 15% MeOH, positive mode.

General protocol for Fmoc-pyrrolidine-CE phosphoramidite modification



(i) DNA synthesis; (ii) Fmoc deprotection; (iii) Reductive amination using 4-pyridine carbaldehyde and $NaCNBH_3$; (iv) Cleavage from solid support and deprotection with aq. NH_4OH.

This material can be reacted with various aldehydes using this protocol. The current modification being tested is 4-pyridine carbaldehyde. The method was reported by Mokhir *et al*. (2001).[46]

Modified TG_4T (5 µmol) was synthesised using semi-manual addition of the pyrrolidine-CE phosphoramidite instead of the 5'-thymidine (Table 7.2). 64.1 mg (78 µmol/g) of preloaded dT CPG was loaded into the column and synthesis was carried out as described in Chapter 2. Synthesis was paused after detritylation prior to the addition of modification. Pyrrolidine-CE phosphoramidite (33.7 mg, final concentration: 0.1M) was weighed onto the column. The activator (400 µL, 0.45 $molL^{-1}$ ETT in acetonitrile) was added (400 µL) and the vacuum was held for thirty seconds to allow for phosphoramidite to be dissolved using a pipette. Coupling time for modified phosphoramidite was increased to 10 mins to ensure the reaction went to completion. Synthesis was then completed as previously described for unmodified phosphoramidites.

The column was removed and attached to a vacuum. Several solution were passed through the solid support. It was first treated with 10% diethylamine/acetonitrile solution (1 mL) for 10 mins to cleave the cyanoethyl protecting groups. The support was then treated with 20% piperidine/DMF (2 mL) for 10 mins to cleave the Fmoc protecting groups. The solid support was then washed with DMF and acetonitrile to remove any remaining reactants.

The solid support was suspended in 100 µL sodium acetate/acetic acid buffer (pH 6.0, 0.05 $molL^{-1}$). A solution of 7.5 µL of 4-pyridine carbaldehyde in 300 µL of DMF was added and this mixture was stirred at room temperature for 10 mins. A 100 µL solution of $NaCNBH_3$ (5 mg, 0.8 $molL^{-1}$) in DMF was added and the reaction mixture was stirred for a further 30 mins. The solid support was filtered and washed with DMF and methanol. DNA was cleaved with 28% aq. NH_4OH. Ammonia and water were evaporated using a speed-vac (Eppendorf Concentrator Plus), and the residue was suspended in Milli-Q H_2O (500 µL). This mixture was centrifuged (5000 rpm, 5 mins) to remove insoluble protecting

groups, and oligonucleotides were purified using RP-HPLC. The fractions were desalted with NAP-5 size exclusion columns, and analysed with ESI-MS (Table 7.2).

Table 7.2 Calculated and Experimental Results for Pyrrolidine Modified DNA sequences

Sequence	Modification	RP-HPLC R_t (min)	Expected M_w (Da)	Experimental M_w (Da)
XG_4T	X = 4-pyridine	12.377	1828	1828.35

RP-HPLC Conditions: 100 mmolL-1 TEAA buffer, pH 7.0 in Acetonitrile. 0 – 25%, 2 - 20 mins, 25 – 80%, 20 – 25 mins., Mass Spectrometry: 20 µmolL^{-1} DNA, 15% MeOH, positive mode.

8. References

1. Watson, J.D., Crick, F.H.C. (1953). Molecular Structure of Nucleic Acids: A Structure of Deoxyribose Nucleic Acid. *Nature*, 171, 737 – 738.
2. Hoogsteen, K. (1963). The Crystal and Molecular Structure of a Hydrogen-Bonded Complex Between 1-Methylthymine and 9-Methyladenine. *Acta. Cryst.*, 16, 907 – 916.
3. (a) Neidle, S., Balasubramanian, S. (2006) *Quadruplex Nucleic Acids*. Cambridge: Royal Society of Chemistry. (b) Petracone, L., Erra, E., Duro, I., Esposito, V., Randazzo, A., Mayol, L., Matia, C.A., Barone, G., Giancola, C. (2005). Relative Stability of Quadruplexes Containing Different Number of G-tetrads. *Nucleosides, Nucleotides & Nucl. Acids.*, 24, 757 – 760. (c) Nutiu, R., Li, Y.F. (2003). Structure-Switching Signalling Aptamers. *J. Am. Chem. Soc.*, 125, 4771.
4. (a) PDB ID: 148D.
   Schultze, P., Macaya, R.F., Feigon, J. (1994). Three-dimensional Solution Structure of the Thrombin Binding Aptamer d(GGTTGGTGTGGTTGG). *J. Mol. Bio.*, 235(5), 1532 – 1547.
   (b) PDB ID: 2GWQ.
   Lee, M.P.H., Haider, S., Parkinson, G.N., Neidle, S. Crystal structure of D(G4T4G4) with four and six quadruplexes in the asymmetric unit.
   (c) PDB ID: 2O4F.
   Creze, C., Rinaldi, B., Haser, R., Bouvet, P., Gouet, P. (2007). Structure of a d(TGGGGT) quadruplex crystallized in the presence of Li+ ions. *Acta. Crystallogr. Sect. D*, 63(6), 682 – 688.
5. Sengar, A., Heddi, B., Phan, A.T. (2014). Formation of G-quadruplexes in Poly-G Sequences: Structure of a Propeller-Type Parallel-Stranded G-quadruplex Formed by a G_{15} Stretch. *Biochemistry*, 53(49), 7718 – 7723.
6. Hou, X., Wu, W., Duan, X.. Liu, N., Li, H., Fu, J., Dou, S., Li, M., Xi, G. (2015). Molecular Mechanism of G-quadruplex Unwinding Helicase: Sequential and Repetitive Unfolding of G-quadruplex by Pif1 Helicase. *Biochem. J.*, 466, 189 – 199.
7. Jiang, H., Ju, Z., Rudolph, K.L. (2007). Telomere Shortening and Ageing. *Z. Gerontol. Geriatr.*, 40(5), 314 – 324.
8. Runge, K. W., Zakian, V. A. (1989). Introduction of Extra Telomeric DNA Sequences into *Saccharomyces Cerevisae* Results in Telomere Elongation. *Mol. Cell. Biol.*, 9(4), 1488 – 1497.
9. Zhou, J.-Q., Monson, E.K., Teng, S.-C., Schulz, V.P., Zakian, V.A. (2000). Pif1p Helicase, a Catalytic Inhibitor of Telomerase in Yeast. *Science*, 289(5480), 771 – 774.
10. Qi, H., Lin, C-P., Fu, X., Wood, L. M., Liu, A. A., Tsai, Y-C., Chen, Y., Barbieri, C. M., Pilch, D. S., Liu, L. F. (2006). G-Quadruplexes Induce Apoptosis in Tumor cells. *Cancer. Res.*, 66(24), 11808 – 11816.
11. Gagou, M., Gnaesh, A., Thompson, R., Phear, G., Sanders, C., Meuth, M. (2011). Suppression of Apoptosis by Pif1 Helicase in Human Tumor Cells. *Cancer Res.*, 71(14), 4998 – 5008.
12. Han, H., Hurley, L.H. (2000). G-quadruplex DNA: a Potential Target for Anti-cancer Drug Design. *Trends Pharmacol. Sci.*, 21(4), 136 – 42.
13. Métifiot, M., Amrane S., Litvak, S., Andreola, M-L. (2014). G-Quadruplexes in Viruses: Function and Potential Therapeutic Applications. *Nucleic Acids Res.*, 42(20), 12352 – 13266.
14. Fleming, A.M., Ding, Y., Alenko, A., Burrow, C.J. (2016). Zika Virus Genomic RNA Possesses Conserved G-Quadruplexes Characteristic of the Falviciradae Family. *ACS. Infect. Dis.*¸2, 674 – 681.

15. Okamoto, A., Ochi, Y., Saito, I. (2005). Fluorometric Sensing of the salt-induced B-Z DNA Transition by Combination of two Pyrene-Labeled Nucleobases. *Chem. Comm.*, 7(9), 1128 – 1130.
16. Tainaka, K., Tanaka, K., Ikeda, S., Nishize, K., Unzai, T., Fujiwara, Y., Saito, I., Okamoto, A. (2007). PRODAN-Conjugated DNA: Synthesis and Photochemical Properties. *J. Am. Chem. Soc.*, 129(15), 4776 – 4784.
17. Prestinari, C., Richert, C. (2011). Intrastrand Locks Increase Duplex Stability and Base Pairing Selectivity. *Chem. Comm.*, 47 (38), 10824 – 10826.
18. DeCorte, B. L., Tsarouhtsis, D., Kuchimanchi, S., Cooper, M. D., Horton, P., Harris, C. M., Harris, T.M. (1996). Improved Strategies for Postoligomerization Synthesis of Oligodeoxynucleotides Bearing Structurally Defined Adducts at the N2 Position of Deoxyguanosine. *Chem. Res. Toxicol.*, 9(3), 630 – 637.
19. Engelhard, D.M., Stratmann, L.M., Clever, G.H. (2018). Structure-Property Relationships of Cu^{II}-Binding Tetramolecular G-Quadruplex DNA. *Chem. Eur. J.*, 24, 2117 – 2125.
20. Engelhard, D.M., Pievo, R., Clever, G.H. (2013). Reversible Stabilization of Transition-Metal-Binding DNA G-Quadruplexes. *Angew. Chem. Int. Ed.*, 52, 12843 – 12847.
21. Engelhard, D.M., Nowack, J., Clever, G. (2017). Copper Induced Topology Switching and Thrombin Inhibition with Telomeric DNA G-Quadruplexes. *Angew. Chem. Int. Ed.*, 56, 11640-11644.
22. Largy, E., Mergny, J-L., (2014). Shape Matters: Size-exclusion HPLC for the Study of Nucleic Acid Structural Polymorphism. *Nucleic Acids Res.*, 42(19), e149.
23. Masiero, S., Trotta, R., Pieraccini, S., De Tito, S., Perone, R., Randazzo, A., Spada, G. P. (2010) A non-empirical chromophoric interpretation of CD spectra of DNA G-quadruplex structure. *Org. Biomol. Chem.*, 8, 2683 – 2692.
24. (a) Paramasivan, S., Rujan, I., Bolton, P.H. (2007) Cicrular Dichroism of Quadruplex DNAs: Applications to Structure, Cation Effects and Ligand Binding. *Methods*, 43, 324 – 331. (b) Kypr, J., Kejnovská, I., Renciuk, D., Vorlícková, M. (2009). Cicular Dichroism and Conformational Polymorphism of DNA. *Nucleic Acids Res.*, 37(6), 1713 – 1725.
25. Mergny, J. L., De Cian, A., Ghelab, A., Sacca, B., Lacroix, L. (2005) Kinetics of tetramolecular quadruplexes, *Nucleic Acids Res.*, 33, 81-94.
26. Collie, G. W., Parkinson, G. N. (2011) The Application of DNA and RNA G-quadruplexes to Therapeutic Medicines, *Chem. Soc. Rev.*, 40, 5867 – 5892.
27. Gannett, P. M., Sura, T. P. (1993) An Improved Synthesis of 8-Bromo-2'-deoxyguanosine. *Syn. Comm.*, 23, 1611 – 1615.
28. Silerme, S., Bobyk, l., Taverna-Porro, M., Cuier, C., Saint-Pierre, C., Ravanat, J-L. (2014). DNA-Polyamine Cross-links Generated upon One Electron Oxidation of DNA. *Chem. Res. Toxicol.*, 27, 1011 – 1018.
29. Münzel, M., Szeibert, C., Glas, A., Globisch, D., Carrell, T. (2011). Discovery and Synthesis of new UV-induced Intrastrand C(8-4)G and G(8-4)C Photolesions. *J. Am. Chem. Soc.*, 133(14), 5186 – 5189.
30. Iannazzo, L., Soroka, D., Triboulet, S., Fonvielle, M., Compain, F., Dubée, V., Mainardi, J-L., Hugonnet, J-E., Braud, E., Arthur, M., Etheve-Quelquejeu, M. (2016) Routes of Carbapenems for Optimizing Both Inactivation of L,D-Transpeptidase LdtMt1 of Mycobacterium tuberculosis and the Stability toward Hydrolysis of β-Lactamase BlaC. *J. Med. Chem.*, 59, 3427 – 3438.
31. Held, H., Roychowdhury, A., Benner, S. (2003). C-5 Modified Nucleosides: Direct Insertion of Alkynyl-Thio functionality in Pyrimidines. *Nucleosides, Nucleotides & Nucl. Acids.*, 22, 391 – 404.

32. Petricci, E., Radi, M., Corelli, F., Botta, M. (2003). Microwave-Enhanced Sonogashira Coupling Reaction of Substituted Pyrimidinones and Pyrimidine Nucleosides. *Tetrahedron Lett.*, 44, 9181 – 9184.
33. Zhang, W., Gao, Q., Wei, S., Fu, B., Yang, Q., Ming, X. (2018). Synthesis of 8-Substituted 2'-Deoxyisoguanosine via Unprotected 8-Brominated 2-Amino-2'-Deoxyadenosine. *Chem. Biodivers.*, 15, e1700335.
34. Garg, N. K., Woodroofe, C. C., Lacenere, C. J., Quake, S. R., Stoltz, B. M. (2005). A Ligand-free Solid-supported System for Sonogashira Couplings: Applications in Nucleoside Chemistry. *Chem. Commun.*, 36, 4551 – 4553.
35. Guermont, J. (1953). Preparation and catalytic hydrogenation of acetylenic amino ether. *Bulletin De La Societe Chimique De France*, 386 – 390.
36. Daniel, D., Middleton, R., Henry, H., Okamura, W. (1996). Inhibitors of 25-Hydroxyvitamin D_3-1α-Hydroxylase: A-Ring Oxa Analogs of 25-Hydroxyvitamin D_3[1]. *J. Org. Chem.*, 61(16), 5617 – 5625.
37. Benylles, A., Cairns, D., Cox, P.J., Kay, G. (2013). Three Salts from the Reactions of Cysteamine and Cystamine with L-(+)-Tartaric Acid. *Acta. Crys. C.*, 69, 658 – 664.
38. Aufort, M., Gonera, M., Le Gal, J., Czarny, B., Le Clainche., Thai, R., Dugace, C. (2011). Oxorhenium-Mediated Assembly of Noncyclic Selective Integrin Antagonists: a Combinatorial Approach. *ChemBioChem*, 12(4), 583 – 592.
39. Akaji, K., Fujino, K., Tatsumi, T., Kiso, Y. (1993). Total Synthesis of Human Insulin by Regioselective Disulfide Formation Using the Silyl Chloride-Sulfoxide Method. *J. Am. Chem. Soc.*, 115, 11384 – 11392.
40. Andreu, D., Albericio, F., Solé, N. A., Munson, M. C., Ferrer, M., Barany, G. (1994). Formation of Disulfide Bonds in Synthetic Peptides and Proteins. *Methods. Mol. Biol.*, 35, 91 – 169.
41. Annis, I., Hargittai, B., Barany, G., (1997). Disulfide Bond Formation in Peptides. *Methods Enzymol.*, 289, 198 – 221.
42. Cline, D. J., Thorpe, C., Schneider, J. P. (2004). General Methods for Facile Intramolecular Disulfide Formation in Synthetic Peptides. *Anal. Biochem.*, 335, 168 – 170.
43. Zhou, J., Rosu, F., Amrane, S., Korkut, D. N., Gabelica, V., Mergny, J. L. (2014) Assembly of chemically modified G-rich sequences into tetramolecular DNA G-quadruplexes and higher order structures, *Methods (Amsterdam, Neth.) 67*, 159 – 168.
44. Miyoshi, D., Karimata, H., Wang, Z-M., Koumoto, K. Sugimoto, N. (2007). Artificial G-Wire Switch with 2,2'-Bipyridine Units Responsive to Divalent Metal Ions. *J. Am. Chem. Soc.*, 129, 5919 – 5925.
45. Müller, J. (2008). Metal-Ion-Mediated Base Pairs in Nucleic Acids. *Eur. J. Inorg. Chem.*, 3749-3763.
46. Mokhir, A. A., Tetzlaff, C. N., Herzberger, S., Mosbacher, A., Richert, C. (2001). Monitored Selection of DNA-Hybrids Forming Duplexes with Capped Terminal C:G Base Pairs. *J. Comb. Chem.*, *3*, 374 – 386.
47. Trevani, L. N., Roberts, J. C., & Tremaine, P. R. (2001). Copper (II)–Ammonia Complexation Equilibria in Aqueous Solutions at Temperatures from 30 to 250° C by Visible Spectroscopy. *J. Sol. Chem.*, *30*(7), 585 – 622.
48. Rorabacher, D. B., & Melendez-Cepeda, C. A. (1971). Steric effects on the kinetics and equilibriums of nickel (II)-alkylamine reactions in aqueous solution. *J. Am. Chem. Soc.*, 93(23), 6071-6076.
49. Rist, M., Amann, N., Wagenknecht, H-A. (2003). Preparation of 1-Ethynylpyrene-Modified DNA via Sonogashira-Type Solid-Phase Couplings and Characterization of the Fluorescence Properties for Electron-Transfer Studies. *Eur. J. Org. Chem.*, 2498 – 2504.

50. Stephenson, A., Partridge, A., Filichev, V. (2011). Synthesis of b-Pyrrolic-Modified Porphyrins and Their Incorporation into DNA. *Chem. Eur. J.*, 17, 6227 – 6238.
51. Cadena-Amaro, C., Delepierre, M., Pochet, S. (2005). Synthesis and Incorporation into DNA Fragments of the artificial Nucleobase, 2-amino-8-oxopurine. *Bioorg. Med. Chem. Lett.*, 15, 1069 – 1073.
52. Hwu, J., Huang, W., Lin, S., Tan, K., Hu, Y., Shieh, F., Bachurin, S., Ustyugov, A., Tsay, S. (2019). Chikungunya virus inhibition by synthetic coumarin–guanosine conjugates. *Eur. J. Med. Chem.*, 166, 136 – 143.
53. Kenta, N., Yasahiro, T. (2010). Method of Producing Pyridine Ethanol Derivative. *Japanese Patent No. JP 2010270008A*.
54. Chatterjee, S., Mallick, S., Buzzetti, F., Fiorillo, G., Syeda, T., Lombardi, P., Saha, K., Kumar, G.S. (2015). New 13-Pyridinealkyl Berberine Analogues Intercalate to DNA and Induce Apoptosis in HepG2 and MCF-7 cells Through ROS Mediated p53 Dependent Pathway: Biophysical, Biochemical and Molecular Modelling Studies. *RSC Adv.*, 5(110), 90632 – 90644.
55. Tatulchenkov M.Y., Prokhorenko, I.A., Kvach, M.V., Navakouski, M.E., Stepanova, I.A., Pilchenko, N.V., Gontarev, S.V., Sharko, O.L., Korchun, V.A, Shmanai, V.V. (2017). Phosphoramidite Reagents and Solid-Phase Supports Based on Hydroxyprolinol for the Synthesis of Modified Oligonucleotides. *Rus. J. Bioorg. Chem.*, 43(4), 377 – 387.
56. Jackson, W., Perlmutter, P., Smallridge, A. (1988). The Stereochemistry of Organometallic Compounds. XXXII. Hydrocyanation of Derivatives of Amino Alkynes. *Aust. J. Chem.*, 41(8), 1201 – 1208.
57. Kim, J.H., Shin, H.S., Kim, S.B., Hasegawa, T. (2004). Hydrogen-Bonding Networks of Dialkyl Disulfides Containing the Urea Moiety in Self-Assembled Monolayers. *Langmuir*, 20, 1674 – 1679.
58. Potta, T., Chun, C., Song, S-C. (2009) Chemically Cross-linkable Thermosensitive Polyphosphazene Gels as Injectable Materials for Biomedical Applications. *Biomaterials*, 30, 6178 – 6192.
59. Yuan, J., Liu, C., Lei, A. (2015). Oxidative Cross S-H/S-H Coupling: Selective Synthesis of Unsymmetrical Aryl *tert*-Alkyl Disulfides. *Org. Chem. Front.*, 2(6), 677 – 680.