UNIVERSITY OF
Southampton
School of Biological Sciences

Faculty of Medicine, Health and Life Sciences

School of Biological Sciences

Effect of Sequence on the Folding and Stability of DNA G-quadruplexes

Phillip Anthony Rachwal

Thesis submitted for the award of Doctor of Philosophy

May 2008

University of Southampton

ABSTRACT

Faculty of Medicine, Health and Life Sciences
School of Biological Sciences
Doctor of Philosophy

Effect of sequence on the folding and stability of DNA G-quadruplexes

Phillip A. Rachwal

Oligonucleotide sequences rich in guanines are known to fold into four-stranded structures that are based on stacks of hydrogen-bonded G-quartets. Sequences with the potential to adopt these structures are found at the end of chromosomes in telomeric DNA, as well as in a number of biologically significant genomic locations, including gene promoter regions. There is considerable interest in establishing whether G-quadruplexes have a natural regulatory role and also whether they could be targets for therapeutic intervention. G-rich sequences can form an extremely diverse range of quadruplex structures, which may vary in terms of the number of strands, the strand polarity and the conformation of the loop regions that join the G-tracts. Despite the frequency with which potential quadruplex-forming sequences occur within the genome, there is presently a limited understanding of the rules that govern the formation of these structures and their stability. This work has focused on investigating the effect of sequence on the formation and stability of DNA G-quadruplexes.

Loop length is known to be an important criterion in determining quadruplex stability and topology. This work first examines the properties of a series of model quadruplex-forming sequences that contain short loops, long loops or combinations of the two, investigating the resultant effects on quadruplex folding, stability and kinetics. Utilising a variety of biophysical techniques, the results highlight the importance of single nucleotide loops in determining quadruplex topology. In the sequences studied, the presence of one single-thymidine loop was sufficient to promote the other loops into an identical conformation, resulting in the formation of parallel-stranded structures. This may be significant given the frequency with which single-nucleotide loops are observed amongst genomic quadruplex-forming sequences.

Besides loop-length, loop sequence can also moderate quadruplex stability. The sequence effects of single nucleotide loops have been examined in both model and biologically relevant promoter sequences. The results show that quadruplex stability is sensitive to changes in single-nucleotide loop identity, with adenines significantly disfavoured over pyrimidine loops.

Finally, the importance of the loop regions on quadruplex folding is well documented, however little is known regarding the length of the G-tract. The properties of intramolecular G-quadruplexes that are formed by sequences with increasing G-tract lengths have been examined. The results reveal that there is no simple relationship between quadruplex stability and the length of the G-tracts, and that sequences containing longer G-tracts are likely to form heterogeneous populations of folded structures. When challenged with their complementary strand, several G-rich sequences preferentially form quadruplex over duplex.

Preface

Some of the results from this thesis, carried out between October 2004 and September 2007, have been presented in the following publications:

Rachwal, P.A. and Fox. K.R. (2007) Quadruplex melting. *Methods*. **43**, 291-301

Rachwal, P.A., Findlow, I.S., Werner, J.M., Brown, T. and Fox, K.R. (2007) Intramolecular DNA quadruplexes with different arrangements of short and long loops. *Nucleic Acids Res*. **35,** 4214-4122

Rachwal, P.A., Brown, T. and Fox, K.R. (2007) Sequence effects of single base loops in intramolecular quadruplex DNA. *FEBS Lett*. **581**, 1657-1660

Rachwal P.A., Brown, T. and Fox, K.R. (2007) Effect of G-tract length on the topology and stability of intramolecular DNA quadruplexes. *Biochemistry*. **46**, 3036-3044

Brown, N.M., **Rachwal, P.A.**, Brown, T. and Fox, K.R. (2005) Exceptionally slow kinetics of the intramolecular quadruplex formed by the *Oxytricha* telomeric repeat. *Org. Biomol. Chem*. **3**, 4153-4157

Acknowledgements

I am extremely grateful to Prof. Keith Fox for his help, encouragement and support over the past three years and also Prof. Tom Brown, Dr Howard Barton, Dr Joern Werner and Dr Stuart Findlow for their help, discussion and technical assistance throughout the course of this work.

In the lab, a huge thanks to the Daves, Rusling and Norris, as well as Andrew (and Percy!), Vicki, Annie, Magdy, Alex and numerous others who have made the Foxlab an enjoyable place to work. Finally I am indebted to Gemma for her patience, support and for putting up with the late nights, and yes, I can now finally say it is finished.

Funding for this research was provided by the BBSRC.

Table of Contents

Chapter 6. Effect of G-tract length on the stability and topology of intramolecular G-quadruplexes

List of Figures

List of Tables

Abbreviations

A	Adenosine
AFM	Atomic force microscopy
bp	Base pairs
BSA	Bovine serum albumin
C	Cytosine
CD	Circular dichroism
cps	Counts per second
DEPC	Diethylpyrocarbonate
DMEM	Dulbeccos modified eagles medium
DNA	Deoxyribonucleic acid
DTT	Dithiothreitol
EDTA	Ethylenediaminetetraacetic acid
FCS	Foetal calf serum
FRET	Fluorescence resonance energy transfer
G	Guanosine
G4P	Quadruplex-forming potential
GQN1	G-quartet nuclease 1
I	Inosine
kb	Kilobases
LNA	Locked nucleic acid
NHE	Nuclease hypersensitivity element
NMR	Nuclear magnetic resonance
NHE	Nuclease hypersensitivity element
nt	Nucleotides
OD	Optical density
PAGE	Polyacrylamide gel electrophoresis
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PNA	Peptide nucleic acid
PNK	Polynucleotide kinase
POT1	Protection of telomeres 1

PQS	Potential quadruplex-forming sequence
RNA	Ribonucleic acid
RNAi	Ribonucleic acid interference
rpm	Rotations per minute
SAP	Shrimp alkaline phosphatase
SPR	Surface plasmon resonance
T	Thymidine
TAMRA	Tetramethylrhodamine
TBE	Tris-boric acid-EDTA gel running buffer
TEBP	Telomere-end binding protein
TEMED	N,N,N,N'-Tetramethylethylenediamine
T_m	Melting temperature
TSS	Transcription start site
UTR	Untranslated region
UV	Ultraviolet

CHAPTER 1

Introduction

1.1 G-quadruplex nucleic acids

The vast majority of chromosomal DNA exists in the double-helical B-form, bound by Watson-Crick base pairs, however both DNA and RNA have the potential to adopt higher-order assemblies that involve alternative base-pairing arrangements. Oligonucleotide sequences rich in guanine are able to form stable four-stranded structures, known as guanine (G)-quadruplexes, that are based on a cyclic arrangement of hydrogen-bonded guanines. Topologically they are highly polymorphic, exhibiting a diverse array of folded structures, and mounting evidence suggests these four-stranded structures have a functional role *in vivo*, as they have been implicated in a number of key biological processes.

The ability of guanine to form higher order complexes has been known for nearly a hundred years (Bang, 1910). Analysis of diffraction patterns of gels produced by guanylic acid first led Gellert and co-workers to suggest guanines could associate in a cyclic, hydrogen-bonded array (Gellert *et al.*, 1962). This tetrameric arrangement was also used to interpret the diffraction patterns produced by poly(dG) and poly(dI) homopolymers (Arnott *et al.*, 1974; Zimmerman *et al.*, 1975), with stacked tetrads forming the basis of four-stranded, right-handed helices. Their formation and stability were uniquely dependent on the presence and identity of cations, postulated to sit between the planes of hydrogen-bonded tetramers (Miles and Frazier, 1978; Howard and Miles, 1982). In the late eighties, these four-stranded structures were first implicated in biology; natural DNA sequences from chromosomal telomeres and immunoglobulin switch regions were shown *in vitro* to form tetrahelical structures composed of guanine quartets (Sen and Gilbert, 1988; Sundquist and Klug, 1989; Williamson *et al.*, 1989; Sen and Gilbert, 1990). Until recently, guanine-rich telomeric regions of DNA have provided the focus of quadruplex research both in terms of the structures formed and their biological role. However increasing evidence suggests that quadruplex structures can form, or be induced to form in a number of other genomic

locations. In addition, a significant number of proteins have now been identified that are able to promote, stabilise or resolve G-quadruplex structures, providing further evidence of biological significance (Fry, 2007).

The implication of G-quadruplexes in several disease-associated sequences, together with their unique three-dimensional geometry, have highlighted these structures as targets for drug design. A number of small molecule ligands have now been uncovered which show a high level of selectively for quadruplex DNA. These compounds represent potential anti-cancer drugs, through quadruplex-mediated anti-telomerase effects and the regulation of oncogenic gene expression. Besides representing a target for drug molecules, G-quadruplexes may have therapeutic potential themselves, as several aptamers isolated by *in vitro* selection techniques have been shown to adopt a quadruplex structure (Bock *et al.*, 1992; Mazumder *et al.*, 1996; Bates *et al.*, 1999). Further to their biological potential, G-quadruplexes also represent an excellent module for the design of DNA-based nanotechnological devices and supramolecular assemblies (Alberti *et al.*, 2006), via their ability to self-associate into highly ordered complexes.

1.2 G-quadruplex structure

1.2.1 General features

The basic building block of the guanine quadruplex is the G-quartet. G-quartets form via the association of four guanines in a square planar array, in which each guanine forms two hydrogen bonds with its neighbouring bases (Figure 1.1a). This arrangement leads to an aggregation of negative charge in the central cavity, caused by the inward facing O^6 guanine carbonyls. The co-ordination of a centrally located cation is therefore essential for the stabilisation of these structures, as without it this cyclic arrangement would be energetically unfavourable. G-quartets then form effective stacking units, linked by the sugar-phosphate backbone, in a G-quadruplex structure (Figure 1.1a). These are not perfectly stacked, but exhibit a right handed helical twist of 30 °.

A G-quadruplex can be defined as a structure containing a core of at least two stacked G-quartets. Quadruplexes can adopt an array of different topologies, dependent on a number of intrinsic and extrinsic factors. These include the number of strands that associate (which can be one, two or four), the orientation of the strands (which may be

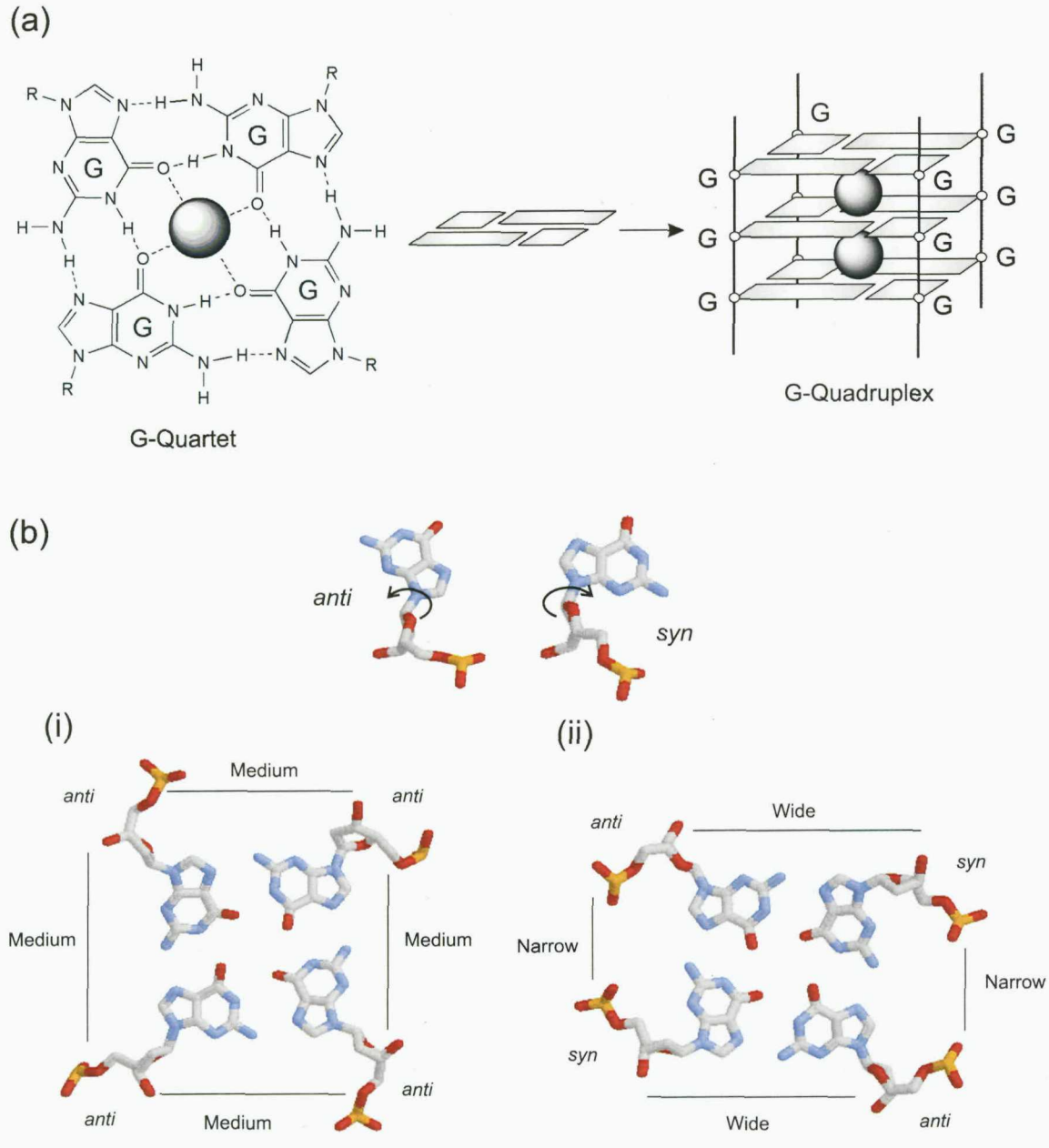


Figure 1.1 (a) The cyclic arrangement of a guanine quartet. G-quartets stack on top of each other, co-ordinated by the sugar-phosphate backbone and stabilised by centrally located cations, forming a G-quadruplex. **(b)** Rotation around the glycosilic bond allows guanine to form both *anti* and *syn* conformations. The distribution of glycosilic torsion angles and strand polarities gives rise to grooves of different width and depth. (i) Quartets with an all parallel strand arrangement have all *anti* glycosilic conformations, resulting in four medium groove of equal width. (ii) Quartets with exclusively alternating antiparallel strands result in two wide and two narrow grooves. Alternative antiparallel topologies can result in structures containing combinations of narrow, medium and wide grooves.

parallel or antiparallel) and the glycosilic bond angles (either *syn* or *anti*) (Figure 1.1b). Additionally, neighbouring strands within the quadruplex stem can be linked by loop regions, which can also adopt different conformations.

A further feature of G-quadruplexes is the presence of four grooves along each of the outer faces, which are of variable width and depth, depending on the topology of the quadruplex (Figure 1.1b). The grooves of quadruplexes containing all parallel strands are of equal width, however for quadruplexes containing antiparallel strands, groove widths can be variable (narrow, medium or wide), depending on the glycosilic distribution of the G-quartets. Accessibility to the grooves is dependent on the presence and conformation of loop regions. Loops are nucleotide sequences which link together the guanine tracts that participate in G-quartet formation, and these can be classified into three main groups (Figure 1.2); (i) Edgewise loops link two adjacent strands in a hairpin turn. (ii) Diagonal loops connect two opposing antiparallel strands. (iii) Double-chain reversal or propeller type loops connect the top and bottom of adjacent parallel strands. These loops project from the side of the quadruplex stem and can form additional interactions in the groove.

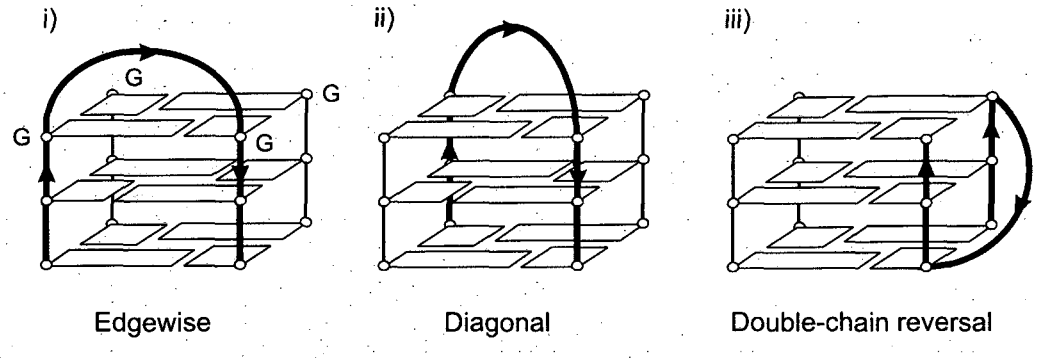


Figure 1.2 Quadruplex loop conformations.

1.2.2 Cation binding

G-quadruplex formation requires the presence of cations. Stacked G-quartets produce a regular geometry with a number of metal ion co-ordination sites within the central channel. Cations can interact with either four O^6 carbonyl oxygens, in the plane of a G-quartet, or with eight O^6 atoms, sandwiched between two tetrads. Quadruplex stability is highly dependent on the identity of the cation, which is an unusual trait amongst nucleic acid structures. The precise order of stability can differ, depending on the particular sequence, though in general for monovalent cations, $K^+ > Rb^+ \geq Na^+ > NH_4^+$

>> Li^+ (Venczel and Sen, 1993; Wlodarczyk *et al.*, 2005). Co-ordination of potassium ions produces more stable structures than sodium ions, and these two are the most commonly studied, due to their presence in biological milieu. For the intermolecular quadruplex d(TGGGT), substitution of sodium ions for potassium ions results in an increase in stability ($T_{1/2}$) of over 30 °C (Mergny *et al.*, 2005). A similar increase (20 °C) is also observed when substituting sodium for potassium ions in the intramolecular quadruplex d(TGGGTGGGTGGGTGGGT) (Rachwal *et al.*, 2007a). The preference for potassium over sodium ions was initially attributed to the better fit of the potassium ions within the central cavity. However consideration of the thermodynamics of ion binding have shown that the hydration energy of the cations is the major determinant of selectivity (Hud *et al.*, 1996). There is a greater energetic penalty in dehydrating Na^+ ions, therefore quadruplexes tend to favour potassium co-ordination. The ionic radii of the metal ions determine the co-ordination geometry; cations with a small ionic radius co-ordinate octahedrally with the oxygen atoms, in plane with the G-quartets. Cations with a larger ionic radius lie between adjacent quartets, with a bipyramidal co-ordination geometry. This has been established through high-resolution crystal and NMR structures, which revealed sodium ions (ionic radius 0.95 Å) are small enough to co-ordinate planar with the guanine bases (Laughlan *et al.*, 1994; Phillips *et al.*, 1997), whereas larger potassium and ammonium ions (1.33 Å and 1.43 Å) lie between two G-quartets (Halder *et al*, 2002; Parkinson *et al*, 2002; Sket and Plavec, 2007) (Figure 1.3). G-quadruplexes are also able to bind a number of divalent cations, including Pb^{2+}, $Ba^{2+}$and Sr^{2+}, (Venczel and Sen, 1997; Smirnov and Shafer, 2000b) although are destabilised by millimolar concentrations of Ni^{2+}, Co^{2+} and Mn^{2+} (Blume *et al.*, 1997).

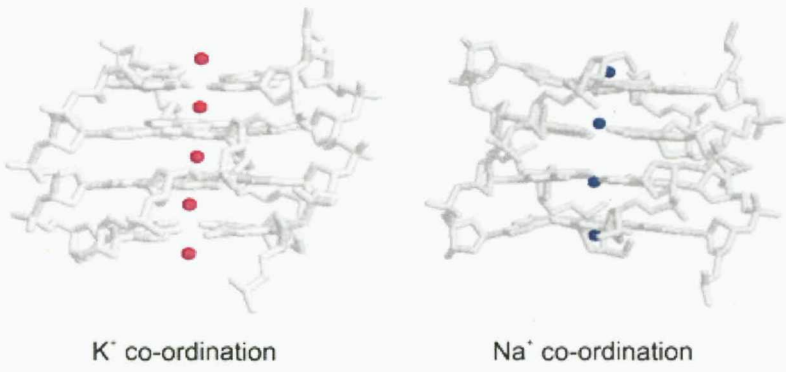


Figure 1.3 Cation co-ordination within G-quadruplexes as determined by X-ray crystallography. Potassium ions co-ordinate exclusively between the planes of two adjacent G-quartets. Sodium ion co-ordination can vary from in plane with a single quartet to equidistant between two tetrads, depending on electrostatic repulsion within the central channel. Crystal structures shown are the sodium form (PDB ID 1JB7) and potassium form (PDB ID 2GWQ) of the bimolecular quadruplex $d(GGGGTTTTGGGG)_2$.

1.2.3 Stoichiometry, folding and topology

G-quadruplexes assemble via the association of one, two or four oligonucleotide strands (Table 1.1). Tetramolecular quadruplexes represent the simplest category of quadruplex structures (Figure 1.4a), as these require the association of four individual strands, each containing a run of guanines. The sequence requirements can be represented as $N_aG_xN_b$, in which N_a and N_b are any non-guanine nucleotide of length a and b (although these can be zero), and G_x is any number of guanines participating in a G-quartet stack of length x. Quadruplexes of this type associate in a parallel stranded arrangement, with all sugar-phosphate backbones oriented in the same direction. High resolution crystal (Laughlan, *et al.*, 1994; Phillips *et al.*, 1997) and NMR (Aboul-ela *et al.*, 1992) studies of the sequence $d(TGGGGT)_4$ have highlighted this parallel organisation, and that all guanines adopt an *anti* glycosilic torsion angle. Unlike quadruplexes of higher molecularity, formation of tetramolecular quadruplexes is largely independent of the cation identity, with similar structures identified in the presence of Na^+, Li^+, Tl^+ and Ca^{2+} ions (Cáceres *et al.*, 2004; Creze *et al.*, 2007; Lee *et al.*, 2007). This motif is not limited to DNA, as the RNA hexanucleotide $r(UGGGGU)_4$ also forms an identical topology (Cheong and Moore, 1992). Tetramolecular quadruplexes can adopt more complex arrangements, as shown by the sequence d(GGGT) which lacks a 5' flanking nucleotide (Krishnan-Ghosh *et al.*, 2004). All strands are aligned in parallel, however eight strands assemble into an interlocked dimer, which also contains two terminal G•T•G•T tetrads.

Type	Sequence Requirements	Description
Tetramolecular	$(N_aG_xN_b)_4$	The association of four oligonucleotide strands, each containing a run of guanines of length x. N represents any non-guanine nucleotide of length a or b, although can be equal to zero.
Bimolecular	$(N_aG_xN_bG_yN_c)_2$	Two oligonucleotide strands containing two separate G-tracts of length x and y. Flanking nucleotides N_a and N_c may or may not be equal to zero. N_b represents any nucleotide sequence of length b involved in loop formation.
Unimolecular	$G_wN_aG_xN_bG_yN_cG_z$	A single oligonucleotide strand containing four blocks of guanine. N_a, N_b and N_c represent loop sequences of any length and identity.

Table 1.1 Sequence requirements for G-quadruplex formation

Bimolecular quadruplexes form through the association of two oligonucleotide strands. These form more diverse topologies than tetramolecular structures, many of which have been characterised by crystallography and NMR studies (Smith *et al.*, 1994; Haider *et al.*, 2002; Crnugelj *et al.*, 2003; Phan and Patel, 2003; Phan *et al.*, 2004). The topology assumed is largely dependent on the nature of the loop regions and the number of guanines in each G-tract. The sequence requirements for bimolecular quadruplexes can be defined as $N_aG_xN_bG_yN_c$, where N_a and N_c represent flanking nucleotides (which may or may not be present) and N_b represents any nucleotide of length b participating in loop formation (Table 1.1). Bimolecular quadruplexes are able to associate in several ways, outlined in Figure 1.4b. Edgewise 'hairpin' loops result in (i) head-to-head, or (ii) head-to-tail dimers in which alternate strands are antiparallel. In order to accommodate this topology, the guanines must either form alternate *anti-syn* alignments along the length of the strand, or all *anti* and all *syn* along adjacent connected strands, to retain an appropriate hydrogen-bonding alignment for G-quartet formation. (iii) Diagonal loops also connect opposing antiparallel strands. (iv) Bimolecular quadruplexes can adopt an all-parallel stranded configuration in which adjacent strands are connected by double chain reversal loops. There is also a further bimolecular topology which does not conform to the descriptor above (v), in which the quadruplex is composed of three G-tracts from one strand, with the remaining G-tract provided by a separate stand. This has been termed a 3+1 hybrid topology (Zhang *et al.*, 2005).

Unimolecular quadruplexes are comprised of a single folded strand of DNA or RNA, and topologically are the most complex. Four runs of guanine are linked by three loop regions which can vary in both length and sequence. The general requirements for the folding of an intramolecular quadruplex can be represented as $G_wN_aG_xN_bG_yN_cG_z$, with $N_{a\text{-}c}$ representing the length and identity of the loop regions. Unimolecular quadruplexes are typified by four particular topologies (Figure 1.4c); Antiparallel 'chair-type', in which all three loops are edgewise, antiparallel 'basket-type' where the central loop is a diagonal, parallel-stranded 'propeller-type' in which all the loops are double chain reversals, and the hybrid 3+1 topology, in which three strands are in one orientation, with the fourth stand in the opposite polarity. This strand alignment requires two edgewise and one double-chain reversal loop. For a sequence containing four equal runs of guanine, there are (at least) 26 possible unimolecular topologies available via the various loop combinations (Webba de Silva *et al.*, 2007), however the length and sequence of the loops can often impose restrictions on the final fold.

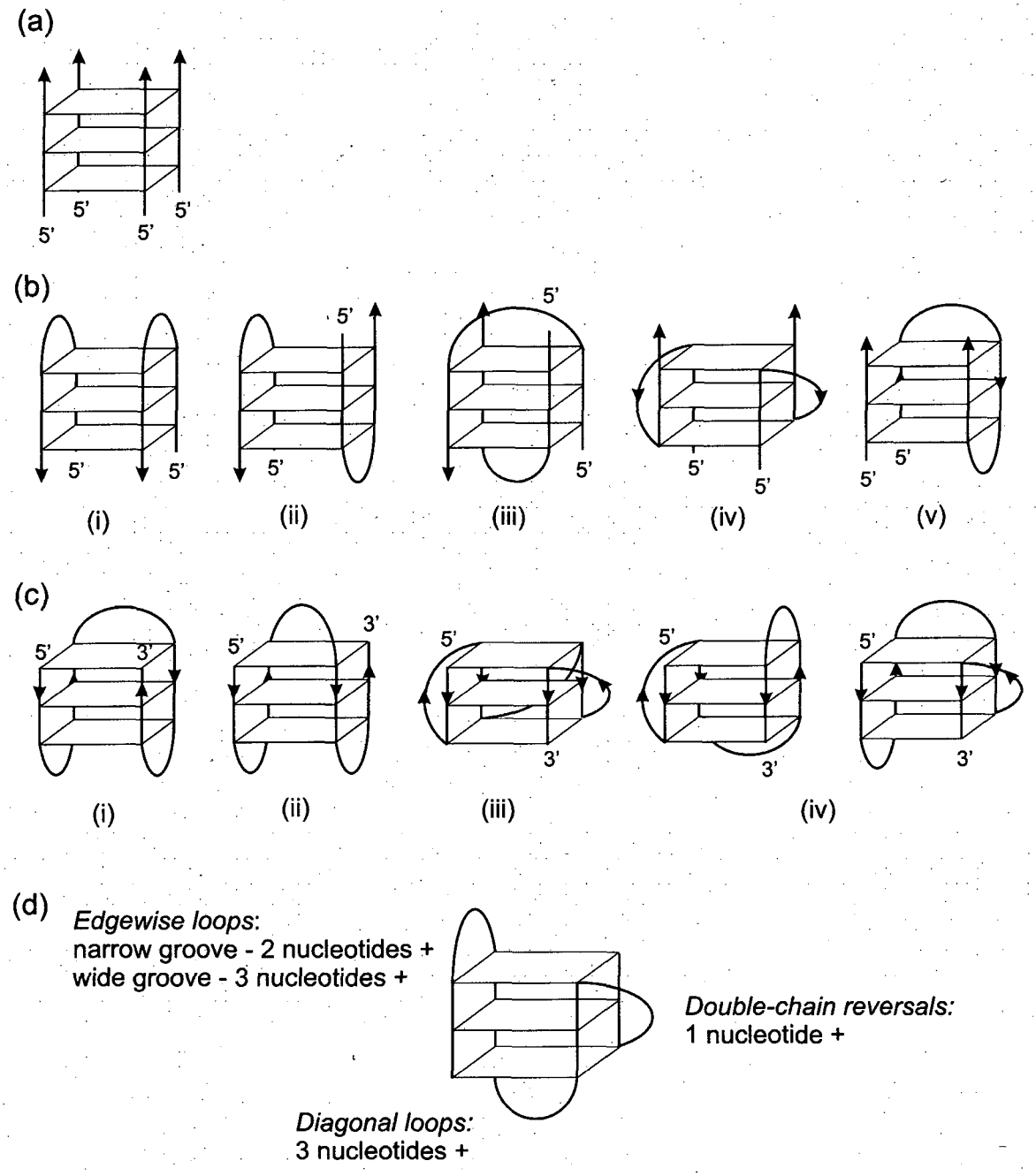


Figure 1.4 Quadruplex stoichiometry and topology. **(a)** Tetramolecular; all parallel strand alignment **(b)** Bimolecular topologies (i) Antiparallel head-to-head (ii) Head-to-tail (iii) Antiparallel diagonal (iv) All parallel (v) 3+1 hybrid **(c)** Intramolecular folds (i) Antiparallel 'chair-type' (ii) Antiparallel 'basket-type' (iii) Parallel 'propeller-type' (iv) 3+1 and 1+3 hybrid topologies. Although many others are possible, these represent the common folds observed to date by crystal or NMR studies. **(d)** The length of the loops determine which looping topologies quadruplexes are able to adopt; Edgewise loops require two nucleotides to span a narrow or medium groove, or three nucleotides to span a wide groove. Diagonal loops require at least three nucleotides, and double-chain reversals can form from a single nucleotide.

1.2.4 Quadruplex loops

1.2.4.1 Loop length

The topology adopted by a quadruplex-forming sequence can be dependent on the length of the loops. A combination of high-resolution structural analysis, together with biophysical and modelling studies has allowed first generation folding rules to be formulated. The first reports investigating loop-length-dependent folding involved a series of sequences related to the consensus family d($G_2N_nG_2N_nG_2N_nG_2$), the smallest of the potential unimolecular folds. The first to be structurally characterised was the thrombin-binding aptamer d(GGTTGGTGTGGTTGG), which forms two stacked G-quartets with an *anti-syn-anti-syn* glycosilic alignment (Macaya *et al.*, 1993; Wang *et al.*, 1993). Both the dinucleotide (T_2) peripheral loops and the central (TGT) trinucleotide loop formed edgewise linkers, resulting in an antiparallel 'chair-type' topology (Figure 1.4ci). The solution structure of the sequence d($G_2T_4G_2CAG_2GT_4G_2T$) which contains longer peripheral loops and a shorter central loop revealed a completely different unimolecular fold (Kuryavyi *et al.*, 2001). Still based on a two G-quartet stem, the first (T_4) and third (GT_4) loops were both diagonal, and the shorter second loop formed a double-chain reversal. Marathias and Bolton performed a more systematic study on a number of sequences in this family, in which the loops were replaced with between one and four nucleotides (Marathias and Bolton, 1999). It was concluded that short loops favoured a parallel-stranded fold and longer loops an antiparallel arrangement, although the precise topology was dependent on the identity of the cation.

More recently, studies using sequences of the type d($TG_3T_{1-7}G_3T_{1-7}G_3T_{1-7}G_3T$), in which loops were restricted to between one and seven thymidine residues, have allowed a greater insight into loop-structure relationships. Hazel and co-workers demonstrated using CD measurements and molecular modelling that a sequence containing all single thymidine loops constrains the topology to an all-parallel conformation (Hazel *et al*, 2005). A single nucleotide is unable to span across the face of a G-quartet either laterally or diagonally, therefore restricting these loop lengths to double-chain reversals (Figure 1.4d). Dinucleotide linkers are topologically less constrained as they can form edgewise, but not diagonal loops and therefore allow parallel or antiparallel conformations. Longer loops are able to form any loop type, although are more sensitive to ionic conditions.

Predicting which topology a quadruplex-forming sequence will adopt based on primary sequence alone is extremely complex, and expanding the current understanding of the loop-dependency of quadruplex folding is one of the aims of this work. Several recent publications have highlighted the importance of short loops in determining the folded topology; Single nucleotide loops are known to enforce a parallel-stranded arrangement, whereas longer loops allow more topological freedom. When both short and long loops are present in combination, single-nucleotide loops can induce the remaining loop(s) into a double-chain reversal, resulting in the formation of parallel-stranded structures (Rachwal *et al.*, 2007c (Chapter 4); Bugaut and Balasubramanian, 2008).

1.2.4.2 Thermal Stability

Besides influencing topology, loops can also play an important role in quadruplex stability. Core quadruplex stability is generated by hydrogen bonding between guanines and from stacking of the hydrophobic G-quartets, with longer G-runs forming more stable structures (Mergny *et al.*, 2005; Rachwal *et al.*, 2007a). However loop regions can also contribute via their ability to form stacking interactions between each other, with the G-quartets, and through the formation of intra-loop hydrogen bonds (Keniry *et al*, 1997). Quadruplex stability can also be related to the length of the loops. Short, single nucleotide loops form the most stable structures; the quadruplex formed by the sequence $dT(G_3T)_4$, with all single thymidine loops, exhibits a melting temperature (T_m) in excess of 90 °C under physiological-like ionic conditions (~100 mM K^+) (Risitano and Fox, 2003a; Hazel *et al.*, 2005). Stability decreases as the length of the loops increases although sequences containing equal length loops of 1-5 thymidines displayed T_m values in excess of 37 °C, suggesting they too may be stable under physiological conditions (Hazel *et al.*, 2005). Reports of quadruplexes with longer loops are less common due their low thermal stability, although one long loop can be compensated for by the presence of two short loops. Bourdoncle and co-workers have reported the formation of a stable quadruplex containing a 21-nucleotide central loop (Bourdoncle *et al.*, 2006), achieved by fixing the peripheral loops as d(TTA). Quadruplexes with loops this long are less likely to be of physiological relevance due to structural competition with the complementary strand, however they have been utilised as quadruplex-based devices for oligonucleotide detection (Bourdoncle *et al.*, 2006). The formation of inter and intra-loop Watson-Crick duplex regions can also permit the formation of stable structures with longer loops (Risitano and Fox, 2003b; McManus and Li, 2008).

In some instances it is possible to use the thermodynamic profile for the formation of the G-quartet stack to estimate the thermodynamic contribution of the loop regions. The loops in the quadruplex-forming thrombin-binding aptamer d(GGTTGGTGTGGTTGG) have been shown to contribute a favourable enthalpy term of between 10 and 15 kcal mol^{-1} (Olsen *et al*, 2006). Substituting nucleobase loops with neutral, non-nucleosidic linkers has also allowed the stabilising contribution of base-stacking interactions to be assessed (Risitano and Fox, 2004). Shorter nucleobase loops are more stable than their equivalent length non-nucleosidic linkers, however hexaethylene glycol loops were shown to be more stabilising than d(TTA), despite their similar length. Similar studies with bimolecular quadruplexes of the type $d(G_4\text{-loop-}G_4)_2$ also showed hexaethylene glycol linkers were stabilising when compared with T_4 loops, although loops containing abasic or propanediol loops were destabilising, resulting in a decrease in ΔH of 35–40 $kcal.mol^{-1}$ (Cevec and Plavec, 2005). The key factors for stability are therefore (i) the number of phosphates within in each loop, as the more stable hexaethylene loop contains two phosphates as opposed to four in a TTA loop. (ii) Both the base and the sugar, as loops devoid of these constituents, but containing the same number of phosphates are detrimental to quadruplex stability.

1.2.4.3 Loop sequence

The base composition of the loop regions can also be significant with regards to both stability and topology. Telomeric quadruplex-forming sequences consist of regular repeating units with predominantly thymine loops. This may be significant, as substitution of d(TTA) loops in the human telomeric repeat with d(AAA) results in complete quadruplex destabilisation (Risitano and Fox, 2003a), thought to be a result of the rigidity of poly-adenine tracts. Different bases contribute differently to quadruplex stability. Adenines tend to stack on top of the quadruplex stem, both in edgewise and diagonal loops (Wang and Patel, 1993) and can even do so as part of a double-chain reversal (Parkinson *et al.*, 2007) whereas thymines generally exhibit more rotational freedom (Spackova *et al.*, 1999). Guanine can be tolerated within the loops although its presence often leads to structural ambiguity, especially when adjacent to a G-quartet. Loop guanines can also make intimate contacts with the quadruplex stem that are crucial to the integrity of the structure. In one example, interchanging apparent loop regions within the same quadruplex was shown to result in complete destabilisation (Rankin *et al.*, 2005). However the NMR structure of this sequence revealed an isolated

guanine thought initially to reside within one of the loops, participates in G-quartet formation despite the presence of four three-guanine tracts (Phan *et al.*, 2007).

1.2.5 Quadruplex polymorphism

Many quadruplex-forming sequences adopt a number of different structures in solution. Quadruplexes are extremely sensitive to environmental conditions and the specific fold or folds adopted can depend on the identity of the cation, ionic strength, temperature, molecular crowding and even the method of preparation.

1.2.5.1 Cation-induced polymorphism

In addition to their stabilising role, different cations, in particular sodium and potassium are able to promote the formation of different quadruplex topologies. Bimolecular quadruplexes are less sensitive to cation switching, although some polymorphism has been observed. The sequence $d(GGGGTTTTGGGG)_2$ ($d[G_4T_4G_4]_2$) forms a bimolecular diagonal hairpin structure in the presence of sodium, potassium and ammonium ions (Schultze *et al*, 1994; Smith and Feigon, 1992; Schultze *et al.*, 1999). However the different binding geometry of the cations result in alternative loop-stem interactions. Sodium ions sit in plane with the terminal quartets, which allow interaction with the loop thymine (T3) O^2 carbonyls. Substitution of sodium for potassium (or ammonium) results in a less rigid loop organisation, as the larger co-ordinated ions sit below the plane of the G-quartets preventing loop-ion interactions.

Cation-induced polymorphism is more notable in intramolecular quadruplexes, in which complete topological rearrangements can occur. Sodium-stabilised structures commonly adopt an antiparallel fold (Wang and Patel, 1993; Smith and Feigon, 1992), whereas potassium can generate both parallel (Parkinson *et al.*, 2002) and mixed strand-polarity structures (Wang and Patel, 1994; Ambrus *et al.*, 2006; Luu *et al.*, 2006). Unimolecular quadruplex-forming sequences that are unrestricted by loop length rarely form the same topology in the presence of sodium and potassium (Dapic *et al.*, 2003). The reasons for the dramatic differences between the two are still largely unclear, although changes in the diameter of the G-quartets, together with specific loop-ion interactions are likely contributing influences (Kettani *et al.*, 1998; Bouaziz *et al.*, 1998).

1.2.5.2 Sequence ambiguity

Sequences in which the G-tracts are of unequal length are a further contributor to quadruplex polymorphism. Uneven runs of guanine result in some G's participating in G-quartet formation, with others residing within the loops (Figure 1.5a), and strand slipping can occur between the structural variants. Structural analysis of sequences of this type frequently lead to the identification of multiple folded species, and isolation of a unique fold is only achieved by mutating the additional guanines to restore parity in the lengths of the G-tracts (Ambrus *et al.*, 2005; Dai *et al.*, 2006).

A further consideration is that G-quadruplexes are not limited to sequences containing four blocks of guanine. When presented with a fifth run of guanines (or more), there is considerable scope for polymorphism, as several combinations of guanine blocks are able to participate in G-quartet formation (Figure 1.5b). This situation is frequently observed in genomic DNA regions, resulting in the formation of multiple folded species that co-exist in dynamic equilibrium (Seenisamy *et al.*, 2004; Sun *et al.*, 2005; Dai *et al.*, 2006). The supplementary blocks of guanine can either remain as flanking sequence, or can be looped out in one of the linker regions. Sequences containing multiple G-tracts that are also of unequal length give rise to a huge potential number of structures, although favourable loop lengths may determine the predominant fold.

(a)

(i) 5'-GGG NNN GGGG NNN GGG NNN GGG

(ii) 5'-GGG NNN GGGG NNN GGG NNN GGG

(b)

(i) 5'-GGG NNN GGG NNN GGG NNN GGG NNN GGG NNN GGG

(ii) 5'-GGG NNN GGG NNN GGG NNN GGG NNN GGG NNN GGG

(iii) 5'-GGG NNN GGG NNN GGG NNN GGG NNN GGG NNN GGG

(iv) 5'-GGG NNN GGG NNN GGG NNN GGG NNN GGG NNN GGG

loop

Figure 1.5 (a) Uneven G-tract lengths can result in strand slipping, as additional guanines can form part of the loop, or the G-quartet. **(b)** Extended G-repeats often lead to the formation of multiple folded complexes in solution. N = arbitrary (non-guanine) loop nucleotide. Figure adapted from Burge *et al.*, 2006.

1.2.6 Alternative structures

Quadruplex structures do not necessarily have to consist of exclusive G•G•G•G quartets; the quadruplex fold is both versatile and robust, and other bases can participate in quartet formation. Both guanine discontinuities and mixed sequence tetrads can be tolerated within the structure, and additional triads, pentads and hexad alignments can all contribute to quadruplex formation and stability.

1.2.6.1 Mixed Tetrads

There are several examples of quadruplexes that contain G•C•G•C tetrads. These consist of a pair of Watson-Crick base pairs, aligned through either their major or minor groove edges (Figure 1.6a). Two G•C•G•C tetrads have been identified in the bimolecular hairpin quadruplex formed by the sequence d(GGGCT$_4$GGGC) (Kettani *et al.*, 1998; Bouaziz *et al.*, 1998). These mixed tetrads form the outermost quartets, with the central two made up of all-G-quartet alignments. A similar arrangement has been recognized in a sequence containing the fragile X syndrome triplet repeat d(GCGGT$_3$GCGG), however on this occasion the mixed tetrads form the central quartets, and these are sandwiched between terminal G-quartets (Kettani *et al.*, 1995). The scaffold provided by the G•G•G•G quartets enables additional base tetrad structures to form; with examples of A•T•A•T (Zhang *et al.*, 2001), T•A•A•T (Webba da Silva, 2005), A•A•A•A (Pan *et al.*, 2003), T•T•T•T (Cáceras *et al.*, 2004) and U•U•U•U (Pan *et al.*, 2006) all reported to form either above or below a stem of G-quartets. A comprehensive study of the result of guanine modifications to G-quadruplex stability has revealed G-quartet modifications and substitutions are almost always destabilising (Gros *et al.*, 2007), and their presence is a result of the rigid geometry provided by the neighbouring G-quartets.

1.2.6.2 Pentads, hexads and capping structures

G-tetrads have the capacity to form additional hydrogen bonds via their external edge. NMR studies have revealed bases involved in double-chain reversal loop formation can project into the groove, and interact with the minor groove edge of the G-quartets (Figure 1.6b). Examples of such alignments include A•(G•G•G•G) pentads (Phan *et al.*, 2005), A•(G•G•G•G)•A hexads (Kettani *et al.*, 2000) and even G(A)•G(A)•G(A)•G

heptads, in which three edges of the G-quartet are bound by loop adenines (Matsugami *et al.*, 2001). These interactions are not limited to adenine bases, as thymine loops can also participate in hexad formation (T•(G•G•G•G)•T) (Webba da Silva, 2005; Mergny *et al.*, 2006). It is noticeable that for these extended G-quartets to form, the double-chain reversal loops must span only two tetrads, presumably to achieve the correct base alignment. Guanine can also be engineered to form G-pentads using the non-standard nucleobase iso-guanine (Chaput and Switzer, 1999). Two hydrogen bonds are retained between neighbouring guanines, although the central cavity is larger, favouring Caesium ion co-ordination. A quadruplex stem may also be 'capped' by additional interactions between loop residues. Besides Watson-Crick and mismatch base-pairing, loop triples (Kuryavyi *et al.*, 2001; Dai *et al.*, 2007) and triads (Kuryavyi *et al.*, 2000; Phan *et al.*, 2005) can be essential determinants of the final folded topology.

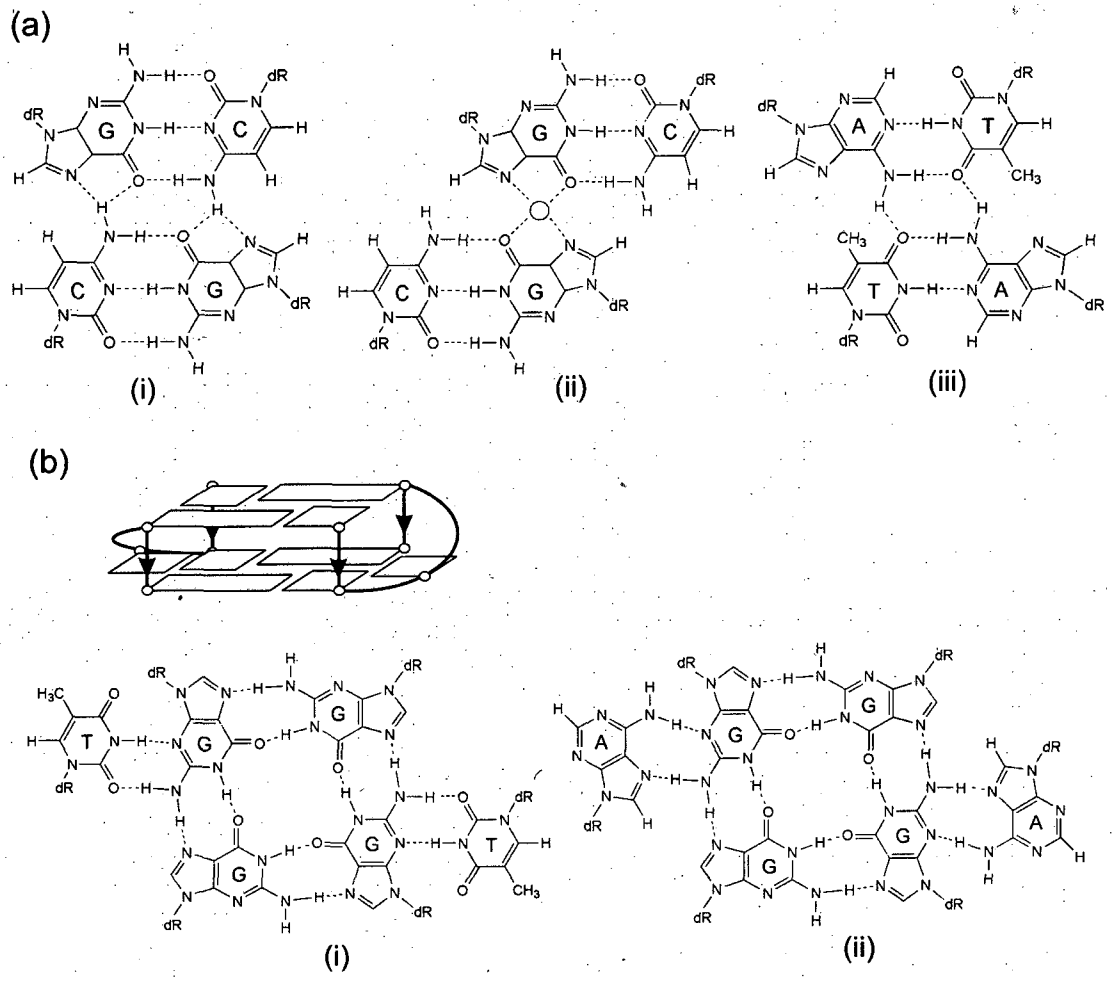


Figure 1.6 (a) Schematic representation of mixed tetrad alignments. G•C•G•C quartets can form with either a direct (i) or slipped (ii) alignment of the major groove edges. The slipped structure requires a co-ordinated cation. (iii) A•T•A•T tetrad formation. **(b)** (i) Adenines or (ii) thymines participating in double-chain reversal loop formation between two tetrads can co-ordinate with the outer face of the G-quartets, forming pentads, hexads or heptads.

1.2.6.3 i-motif

An alternative four-stranded structure can be composed entirely of cytosines. Rather than forming a cyclic array in the manner of a G-quartet, cytosines are able to form series of intercalated $C{\bullet}C^+$ base pairs known as an i-motif (Figure 1.7). Structures of this type were first revealed by NMR analysis of the DNA hexamer d(TCCCCC) (Gehring *et al.*, 1993). At acidic pH, this sequence was shown to form a four-stranded complex composed of two parallel, hemiprotonated duplexes intercalated within each other in a head-to-tail arrangement. Crystal structures of the sequences d(CCCC) and d(CCCT) also confirmed this zipped duplex arrangement (Chen *et al.*, 1994; Kang *et al.*, 1994) which contains two narrow and two wide grooves. I-motifs can form tetra, bi and intramolecular complexes composed of as little as two base-pairs (Leroy and Gueron, 1995), although stability increases with the length of the C-tracts (Mergny *et al.*, 1995). The sequence requirements for i-motif formation are essentially the same as those for G-quadruplexes (switching G for C), therefore the Watson-Crick complement of any quadruplex-forming sequence has the potential to adopt an i-motif. This is rarely observed under physiological conditions, as i-motif structures exhibit only marginal stability at neutral pH due the low pK_a of cytosine; increasing the pH by one unit results in a drop in stability of around 20-25 °C (Mergny *et al.*, 1995). Nonetheless, it has been suggested that some biological processes may lead to localised changes in pH that could stabilise pH-dependent DNA structures such as i-motif (Völker *et al.*, 2001). As such, the simultaneous formation of both G-quadruplex and i-motif structures may represent viable alternatives to Watson-Crick duplex.

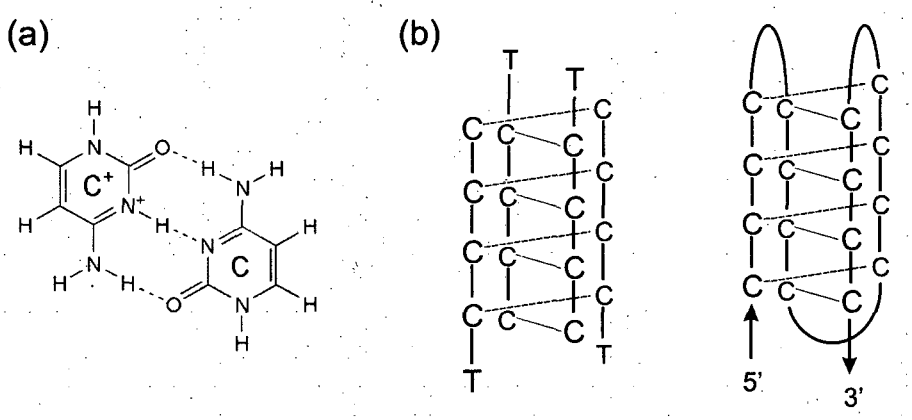


Figure 1.7 (a) $C{\bullet}C^+$ base pair. **(b)** At low pH, C-rich oligonucleotides can associate into four-stranded i-motif formation. Two hemiprotonated parallel duplexes associate in an antiparallel alignment bound by intercalated $C{\bullet}C^+$ base pairs. These can form both inter and intramolecular complexes.

1.3 Quadruplex-duplex competition

The thermal stability of quadruplex structures has prompted several groups to question whether quadruplexes could compete with duplex formation. A number of studies have focused on the structural predominance of human telomeric DNA, composed of d(GGGTTA)/(TAACCC)$_n$ repeats, as both strands have the capacity to form tetrahelical structures. Using equimolar concentrations of the telomeric oligonucleotides d[AGGG(TTAGGG)$_3$] and d[(CCCTAA)$_3$CCCT], Phan and Mergny showed that at near-physiological pH, ionic strength and temperature, the double helix was the predominant form (Phan and Mergny, 2002). However at lower pH or higher temperatures, the G-quadruplex and i-motif were able to effectively compete with the duplex. A structural and thermodynamic study of similar oligonucleotides also revealed the competition is dependent on pH, and the identity and concentration of the cation (Li *et al.*, 2002). Several other quadruplex-forming sequences have been found to preferentially form a quadruplex in the presence of their Watson-Crick complement (Datta and Armitage, 2001; Risitano and Fox, 2003a). These structures generally have shorter loops than the human telomeric repeat, generating more stable quadruplexes and less stable duplexes. The quadruplex formed by the oligonucleotide d[(GGGT)$_3$GGG], which contains all single nucleotide loops, persists even in the presence of a 50-fold excess of its complementary strand (Risitano and Fox, 2003a). Complementary strands composed of PNA or LNA however shift the equilibrium towards the duplex state, forming extremely stable PNA-DNA, or LNA-DNA chimeras (Datta and Armitage, 2001; Kumar and Maiti, 2007). The equilibrium between quadruplex and duplex structures may be further influenced by other intracellular conditions, including supercoiling, protein binding, or molecular crowding. Molecular crowding has been proposed to destabilise telomeric duplex DNA, which may provide an opportunity for quadruplex formation (Miyoshi *et al.*, 2004).

Besides their biological significance, the interconversion between DNA quadruplex and duplex secondary structures has been harnessed to generate simple nanomachines in the field of nanotechnology. Alberti and Mergny described a device capable of extension-contraction movements composed of a 21 nucleotide DNA sequence (Alberti and Mergny, 2003). The nanomachine could be cycled between the quadruplex closed-state and the duplex open-state by the sequential addition of single strands, termed 'C-fuel' and 'G-fuel', with fluctuation between the two states inducing a 5 nm lateral movement.

Studies to further characterise the competition between quadruplex and duplex structures may also help to improve the design and efficiency of these nanodevices.

1.4 Quadruplex kinetics

Besides the secondary structure, the kinetics of folding and unfolding are important aspects in the biological function and therapeutic applications of nucleic acids. The kinetic features of quadruplex association and dissociation are unique in their sequence dependence and metal ion requirements.

1.4.1 Tetramolecular quadruplexes

Tetramolecular parallel-stranded quadruplexes are known to exhibit extremely slow folding and unfolding kinetics. The mechanism by which the strands associate is still not fully understood, with a number of models proposed. The association process has been shown to be fourth order with respect to monomer, but this does not simply imply that a four-strand collision step occurs (Mergny *et al.*, 2005). Wyatt and co-workers suggested that four G-rich single strands are in fast equilibrium with two dimers, which then associate (in the rate limiting step) to form a quadruplex (Wyatt *et al.*, 1996). The identity of this dimer intermediate is as yet unknown, although molecular dynamics simulations suggest it will be present in extremely low levels (Stifl *et al.*, 2003). Other models include the sequential addition of single strands into duplex and triplex intermediates (Bardin and Leroy, 2007), or even the disproportionation of two triplexes, forming a duplex and a quadruplex (Hardin *et al.*, 1997). Tetramolecular quadruplex are extremely stable once formed. Structures containing four of more G-quartets often persist even at 95 °C, therefore obtaining accurate T_m data for such complexes can be difficult. Equivalent RNA structures are even more stable, possibly due to their predisposition for adopting the *anti* glycosilic conformation (Sacca *et al.*, 2005). Several groups have used the quasi-irreversible melting behaviour of tetramolecular structures to examine the effects of various parameters on the association and dissociation processes individually (Mergny *et al.*, 2005; Merkina and Fox, 2005; Petraccone *et al.*, 2005). Based on sequences related to $d(TG_nT)$, it has been shown that longer G-tracts, higher ionic strengths ($Na^+ > K^+ > Li^+$) and higher strand concentrations all accelerate the association process, whereas higher temperatures, flanking nucleotides and guanine base modifications all increase dissociation rates. Single-site guanine substitutions with

other natural and non-natural bases have shown the central quartet is the most important with regards to the nucleation process (Gros *et al.*, 2007). Base modifications at the 5'-end were more deleterious to the association rate than at the 3', suggesting the rate-limiting step involves the 5'-end of the oligonucleotide.

1.4.2 Uni and bimolecular quadruplexes

Unimolecular and bimolecular quadruplexes add a further complexity, as multiple conformations can exist in equilibrium. In comparison to linear four-stranded structures, unimolecular quadruplexes fold more rapidly (Mergny *et al.*, 1998), although there is less data regarding their rates of association and dissociation. Raghuraman and Cech monitored the unfolding of the sequence $d(TTTTGGGG)_4$, which contains four repeats of the *Oxytricha* telomeric DNA sequence, by capturing the released strand with its Watson-Crick complement, reporting half lives of around 4 and 18 h at 37 °C in 50 mM sodium or potassium respectively (Raghuraman and Cech, 1990). A similar approach using PNA to trap the unfolding human telomere repeat quadruplex, gave dissociation half-lives of 30 mins and 40 h at 20 °C in the presence of 100 mM Na and K respectively (Green *et al.*, 2003). G-quadruplex unfolding can also be monitored by surface plasmon resonance (SPR) (Zhao *et al.*, 2004; Halder *et al.*, 2005) and NMR, which enables multiple folded species that may be in a dynamic equilibrium to be monitored simultaneously. Phan and Patel revealed that a bimolecular quadruplex formed from two repeats of the human telomeric sequence d(TAGGGTTAGGGA) can form two distinct bimolecular structures, one parallel and one antiparallel stranded (Phan and Patel., 2003). Temperature-jump, concentration-jump and complementary-strand trapping analysis showed that at physiological temperatures, the antiparallel quadruplex folds faster, but unfolds slower than the parallel stranded structure, although at higher temperatures the parallel form is favoured.

In general, quadruplex folding and unfolding is known to be a slow process, although folding and unfolding times can range from seconds to years at 37 °C, depending on a variety of parameters. Knowledge of the association and dissociation parameters in a physiological context may be invaluable for a variety of applications, including the study of small molecule and protein binding interactions.

1.5 G-quadruplexes in Biology

The readiness with which guanine-rich sequences can form quadruplex structures *in vitro* suggests that they may be able to form *in vivo*, and evidence is mounting to support the intracellular existence of quadruplex DNA. G-rich sequences have been identified in a number of important functional genomic locations, including telomeres, immunoglobulin heavy-chain switch regions, recombination hotspots, regulatory elements within oncogene promoters and enhancers, and also elements involved in triplet repeat expansion disease. *In vitro*, many of these sequences have been shown to form quadruplex structures under physiological-like conditions, although *in vivo*, the majority of genomic DNA is maintained as a Watson-Crick duplex. This would usually prevent quadruplex folding, however duplex denaturation occurs during several key cellular events, including replication, transcription and recombination, and quadruplex formation may occur during the transient denaturation that precedes each of these events.

1.5.1 Quadruplex-binding proteins

The discovery of several proteins in a variety of organisms that are able to promote, stabilise or resolve quadruplex structures has provided further evidence supporting a biological role. At the telomere, the ß-subunit of the *Oxytricha nova* telomere binding protein has been shown to accelerate quadruplex assembly from the single-stranded sequences $d(T_4G_4T_4G_4)$ and $d(T_2G_4T_2G_4)$ (Fang and Cech, 1993a), enhancing the rate of G-quartet formation by 10^5-10^6 fold on addition of the protein subunit (Fang and Cech 1993b). Similarly, in the budding yeast *Saccharomyces cerevisiae*, the telomere-binding protein RAP1 has been reported to promote G-quadruplex assembly in telomeric DNA (Giraldo and Rhodes, 1994; Giraldo *et al.*, 1994b). Human topoisomerase I has been shown to interact with preformed quadruplexes, and also catalyse their formation from single strands (Arimondo *et al.*, 2000). Furthermore, quadruplex-binding could inhibit topoI-mediated DNA cleavage (Marchand *et al.*, 2002).

In addition to proteins that facilitate the formation of G-quadruplexes, several proteins have been identified that destabilise, unwind or nick quadruplex DNA. Once formed, quadruplexes are extremely stable, therefore cells appear to have evolved mechanisms by which to resolve these structures. The human POT1 (protection of telomeres 1) protein binds to telomeric DNA and its absence leads to irregular recombination (Wu *et*

al., 2006). POT1 has been shown to bind and destabilise quadruplexes, ensuring telomere ends are accessible to telomerase. POT1 also stimulates RecQ-family helicase activity directed towards the telomere (Opresko *et al.*, 2005). Several members of the RecQ helicase family have been found to bind quadruplexes with high affinity. These include the SgsIp helicase (Huber *et al.*, 2002) and the human helicases that are products of the Bloom's and Werner's syndrome genes, both of which are capable of unwinding quadruplexes in a magnesium and ATP-dependent manner (Sun *et al.*, 1998; Mohaghegh *et al.*, 2001). The activity of these helicases can also be inhibited by the presence of quadruplex-stabilising compounds (Han *et al.*, 2000).

A number of nucleases specific for quadruplex DNA have also been characterised (Liu and Gilbert 1994; Ghosal and Muniyappa, 2005). G-quartet nuclease-1 (GQN1) is a human endonuclease that cleaves four-stranded structures several nucleotides upstream of the quadruplex stem (Sun *et al.*, 2001). It is also highly specific and will not cleave duplex DNA, ssDNA, Holliday junctions or even quadruplexes composed of RNA. KEM1, like GQN1 is a nuclease that cleaves at a site proximal to quadruplex formation, and homozygous deletion of the KEM1 gene has been shown to cause cellular senescence and telomere shortening (Liu *et al.*, 1995). It has been hypothesised that KEM1 participates in recombination by excising DNA next to quadruplex 'synapses' between homologous chromosomes, and that its activity is required for telomere maintenance (Liu *et al.*, 1995). Despite the range of proteins that show high specificity for quadruplex DNA, there are still no high-resolution structures of a protein bound to a G-quadruplex.

1.5.2 Quadruplexes *in vivo*

One of the strongest lines of evidence demonstrating quadruplexes can form in telomeric DNA *in vivo* has been highlighted in the unicellular eukaryote *Stylonichia lemnae*. These ciliates generate a macronucleus of millions of gene-sized 'nanochromosomes' during specific stages of their cell cycle, each of which has two telomeres consisiting of $d(T_4G_4T_4G_4)$ (Jahn and Klobutcher, 2002). Schaffitzel and co-workers generated fluorescently-labelled, high affinity quadruplex-specific antibodies (K_d 3-5 nM), which they showed could specifically stain quadruplex structures in the macronucleii, thought to from between the telomeres of adjacent nanochromosomes (Schaffitzel *et al.*, 2001). Furthermore, immune fluorescence was absent in replicating

DNA, suggesting these structures were resolved during telomere replication. Subsequent studies have shown telomere end-binding proteins (TEBPs) control the formation of quadruplex structures within this region (Paeschke *et al.*, 2005). The TEBP subunits α and ß act co-operatively to regulate quadruplex assembly and unfolding under the control of the cell-cycle. RNAi-mediated silencing of either subunit eliminated antibody binding, and resulted in the loss of telomeric DNA, ultimately leading to cell death (Paeschke *et al.*, 2005). An alternative strategy to isolate quadruplex structures in telomeric DNA *in vivo* has involved targeting quadruplexes with specific fluorescent (Chang *et al.*, 2004) or radioactive (Granotier *et al.*, 2005) ligands. Both of these approaches revealed significant localisation at the terminal ends of human chromosomes.

Quadruplexes have also been identified in guanine-rich transcription bubbles known as G-loops. Electron microscopy of plasmid DNA has revealed G-loops of several hundred base-pairs form readily during the transcription of G-rich templates, with a DNA-RNA hybrid forming on the C-rich template strand, and G-quadruplexes interspersed along the opposing strand (Duquette *et al.*, 2004). Mutating out RecQ helicase was shown to increase the prevalence of G-loops, and quadruplex formation has been further verified by the observation of tight binding by both GQN1 and the quadruplex-binding protein nucleolin (Duquette *et al.*, 2004).

1.6 Quadruplex structures at the telomere.

1.6.1 Telomeric DNA

The greatest focus of G-quadruplex research to date has concerned structures related to telomeric DNA, due to its repetitive guanine-rich nature and functional importance in maintaining chromosomal integrity. Telomeric DNA in all higher eukaryotes consists of the highly-conserved double-stranded tandem repeat $d(GGGTTA)_n$ which can range from 2 to 10 kb in length (Cech, 2000). At the extreme 3' termini of the telomere lies a single-stranded overhang of between 100-200 nucleotides, which has important implications with regard to cell division. *In vitro*, this sequence readily adopts a quadruplex structure under physiological-like conditions and several roles have since been envisaged, including protection of the 3' overhang from nucleolytic attack, telomere association and recombination, pairing of homologous chromatids during

meiosis or stabilising the lariat t-loop structures found at the end of the chromosome (Sundquist and Klug, 1989; Griffith *et al.*, 1999) (Figure 1.8).

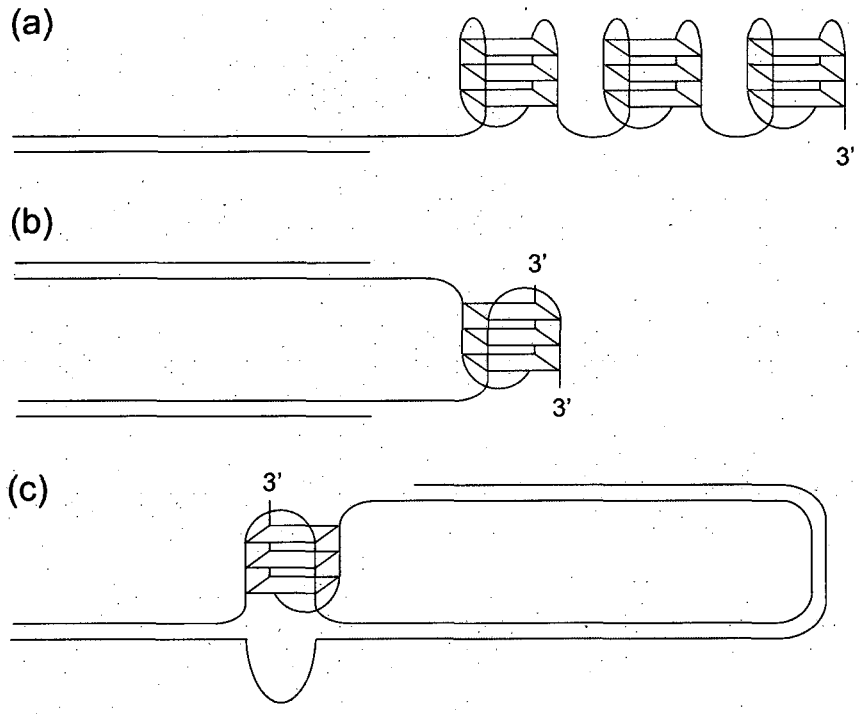


Figure 1.8 Potential quadruplex structures at the telomere. **(a)** Sequential alignment of intramolecular quadruplex structures along the 3' overhang. **(b)** Joining of homologous chromatids via a bimolecular quadruplex interaction. **(c)** Stabilisation of a t-loop via a strand-invading bimolecular quadruplex.

1.6.2 Targeting quadruplexes at the telomere

Quadruplex structures at the telomere have also become novel targets for anti-cancer therapeutics. Telomeres are known to shorten by 50-200 bases during every round of replication, as DNA polymerase is unable to fully replicate the 3' overhang. Telomere shortening continues until a critical length is reached, at which point the cell undergoes apoptosis or programmed cell death (Harley *et al.*, 1990). In both germ-line and stem cells, a mechanism exists to maintain telomere length, in order to preserve the genetic material. These cell-types express the ribonucleoprotein telomerase, which catalyses the addition of d(GGGTTA) repeats to the end of the telomere, counteracting the loss of DNA through replication and stabilising telomere length. Normal somatic cells exhibit little or no telomerase activity, however more than 85% of cancer cells express high levels of telomerase, effectively immortalising them from telomere-induced senescence

(Kim *et al.*, 1994). Telomerase activity is dependent on the availability of the telomeric 3'-overhang. In a key discovery, the folding and stabilisation of quadruplex structures within this region has been shown to inhibit telomerase activity (Zahler *et al.*, 1991). By sequestering the telomerase primer, quadruplex formation has been shown to be a negative regulator of telomere elongation, and as such has become a target for therapeutic intervention.

From a drug targeting perspective, extensive efforts have been made to find specific quadruplex-stabilising compounds that could enhance the anti-telomerase effects and limit cell proliferation (recently reviewed by De Cian *et al.*, 2008; Monchaud and Teulade-Fichou, 2008). Selective targeting of G-quadruplexes requires a precise knowledge of the structure, therefore significant efforts have been made to characterise the quadruplex structures formed at the telomere.

1.6.3 Human telomere folding topologies

The sequence d(AGGGTTAGGGTTAGGGTTAGGG) has been commonly used in X-ray, NMR and biophysical studies as a model for structural characterisation of human telomeric DNA. This sequence has been shown to exhibit a high degree of structural polymorphism, and has required significant efforts to isolate the physiologically relevant structures. The first solution structure of this sequence was obtained in the presence of sodium, revealing an antiparallel 'basket-type' topology which contained three central stacked G-quartets, linked by two edgewise and one diagonal loop of d(TTA) (Wang and Patel, 1993) (Figure 1.9a). Biophysical and chemical probing of a similar sequence suggested this was not the only possible structure, due to differences in reactivity of the loop adenines to diethylpyrocarbonate (DEPC) in sodium and potassium buffers (Balagurumoorthy and Brahmachari, 1994). The crystal structure of the same 22mer oligonucleotide grown in potassium confirmed this sequence could fold into an alternative structure, although with an unanticipated topology (Parkinson *et al.*, 2002). The structure maintains the three-quartet central stack, but each of the trinucleotide loops form double-chain reversals, resulting in strands of all the same polarity (Figure 1.9b). Each of the guanines adopts the *anti* conformation, and the 5' and 3' ends project from opposing faces. This propeller-type fold results in exposed G-quartet faces at the top and bottom of the quadruplex and loops that project 10Å from the core, both of which represent potential drug recognition sites. It was proposed that

this topology could also allow facile oligomerisation of multiple telomeric quadruplex units (Parkinson *et al.*, 2002), as a 200 bp 3' overhang could in principle form up to 8 stacked quadruplexes. This arrangement would also present an ideal cavity for small-molecule intercalation between stacked quadruplex units.

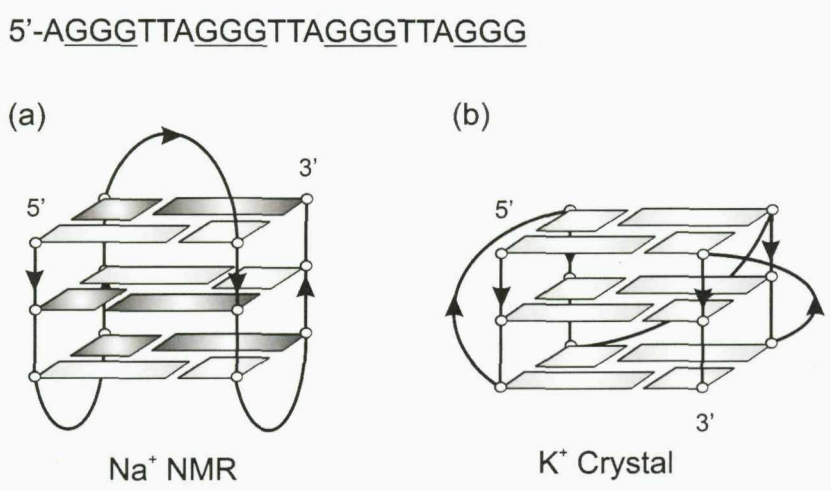


Figure 1.9 Human telomere four-repeat folding topologies. **(a)** Solution structure in sodium. **(b)** Crystal structure in potassium. Bases in light grey are in the *anti* conformation; dark grey represents bases in *syn*.

Given the presence of three trinucleotide loops, which are able to form topologically distinct conformations, dependant on the environmental conditions, subsequent studies have questioned whether the structure obtained in crystal is physiologically relevant, and whether the crystal environment may have selected for a particular conformation (Li *et al.*, 2005). NMR studies of the same oligonucleotide in potassium do not give a unique folding pattern in solution (Phan and Patel, 2003). Several groups have attempted to characterise the structures present in potassium solution via a range of biophysical and chemical footprinting methods (Redon *et al.*, 2003; He *et al.*, 2004; Risitano and Fox, 2005; Li *et al.*, 2005; Qi and Shafer, 2007). Single molecule FRET (Ying *et al.*, 2003; Lee *et al.*, 2005) and molecular dynamics simulations (Hazel *et al.*, 2005) have shown that both parallel and antiparallel topologies could co-exist in solution, and there is only a small free energy difference between them. The sensitivity of the equilibrium which is likely to exist between the folded forms, may be influenced by particular experimental conditions.

1.6.4 Telomeric hybrid structures

The problem of conformational heterogeneity exhibited by the human telomeric repeat 22mer in potassium has recently been approached in a novel way, which has identified a third topology for this sequence (Xu *et al.*, 2006). By selective substitution of dG residues with 8-bromo-dG, thereby locking specific nucleotides in the *syn* conformation, Xu and co-workers proposed that two conformations were present in the unmodified sequence; one antiparallel (chair-type), and a second containing a mixed parallel/antiparallel-stranded alignment. A similar study using ribo-G modifications to lock selected nucleotides in the *anti* conformation was also able to isolate a folded complex with a mixed parallel-antiparallel strand arrangement (Qi and Shafer, 2007).

Three independent NMR studies have since verified the presence of a conformer with a mixed stand alignment, either by altering the nucleotides flanking the four G-tracts (Ambrus *et al.*, 2006; Luu *et al.*, 2006), or by 8-^{Br}G modifications (Matsugami *et al.*, 2007). By screening series of sequences with terminal modifications, the groups of Dinshaw Patel and Danzhou Yang were able to isolate a single folded form with the hybrid-type parallel/antiparallel conformation). Ambrus and co-workers used the 26nt sequence d(AAAGGG(TTAGGG)$_3$AA) whereas Luu and co-workers the 24nt sequence d(TTGGG(TTAGGG)$_3$A). Both groups reported a structure containing two edgewise loops and one double-chain reversal, resulting in a three up one down (3+1) strand alignment. There is also an unusual alternation of *syn* and *anti* conformations, with one *anti-syn-syn-syn* and two *syn-anti-anti-anti* G-quartets. The presence of the (non-telomeric) terminal nucleotides provide additional stability via Watson-Crick base pairing and adenine-triple capping interactions, which preferentially stabilise one conformer (Dai *et al.*, 2007). Subtle adjustment of the terminal nucleotides to d(TAGGGTTAGGGTTAGGGTTAGGGTT) have resulted in the elucidation of a second new topology, also with a 3+1 strand alignment but with a different loop arrangement (Phan *et al.*, 2006). This latter structure contains the double-chain reversal as the third loop (termed hybrid-2) as compared to the former in which this loop is in the first position (hybrid-1) (Figure 1.10).

The biologically native, unmodified sequence has now been shown to form a combination of both hybrid-1 and hybrid-2 conformers (Dai *et al.*, 2007b; Phan *et al.*, 2007b). Distinct capping structures appear to determine the specific hybrid structure

present, although the energy barrier between the two hybrid structures is small (Dai *et al.*, 2007b).

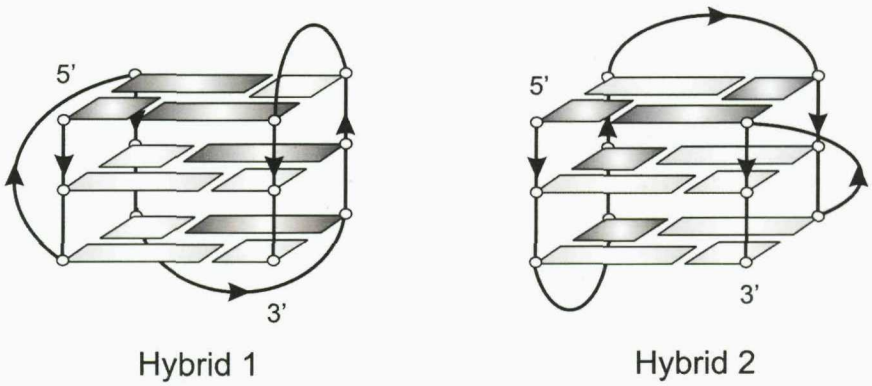


Figure 1.10 New models for quadruplex folding at the telomere. NMR studies in the presence of potassium have isolated two different structures with a 3+1 stand alignment. Both contain two edgewise and one double-chain reversal loop, although the order in which the loops appear differs between the two structures.

1.6.5 Quadruplex multimers

What is as yet unclear is how the presence of other quadruplexes may influence this equilibrium, as at present most studies have examined short telomeric DNAs of repeat length not more than four. The hybrid structures, like the original crystal structure may allow end-on-end stacking interactions, however current evidence suggests that longer telomeric repeat lengths do not stack in this manner. Thermodynamic profiles of telomeric repeats totalling between one and four complete intramolecular quadruplexes, revealed a destabilisation of the complex on increasing the repeat number (Vorlickova *et al.*, 2005). Current models suggest that rather than concomitant G-tetrad stacking between quadruplex units, there is more likely to be a beads-on-a-string composition in longer telomeric DNAs (Yu *et al.*, 2006).

1.6.6 Other telomeric G-quadruplex topologies

1.6.6.1 Tetrahymena

In addition to the human telomeric repeat sequence, a number of other organisms with different telomeric repeat sequences have provided useful structural insights into quadruplex architecture. The telomeric repeat sequence of the ciliate protozoa

Tetrahymena differs from the human telomeric repeat by a single A to G substitution ($d(TTGGGG)_n$). Despite the potential for an additional G-quartet, the four-repeat sequence d(TTGGGGTTGGGGTTGGGGTTGGGG) forms a three G-quartet stack with three guanines participating within the loops (Wang and Patel, 1994). The three loops are unequal in length; d(GTT) and d(GTTG) form edgewise loops whereas the shorter d(TT) adopts a double-chain reversal, resulting in a three up, one down strand arrangement. This was the first example of a hybrid 3+1 topology and was thought to be an isolated example, although it now appears there is a distinct structural similarity between this fold and that of the recently elucidated human telomeric repeat.

1.6.6.2 Oxytricha nova

The third telomeric sequence to have been extensively studied is that of the marine ciliate *Oxytricha nova* $d(GGGGTTTT)_n$. Several groups have resolved the solution structure of the four-repeat sequence in sodium, with the consensus showing a four-quartet stack with two edgewise and a central diagonal loop (Smith and Feigon, 1992; Wang and Patel, 1995). No structure has been isolated for this sequence in potassium, although CD measurements suggest multiple conformations are likely (Dapic *et al.*, 2003).

The two repeat structure d(GGGGTTTTGGGG) has been the subject of numerous structural studies, as subtle alterations to the base composition have been shown to result in significant topological rearrangements. NMR (Schultze *et al.*, 1999) and X-ray (Hovarth *et al.*, 2001; Haider *et al.*, 2002) structures have shown that in both potassium and sodium, $d(G_4T_4G_4)_2$ forms a bimolecular structure with diagonal T_4 loops. This structure is maintained in the related sequence $d(G_3T_4G_3)_2$ (Smith *et al.*, 1994) but $d(G_4T_3G_4)_2$ and a brominated analogue $d(G_4{}^{Br}UTTG_4)_2$ both form alternative structures containing lateral loops (Hazel *et al.*, 2006). This suggests that bimolecular quadruplexes containing four thymidine loops favour the formation of diagonal loops, whereas those containing three (or two) thymidines adopt a lateral loop conformation. Removing a guanine from the 5' or 3' to create asymmetry in the G-tracts results in differing conformations. $d(G_3T_4G_4)_2$, which lacks a 5' guanine, revealed a structure containing several novel features (Crnugelj *et al*, 2002). It contained both edgewise and diagonal loops, resulting in one antiparallel and three parallel strands, an unusual alignment in bimolecular complexes. A further unusual element was a single phosphate

spanning across the central G-quartet in a chain reversal. The sequence isomer $d(G_4T_4G_3)_2$ formed a structure closer to the *Oxytricha* parent sequence (Crnugelj *et al*, 2003), with two diagonal loops connecting a stem of three G-quartets. The additional guanines participate within the loop and as part of a 5' overhang. These telomeric variant sequences have provided key insights into the variability of bimolecular quadruplex topology, although at present there is still not enough structural and energetic data to define specific folding rules.

1.7 Quadruplexes in the genome

Besides at the telomere, many genomic DNA regions carry sequences that readily form G-quadruplexes *in vitro*, and this has led to considerable interest in investigating potential *in vivo* roles, and the structures they adopt. Unlike at the telomere where the DNA sequence is highly repetitive, genomic quadruplex topologies are far more varied due to the high level of sequence variation.

G-rich chromosomal domains are prevalent in a number of functional repetitive DNA regions (as well as telomeres) including minisatellites (Wietzmann *et al.*, 1997), immunoglobulin heavy-chain switch regions (Sen and Gilbert, 1988; Catasti *et al.*, 1996) and ribosomal DNA (rDNA) repeats (Hanakahi *et al.*, 1999, Maizels, 2006). The knowledge that transcriptional control elements can form unusual structural motifs and are frequently G-rich has also led to speculation that quadruplex formation (or resolution) may act as a regulatory signal in transcription (Woodford *et al.*, 1994; Simonsson *et al.*, 1998). Unlike at the telomere, for these G4 structures to be realised, the G-strand at some point must be released from its complementary strand. Duplex denaturation occurs during several key cellular events, including replication, transcription and recombination, and quadruplex formation may occur during the transient denaturation that precedes each of these events (Duquette *et al.*, 2007). This equilibrium may also be influenced by ligands, and the possibility that quadruplex formation may alter gene transcription has prompted the design of therapeutics targeted towards these structures with the aim of down-regulating oncogene expression (Hurley, 2002; Hurley *et al.*, 2006). The biological specificity of these compounds would require selectivity between different quadruplex scaffolds.

1.8 Quadruplexes as gene regulatory elements

1.8.1 ß-Globin

The earliest evidence for a relationship between quadruplex formation and transcriptional control was based on studies of the chicken ß-globin gene. Woodford and co-workers described a potassium dependent DNA synthesis arrest site in a G-rich region of the gene promoter (Woodford *et al.*, 1994). This site was located within a nuclease hypersensitive region, sequences known to adopt atypical DNA secondary structures. Through the use of a primer extension assay, this sequence was mapped to be $d(G_{16}CG(GGT)_2GG)$. Although originally proposed to form a complex triple helical conformation (Kohwi, 1989), further analysis revealed properties consistent with the formation of an intramolecular tetraplex; formation was template strand independent, specific in the requirement for potassium ions and involved non Watson-Crick base interactions between guanines. Based on chemical probing experiments, the authors composed a model structure consisting of five tetrads with two G-G-G-G quartets and three incomplete tetrads of G-G-G-T (2) and G-G-G-C (Howell *et al.*, 1996). It was proposed that the unconventional G-quadruplex framework was stabilised by the 5' and 3' flanking sequences via the formation of a molecular cinch.

1.8.2 C-myc

The most well-documented evidence concerning G-quadruplex formation as a transcriptional regulatory element is in the oncogene c-myc, the protein product of which is known to play a vital role in cellular growth and differentiation (Henriksson and Luscher, 1996). C-myc activity in normal cells is tightly regulated by a major control element situated 115 bp upstream of the P1 promoter, termed the nuclease hypersensistive element (NHE) III_1 (Siebenlist *et al.*, 1984). This 27 bp region accounts for ~85% of the transcriptional activation of the gene (Berberich and Postel, 1995) and has unusual asymmetry, in that one strand is almost all pyrimidines and the other (almost) all purines. Disruption of c-myc control has been associated with a number of cancers, and has therefore become an attractive target for anticancer therapies.

In vitro, the purine-rich strand of the NHE III_1 has been shown to readily fold into G-quadruplex structures (Simonsson *et al.*, 1998), and even in the presence of its

complement, small molecule 'drivers' could induce the formation of a quadruplex (Rangan *et al.*, 2001). It was suggested that quadruplex folding in this region could act as a gene silencing mechanism, able to repress c-myc transcription (Simonsson *et al.*, 1998; Grand *et al.*, 2002). To test this hypothesis, a series of expression constructs to assay promoter activity were designed, in which quadruplex-abolishing mutations were introduced (Siddiqui-Jain et al., 2002). Single G-to-A point mutations were found to increase basal c-myc transcription levels 3-fold. Furthermore, addition of the quadruplex-stabilising agent TMPyP4, a cationic porphyrin, was shown to further repress the promoter activity, providing compelling evidence that a G-quadruplex forms in this region. Further analysis also revealed that multiple quadruplex conformations were likely, not all of which were biologically active (Siddiqui-Jain *et al.*, 2002).

1.8.2.1 Topology and structure

Significant efforts have now been made to characterise the structure of the quadruplex elements present within this region. However the c-myc NHE III_1 exemplifies the difficulties in defining a precise topology for many genomic quadruplex-forming sequences. The 27 nt sequence comprises six individual G-tracts which are unequal in length, with G-tracts of four, three and two guanines (Table 1.2). The standard requirement for only four G-tracts in a unimolecular structure mean that in principle, any combination of the 6 tracts could participate in G-tetrad formation. Early structural assignments based on chemical probing (Simonsson *et al.*, 1998) and FRET analysis (Simmonson and Sjoback, 1999) suggested an antiparallel topology containing three G-quartets derived from G-tracts 1, 2, 4 and 5. However further examination of this sequence revealed a dynamic mixture of conformations which could interconvert in solution, and based on mutagenesis and transfection assays, not all appeared to be biologically relevant (Siddiqui-Jain *et al.*, 2002; Seenisamy *et al.*, 2004).

In an similar manner to the human telomeric repeat, high resolution NMR analysis of this sequence has required modifications and truncations from the native sequence to achieve well resolved spectra (Table 1.2). Specific modifications have resulted in alternative topologies, although general features of the native sequence can be identified. Myc-2345 and Myc-1245 each contain four G-tracts and, like the human telomere repeat in crystal, form parallel-stranded structures with all double-chain reversal loops (Figure 1.11a) (Phan *et al.*, 2004). The peripheral loops of each sequence

are single nucleotides, with Myc-2345 containing a central dinucleotide (GA) loop and Myc-1245 a longer (TTTTTA) loop. Myc-1245 with the longer loop is noticeably less stable (16 °C) than its sequence isomer. Similar modifications to remove ambiguity in the guanines participating in G-quartet formation (Myc22) via G-to-T mutations resulted in a similar all parallel topology (Ambrus *et al.*, 2005).

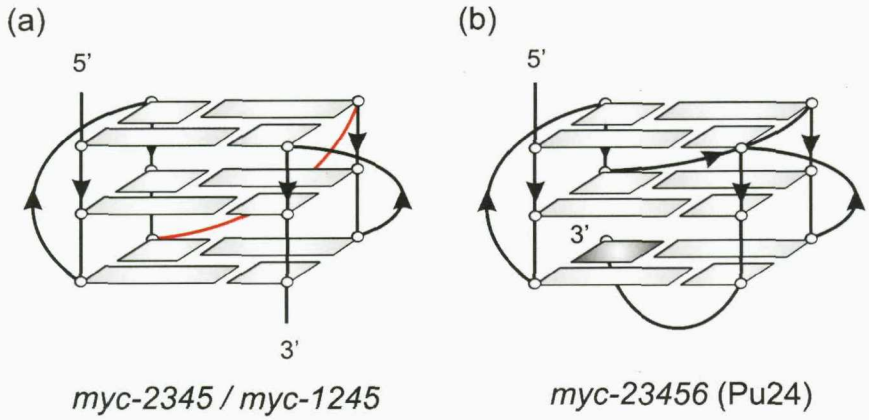


Figure 1.11 Schematic representation of c-myc NHE quadruplex folding topologies. **(a)** myc-2345 and 1245 both exhibit a parallel-stranded topology with three double-chain reversal loops. The central loop (red) is two nucleotides in myc-2345, and six nucleotides in myc-1245. **(b)** Pu24 is also parallel-stranded, although contains an unusual snap-back motif to incorporate a fifth guanine-tract into the fold.

Name	Sequence
	1 2 3 4 5 6
c-myc NHE	5′-TGGGGAGGGTGGGGAGGGTGGGGAAGG-3′
Myc-2345[a]	**T**GAGGGTGGGGAGGGTGGGGAA
Myc-1245[a]	TGGGGAGGGT**TTTT**AGGGTGGGGA
Myc22-G14T/G23T[b]	**T**GAGGGTGGG**T**AGGGTGGG**T**AA
Pu24[c]	**T**GAGGGTGGGGAGGGTGGGGAAGG
Pu24I[c]	**T**GAGGGTGG**I**GAGGGTGGGGAAGG

Table 1.2 C-myc NHE and sequence variants used to define the structural elements present within this region. Modifications required for structural characterisation highlighted in bold. Guanines participating in G-quartet formation are underlined. [a] Phan *et al.*, 2004, [b] Ambrus *et al.*, 2005, [c] Phan *et al*, 2005,

A third NMR structure, containing five contiguous G-tracts (Pu24I) isolated a third distinct topology containing a novel DNA fold (Figure 1.11b) (Phan *et al.*, 2005). Like the previous two, the structure contains three G-quartets with all parallel strands linked

by double chain reversals. But in an unusual feature, the extreme 3' guanine snaps back to participate in the G-tetrad core, displacing a guanine from one of the G-tracts. The tetrad stack is further stabilised by an A•A pair and a G•AG triad capping the top and bottom respectively. This previously unseen folding topology may represent a viable mechanism for sequences that contain more than four G-tracts to fold into intramolecular quadruplex structures.

1.8.3 Other promoter quadruplex sequences.

The seminal work carried out on the *c-myc* promoter has resulted in the search for and characterisation of other oncogenic gene promoters which carry a quadruplex-forming motif. *In vitro*, numerous other examples of quadruplex formation have been reported in various oncogenic promoter sequences, including bcl-2 (Dexheimer *et al.*, 2006), RET (Guo *et al.*, 2007), VEGF (Sun *et al.*, 2005), HIF-1α (De Armond *et al.*, 2005), KRAS (Cogoi and Xodo, 2006) and c-kit, (Rankin et al., 2005; Fernando *et al.*, 2006) (Table 1.3), although mainly by biophysical analysis of the single-stranded G-template. Biological evidence to support a claim for physiological relevance has yet to be reported in each of these cases, although the quadruplex-forming sequences in bcl-2 and c-kit promoters have been structurally characterised by NMR.

1.8.3.1 bcl-2

Bcl-2 is potent oncogene and plays an essential role in cell survival through the inhibition of apoptosis (Adams and Corey, 1998; Chao and Korsmeyer, 1998). Deregulation of the bcl-2 gene leads to overexpression in a number of tumours and therefore represents an attractive target for therapeutic intervention. The major transcriptional control element of bcl-2 transcription is the P1 promoter, the 5'-end of which is guanine-rich, and this region has been implicated as a major regulatory element (Tsujimoto and Croce, 1986; Young and Korsmeyer, 1993). This 39 bp sequence contains six individual runs of three guanines or more (Table 1.3), and NMR and chemical footprinting methods have confirmed it is able to form a stable tetrahelical structure in the presence of potassium (Dai *et al.*, 2006; Dexheimer *et al.*, 2006). Like c-myc, it is likely that multiple folded forms are possible due the extended number of G-tracts. A high resolution NMR study of a modified version of this sequence so as to eliminate heterogeneity, has shown one of the structures contains a three-quartet stem

with a parallel/antiparallel 3+1 topology (Dai *et al.*, 2006b). Mutational analysis of the loop nucleotides revealed several were crucial in stabilising the correct fold.

Name	Sequence
bcl-2	5′-AGGGGCGGGCGCGGGAGGAAGGGGGCGGGAGCGGGGCTG
c-kit87up	5′-AGGGAGGGCGCTGGGAGGAGGGG
c-kit21	5′-CGGGCGGGCGCGAGGGAGGGG
RET	5′-AGCGGGTAGGGGCGGGGCGGGGCGGGGG
VEGF	5′-GGGGCGGGCCGGGGGCGGGGTCCCGGCGGGGCGG
KRAS	5′-GGGAGGGAGGGAAGGAGGGAGGGAGGGA
HIF-1α	5′-GCGCGGGAGGGGAGAGGGGGCGGGAGCGCG

Table 1.3 Guanine-rich regions identified as potential quadruplex-forming motifs within oncogenic promoter sequences. Guanine tracts highlighted in red.

1.8.3.2 c-kit

The third well-characterised promoter quadruplex has been isolated from within the *c-kit* kinase gene, an important target in gastrointestinal tumour therapies (Tuveson *et al.*, 2001). There are two separate G-rich stretches in the c-kit promoter, both of which are highly conserved across vertebrate species (Rankin *et al.*, 2005). Of the two G-rich regions, one has been found to adopt a single folded quadruplex in the presence of potassium (c-kit87up) (Rankin *et al.*, 2005), although the second (c-kit21) required sequence mutations to remove G-tract ambiguity before a single folded complex could be isolated. Despite c-kit87up containing three equal runs of guanine (Table 1.3), an unusual topology was anticipated, as biophysical analysis had shown interchanging the loop regions, so that the net loop length remained the same, but with the linkers in a different order, resulted in quadruplex destabilisation (Rankin *et al.*, 2005). The NMR solution structure of this sequence revealed an unprecedented fold, in which a guanine thought previously to form part of the central loop was actually involved in G-quartet formation, displacing one of the guanines in the fourth G-tract (Phan *et al.*, 2007). The quadruplex was therefore composed of three G-quartets and four loops; two double-chain reversals, a linker connecting two adjacent corners and an unusual fourth loop inserting the terminal guanine resides back into the G-quartet stem to complete the G-quartets (Figure 1.12). This discovery has had implications with regards to both what

constitutes a quadruplex forming sequence and also quadruplex structural predictions. Rather than a 1-4-4 loop series, as would be predicted from the sequence, there are actually four loops making up a 1-1-2-5 loop arrangement, adding a further level of complexity to potential loop arrangements. In addition, the inability of the c-kit87up loop isomers to form a stable quadruplex suggests loop sequence may have a more significant impact on quadruplex folding than previously thought.

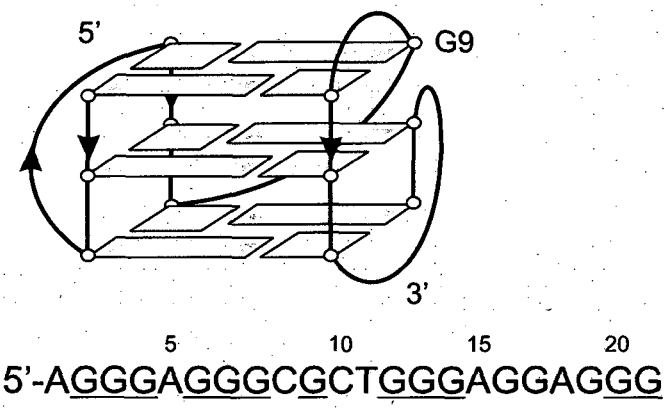


Figure 1.12 NMR-defined folding topology of the c-kit87up sequence in the presence of potassium. Guanines participating in G-quartet formation are underlined.

The structural diversity exhibited by promoter quadruplexes make these regions attractive targets for specific and selective drug targeting. Of the small sample base examined thus far, there are generic traits, but also unique features. Sequence analysis of the c-kit87up quadruplex-forming region has found only a few closely related sequences that correspond to the same stable tertiary structure within the genome (Todd *et al.*, 2007). It was concluded that the likelihood of any of these playing a role in gene expression was also very low. This feature, together with its distinct topological arrangement, suggests c-kit87up represents a good target for drug design and that compounds selective towards this structure would act only on this pathway.

1.9 Bioinformatics

Reports of promoter regions with quadruplex-forming potential have become increasingly common, as groups have actively searched important, often oncogenic, genes for the presence of these motifs. In the early *c-myc* studies of Simonsson and co-workers, a generic requirement for an intramolecular quadruplex folding motif was outlined; $G_xN_{y1}G_xN_{y2}G_xN_{y3}G_x$ where x represents the length of the G-tracts and y1-3

the length of the loops (Simonsson *et al.*, 1998). It was noted that not only did the c-myc NHE comply with these sequence requirements, but that the promoter regions of numerous other oncogenes also contained this intramolecular fold-back motif. More extensive genomic searches over the last few years have adopted a similar search criterion extended over the whole genome. These have revealed that the frequency with which reports of promoter quadruplexes have arisen is unsurprising, as quadruplex-forming motifs are significantly over-represented in gene promoters (Huppert and Balasubramanian, 2007). Further quadruplex-motif 'hotspots' have included the first intron of genes (Eddy and Maizels, 2008) and RNA 5'-untranslated regions (UTRs) (Kumari *et al.*, 2007). The use of bioinformatics as a quadruplex search tool has since yielded a wealth of data regarding the genomic frequency, sequence composition and location of these motifs.

1.9.1 Prevalence of quadruplex motifs in the genome

Several bioinformatic studies have now been performed with the aim of establishing the number of quadruplex-forming motifs that are present within the genomes of a range of organisms (Todd *et al.*, 2005; Huppert and Balasubramaniann, 2005; Rawal *et al.*, 2006; Du *et al.*, 2007; Hershmann *et al.*, 2007; Yadav *et al.*, 2007). The task of predicting the presence of a three dimensional structure from a linear oligonucleotide has required the definition of an algorithm or folding rule that describes a quadruplex-forming sequence, and different groups have applied their own interpretations. Two independent groups have systematically searched the entire human genome using similar search criteria: Todd and co-workers defined a general unimolecular quadruplex-forming sequence as $G_{3-5}N_{1-7}G_{3-5}N_{1-7}G_{3-5}N_{1-7}G_{3-5}$ (Todd *et al.*, 2005). Loops lengths were restricted to between 1 and 7 nucleotides, partly because shorter loops form more stable structures, but also to limit the potential number of hits. When every possible combination was considered, 5,713,900 potential quadruplex hits were isolated within the genome (Todd *et al.*, 2005), however the number of distinct, non-overlapping quadruplexes that could potentially form at once was 375,157. This value was corroborated by Huppert and Balasubramanian, who using a similar folding rule ($G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}N_{1-7}G_{3+}$) with no restraint on the length of the G-tracts, obtained a value of 376,446 potential quadruplex-forming sequences (PQS) (Huppert and Balasubramanian, 2005). This number by no means implies there are this many quadruplexes formed at once, (if any at all), as

quadruplex formation is a dynamic process which involves inter-conversion between quadruplex and duplex forms.

1.9.2 Loop length, sequence and distribution patterns

Further to establishing how many motifs are present, these data also revealed several trends with regards to the putative loop regions. The length of the PQS loops has been found to be distinctly non-random; short, single nucleotide loops are strongly favoured, whereas long loops are disfavoured (Huppert and Balasubramanian, 2005). The most frequently observed set of loop lengths is (1,1,1), making up around 8 % of the entire number of hits, and 15 of the top 20 most common sets of loop lengths contain two single nucleotide loops (Table 1.4).

(a)

Most common loop lengths			
Loop 1	*Loop 2*	*Loop 3*	*Frequency*
1	1	1	47,475
1	4	1	11,328
1	2	1	10,656
1	1	2	10,415
2	1	1	10,040
2	2	2	9411
1	3	1	9127
1	5	1	7799
5	1	1	7379
1	1	5	7337
3	3	3	6827
3	1	1	6458
1	1	3	6403
1	1	4	6196
4	1	1	6189
2	2	1	5123
1	2	2	5046
2	1	2	4780
1	6	1	4556
6	1	1	4462

(b)

Most common loop sequences				
Sequence	*Frequency*	*Loop 1*	*Loop 2*	*Loop 3*
A	193,756	51,361	63,872	78,523
T	121,406	53,234	37,657	30,515
C	44,020	14,983	14,907	14,130
AA	40,026	12,778	13,717	13,531
CT	32,472	11,637	10,554	40,281
CA	32,070	10,781	10,846	10,443
G	29,623	7183	8375	14,065
AT	19,957	6789	7242	5926
AGA	19,144	5377	6919	6848
TT	17,089	7437	5530	4122
TA	12,641	4744	4329	3568
CC	10,955	3646	3726	3583
AGT	9869	2767	4447	2682
AGGA	9463	1932	3559	3972
AGGT	9434	1516	6448	1470
TGA	9237	3006	2849	3382
AAA	7839	2393	2970	2476
CCT	7151	2540	2298	2313
TGT	6619	2530	2307	1782
CCA	6269	2105	2048	2116

Table 1.4 (a) The 20 most common sets of loop-lengths in PQS (Huppert and Balasubramanian, 2005). **(b)** The 20 most common loop sequences, and their loop positional frequency (Todd and Neidle, 2005).

Differences from a normal distribution can be interpreted as selective pressure for or against a particular motif, therefore it is intriguing to note that sequences which form the most stable quadruplexes, those containing single-nucleotide loops, are by far the most common. Biophysical analysis of sequences of this type have also suggested they adopt a common folded topology (Rachwal *et al.*, 2007d; Bugaut and Balasubramanian, 2008). Although longer loops occur less frequently, sequence analysis has identified a number of notable exceptions, with CCTGTT and TAGCATT found to be over-represented amongst longer length loops (Todd *et al.*, 2005; Burge *et al.*, 2006). These loops also show a strong bias for loop position, with CCTGTT found 18 and 9 times in the second and third loops respectively, but 1266 times in loop one (Todd *et al.*, 2005). This loop position bias may have structural consequences, and the specific base composition of the loop may be important in determining a particular fold.

1.9.3 Quadruplex motif distribution

Early search functions were able to enumerate genomic quadruplex-forming sequences, without being able to correlate these with specific gene locations and functions. However it was noted that potential quadruplex-forming sequences (PQS) were disfavoured in exonic regions of DNA, which would suppress the number of the quadruplex-forming sequences in coding mRNA (Huppert and Balasubramanian, 2005). More refined searches have shown that PQS are significantly enriched in gene promoters, with around 40 % of human gene promoters carrying at least one (Huppert and Balasubramanian, 2007). The highest density of PQS were observed within the first 100 bases upstream of the transcription start site (TSS) and these sites also strongly associate with regions of nuclease hypersensitivity, leading to further speculation that G-quadruplex formation may be directly involved in gene expression. Furthermore, quadruplex-forming potential (termed G4P) has been correlated with particular functional classes of genes; tumour suppressor genes have been found to have a low G4P, and proto-oncogenes high G4P (Eddy and Maizels, 2006). These data suggest there may be evolutionary selective pressure towards concentrating quadruplex-forming regions within gene promoters, and selection based on G4P maybe specific for certain genes. It could be imagined that the low G4P in tumour-suppressor genes prevents the possibility of impaired gene function caused by the formation of a G-quadruplex, whereas the high G4P within proto-oncogenes may make them targets for transcription-induced destabilisation (Eddy and Maizels, 2006).

Promoter PQSs are even more selective with regards to loop length. Almost 80 % of promoter PQSs in the first 100 bases upstream of the TSS contain at least one single nucleotide loop, however this percentage decreases as the distance from the TSS increases (Huppert and Balasubramanian, 2007). This would imply that there could even be selective pressure in favour of the most thermodynamically stable quadruplexes in regions close to the TSS.

It is important to note that factors other than G-quadruplex formation may account for the prevalence of quadruplex-forming motifs in the genome, and in particular gene promoters. Many transcription factor recognition sequences are G-rich, and multiple contiguous binding sites would generate a hit on most bioinformatic searches. In fact the similarity in the distribution of PQS and Sp1 binding sites has recently been reported (Todd and Neidle, 2008). The dynamic conversion of quadruplex and duplex structural forms may therefore represent a mechanism for modulating gene expression by steric inhibition of transcription factor binding. In addition to gene promoters, a second region of G-richness has recently been uncovered downstream of the TSS in the first intron of genes (Eddy and Maizels, 2008). This conserved trait is notable in its strand bias, containing a G-rich non-template strand only. These elements could therefore be recognised as DNA or RNA, and be involved in gene regulation or mRNA processing.

1.9.4 Quadruplexes in RNA

RNA quadruplexes are far less well characterised than their DNA counterparts, although have been found to exhibit superior thermodynamic stability (Mergny *et al.*, 2005). Their single-stranded nature may also make them more likely candidates than DNA to adopt a quadruplex structure. RNA G-quadruplexes have been implicated in several post-transcriptional regulatory roles, including translational repression, splicing, mRNA turnover and protein binding (Kumari *et al.*, 2007; Wieland and Hartig, 2007; Gomez *et al.*, 2004; Bagga *et al.*, 1998; Darnell *et al.*, 2001). In one notable example, quadruplex formation in the 5' untranslated region (UTR) of the oncogene NRAS has been shown to repress translation, and bioinformatic searches for similar motifs have identified close to 3000 other 5'UTRs with quadruplex forming potential (Kumari *et al.*, 2007). Are the folding rules different for RNA quadruplexes? Current searches for quadruplex motifs in RNA have used the same algorithms as those applied for DNA

quadruplexes (Kumari *et al.*, 2007; Kikin *et al.*, 2007). These generic folding rules are likely to be universal, however there are examples of the same DNA and RNA sequences adopting different quadruplex topologies (Liu *et al.*, 2002; Qi and Shafer, 2007).

1.9.5 Sequence-structure relationships

The rules defined for the bioinformatic genome searches are approximations of what constitutes a quadruplex-forming sequence based on the current knowledge of quadruplex folding. They by no means represent an absolute descriptor, as several sequences which fall within these criteria are now known not to form a quadruplex (Rankin *et al.*, 2005) but equally there are many sequences which fall outside these folding rules which can (Bourdoncle *et al.*, 2006; McManus and Li, 2008). While these searches can reveal the presence, prevalence and location of these motifs, they reveal nothing about structure and topology, and little about stability, which will be of critical importance in the targeting of different quadruplex structures.

1.10 Aims of this research

The aim of this work is to investigate the role of sequence in the folding and stability of DNA G-quadruplexes. At present, predicting quadruplex stability and topology from the primary sequence alone is extremely difficult, due to the range of intrinsic and extrinsic factors that can influence quadruplex structure. This work aims to expand the knowledge of quadruplex sequence–structure relationships using series of model quadruplex-forming sequences.

Using a variety of biophysical techniques, including melting studies, circular dichroism and NMR, this work first examines the role of loop length in quadruplex folding. The role of loop sequence is then examined, both in model and biologically-relevant promoter sequences, to investigate how single nucleotide substitutions can alter quadruplex stability and also promoter activity. The final chapter systematically explores the effects of increasing the length of G-tracts on the topology and stability of series of model quadruplex-forming sequences.

CHAPTER 2

Materials and Methods

2.1 Materials

2.1.1 Chemicals, enzymes and reagents

All chemicals for use in standard buffers were purchased from Sigma Aldrich (Poole, UK). Restriction enzymes, shrimp alkaline phosphatase (SAP) and T4 DNA ligase were purchased form Promega (Southampton, UK) and stored at -20 °C. Taq DNA polymerase was purchased from New England Biolabs (Hitchen, UK). pGL3 luciferase expression vectors, sequencing primers and luciferase assay system kit were also purchased from Promega. Plasmid purification kits (Miniprep and Maxi kits) were purchased from Qiagen. T7 dideoxy sequencing kits were purchased from USB corporation (Cleveland, USA). Redivue radioactive [α-^{32}P] and [γ-^{32}P] dATP were purchased from Amersham Biosciences (Little Chalfont, UK) with an initial activity of 3000 Ci/mmol. Polyacrylamide concentrates for gel electrophoresis as Accugel (40% (w/v) 19:1 acrylamide:bis-acrylamide solution) and Sequagel (25% (w/v) 19:1 acrylamide:bis-acrylamide solution containing 8 M urea) were purchased from National Diagnostics (Hull, UK). All other items will be described where appropriate.

2.1.2 Oligonucleotides

Oligonucleotides were synthesised on an Applied Biosystems 394 DNA/RNA synthesiser on either the 0.2 or 1.0 μM scale and were provided by Prof. T. Brown (Department of Chemistry, University of Southampton) and stored at -20 °C. Phosphoramidite monomers and other reagents were purchased from Applied Biosystems, Proligo and Link Technologies. For the fluorescence melting experiments, oligonucleotides were labelled at the 5'-end with 6-amidohexylfluorescein (FAM) and at the 3'-end with dabcyl using C7 dabcyl cpg (Link technologies). All oligonucleotides were purified by gel filtration using Nap10 columns (GE Healthcare) and analysed by gel electrophoresis. Oligonucleotide sequences are listed in the relevant chapters.

2.1.3 Buffers and solutions

Lithium phosphate buffer	10 mM lithium hydroxide, adjusted to pH 7.4 with phosphoric acid
Potassium phosphate buffer	80.2 % K_2HPO_4, 19.8 % KH_2PO_4, pH 7.4
Sodium phosphate buffer	77.4 % Na_2HPO_4, 22.6 % NaH_2PO_4, pH 7.4
5x TBE	108 g Tris, 55 g Boric acid, 9.4 g EDTA (2 L)
Diluent	50 % (w/v) urea
Formamide stop solution	0.3 % bromophenol blue, 0.3 % xylene cyanol, 10 mM EDTA pH 7.5, 97.5 % deionised formamide.
Loading dye	20 % (w/v) ficoll
DNA elution buffer	10 mM Tris-HCl, 10 mM EDTA, pH 7.4
Transformation buffer	50 mM $CaCl_2$, 10 mM Tris-HCl, pH 7.4
Polymerase stop reaction buffer	10 mM Tris-HCl pH 7.4, 10 mM $MgCl_2$, 0.5 mM DTT, 0.1 mM EDTA.

2.2 Methods

2.2.1 Spectroscopic studies

2.2.1.1 Circular Dichroism

CD measurements were carried out on a Jasco J-720 spectropolarimeter. Oligonucleotide solutions (5 μM) were prepared in 10 mM lithium phosphate buffer, supplemented with varying concentrations of either potassium chloride or sodium chloride. All samples were first heated to 95 °C and annealed by slowly cooling to 15 °C over a period of 15 hours to ensure correct folding. Spectra were recorded between 220 – 320 nm in 5 mm path length quartz cuvettes. Spectra were averaged over 10 - 16 scans, which were recorded at 100 $nm.min^{-1}$ with a response time of 1 s and 1 nm bandwidth. A buffer baseline was subtracted from each spectrum and the spectra normalized to have zero ellipticity at 320 nm.

2.2.1.2 1D imino proton NMR

One-dimensional ^{1}H NMR experiments were performed on a Varian 600 MHz spectrometer. Oligonucleotides were prepared in a 100 mM potassium phosphate buffer (pH 7.4) and annealed by heating to 95 °C before slowly cooling to 15 °C over a period of 15 hours. 300 µl of oligonucleotide sample was mixed with 20 µl D_2O and placed in a Shigemi NMR tube. The final strand concentration was 100 µM. 1D proton NMR spectra were recorded at 25 °C with a sweep width of 25 p.p.m, WATERGATE water suppression, an acquisition time of 0.5 s and 32 k scans. Data were processed using VNMR software (Varian Inc.) with zero filling and resolution enhancement.

2.2.1.3 UV melting studies

Quadruplex thermal stability can be monitored by measuring changes in absorbance as a function of temperature. As the structure is heated, the unstacking of the bases leads to a change in absorbance, which can be monitored at an appropriate wavelength. The midpoint of the transition, or T_m, is the point at which half the complex is in the folded state and half is unfolded, and this value serves as an indication of the stability of the quadruplex structure.

UV melting studies were carried out using a Varian Cary 400 Scan UV-Visible spectrophotometer. Absorbance values were determined in quartz cuvettes with a 10 mm pathlength. Measurements were taken at a wavelength of 295 nm, as opposed to 260 nm which is the standard for most nucleic acid structures, as G-quadruplexes display a greater absorbance change and flatter baselines at this wavelength (Mergny *et al.*, 1998). At 295 nm, the transition from folded to unfolded results in a decrease in absorbance. Melting experiments were carried out in a final reaction volume of 1 ml containing 5 µM oligonucleotide, 10 mM lithium phosphate buffer and varying concentrations of potassium chloride or sodium chloride. All samples were first heated to 90 °C and held for 5 minutes to fully denature the complex. Samples were then cooled to 20 °C at a rate of 0.2 °C.min^{-1} and held again for a further 5 minutes to allow the samples to equilibrate. Samples were then finally heated to 90 °C at the same rate of change. Measurements were taken on both the melting and annealing transition stages to assess the reversibility of complex formation. T_m values were determined from the first derivative of the melting profiles.

2.2.1.4 Fluorescence melting studies

A second method for determining the thermal stability of quadruplex structures is by the use of a molecular beacons approach (Darby *et al.*, 2002). A fluorophore and quencher are located within the oligonucleotide sequence, such that when a melting transition occurs, a change in fluorescence is observed. For quadruplex-forming sequences, a fluorophore (fluorescein) is attached at the 5' end of the oligonucleotide and a quencher (dabcyl) attached at the 3' end. When the quadruplex is folded, the fluorophore and quencher are in close proximity and collisional quenching of the fluorescence occurs. As the complex melts, the fluorophore and quencher separate, which results in a large increase in fluorescence (Figure 2.1a). This approach can also be used to monitor the equilibrium between quadruplex and duplex structures. Addition of the (non-fluorescently-labelled) DNA strand complementary to the quadruplex-forming strand results in competition between the two DNA structural forms. In the quadruplex form, fluorescence emission is low, with fluorescence levels of the random coil much higher. However when a duplex is formed, fluorescence levels are higher still than random coil, as the time-averaged distance between the fluorophore and quencher is greatest when locked in the double helical form (Figure 2.1b).

Fluorescence melting curves were determined in a Roche LightCycler version 1 or 1.5 in a total reaction volume of 20 µl. Oligonucleotides (final concentration 0.25 µM) were prepared in 10 mM lithium phosphate buffer, which was supplemented with various concentrations of potassium chloride or sodium chloride. All samples were capped with 2 µl of mineral oil to prevent evaporation at high temperatures, necessary for experiments at very slow rates of heating. The LightCycler has one excitation source (488 nm) and the changes in fluorescence were measured at 520 nm. In a typical experiment, the oligonucleotides were first denatured by heating to 95 °C for 5 minutes. Samples were then annealed by cooling to 30 °C and then melted by heating to 95 °C at the same rate. The fluorescence was recorded during both the annealing and melting transitions to check the reversibility of the melting profiles. The rate of heating and cooling is specified for each individual set of oligonucleotides, although was usually 0.2 $^{\circ}C.min^{-1}$.

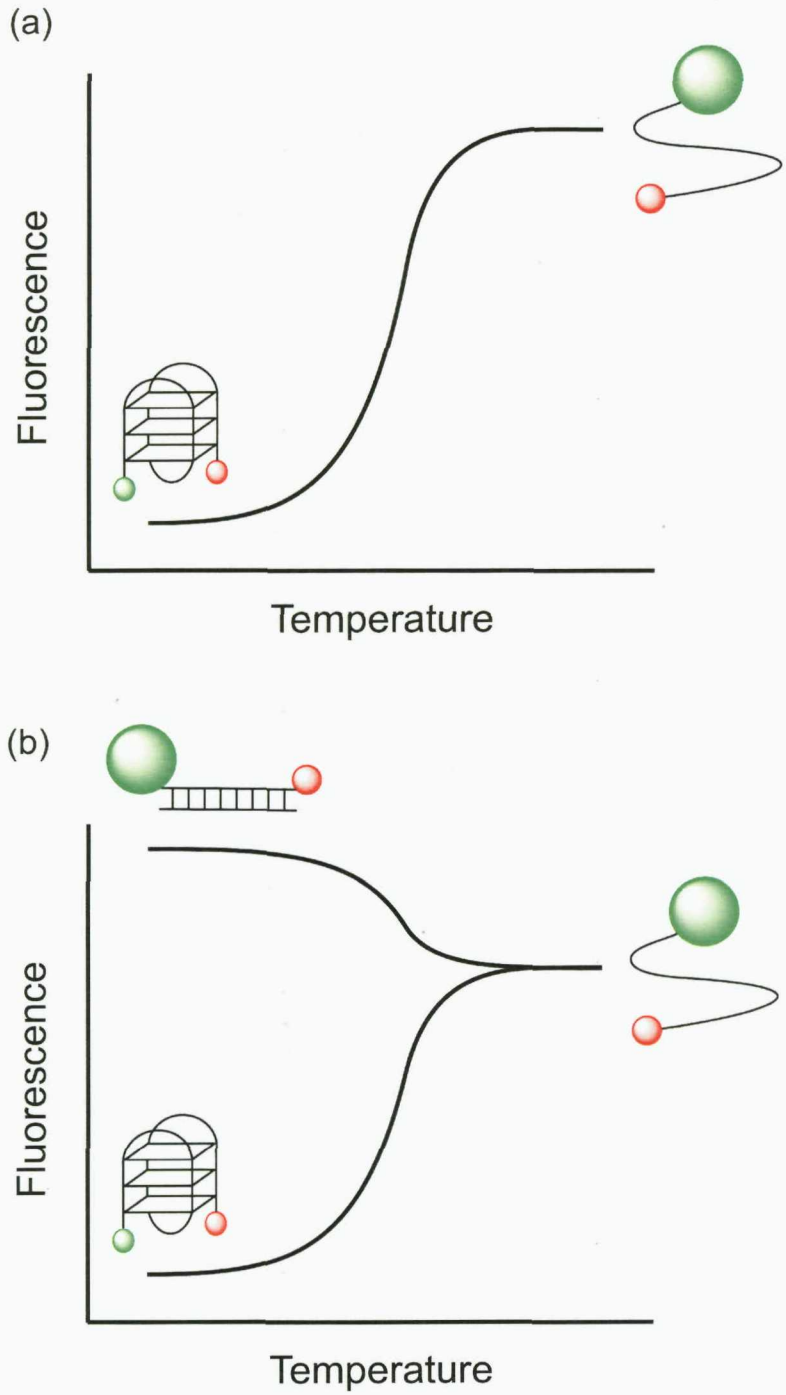


Figure 2.1 (a) Schematic representation of the melting of a fluorescently-labelled quadruplex-forming oligonucleotide. **(b)** Addition of the unlabelled complementary strand results in competition between duplex and quadruplex. The duplex exhibits the highest fluorescent signal as the time-averaged distance between the fluorophore and quencher is greatest when locked at opposing ends of a double helix.

2.2.2 Data analysis

T_m values were obtained from the maxima of the first derivatives of the melting profiles using the LightCycler software or, together with ΔH, from van't Hoff analysis of the melting profiles. In some instances, the melting curves showed a linear change in fluorescence with temperature in regions outside the melting transition. This was accounted for by fitting a straight line to the first and last portions of the fluorescence curve. Correct baseline determination is essential in order to obtain accurate thermodynamic parameters from melting curves (Mergny and Lacroix, 2003), therefore thermodynamic analysis was restricted to values between the temperature range of 35 °C – 85 °C. The fraction folded (α) can then be calculated as below, where F_L is the low temperature baseline, F_U the high temperature base line and F represents the measured fluorescence value. (Figure 2.2).

$$\text{Fraction folded } (\alpha) = \frac{(F - F_L)}{(F_U - F_L)}$$

Assuming the folding/unfolding is a simple unimolecular reaction and the unfolding is a single step, there will only be two species in equilibrium and the relative proportions at each temperature will be determined by the equilibrium constant:

$$K_{eq} = \text{folded/unfolded} = \alpha/1\text{-}\alpha$$

Since $\Delta G = -RT \ln(K_{eq}) = \Delta H - T.\Delta S$

$$\ln(K_{eq}) = \Delta H/R.\ 1/T + \Delta S/R$$

ΔH could therefore be obtained either by plotting ln(K) vs 1/T, or by fitting the data with a theoretical curve using FigP for Windows software. All reactions were performed at least twice and calculated T_m values usually differed by < 0.5 °C with a 5 % variation in ΔH.

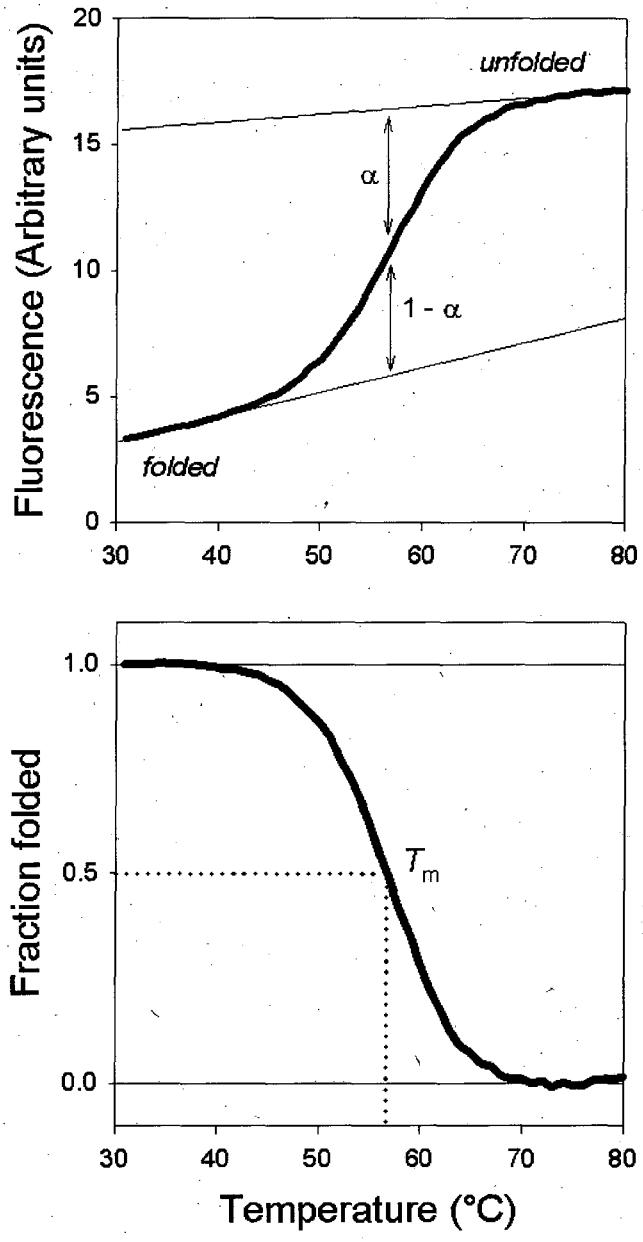


Figure 2.2 Conversion of a fluorescence melting profile to a fraction folded plot. Sloping start and end regions are corrected for by fitting upper and lower baselines to the melting profile (top panel), which can then be converted to a fraction folded plot (bottom panel). The temperature at which α is equal to 0.5 is defined as the melting temperature, T_m.

2.2.3 Salt concentration dependence

Quadruplex stability is known to be strongly dependent on ionic strength, which is likely to be a result of specific ion binding sites within the quadruplex, rather than ion-screening effects (Jing *et al.*, 1997). This ionic dependency can be used to determine the stoichiometry of ion binding as previously described (Jing *et al.*, 1997; Cantor and Schimmel, 1980).

The assumed melting transition is

$$\text{folded quadruplex} \leftrightarrow \text{single strand} + \Delta n\,[M^+]$$

where Δn represents the number of specifically bound K^+ or Na^+ released on melting. Δn can be calculated from the rate of change in ΔG as a function of $\ln[M^+]$.

$$\Delta n = \frac{d \ln K}{d \ln [M^+]} = \frac{\Delta\Delta G}{2.3\ RT\ \Delta \log [M^+]}$$

ΔH values were calculated at a range of sodium and potassium ion concentrations as described previously. Since $\Delta G = 0$ at the T_m, ΔS was estimated as $\Delta H/T_m$. Values for ΔG at 37 °C were then estimated from $\Delta G = \Delta H - T\Delta S$ (assuming that $\Delta Cp = 0$) and plotted as a function of $[M^+]$. The slope of the line could then be used to estimate values for Δn. This analysis assumes that there is a two-state equilibrium between the folded and unfolded forms and that ΔH is independent of temperature (i.e. $\Delta Cp = 0$).

2.2.4 Calculating kinetic parameters

2.2.4.1 Temperature-jump relaxation kinetics

The dissociation kinetics of intramolecular quadruplexes can be determined by a method analogous to that of temperature-jump relaxation kinetics (Callender and Dyer, 2002). Using the Roche LightCycler, samples were subjected to a rapid increase in temperature, causing the system to be perturbed from its equilibrium position. The subsequent rate of relaxation to its new equilibrium can then be measured (Figure 2.3).

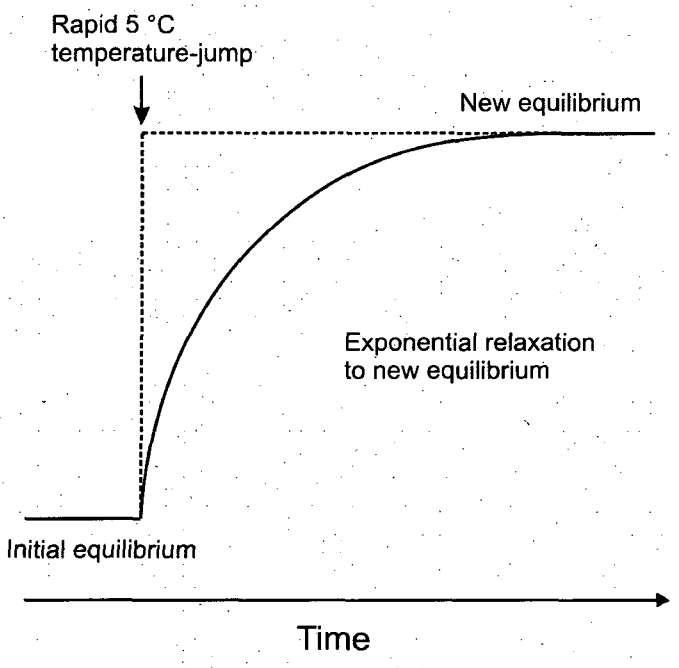


Figure 2.3 Schematic representation of a temperature-jump profile (Figure adapted from Fersht, 1977).

In a typical experiment, samples (0.25 μM oligonucleotide, 10 mM lithium phosphate buffer, 20 mM potassium chloride) were first heated to 95 °C to fully denature, before slowly cooling to 30 °C over a period of 10 hours to ensure correct folding. Following an equilibration period of 10 minutes, the temperature was rapidly increased by 5 °C at the fastest rate of heating (20 °C.s^{-1}). This increase in temperature causes dissociation of some of the quadruplex and therefore an increase in fluorescence. Samples were then allowed to equilibrate for between 3 and 10 minutes depending on the sequence (indicated where appropriate) and the fluorescence change recorded as the system relaxed to a new equilibrium. Although the theoretical dead-time for a 5 °C jump is only 0.25 s, all fluorescence changes that occurred in the first 2 s of the relaxation trace were disregarded. Successive temperature-jumps were carried out on the same sample by increasing the temperature by a further 5 °C, and temperature-jump steps continued until the quadruplex had fully dissociated.

The time dependent changes in fluorescence were then fitted with an exponential function, generating the relaxation constant (k) at each temperature. For unimolecular quadruplexes, the relaxation rate constant is equal to the sum of the folding (k_1) and unfolding (k_{-1}) rate constants (Czerlinski, 1966), although it is not possible to determine these rate constants individually.

2.2.4.2 Non-equilibrium melting curves

When the rate of heating and cooling is faster than the rate at which the quadruplex folds and / or unfolds, hysteresis is observed. As a result, the T_m for the melting transition is overestimated and the T_m of the annealing transition is underestimated. (Figure 2.4). Provided the reaction is unimolecular and concentration independent, individual kinetic parameters for association and dissociation can be extracted via analysis of this hysteresis as previously described (Rougee *et al.*, 1992; Bernal-Mendez and Leumann, 2002; Mergny and Lacroix, 2003).

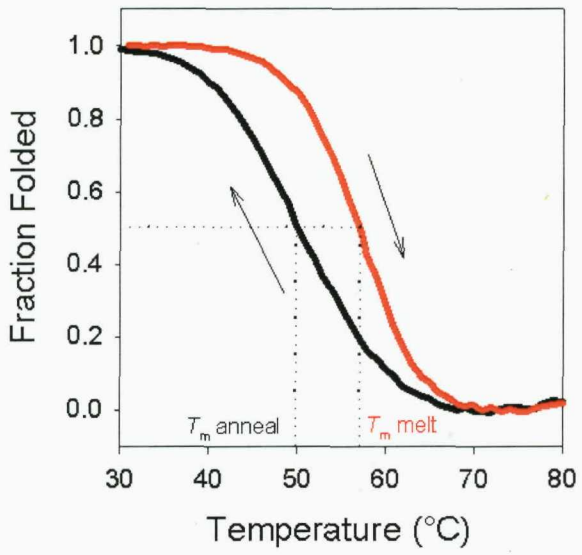


Figure 2.4 Representative non-equilibrium melting curves. When the rate of heating is faster than the rate of folding or unfolding, hysteresis is observed, whereby the transitions are not superimposable.

If α_c and α_h are the fractions of the folded quadruplex at any temperature in the cooling and heating profiles respectively, and t and T are the time and temperature, then:

$$d(\alpha_c)/dt = d(\alpha_c)/dt \times (dT/dt)^{-1}$$

and $$d(\alpha_h)/dt = d(\alpha_h)/dt \times (dT/dt)^{-1}$$

If k_1 and k_{-1} are the association and dissociation rate constants for quadruplex folding:

$$d(\alpha_c)/dt = k_1(1-\alpha_c) - k_{-1}\,\alpha_c$$

and $$d(\alpha_h)/dt = k_1(1-\alpha_h) - k_{-1}\,\alpha_h$$

By measuring $d(\alpha_c)/dT$, $d(\alpha_h)/dT$, α_c and α_h, the individual rate constants for association and dissociation can be estimated at a range of temperatures. This analysis requires that there is sufficient difference between the melting and annealing curves, but also that there is some overlap between the two profiles; for both heating and cooling curves there must be temperatures for which $0 < \alpha < 1$. This analysis can be checked by confirming that the estimated rate constants are independent of the rate of healing and cooling.

2.2.5 Polyacrylamide gel electrophoresis

For non-denaturing gel electrophoresis, oligonucleotide samples were prepared at a strand concentration of 20 μM containing an appropriate concentration of sodium or potassium phosphate (usually 50 mM). Each sample was heated to 95 °C and allowed to cool slowly to ensure correct annealing. Samples were run on 40 cm, 16 % polyacrylamide gels that were supplemented with 50 mM NaCl or KCl and run in TBE buffer supplemented with 50 mM NaCl or KCl. Gels were run at 500 V, 15 W for around 4 hrs at room temperature and then transferred on to Saran wrap to be examined under UV light. For denaturing electrophoresis, samples were boiled for 3 minutes prior to loading, and were performed using 14 % polyacrylamide gels containing 8 M urea for around 2 hours at 1500 V.

2.2.6 Polymerase arrest studies

2.2.6.1 γ-^{32}P Radiolabelling

For the polymerase arrest studies, the DNA primer was first 5'-end labelled by incubating 2 μl (5 μM) DNA with 1 μl (10 units) polynucleotide kinase (PNK), 2 μl PNK buffer, 2 μl [γ-^{32}P] dATP and 13 μl H_2O for 1 hour at 37 °C. 10 μl formamide stop solution was then added and the sample boiled for 3 minutes. The end-labelled DNA was then run on a 14 % polyacrylamide gel containing 8 M urea at 1500 V. After 2 hours, the gel was exposed to X-ray film to locate the position of the radiolabelled oligonucleotide (developed using an X-ograph Compact X2), and the radiolabelled fragment excised from the gel. The DNA was eluted into 300 μl elution buffer in a shaker for several hours. The DNA was then purified by ethanol precipitation by adding 30 μl 3 M NaOAc, 900 μl EtOH and leaving on dry ice for at least 20 minutes. Samples

were then spun at 13,000 rpm for 10 minutes and the supernatant removed. The DNA pellet was then washed by adding 100 µl 70 % EtOH before spinning again at 13,000 for 2 minutes. The supernatant was again removed, with any residual EtOH removed in a SpeedVac and the DNA resuspended in 10 µl Tris-Na buffer (10 mM Tris, 10 mM NaCl, pH 7.4). The labelled primer and 5 µl template DNA (5 µM) were then annealed by heating to 95 °C and cooling slowly to room temperature. The template-primer complex was then purified on an 8 % native polyacrylamide gel and eluted as above. The purified DNA was then diluted to 1 cps/µl as determined on a handheld Geiger counter.

2.2.6.2 Primer extension

Following purification, 2 µl of the primer-template complex was added to 2 µl polymerase stop reaction buffer and 2µl 0.5 mM dNTPs. Where appropriate, 2 µl of various concentrations of KCl or NaCl (0 - 100 mM final) were added and the samples incubated at room temperature for 30 minutes. 1 µl (5 units) Taq DNA polymerase was added following the initial incubation, and the primer allowed to extend at 37 °C for 15 minutes. The reaction was stopped by adding an equal volume of formamide stop solution and the samples boiled for 3 minutes. A guanine marker lane was also run on each gel. This was prepared by annealing unlabelled primer to the template and sequencing (using the protocol outlined below), incorporating dideoxy-G to highlight the position of all guanines in the sequence. Extension products were resolved on a 9 % polyacylamide gel containing 8 M urea for 1 hour at 1500 V, and then fixed in 10 % (v/v) acetic acid. Gels were then transferred to Whatman 3MM paper and dried under vacuum for at least 1 hour. Dried gels were subjected to phosphorimaging using a Molecular Dynamics Storm phosphorimager and visualised in ImageQuant 5.0.

2.2.7 Transfection assays

2.2.7.1 Luciferase reporter assay constructs

For studying the role of quadruplex formation on gene expression, the pGL3 luciferase reporter vector system (Promega) was used. Sequences of interest were cloned into the promoter-less pGL3-Basic vector, with the pGL3-Control vector used as an internal standard (Figure 2.5).

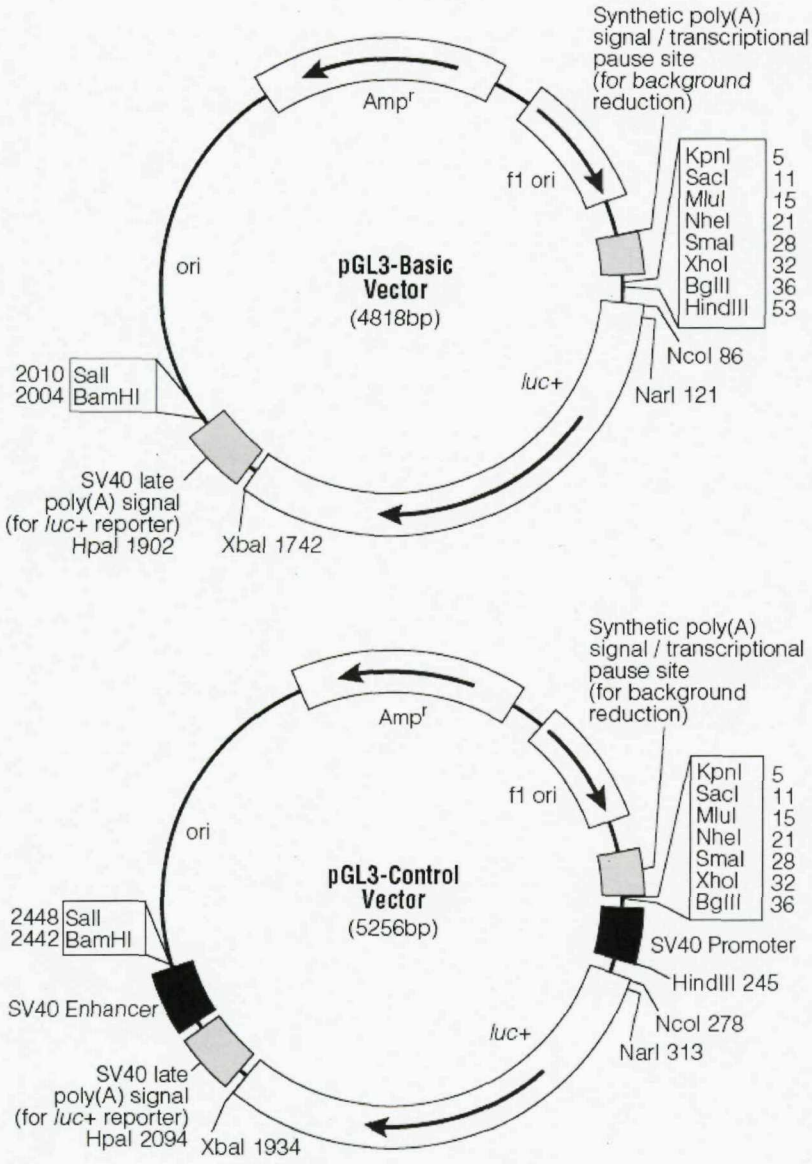


Figure 2.5 pGL3-Basic and Control promoter expression vectors (Promega).

2.2.7.2 Cloning

Promoter sequences to be cloned into the reporter plasmid were divided into three cassettes, allowing different length (and sequence) constructs to be assayed for promoter activity. For full details of these cassettes, including the oligonucleotide sequences see Chapter 5 p133.

Promoter sequences were initially diluted to 5 μM and annealed by heating to 95 °C before cooling to room temperature over several hours. Insert sequences were cloned

into either the KpnI / BglII site (3 cassette constructs) or MluI / BglII site (2 cassette constructs) (Figure 2.5). To open the plasmid, 1 µl pGL3 plasmid, 2 µl enzyme buffer, 1 µl (10 units) of each restriction enzyme and 15 µl H_2O were combined and allowed to incubate for 1 hour. 2 µl (2 units) shrimp alkaline phosphatase (SAP) and 2 µl SAP buffer were then added and incubated at 37 °C for a further hour, before the phosphatase was inactivated by heating to 65 °C for 15 mins. Samples were then ethanol precipitated and resuspended in 10 µl H_2O. To ligate in the inserts, 5 µl of each insert were combined with the purified 10 µl pGL3 vector, together with 3 µl (30 units) T4 DNA ligase and 2µl ligase buffer before leaving at room temperature for 3 hours. Cells were then transformed (see below) and grown on carbenicillin-agar plates (100 µg/ml). Several colonies were picked and sequenced to ensure the plasmids contained the correct insert.

2.2.7.3 Competent cell preparation

E.coli TG2 cells were grown in 5 ml 2YT media (16 g tryptone, 10 g yeast extract, 5 g NaCl / L) in a 37 °C shaker overnight. 50 ml 2YT media was then inoculated with 0.5 ml of the overnight culture and allowed to grow to an OD of between 0.5 and 0.8 at 600 nm. Media were then transferred to a centrifuge tube and spun at 5000 rpm at 4 °C for 5 mins. The supernatant was removed and the pellet resuspended in 20 ml transformation buffer and left on ice for 30 mins. Cells were then spun again at 5000 rpm for 5 mins, the supernatant removed and the pellet resuspended in 4 ml transformation buffer. Competent cells were stored at 4 °C for a maximum of 2 weeks.

2.2.7.4 Transformation

5 µl ligated plasmid DNA was added to 200 µl competent TG2 cells and placed on ice for 30 minutes. Cells were then heat-shocked for 1 minute in a 45 °C water bath and returned to ice. Cells were then spread on carbenicillin-agar plates (100 µg/ml), grown overnight at 37 °C and the plates stored at 4 °C.

2.2.7.5 Plasmid purification

Following transformation, a single colony was picked and grown overnight in 5 ml 2YT media. For sequencing, plasmids were purified using a Miniprep spin kit (Qiagen) as

described in the QIAprep handbook. DNA was eluted into 50 µl buffer (10 mM Tris, pH 8.5) and stored at -20 °C. For transfection experiments where higher yields were required, 100 ml 2YT media was inoculated with 1 ml of the initial 5 ml culture and grown overnight. Plasmid DNA was then purified using a Plasmid Maxi kit (Qiagen) as described in the handbook, diluted to 1 µg/µl in deionised water and stored at -20 °C.

2.2.7.6 Dideoxy Sequencing

40 µl of plasmid DNA prepared as described above was first denatured by adding 10 µl of 2 M NaOH. Samples were then allowed to stand at room temperature for 10 minutes. The DNA was purified by ethanol precipitation as described previously and the samples were resuspended in 10 µl deionised water. The sequencing procedure was carried out using a T7 Sequencing kit (USB). To each 10 µl sample, 2 µl annealing buffer (from T7 kit) and 2 µl (5pmol/µl) of RVprimer 3 (Promega) were added and allowed to incubate at 37 °C for 20 minutes and then at least 10 minutes at room temperature. For each sample, four microcentrifuge tubes were set up containing 2.5 µl of the dideoxy mixes G-short, C-short, T-short or A-short and placed in a 37 °C heat block. A polymerase mix, containing 12 µl label mix A, 7 µl deionised water and 1 µl [α-^{32}P] dATP was added to an enzyme mix, containing 6.5 µl enzyme buffer and 1.5 µl T7 polymerase, which was then left on ice. 6 µl of the polymerase-enzyme mix was added to each plasmid DNA sample and left at room temperature for 5 minutes. For the sequencing reaction, 4.5 µl of the reaction mixture was added to each of the dideoxy mixes and allowed to polymerise for 5 minutes at 37 °C. The reactions were terminated by adding 5 µl formamide stop solution and boiled for 3 minutes. Samples were resolved on a 9 % polyacylamide gel containing 8 M urea for around 2 hours at 1500 V, and then fixed in 10 % (v/v) acetic acid. Gels were then transferred to Whatman 3MM paper and dried under vacuum for at least 1 hour. Dried gels were subjected to phosphorimaging using a Molecular Dynamics Storm phosphorimager and visualised in ImageQuant 5.0.

2.2.7.7 Transfection

Raw264.7 macrophage cells for use in all transfection studies were maintained in Nunclon flasks containing 25 ml growth media (Dulbeccos modified eagles medium (DMEM) containing 10 % foetal calf serum (FCS)) in a 5 % CO_2 incubator. Cells were harvested by gentle scraping, transferred to a sterile polypropylene tube and cell

numbers obtained using a haemocytometer, from a 1:1 dilution of the cell suspension and trypan blue. Cells were diluted to 0.5 x 10^6 /ml with DMEM containing 10 % FCS. 1 ml of the cell suspension was loaded into each well of a 12-well tissue culture plate (Greiner) and allowed to grow overnight. Next day, the media was removed by aspiration and the cells washed twice with DMEM, before 0.5 ml fresh DMEM (lacking FCS) was added to each well. Cells were then returned to the CO_2 incubator prior to transfection.

Transfection complexes were prepared using 1µl of each pGL3 plasmid construct (1 µg/µl) diluted to 100 µl with DMEM (Tube 1). 5 µl of the cationic transfection reagent Metafectene™ (Biontex) was then added to a further 100 µl of DMEM (Tube 2). Tubes 1 and 2 were then combined and incubated at room temperature for 20 minutes. After incubation, 60 µl of the plasmid-Metafectene complexes were dispensed into each of the three replicate wells, and the plate returned to the incubator for four hours. Following this period, 0.5 ml 20 % FCS-DMEM was loaded to each well (10 % FCS final) and returned to the incubator overnight.

2.2.7.8 Luciferase Assay

Cells were harvested according to the manufacturers instructions (Promega luciferase assay system). Briefly, the growth medium was carefully removed and cells were rinsed with phosphate buffered saline (PBS). 50 µl lysis buffer was added to each well and the cells scraped from the plate. The cells and all the buffer were then transferred to microcentrifuge tubes and spun at 13,000 rpm for 5 mins, before the supernatant was transferred to a microtiter plate and stored on ice. Luciferase assay reagent was prepared by adding 10 ml luciferase assay buffer to the lyophilized substrate, and the stock stored at -80 °C. To carry out the luciferase assay, 50 µl luciferase reagent was loaded into luminometer tubes, followed by 5 µl of the cell lysate. Luminescence measurements were taken on a TD-20/20 luminometer, with each sample measured in triplicate and each plasmid transfection repeated three times per 12 well plate.

Luciferase expression was normalised to account for the protein concentration within the supernatant, determined using a Bio-Rad protein assay. On a 96 well microtiter plate, 0 to 11 µl of 4 mg/ml bovine serum albumin (BSA) was pipetted out to act as the standard, alongside 2.5 µl of each cell lysate. Dye was prepared by diluting 1 part dye

reagent concentrate with 4 parts distilled water before loading 250 µl to each well. Absorbance was measured at 570 nm on a Dynex Technologies microtiter plate reader and a standard curve constructed from these data. Luciferase expression could then be calculated as mean LU/µg protein.

CHAPTER 3

Fluorescence Melting as a Tool for Measuring G-Quadruplex Folding and Stability

3.1 Introduction

The conversion of a single-stranded DNA sequence into a folded G-quadruplex is a large structural transition, which results in the 5' and 3'- ends lying in close proximity. The conjugation of fluorescent probes to the termini of quadruplex-forming sequences can allow quadruplex formation to be monitored by means of fluorescence resonance energy transfer (FRET), or Dexter (collisional) quenching. The use of FRET for studying quadruplex formation was first reported using a 22nt region of the c-myc nuclease hypersensitivity element (NHE) (Simonsson and Sjöback, 1999), and has since been employed to study a variety of quadruplex properties, including their structure and kinetics (Ying *et al.*, 2003; Green *et al.*, 2003; Brown *et al.*, 2005), equilibrium with duplexes (Risitano and Fox, 2004; Kumar and Maiti, 2004) and metal ion complexation (Nagatoishi *et al.*, 2006). Fluorescence-based melting assays are also an extremely valuable tool for screening libraries of potential quadruplex-stabilising ligands (Mergny *et al.*, 2001; De Cian *et al.*, 2006). This chapter introduces the use of a fluorescence-melting assay for monitoring the folding and stability of different intramolecular quadruplexes, a technique used extensively in later chapters.

A high-throughput method for determining the stability of nucleic acid structures using a Roche LightCycler via the attachment of molecular beacons has been developed (Darby *et al.*, 2002) and applied to the study of intramolecular quadruplex formation. This method differs from FRET in the choice of the fluorescent probes; FRET involves energy transfer between an excited fluorophore and an acceptor (typically FAM and TAMRA) such that following excitation of the donor, energy is transferred to the acceptor, which emits this energy at its characteristic emission wavelength (Figure 3.1a). This transition can therefore be monitored as either the decrease in donor emission or increase in acceptor emission. The fluorescence melting approach instead

makes use of a fluorophore and a quencher pair (FAM and dabcyl) so that when both are in close proximity, collisional (Dexter) quenching occurs (Figure 3.1b).

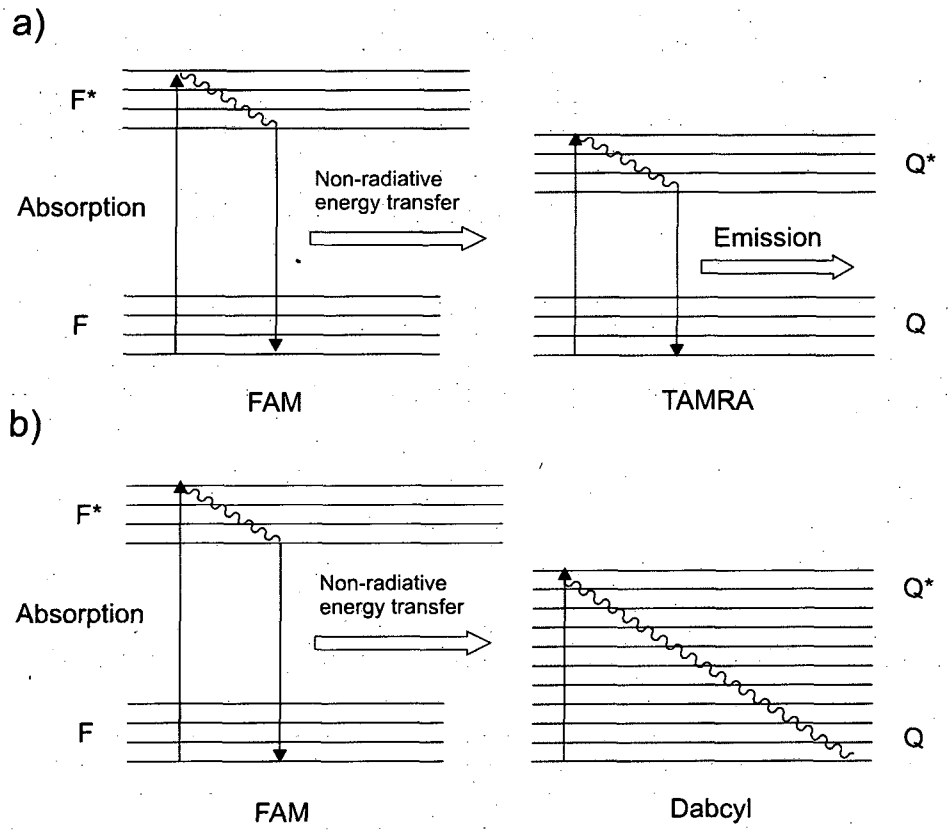


Figure 3.1 Jablonski energy level diagram depicting **(a)** FRET and **(b)** collsional (Dexter) quenching.

The fluorescence melting technique is beneficial with respect to monitoring intramolecular quadruplex formation, as the distance between the fluorescent groups in the folded complex can be too short for FRET, where the decrease in donor emission is often the result of collisional quenching (Nagatoishi *et al.*, 2006). Following the emission of the fluorophore over a temperature range typically between 25 - 95 °C allows the structural transition from quadruplex to random coil to be monitored. This technique therefore enables comparisons to be made between the stability of different quadruplex-forming oligonucleotides.

3.1.1 Fluorescence verses Absorbance

The fluorescence melting technique has several advantages over conventional UV absorbance studies; (i) Smaller volumes are required (20 µl) compared to 1-3 ml for UV studies. (ii) Much lower DNA concentrations are also needed (typically 0.25 µM),

which favour the formation of *intra* rather than *inter*molecular quadruplexes. (iii) Using a Roche LightCycler real-time PCR machine, up to 32 samples can be analysed simultaneously compared to only 6 samples for most UV spectrophotometers.

3.2 Experimental Design

Although fluorescence melting is a convenient and simple technique, it requires the attachment of reporter molecules to the oligonucleotide, which may influence quadruplex structure and/or stability. Quadruplexes are also known to be extremely polymorphic, with some sequences able to adopt a number of metastable forms. It is therefore important that the attachment of these dyes results in minimal disruption to the distribution of folded complexes. This preliminary chapter examines several aspects of fluorescence melting which are important when applying this technique for the study of G-quadruplexes. These include the choice of oligonucleotide, the position and attachment of the fluorescent probes, the rate of heating and the effects of different cations. These parameters have been assessed using a sequence based on three and a half repeats of the human telomeric sequence $d[(G_3TTA)_3G_3]$. The properties of the structures formed by these fluorescent- and non-fluorescently labelled oligonucleotides have been probed by comparing their fluorescence and absorbance melting profiles, and also through comparison of their CD spectra.

3.3 Results

3.3.1 Oligonucleotide design

When designing quadruplex-forming oligonucleotides containing terminal fluorophores and quenchers, there are a number of important considerations regarding the exact location and attachment of these groups:

(i) The chemical linkage of the fluorescent probes to the oligonucleotide
(ii) Presence/absence and identity of spacer nucleotides
(iii) The orientation of the fluorescent dyes (5'-FAM or 3'-FAM?)

The first requirement is that the fluorescent probes must not stack or interact with the quadruplex. Several fluorescent probes are similar in structure to those of multi-ring

aromatic compounds designed to stabilise quadruplexes by end-stacking on the terminal quartets. FAM and dabcyl have therefore been attached to the oligonucleotides via 6-carbon aliphatic linkers to minimise fluorophore-DNA interference, while allowing fluorophore-quencher interaction (Figure 3.2a). These linkers are relatively long, and there is therefore no significant difference in quenching efficiency between different folded quadruplex forms (Figure 3.2b). As a result, this technique is unable to distinguish between the various quadruplex topologies.

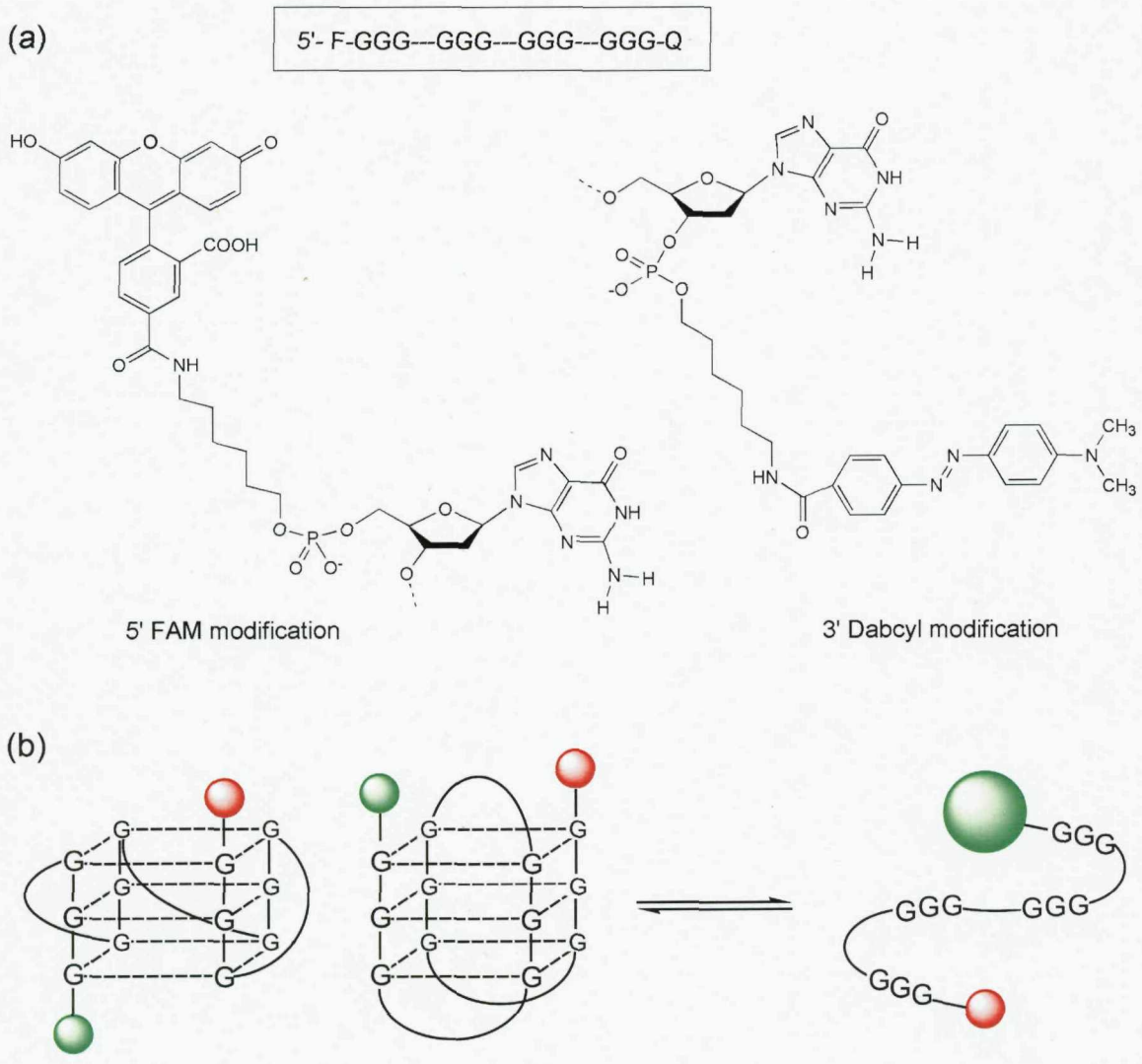


Figure 3.2 (a) Chemical linkage of the fluorescent probes 6-amidohexylfluorescein (FAM), and dabcyl to the 5' and 3'-ends respectively of a quadruplex-forming oligonucleotide. **(b)** Schematic representation of the melting of a quadruplex-forming oligonucleotide that has been labelled with a fluorophore (green) and quencher (red). The folding of a parallel or antiparallel-stranded structure results in fluorescence quenching of similar efficiency.

A second consideration is whether the fluorescent groups should be attached immediately adjacent to the G-tetrads, or separated from the G-stack by other nucleotides, and if so, which ones? Placing the fluorophore immediately adjacent to a guanine base can result in fluorescence quenching by G-base stacking (Marras *et al.*, 2002; Maruyama *et al.*, 2005; Saito *et al.*, 2007) therefore ideally one or more bases should separate the reporter groups from the quadruplex. This also adds further distance between the fluorophore and quencher and the stacked G-quartets. For studies with the human telomeric sequence, the 5' and 3'-terminal bases should be A and T respectively, but for other model quadruplexes the choice is less clear. For any series of related quadruplexes (such as when varying the length of the loop regions) these bases should be kept constant. Table 3.1 shows the effect of altering the flanking bases on the stability of the human telomeric repeat sequence. It can be seen that the sequence with two thymidines at the 5' and 3' ends forms the least stable structure, and quadruplex stability increases as the flanking nucleotides are removed. The identity of these bases is also significant; the presence of a 5'-A, which is found in the native telomeric sequence, lowers the stability more than a 5'-T. The fluorophore is usually positioned at the 5'-end, with the quencher at the 3'-end, as this is the simplest and cheapest method for synthesis. The position of these fluorescent groups can be reversed, however the effects on stability can be significant (Table 3.1). Placing methyl red at the 5' and fluorescein at the 3'-end increased the stability of the human telomeric repeat by almost 10 °C compared to the same sequence with the fluorescent reporter groups the other way round. It is clear that any comparison between different quadruplex-forming sequences must use the same orientation of fluorophore and quencher moieties.

Oligonucleotide sequence	T_m
5'- FAM - GGGTTAGGGTTAGGGTTAGGG - dabcyl	59.4
5'- FAM - **T** GGGTTAGGGTTAGGGTTAGGG **T** - dabcyl	58.1
5'- FAM - **A** GGGTTAGGGTTAGGGTTAGGG **T** - dabcyl	55.9
5'- FAM - **TT** GGGTTAGGGTTAGGGTTAGGG **TT** - dabcyl	52.8
5'- Methyl red - **A** GGGTTAGGGTTAGGGTTAGGG **T** - FAM	64.8

Table 3.1 Melting temperatures (°C) of the quadruplex forming oligonucleotides containing different nucleotides between the terminal guanosine and the fluorophore and quencher. Experiments were performed in 10 mM lithium phosphate buffer, pH 7.4 containing 50 mM KCl. Samples were heated at a rate of 0.2 °C.min^{-1} and no hysteresis was observed. Flanking bases are shown in bold.

3.3.2 Fluorescence melting verses UV melting

Fluorescent dyes have been shown to have relatively little effect on duplex (Moriera *et al.*, 2005) and triplex (Wang *et al.*, 2005) stability, however studies with fluorescently-labelled quadruplex (Mergny and Maurizot, 2001) and i-motif (Mergny, 1999) structures have shown these groups can have a more pronounced effect. To assess the effect of the FAM-dabcyl combination on quadruplex stability, the melting profiles of the human telomeric repeat sequence have been determined in the presence sodium and potassium ions, and compared with those of the unlabelled oligonucleotide. Figure 3.3 compares the fluorescence melting profile of F-GGGTTAGGGTTAGGGTTAGGG-Q with the UV melting profile for d(GGGTTAGGGTTAGGGTTAGGG) and the results are summarised in Table 3.2. It can be seen that the two methods give very similar profiles, though addition of the fluorophore decreases the T_m by a few degrees. The effect is similar in sodium (in which the quadruplex adopts a different topology) with the T_m value of the labelled oligonucleotide 53.6, and the T_m of the unlabelled sequence 50.8 °C. It is also important to note that the T_m values derived from the fluorescence melting experiments are in close agreement with the values derived from UV melting of the same fluorescently-labelled sequence. This suggests the transition observed in fluorescence melting experiments accurately represent the folding/unfolding of the quadruplex. The destabilisation of quadruplex structures by the attachment of fluorescent groups has been previously reported (Mergny and Maurizot, 2001; Green *et al.*, 2003) and the thermodynamic effects can be both dye specific and dependent on whether the oligonucleotide is single or dual labelled. This is considered further in the Discussion.

Oligonucleotide sequence	KCl	NaCl
5'- d(GGGTTAGGGTTAGGGTTAGGG)	63.3	53.6
5'- F-GGGTTAGGGTTAGGGTTAGGG-Q	59.4	50.8

Table 3.2 T_m values of the 3.5 repeat human telomeric sequence derived from UV melting or fluorescence melting profiles in the presence of 50 mM KCl or NaCl. F = FAM, Q = Dabcyl.

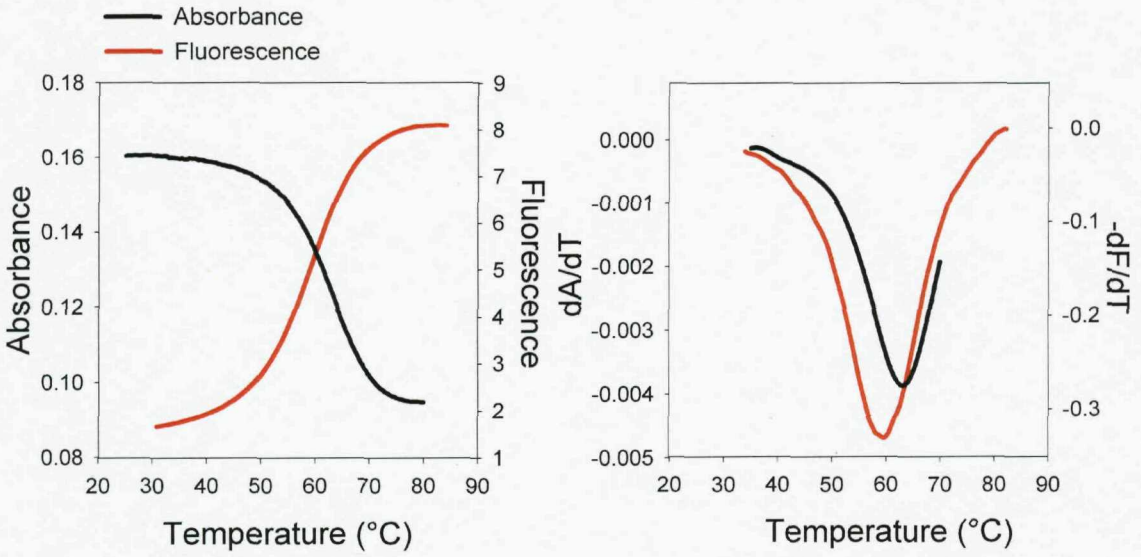


Figure 3.3 (a) Comparison of the melting curves obtained for d(GGGTTAGGGTTAGGGTTAGGG), determined by changes in absorbance at 295 nm, and F-GGGTTAGGGTTAGGGTTAGGG-Q (F = FAM and Q = dabcyl) determined by changes in the fluorescence, measured in 10 mM lithium phosphate buffer, pH 7.4, containing 50 mM KCl. **(b)** First derivative plots of the data shown in (a).

3.3.3 Fluorophore effects on topology

The simplest method of comparing the global topologies assumed by the quadruplexes formed from labelled and non-labelled oligonucleotides is by circular dichroism (CD). CD spectra can give an indication of quadruplex topology; Quadruplexes in which all the strands are in a parallel orientation typically display a positive peak at 265 nm, whereas structures containing antiparallel strands display a positive peak at 295 nm. Although these spectra are of low structural resolution, they allow a rapid comparison to be drawn between global folds. Figure 3.4 shows the CD spectra of the 21mer human telomeric repeat with and without fluorescent groups attached, in the presence of potassium or sodium. These results show that the spectral peaks of each oligonucleotide are in the same position, indicating that under these conditions, the labelled and unlabelled oligonucleotides adopt the same topology. The fluorescently-labelled oligonucleotide exhibits a slightly lower signal intensity than the unlabelled sequence, although this may be a result of the lower stability of the labelled structure. The human telomeric repeat sequence is known to form different folded structures in the presence of sodium (antiparallel) (Wang and Patel, 1993) and potassium (mixed strand polarity), (Luu *et al.*, 2006; Ambrus *et al.*, 2006) and these differences are reflected in their CD spectra.

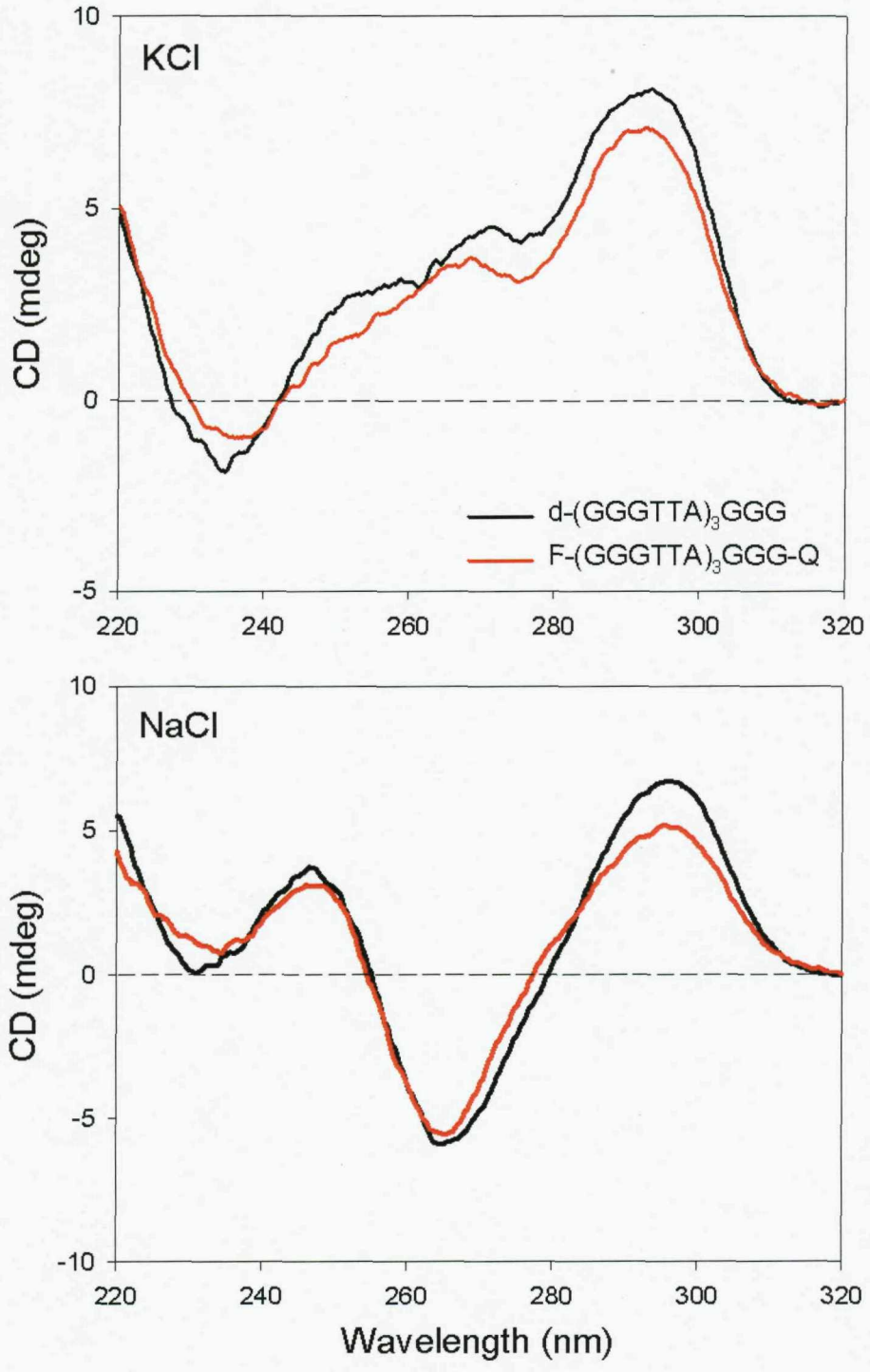


Figure 3.4 CD spectra of the 21mer human telomeric repeat with fluorescent groups attached (red) and removed (black). Oligonucleotides (5 μM) were dissolved in 10 mM lithium phosphate buffer, pH 7.4, containing 50 mM potassium chloride (top) or 50 mM sodium chloride (bottom). Spectra are buffer corrected and normalised to zero at 320 nm.

The fluorescently-labelled oligonucleotides display similar spectra to their non-labelled counterparts in both sodium and potassium, suggesting the attached dyes hinder neither topological conformation. This is in contrast to changes seen with an analogous FAM-TAMRA duel-labelled oligonucleotide, in which the fluorescent dyes caused a significant change in the CD spectrum (Mergny and Maurizot, 2001). However this change was though not to be a result of disruption to the quadruplex structure itself, but due to interaction between the two dyes, which both absorb light in this region (Mergny and Maurizot, 2001).

3.3.4 Rate of temperature change.

Since many quadruplex-forming sequences can adopt several conformations, it is essential that oligonucleotides are properly annealed prior to use. The slowest rate of continuous temperature change that can be achieved using the LightCycler is 6 °C.min^{-1} (0.1 °C.sec^{-1}). At this rate of temperature change, the melting profiles of many intramolecular quadruplex-forming sequences exhibit hysteresis (Risitano and Fox, 2003; Risitano and Fox, 2004), i.e. the temperature gradient is faster than the rate of quadruplex folding or unfolding. Figure 3.5 shows representative fluorescence melting and annealing profiles of three quadruplex-forming sequences at a temperature gradient of 6 °C.min^{-1}; G_3TTA (F-AGGGTTAGGGTTAGGGTTAGGGT-Q), G_4T_4-T-T_4 (F-TGGGGTTTTGGGGTGGGGTTTTGGGGT-Q) and G_4T_4 (F-TGGGGTTTTGGGGT TTTGGGGTTTTGGGGT-Q). It can be seen that these sequences exhibit varying degrees of hysteresis at this temperature gradient, which prevents the determination of equilibrium T_m measurements. This limitation can be overcome by programming the LightCycler to increase the temperature in small steps (1 °C), and recording the fluorescence values at each temperature after a specific equilibration time. The bottom panels of Figure 3.5 show representative melting and annealing profiles of the same three oligonucleotides at a temperature transition rate of 0.2 °C.min^{-1} (i.e. 1 °C every 5 mins). At this slower rate, hysteresis is eliminated for G_3TTA and G_4T_4-T-T_4 and the curves appear to be at thermodynamic equilibrium. G_4T_4, which is known to exhibit exceptionally slow folding/unfolding kinetics (Brown *et al.*, 2005), still exhibits hysteresis and a slower rate would be required to eliminate this.

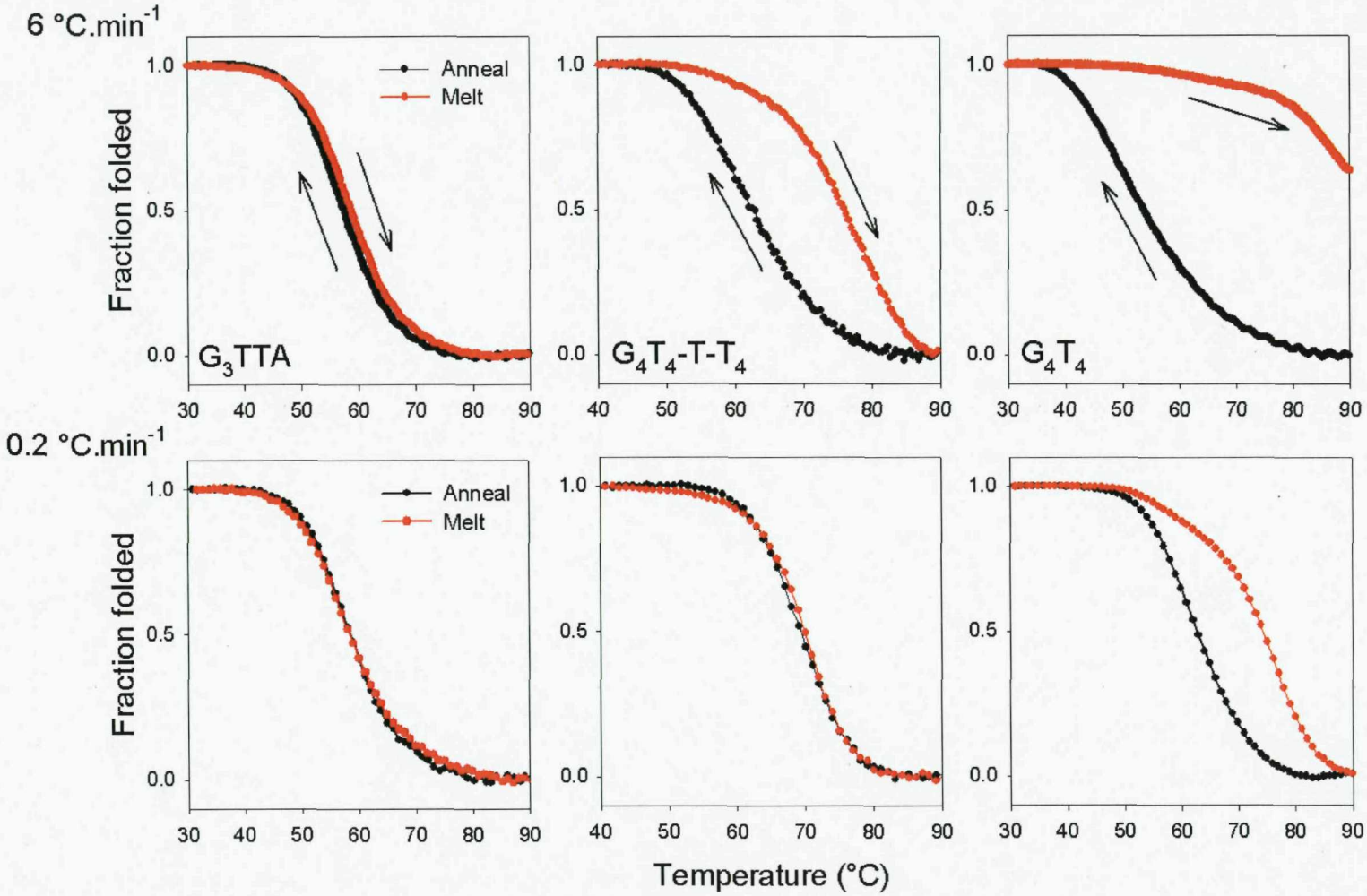


Figure 3.5 Melting and annealing profiles of three different quadruplex-forming oligonucleotides at 6 °C.min^{-1} (top panels) and 0.2 °C.min^{-1} (bottom panels). G_3TTA = F-AGGGTT AGGGTTAGGGTTAGGGT-Q, G_4T_4-T-T_4 = F-TGGGGTTTTGGGGTGGGGTTTTGGGGT-Q and G_4T_4 = F-TGGGGTTTTGGGGTTTTGGGGTTTTGGGGT-Q.

3.4 Discussion

The aim of the work described in this chapter was to introduce the use of fluorescence melting for studying intramolecular G-quadruplex formation and stability. Fluorescence melting is a high throughput assay that can be used to examine the stability of series of related quadruplex-forming sequences and also the interaction of quadruplexes with potential stabilising compounds. The use of a real-time PCR platform allows the fluorescence emission of 32 samples to be monitored simultaneously, over a temperature range typically between 25-95 °C. This represents a far higher throughput than can be achieved by conventional melting techniques such as UV or CD.

3.4.1 The fluorescent oligonucleotide

It has previously been reported that attaching fluorescent probes to quadruplex forming oligonucleotides can affect their biophysical properties (Mergny and Maurizot, 2002). The effects of attaching the fluorophore, FAM and the quencher, dabcyl to the termini of 3.5 repeats of the human telomeric sequence has been examined. Attaching FAM at the 5'-end and dabcyl at the 3'-end of the oligonucleotide via 6-carbon linkers resulted in destabilisation of the quadruplex by around 3-4 °C as compared to the unlabelled oligonucleotide in both sodium and potassium buffers. A similar effect has been observed previously with both FAM-TAMRA (Mergny and Maurizot, 2002), and TAMRA-Cy5 (Green *et al.*, 2003) labelled quadruplex-forming oligonucleotides, although the destabilisation was closer 10 °C in each of these cases. The effects on quadruplex stability are likely to be specific to both the identity of the fluorescent groups attached and the nature of the chemical linkers, although in general they are destabilising relative to the native oligonucleotide. In the case of FAM, the destabilising effect may be a result of an additional negative charge, as additional phosphates at the oligonucleotide termini are known to be destabilising (Mergny and Maurizot, 2002). The consistency of the destabilising influence in both sodium and potassium buffers is significant, as this shows the negative effect on stability of the two fluorescent dyes is non-discriminatory between different structural forms.

The presence of flanking nucleotides adjacent to the quadruplex also results in destabilisation of the human telomeric repeat structure, although these nucleotides further increase the distance between the fluorophore and the terminal quartets, and can

also reduce G-quenching. Their presence will also more accurately reflect the situation *in vivo*, in which nucleotide tails will flank any quadruplexes that form.

3.4.2 CD spectroscopy

Comparing the CD spectra of the labelled and unlabelled sequences has allowed the effect of fluorescent groups on global quadruplex topology to be assessed. In the presence of both sodium and potassium, the CD spectra were a close match, suggesting that the attachment of FAM and dabcyl does not result in significant disruption to the quadruplex structure whether it is an antiparallel fold or a mixed strand polarity structure. However for sequences that fold to form a number of structures of similar topology (i.e. similar loop orientations but in a different order), it cannot be discounted that the equilibrium between these structures is disrupted or altered by the presence of fluorescent groups. Similar studies using alternative sequences, including the *Oxytricha* telomeric repeat G_4T_4 (TGGGGTTTTGGGGTTTTGGGGTTTTGGGGT) and G_3T (TGGGTGGGTGGGTGGGT) also revealed no significant differences between the spectra of the labelled and unlabelled oligonucleotides.

3.4.3 Rate of heating

The rate of quadruplex folding and unfolding can be extremely fast, but equally exceptionally slow, dependent on the nucleotide sequence and the identity of the counterion. The use of a real-time PCR machine for studying quadruplex formation requires designing new temperature gradient profiles, as the standard protocols are too fast to allow equilibrium T_m measurements. The standard temperature gradient used for the majority of the melting studies in later chapters is 0.2 $^{\circ}C.min^{-1}$, although modifications to this rate are indicated where appropriate.

3.4.4 Conclusions

The following chapters make extensive use of the fluorescence melting technique for the determination of thermodynamic and kinetic parameters of series of related quadruplex-forming oligonucleotides. In these experiments, all oligonucleotide sequences contain a 5'-FAM and a 3'-dabcyl, with each sequence containing a single T spacer between the terminal guanine and the fluorescent probes (with one exception).

CHAPTER 4

Effect of Loop Length and Position on the Thermodynamic, Kinetic and Folding Properties of G-Quadruplexes

4.1 Introduction

G-quadruplexes all share a common structural core in the form of stacked G-quartets. Despite this, intramolecular quadruplexes still exhibit a diverse array of folding topologies, and this variation is primarily due to differences in the size and composition of the loop regions. This chapter investigates how loop length and position can affect the biophysical properties of intramolecular DNA G-quadruplexes.

In many cases, the same guanine-rich sequence can fold to form a number of different quadruplex structures, however the biological function may well depend on one particular conformation, especially if this involves interaction with specific proteins. Such an effect has been suggested for the NHE element of the c-myc promoter, which is able to form a number of different folded structures (Siddiqui-Jain *et al.*, 2002). An understanding of how loop regions affect not only quadruplex folding, but also the thermodynamic and kinetic properties would be extremely beneficial for the long-term goal of being able to predict quadruplex properties from sequence alone. Determining high-resolution structures is one way in which the role of loops can be assessed (Hazel *et al.*, 2006), however the scope of these techniques is limited due to the sensitivity of quadruplexes not only to sequence, but also experimental conditions. More commonly, biophysical experiments and molecular dynamics simulations allow a more systematic approach to studying the role of the loops.

For most intramolecular quadruplexes, three loop regions separate the G-tracts. Variation in both loop sequence and length have been shown to alter quadruplex stability (Smirnov and Shafer, 2000), topology (Hazel *et al.*, 2005) and molecularity (Vorlickova *et al.*, 2007) although nucleosidic linkers are not necessarily required to achieve quadruplex formation (Risitano and Fox, 2004; Cevec and Plavec, 2005). It has

been shown that short loops consisting of a single thymine adopt parallel stranded structures which are very stable, whereas longer loops prefer to form an antiparallel strand arrangement (Hazel *et al.*, 2005). These effects have been attributed to base-stacking interactions and hydrogen bonding within the loop regions (Keniry *et al.*, 1997). These studies provided the first general rules regarding quadruplex folding, although further investigation is still required before the role of the loops in influencing topology is properly understood. Many studies to date have used model sequences in which all the loops-lengths are equal; this chapter examines the effect of combining short and longer loops within the same sequence on quadruplex folding. The effects on the thermodynamic and kinetic properties of the structures formed are investigated.

4.2 Experimental Design

The oligonucleotide sequences chosen for this study are based on one very stable intramolecular quadruplex that contains four G_3 tracts separated by single T residues (Risitano and Fox, 2003; Hazel *et al.*, 2005). Each of the single T loop regions has been systematically replaced with T_4 to examine how the length and position of the loops affects quadruplex folding and stability. Single T loops are limited in their potential orientation due to their short span, but T_4 loops are long enough to form all possible loop types. In addition, quadruplexes which contain all single nucleotide T loops, have been shown to fold extremely quickly (Jing *et al.*, 1997) whereas structures containing T_4 loops fold much more slowly (Brown *et al.*, 2005). Combining the two different loop types may therefore give some insight into the mechanism of quadruplex folding. Table 4.1 shows the sequences of the oligonucleotides used in this study.

Name	Oligonucleotide Sequence						
		Loop 1		Loop 2		Loop 3	
G_3T	d-F-TGGG	T	GGG	T	GGG	T	GGGT-Q
G_3T-T_4-T	d-F-TGGG	T	GGG	TTTT	GGG	T	GGGT-Q
G_3T_4-T-T	d-F-TGGG	TTTT	GGG	T	GGG	T	GGGT-Q
G_3T_4-T-T_4	d-F-TGGG	TTTT	GGG	T	GGG	TTTT	GGGT-Q
G_3T_4-T_4-T	d-F-TGGG	TTTT	GGG	TTTT	GGG	T	GGGT-Q
G_3T_4	d-F-TGGG	TTTT	GGG	TTTT	GGG	TTTT	GGGT-Q

Table 4.1 Sequences of the oligonucleotides used in this work. F = FAM, Q = Dabcyl

Sequences in which the loops are all the same length e.g. d[T(GGGT)$_4$] are written in the shorthand nomenclature, G_3T. Sequences in which the loops are different are written G_3 followed by the sequence of loops. This nomenclature is standard throughout this thesis. All oligonucleotides were fluorescently labelled with a 5' FAM and a 3' dabcyl.

4.3 Results

A variety of biophysical techniques have been used to examine the folding, stability and kinetics of the quadruplexes formed by the sequences in Table 4.1. Circular dichroism, gel electrophoresis and 1D NMR were carried out to assess the structure and topology of the folded complexes. Fluorescence melting has been used to study the thermodynamic stability and kinetic properties of each of the quadruplex structures.

4.3.1 Circular Dichroism

Circular dichoism is frequently used to indicate the folding topology of G-quadruplexes (Hardin *et al.*, 1991; Balagurumoorthy *et al.*, 1992; Rujan *et al.*, 2005; Hazel *et al.*, 2005). The strength and position of the spectral peaks can give an indication of the strand alignment of the structure; either parallel, antiparallel or a combination of the two. CD spectra for these oligonucleotides in the presence of sodium or potassium are shown in Figures 4.1 and 4.2. Samples were heated to 95 °C before being slowly cooled to room temperature over a period of 10 hours in a thermal controller. In the presence of potassium, all the sequences except G_3T_4 display a major positive peak at around 265 nm with a minimum around 240 nm, suggesting they adopt a topology in which the strands are in a parallel-stranded arrangement. In contrast G_3T_4 exhibits a positive peak at 295 nm, indicative of an antiparallel arrangement. Quadruplex structures with single nucleotide loops are thought only to form 'propeller-type' fold back loops, generating parallel stranded complexes (Hazel *et al.*, 2004; Rachwal *et al.*, 2007). The longer T_4 loops allow more flexibility and are therefore able to form edgewise, diagonal or double-chain reversal loops. These CD spectra suggest that the presence of only one single T loop is sufficient to induce an all parallel-stranded structure. Only once all three of the single base loops have been replaced with longer loops does the quadruplex adopt an antiparallel topology. In general, these CD spectra were independent of the potassium concentration between 20 and 200 mM, though pronounced changes were observed for G_3T_4 (Figure 4.1b). For this sequence a secondary peak is visible around

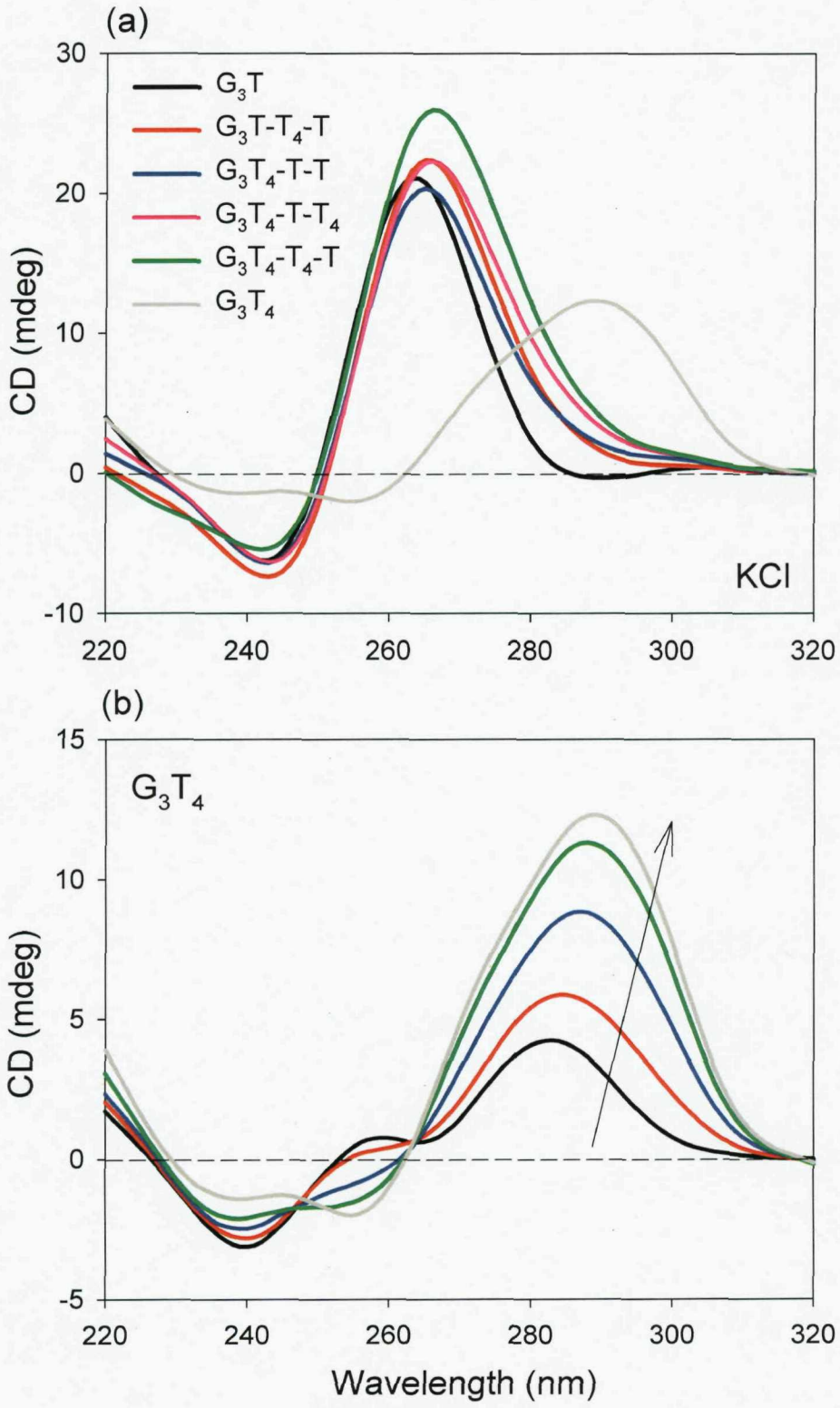


Figure 4.1 (a) CD spectra of the fluorescently-labelled quadruplex-forming oligonucleotides (5μM) in the presence of 10 mM lithium phosphate pH 7.4 containing 200 mM KCl. Spectra are buffer normalised and a zero correction applied at 320 nm. G_3T (black); G_3T-T_4-T (red); G_3T_4-T-T (blue); G_3T_4-T-T_4 (pink); G_3T_4-T_4-T (green); G_3T_4 (grey). **(b)** CD spectrum of G_3T_4 in the presence of increasing concentrations of KCl; black, 1 mM; red, 5 mM; green, 20 mM; blue, 50 mM; pink, 200 mM.

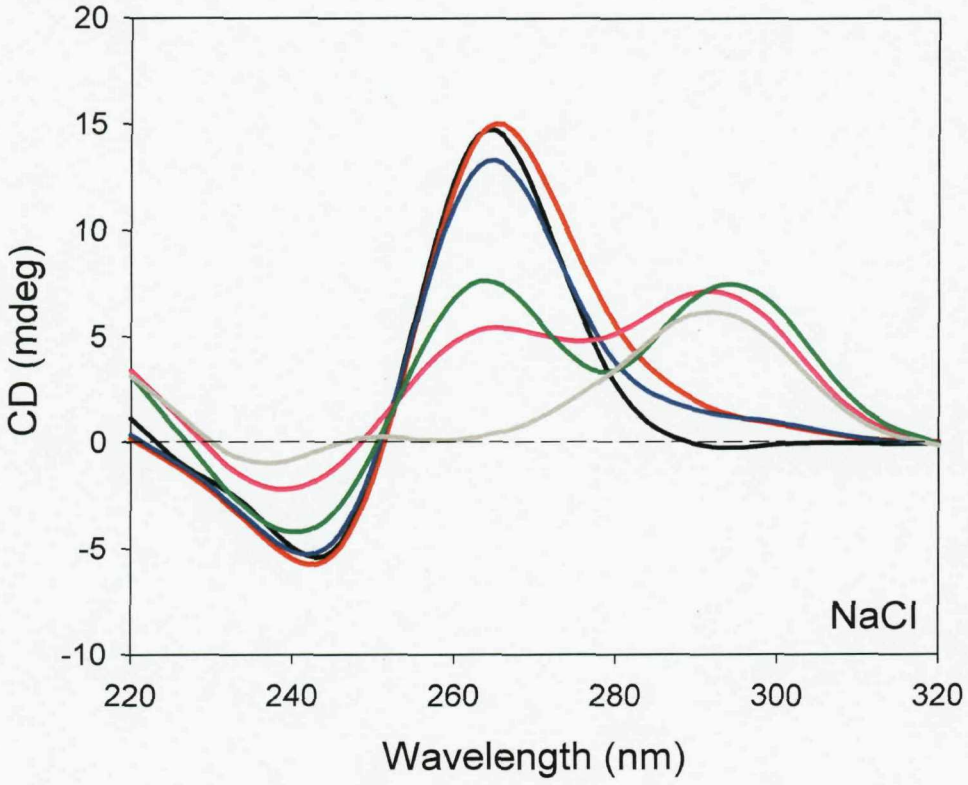


Figure 4.2 CD spectra of the fluorescently-labelled quadruplex-forming oligonucleotides (5μM) in the presence of 10 mM lithium phosphate pH 7.4 containing 200 mM NaCl. Spectra are buffer normalised and a zero correction applied at 320 nm. G_3T (black); G_3T-T_4-T (red); G_3T_4-T-T (blue); G_3T_4-T-T_4 (pink); G_3T_4-T_4-T (green); G_3T_4 (grey).

260 nm at low potassium concentrations, which decreases as the ionic strength is increased; this is accompanied by an increase in the peak intensity at 295 nm. The presence of isoelliptic points in these spectra suggest that this sequence may adopt two distinct structural forms in the presence of low or high potassium ion concentrations.

In the presence of sodium ions, the trends observed are slightly different to those seen in potassium (Figure 4.2). G_3T, G_3T-T_4-T and G_3T_4-T-T are typical of parallel-stranded structures, with a large positive peak at 265 nm. The substitution of a second T_4 loop results in a spectrum with equal sized peaks at both 265 nm and 295 nm. Sodium ions are thought to favour antiparallel quadruplex folds, as has been observed for both the human (Wang and Patel, 1993) and *Oxytricha* (Smith and Feigon, 1992) telomeric repeat sequences. It is therefore possible that the two longer loops are laterally arranged, with the single T loop forming a double-chain reversal. This loop arrangement would result in a '3 + 1' core scaffold, a topology recently observed for the human telomeric repeat (Luu *et al.*, 2006; Ambrus *et al.*, 2006). This conformation would account for the presence of two distinct peaks in the CD spectrum, although equally the co-existence of parallel and antiparallel topologies may also result in such a spectrum. A hybrid 3 + 1 structure containing a combination of loop types would seem more probable, as a quadruplex containing any single nucleotide loops would be unlikely to form a fully antiparallel topology. The spectrum of G_3T_4 is similar in the presence of sodium and potassium ions with a major peak at 295 nm, suggesting an antiparallel topology.

4.3.2 Gel Mobility

To further compare the global structures formed by these sequences, native polyacrylamide gel electrophoresis was carried out (Figure 4.3). To ensure the quadruplexes were in the folded state, the gel was supplemented with 20 mM potassium phosphate. Each of the structures ran as a single band with the exception of G_3T_4 which shows a slight smearing. This may be due the dissociation of the complex within the gel, as G_3T_4 is relatively unstable in 20 mM potassium. G_3T shows the fastest mobility, as would be expected as it has the lowest molecular weight. Surprisingly, G_3T-T_4-T and G_3T_4-T-T have different mobilities despite their apparent similar topology. The same trend is observed for G_3T_4-T-T_4 and G_3T_4-T_4-T, where the former appears to have a slower mobility than the latter. The position of the short and long loops appears to affect

the gel mobility of the structures. When the central loop contains T_4, the quadruplex appears to have a more compact structure than if it were a single nucleotide loop.

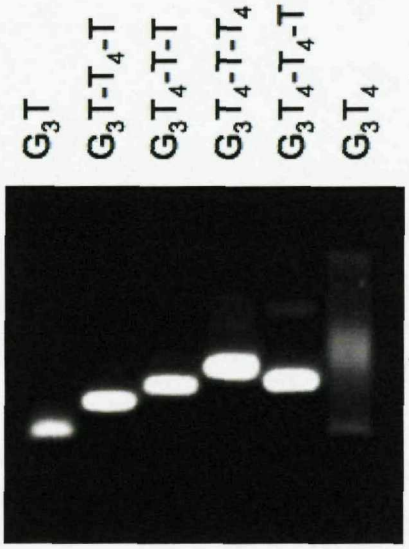


Figure 4.3 Gel mobility of the series of oligonucleotides (20 μM) in a 14 % native polyacrylamide gel supplemented with 20 mM potassium phosphate. Samples were heated to 95 °C and allowed to cool slowly prior to loading.

4.3.3 Imino Proton NMR Spectra

One of the defining features of structures composed of guanine quartets is the appearance of imino proton resonances between 10.5 and 12 ppm (Feigon *et al.*, 1995). This technique is firstly able to confirm the presence of a folded quadruplex conformation, and also can be used to assess whether the sequence adopts a unique structure (Ambrus *et al.*, 2006; Phan *et al.*, 2006; Dai *et al.*, 2007). The presence of multiple or ill-defined peaks is evidence for the existence of multiple folded structures. The imino proton spectra for each of the sequences are shown in Figure 4.4. It can be seen that the spectra of G_3T, G_3T-T_4-T, G_3T_4-T-T, and G_3T_4-T-T_4 display between 10 and 12 well resolved peaks, indicative of well-defined structures. In the cases where 10 or 11 peaks are observed, the intensities indicate that one or two imino protons have degenerate chemical shifts. Hence the number of hydrogen-bonded imino proton resonances is 12, as would be expected for a three G-quartet structure. G_3T_4-T_4-T also shows 12 major peaks, although the spectrum also contains some minor peaks, which may suggest the presence of a small amount of a second structure. In contrast, the imino proton spectrum of G_3T_4 shows multiple peaks confirming that it adopts more than one stable conformation.

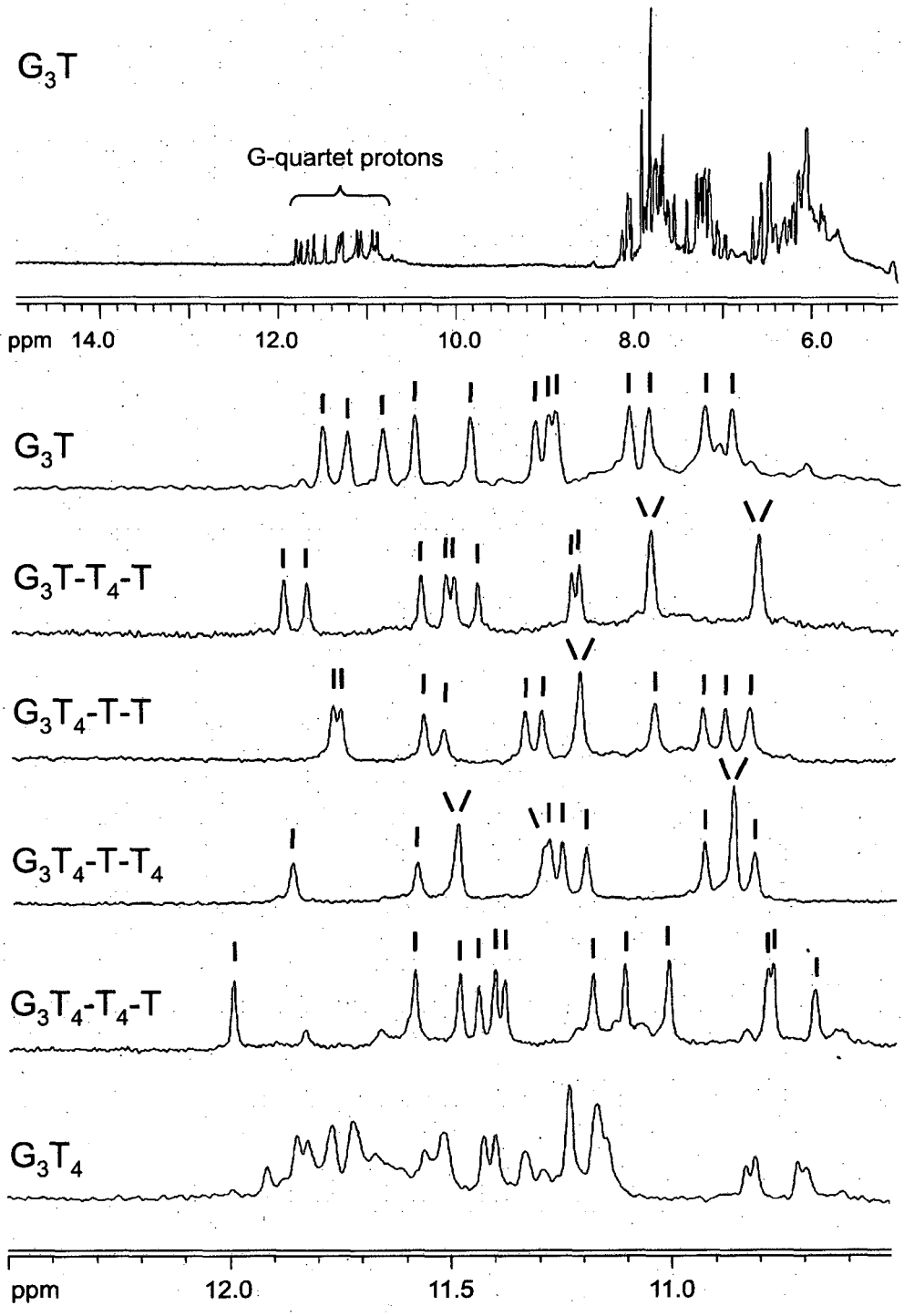


Figure 4.4. 1D imino proton NMR spectra of the quadruplex-forming oligonucleotides. The samples (100 μM) were prepared in 200 mM potassium phosphate pH 7.4. The top panel shows the 1D-NMR spectrum for G_3T between 5 and 15 p.p.m., while the other panels show the imino proton region for each oligonucleotide. The individual peaks are indicated.

4.3.4 Fluorescence Melting Curves

Representative fluorescence melting curves for these sequences are shown in Figure 4.5a. The T_m values at different potassium concentrations together with the calculated values for ΔH are shown in Table 2, and the T_m values in sodium are shown in Table 4.3. All samples were annealed and then melted at a rate of 0.2 °C.min^{-1}; at this rate of temperature change no hysteresis was observed, except for G_3T_4 at low ionic strengths. As expected, all the complexes are more stable in potassium than sodium ions. G_3T is the most stable and the only complex to form in the presence of lithium alone. Substituting one single T loop for a T_4 loop results in a 20 °C decrease in stability, irrespective of whether the replacement is in a central (G_3T-T_4-T) or peripheral (G_3T_4-T-T) loop. G_3T-T_4-T is however more stable than G_3T_4-T-T by around 2-3 °C. Replacing a second T for a T_4 loop causes a further reduction in stability by around 20 °C, with G_3T_4-T_4-T 2-3 °C more stable than G_3T_4-T-T_4. Replacing all three single T loops with T_4 results in a further 10 °C reduction in T_m, although the melting and annealing curves for this sequence show hysteresis at low potassium concentrations and the melting (but not the annealing) profiles are biphasic. In the presence of sodium ions, the relative order of stability is the same, but the effects of replacing the loops are less dramatic (Table 4.3). Substituting a single T with T_4 results in a 15-20 °C decrease in T_m, and a second substitution causes a further 10 °C decrease. As seen with potassium, sequence isomers with a central T_4 loop are slightly more stable than those with a central T loop. G_3T_4 has a similar T_m to the sequences with two T_4 loops, but there is no evidence of hysteresis as is observed in potassium. To ensure that all transitions were due to intramolecular and not intermolecular complex formation, the concentration dependence of the melting profiles was examined (Figure 4.5b). Oligonucleotide concentrations were tested between the range of 0.1 and 10 μM. The results show that the T_m values are independent of oligonucleotide concentration within this range and therefore must form intramolecular complexes.

ΔH values for the quadruplex-single-strand transition in the presence of potassium were derived from these melting profiles by van't Hoff analysis, assuming that the reaction is a two-state equilibrium, and the values are shown in Table 4.2. As previously reported for other quadruplexes, ΔH increases with ionic strength, consistent with the presence of specific cation binding sites within the quadruplex (Jing *et al.*, 1997). The slopes of plots of ΔG against log[M^+] (Figure 4.6) can be used to determine the stoichiometry of

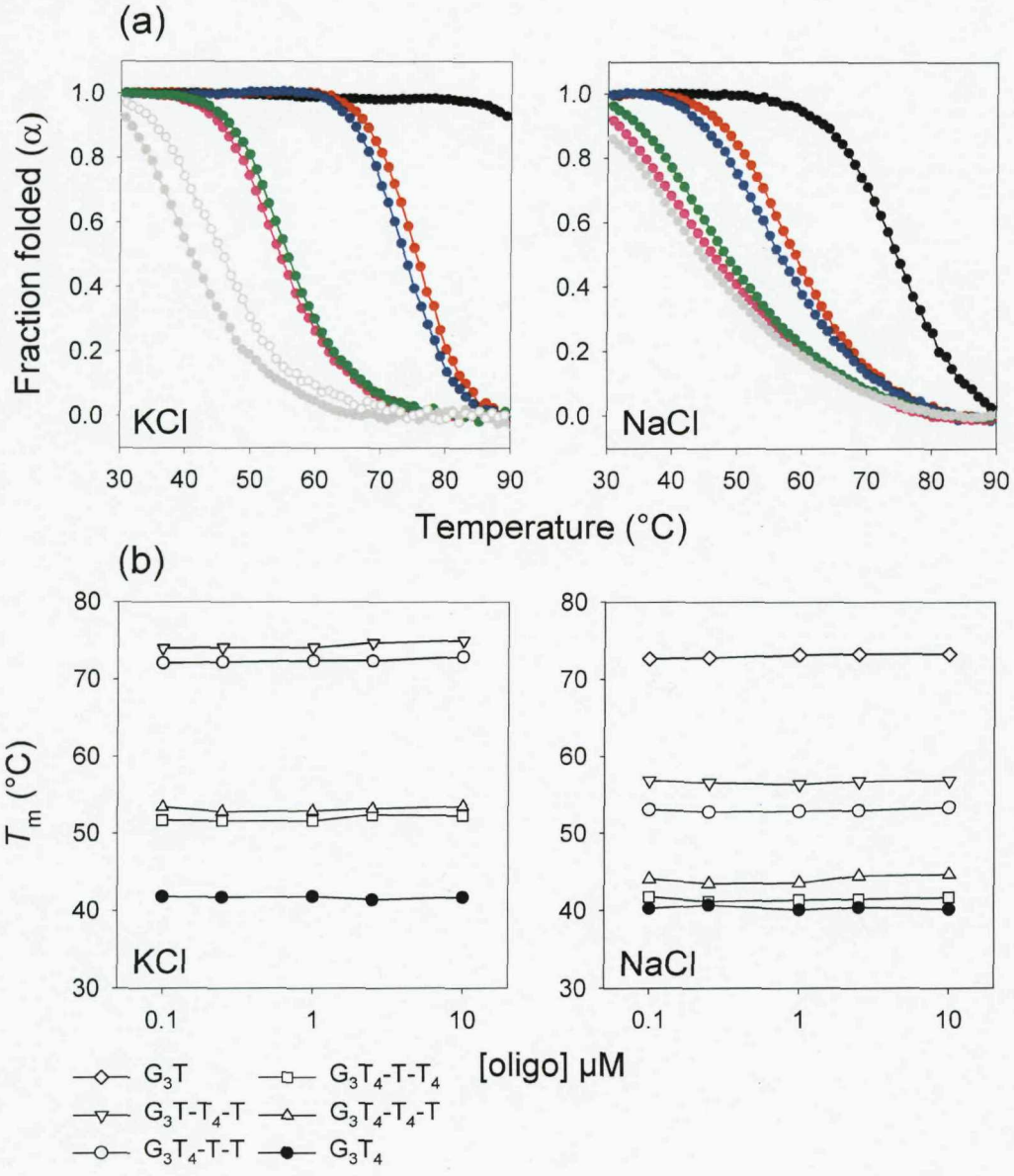


Figure 4.5 **(a)** Fluorescence melting/annealing profiles for each of the quadruplex-forming sequences. Oligonucleotides (0.25 μM) were dissolved in 10 mM lithium phosphate pH 7.4, containing either 20 mM KCl (left-hand panel) or 200 mM NaCl (right-hand panel). Each of the melting and annealing profiles were superimposable at a rate of temperature change of 0.2 °C.min^{-1} (except G_3T_4) therefore only the annealing profiles are presented. Hysteresis is only evident for G_3T_4 in potassium, which also displays a biphasic melting profile (open circles). The curves show the fraction folded (α) as a function of temperature, calculated as described in the Materials and Methods. G_3T (black); G_3T-T_4-T (red); G_3T_4-T-T (blue); G_3T_4-T-T_4 (pink); G_3T_4-T_4-T (green); G_3T_4 (grey). **(b)** Concentration dependence of the T_m values between 0.1 and 10 μM oligonucleotide in 20 mM KCl or 200 mM NaCl.

[KCl] mM	G_3T		G_3T-T_4-T		G_3T_4-T-T		G_3T_4-T-T_4		G_3T_4-T_4-T		G_3T_4	
	T_m	ΔH	T_m	ΔH	T_m	ΔH	T_m	ΔH	T_m	ΔH	T_m	ΔH
0	46.6											
0.1	57.2	-242 ± 4	37.7									
1	73.5	-271 ± 5	53.8	-227 ± 7	51.3	-208 ± 7						
5	85.7	-275 ± 5	65.3	-259 ± 11	63.4	-250 ± 15						
10			70.7	-266 ± 9	67.3	-258 ± 9	46.3	-184 ± 13	48.1	-205 ± 5		
20			75.1	-267 ± 7	73.1	-262 ± 10	52.5	-207 ± 3	54.1	-227 ± 11	44.5 (37.6)*	
50			81.4	-284 ± 7	79.8	-275 ± 8	59.5	-225 ± 13	60.5	-246 ± 8	47.0 (51.0)*	-234 ± 11
100			87.5		84.6		65.3	-247 ± 8	67.1	-269 ± 13	56.3	-266 ± 10
200							72.8	-256 ± 9	74.2	-276 ± 10	63.3	-293 ± 9
Δn	2.13 ± 0.10		2.29 ± 0.11		2.75 ± 0.16		2.75 ± 0.11		2.89 ± 0.12		4.01 ± 0.16	

Table 4.2. T_m (°C) and ΔH (kJ.mol^{-1}) values for the fluorescently labelled quadruplex-forming oligonucleotides, determined in the presence of 10 mM lithium phosphate pH 7.4, containing different concentrations of KCl. The samples were heated and cooled at a rate of 0.2 °C.min^{-1}. * Indicates a biphasic melting profile. All reactions were performed at least twice and the calculated T_m values usually differed by less < 0.5 °C. ΔH values were typically calculated for melting profiles for which the T_m was between 40 °C and 80 °C. Missing values at low concentrations of KCl correspond to complexes for which the T_ms were too low to measure (< 30 °C), while those at high ionic strengths (especially G_3T) were too stable (T_m > 85 °C).

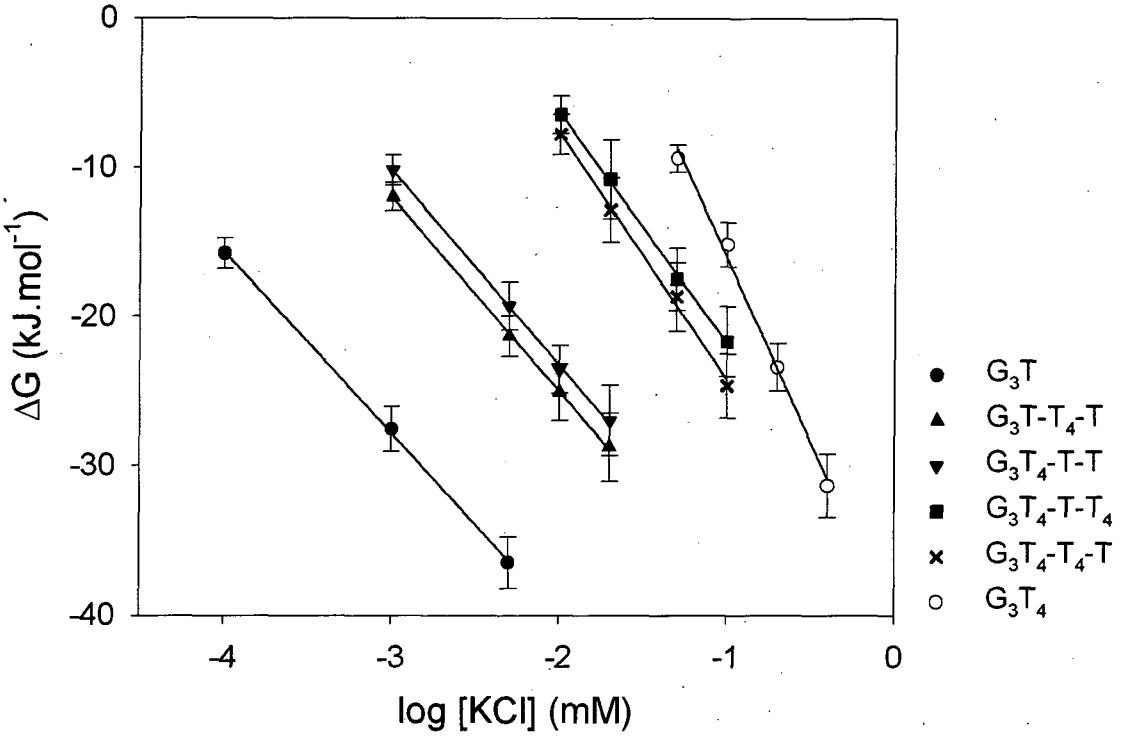


Figure 4.6 Calculated ΔG values plotted against log[KCl]. These data were fitted to a straight line, yielding a slope of ΔΔG / Δlog[KCl], or Δn. ΔG values were calculated as $\Delta H^*(1\text{-}310/T_m)$, with T_m and ΔH values derived from van't Hoff analysis of the fluorescence melting profiles.

	T_m (°C)					
	G_3T	G_3T-T_4-T	G_3T_4-T-T	G_3T_4-T-T_4	G_3T_4-T_4-T	G_3T_4
10 mM Li^+	46.9	-	-	-	-	-
[NaCl] mM						
10	52.6					
50	60.4	43.4	39.4			
100	67.0	49.5	46.3			
200	74.3	58.6	54.9	43.4	46.7	41.5
500	78.4	63.7	61.1	50.6	54.1	51.9
1M				58.8	62.2	61.6

Table 4.3. T_m (°C) values for the fluorescently-labelled quadruplex-forming oligonucleotides, determined in the presence of 10 mM lithium phosphate pH 7.4, containing different concentrations of NaCl. The samples were heated and cooled at a rate of 0.2 °C.min^{-1}. Missing values at low concentrations of NaCl correspond to complexes for which the T_ms were too low to measure.

cation binding (Jing *et al.*, 1997) yielding values of Δn (the difference between the number of ions bound in the folded and unfolded states) and the values of Δn are listed in Table 4.2. For a quadruplex with three G-quartets and antiparallel strands, Δn would be expected to be either two (the number of potassium ions located between the stacked quartets) or four (including two additional ions that may be co-ordinated between the loops and the terminal quartets). A value of two seems more likely for a structure containing single nucleotide, double-chain reversal loops, in which the loops do not interact with the terminal quartets. For all the complexes containing at least one single T loop, the values of Δn are between two and three (Table 4.2), although there is a steady increase in this value with the number of longer T_4 loops. This will be considered further in the Discussion. Δn is larger for G_3T_4, consistent with the suggestion that it adopts an alternative topology, more likely containing lateral or diagonal loops. This value must however be interpreted with caution, as there is some hysteresis in its melting profile at low potassium concentrations and the NMR spectrum suggests that it adopts more than one folded conformation.

4.3.5 Kinetics of Quadruplex Formation

4.3.5.1 Hysteresis.

The fluorescence melting experiments, as illustrated in Figure 4.5 were performed at a rate of temperature change of 0.2 °C.min^{-1} and only G_3T_4 showed hysteresis between the melting and annealing profiles. The hysteresis between these curves arises because the complexes are heated or cooled at a rate that is too fast to allow thermodynamic equilibrium, and indicates the presence of slow steps in the association and/or dissociation reactions. The effect of increasing the rate of heating and cooling on the series of oligonucleotides was then assessed. Representative melting and annealing profiles illustrating the effects of increasing the rate of heating to 2 °C.min^{-1} and 12 °C.min^{-1} on this series of oligonucleotides are shown in Figure 4.7 and the T_m values are summarised in Table 4.4. In order to make a direct comparison between sequences, the salt concentration was maintained at either 20 mM KCl, or 200 mM NaCl. In the presence of 20 mM potassium, G_3T is too stable to compare directly with the other sequences. However at lower ionic strengths (1 mM) G_3T exhibits no hysteresis, even at the fastest rate of heating. At a heating rate of 0.2 °C.min^{-1} (Figure 4.7 top panel), only G_3T_4 is not at thermodynamic equilibrium, with a ΔT_m value (the difference between the

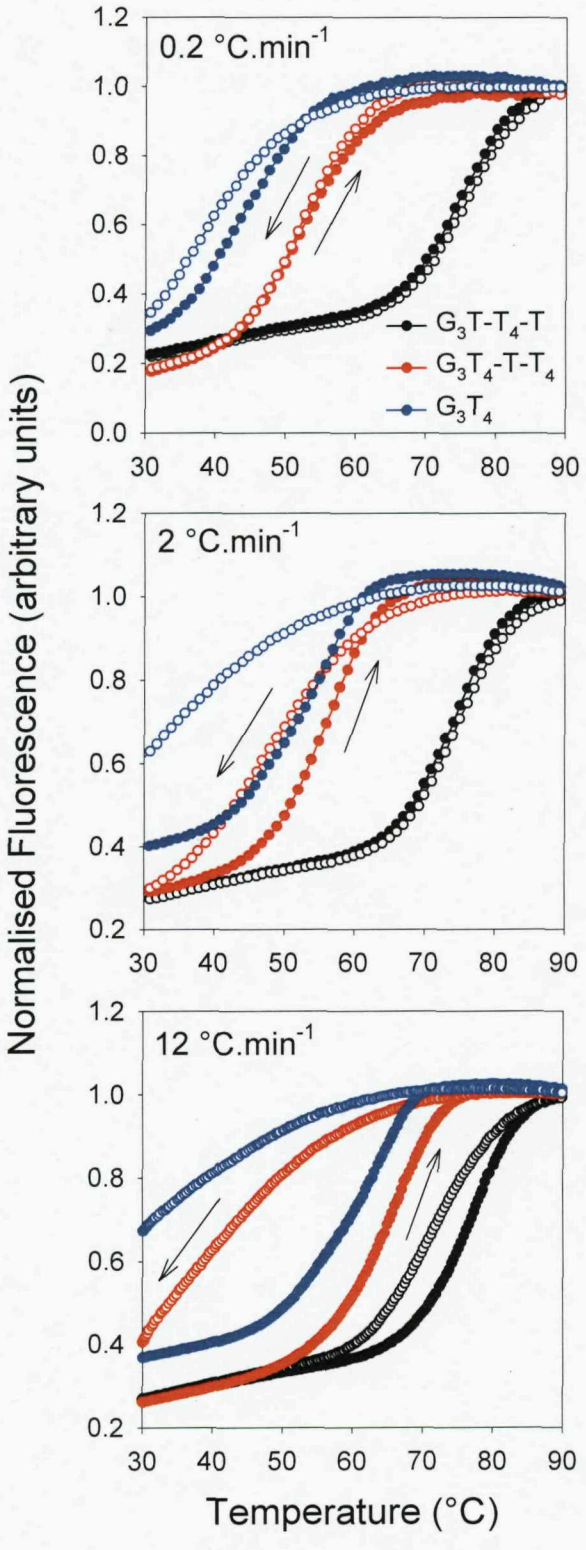


Figure 4.7 Fluorescence melting (closed circles) and annealing (open circles) profiles of the complexes formed by the oligonucleotide sequences G_3T-T_4-T, G_3T_4-T-T_4 and G_3T_4 in the presence of 10 mM lithium phosphate pH 7.4, supplemented with 20 mM KCl at a range of heating and cooling rates. For clarity of presentation, G_3T_4-T-T and G_3T_4-T_4-T are omitted. (G_3T does not display a melting transition under these conditions). The curves have been normalised to the same final fluorescence value.

	T_m values at 20 mM KCl (°C)									200 mM NaCl		
	0.2 °C.min^{-1}			2 °C.min^{-1}			12 °C.min^{-1}			12 °C.min^{-1}		
Sequence	melt	anneal	ΔT_m	melt	anneal	ΔT_m	melt	anneal	ΔT_m	melt	anneal	ΔT_m
G_3T	> 95	> 95	-	-	-	-	-	-	-	75.4	75.7	-0.3
G_3T-T_4-T	75.2	75.0	0.2	74.1	73.8	0.3	**78.3**	**70.9**	**7.4**	58.6	58.5	0.1
G_3T_4-T-T	72.9	73.2	-0.3	72.1	71.7	0.4	**76.0**	**67.3**	**8.7**	55.0	54.8	0.2
G_3T_4-T-T_4	52.7	52.3	0.4	**56.2**	**46.6**	**9.6**	66.6	< 30	> 36	43.7	43.1	0.6
G_3T_4-T_4-T	53.9	54.2	-0.3	**56.8**	**49.2**	**7.6**	66.1	< 30	> 36	46.5	46.9	-0.4
G_3T_4	**44.5***	**37.6**	**6.9**	46.7	< 30	> 16.7		No data		41.8	41.3	0.5

Table 4.4. T_m and ΔT_m values derived from the hysteresis between the fluorescence melting and annealing profiles of the series of oligonucleotides at different rates of heating. Samples were prepared in 10 mM lithium phosphate pH 7.4, containing either 20 mM potassium chloride or 200 mM sodium chloride. * Indicates a biphasic melting profile. Values in bold indicate the rates chosen for the thermodynamic analysis of folding for each sequence.

melting and annealing T_m) of around 7 °C. As the heating rate is increased to 2 °C.min^{-1} (Figure 4.7 middle panel), this value rises to in excess of 16 °C, and hysteresis is also observed in the melting and annealing profiles of the complexes containing two T_4 loops (G_3T_4-T-T_4 and G_3T_4-T_4-T; ΔT_m 7–10 °C). The sequences containing a single T_4 loop still display equilibrium melting curves at this rate. Only once the rate of heating has been increased further to 12 °C.min^{-1} do the sequences containing a single T_4 loop (G_3T-T_4-T and G_3T_4-T-T) show hysteresis (Figure 4.7 bottom panel) with a ΔT_m value of around 8 °C. At this rate G_3T_4-T-T_4 and G_3T_4-T_4-T exhibit a ΔT_m in excess of 36 °C, emphasising the slower kinetics of the complexes containing two T_4 loops as compared to one. In the presence of sodium no hysteresis was observed for any of the sequences, even at the fastest rate of heating and cooling.

Thermal melting profiles in which the heating and cooling curves are non-superimposable can allow the determination of the folding (k_1) and unfolding (k_{-1}) rate constants as a function of temperature. This analysis has previously been described for DNA triplexes (Rougée *et al.*, 1992), i-motif (Mergny and Lacroix, 1998) as well as G-quadruplexes (Brown *et al.*, 2005). Kinetic analysis of intramolecular quadruplexes by this method is however often difficult, since they fold much more rapidly than their intermolecular counterparts, which become quasi-irreversibly dissociated once melted (Wyatt *et al.*, 1996; Mergny *et al.*, 2005; Merkina and Fox, 2005). This problem has been overcome by changing the rate at which the sequences are melted and annealed (either 0.2 °C.min^{-1}, 2 °C.min^{-1} or 12 °C.min^{-1}), and then performing the kinetic analysis at the rate at which hysteresis is observed (Table 4.4). In this way, a full analysis can be performed for each sequence at the same ionic strength. At each rate, successive heating and cooling cycles were carried out to ensure the individual curves were superimposable (i.e. the amount of hysteresis was reproducible) and also oligonucleotide concentration independent. Since none of the sequences in the presence of sodium displayed any hysteresis, this analysis was limited only to the potassium-containing samples.

Folding and unfolding rate constants were calculated between a temperature range ± 10 °C of the T_m of the quadruplex for each of the sequences in 20 mM KCl (described in detail in the Materials and Methods section). This data range was chosen as it is the largest window in which the values could be calculated with confidence; the data range is limited by the size of the hysteresis, which in turn is dependent on the rate of heating. Figure 4.8 (left-hand panels) shows the melting and annealing profiles for the single T_4

substituted sequences (G_3T-T_4-T and G_3T_4-T-T) at a heating and cooling rate of 12 °C.min^{-1}. Figure 4.9 (left-hand panels) shows the equivalent profiles for the sequences containing two T_4 loops (G_3T_4-T-T_4 and G_3T_4-T_4-T) at 2 °C.min^{-1}. Figure 4.9 shows the Arrhenius plots for the folding and unfolding rates constructed from these data, with the kinetic parameters derived from these Arrhenius plots presented in Table 4.5.

Several factors are apparent from the kinetic data. Firstly, the association reactions show an unusual temperature dependence, with an apparent negative activation energy, i.e. the reaction is faster at lower temperatures. This phenomenon has been reported previously for a number of DNA structures (Mergny and Lacroix, 1998; Wallace *et al.*, 2001; Alberti *et al.*, 2002), and is explained by the suggestion that these complexes associate via a nucleation-zipper mechanism, whereby a folding intermediate is stabilised at lower temperatures. This feature results in a characteristic V or X shaped Arrhenius plot where the intersection of the straight lines ($k_1 = k_{-1}$) gives the T_m value. Secondly, the data for G_3T_4-T-T_4 are very similar to G_3T_4-T_4-T and G_3T-T_4-T is similar to G_3T_4-T-T, suggesting that the distribution of the different loops is less important than their length. The complexes containing a single T_4 loop, in either the central or peripheral position, have similar dissociation half-lives of around 10 hours at 37 °C. For the sequences with two T_4 loops the half-lives are even more closely matched, both with values around 1.5 hours. Thirdly, the unfolding parameters are very similar for all four oligonucleotides, while the folding parameters vary according to the loop lengths. For the association reaction both ln(A) and E_a are less negative for the complexes with longer loops. Subtracting the value for E_a (k_1) from E_a (k_{-1}) for each of the sequences gives ΔH values that are in close agreement with those calculated from van't Hoff analysis (Table 4.5). This kinetic analysis was not performed for G_3T as the melting and annealing curves were superimposable, even at the fastest rate of temperature change. This analysis was also not performed for G_3T_4, due to the biphasic nature of the melting curve.

4.3.5.2 Temperature-jump kinetics

In order to confirm the kinetic data obtained by the hysteresis method of analysis, a second method of determining kinetic parameters was employed. Temperature-jump relaxation involves the rapid heating of a sample, followed by the monitoring of the time-dependent fluorescence change as the reaction relaxes to a new equilibrium. This

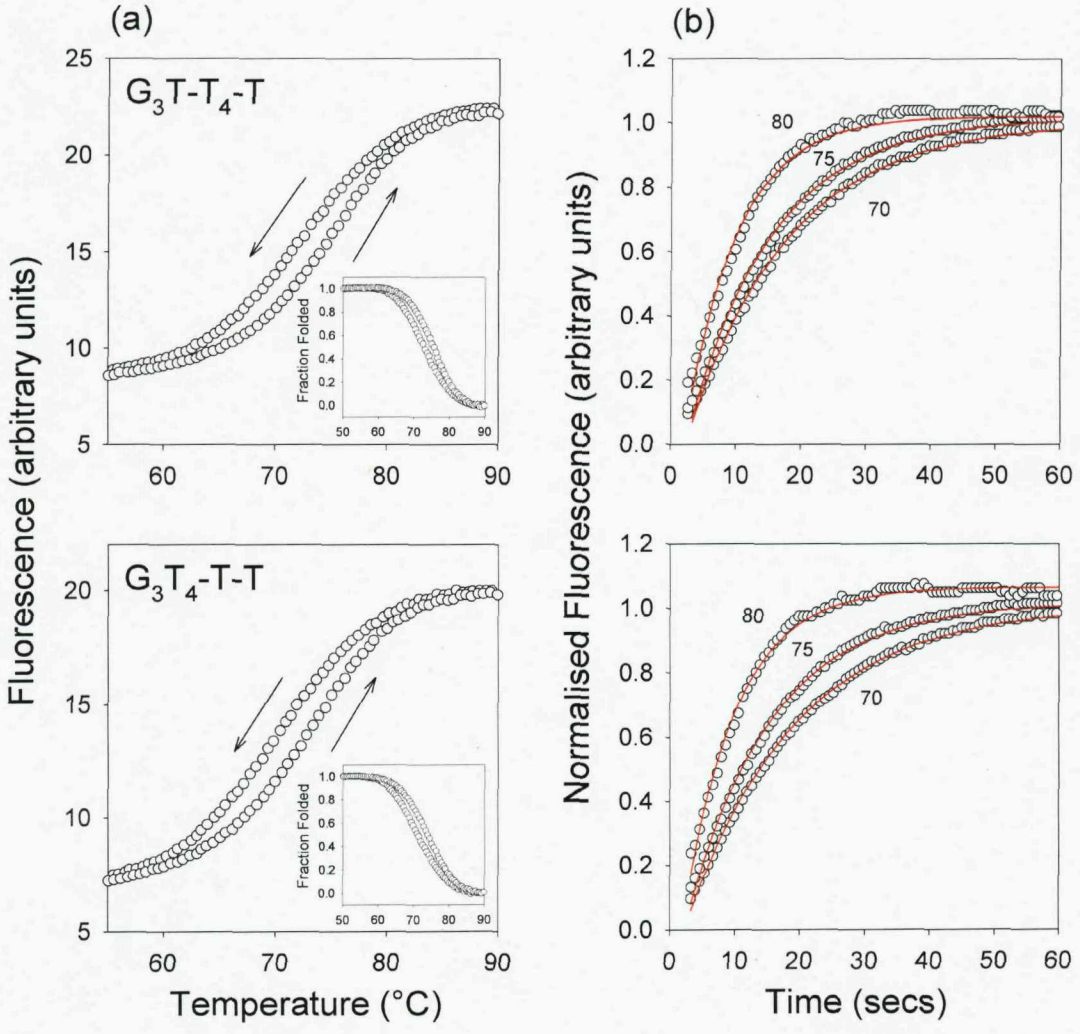


Figure 4.8 (a) Hysteresis between the fluorescence melting and annealing profiles for the sequences G_3T-T_4-T and G_3T_4-T-T, at a rate of temperature change of 12 °C.min^{-1}, in the presence of 10 mM lithium phosphate pH 7.4 containing 20 mM KCl. These curves have then been converted to fraction folded plots (inset) **(b)** Temperature-jump relaxation profiles for the sequences G_3T-T_4-T (top) and G_3T_4-T-T (bottom). Traces show the rate of approach to a new equilibrium following a rapid 5 °C increase in temperature to the value shown. The profiles have been normalised to show the fractional change in fluorescence with time and fitted with a single exponential curve.

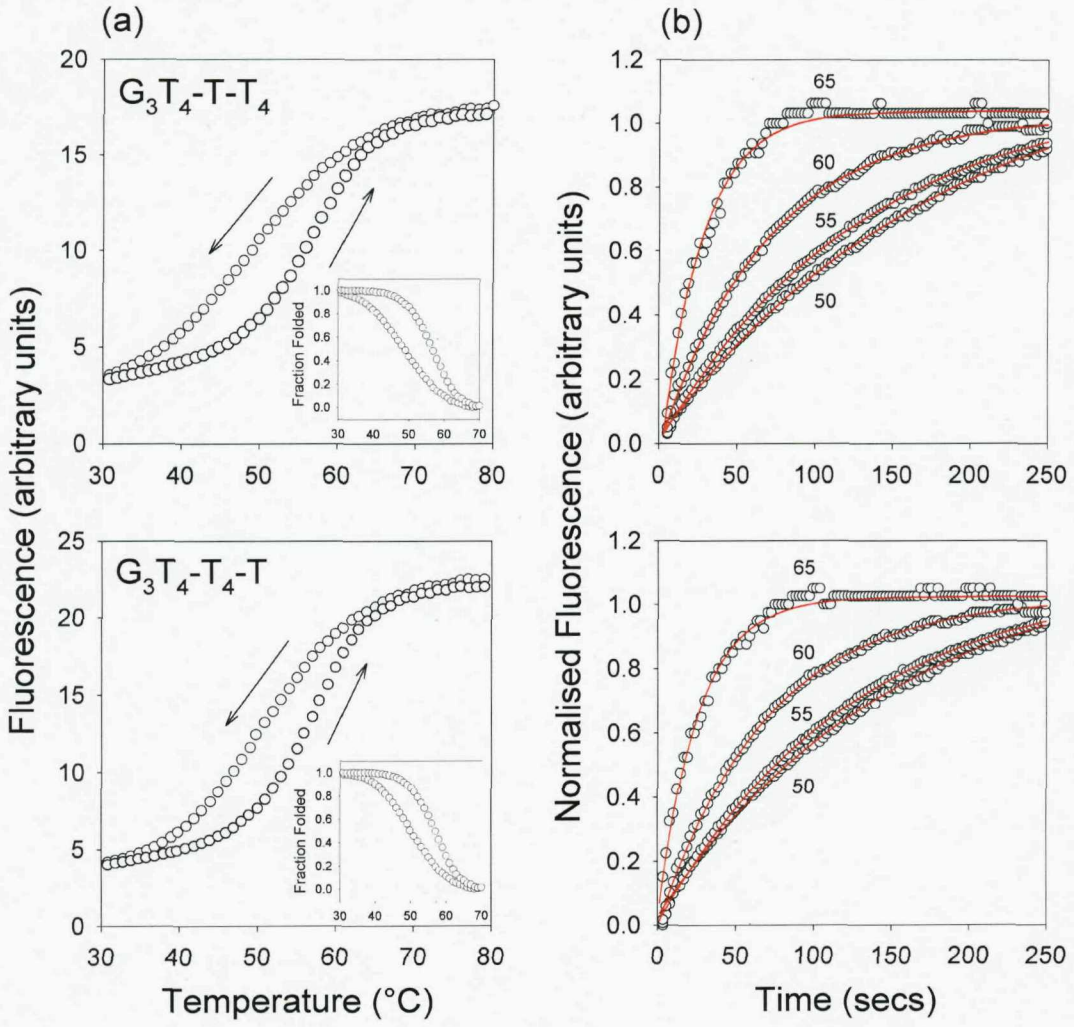


Figure 4.9 (a) Hysteresis between the fluorescence melting and annealing profiles for the sequences G_3T_4-T-T_4 and G_3T_4-T_4-T, at a rate of temperature change of 2 °C.min^{-1}, in the presence of 10 mM lithium phosphate pH 7.4 containing 20 mM KCl. These curves have then been converted to fraction folded plots (inset) **(b)** Temperature-jump relaxation profiles for the sequences G_3T_4-T-T_4 (top) and G_3T_4-T_4-T (bottom). Traces show the rate of approach to a new equilibrium following a rapid 5 °C increase in temperature to the value shown. The profiles have been normalised to show the fractional change in fluorescence with time and fitted with a single exponential curve.

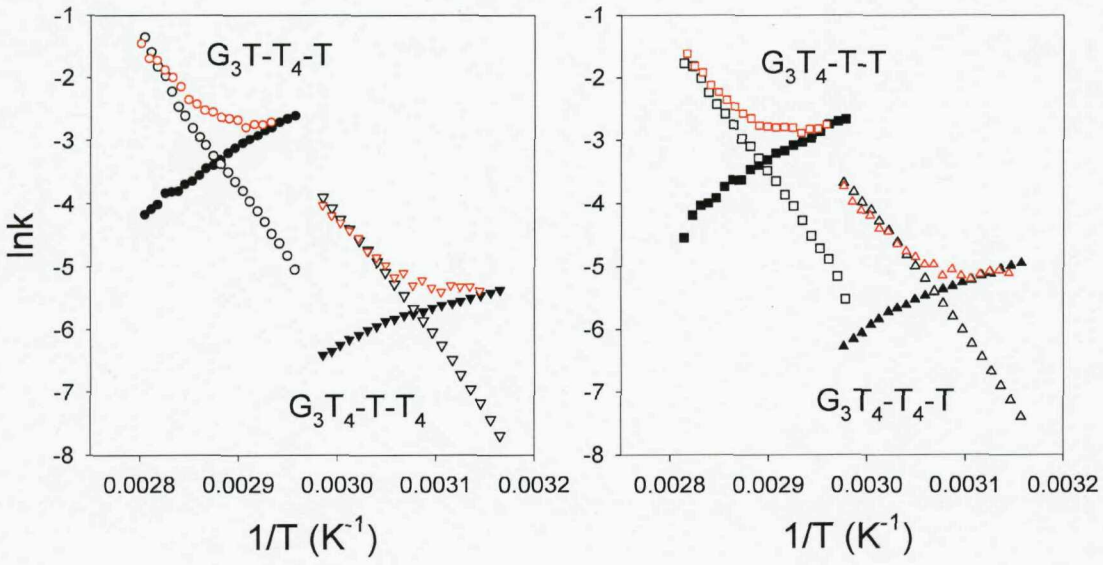


Figure 4.10 (a) Arrhenius plots derived from the hysteresis and temperature-jump data, showing the temperature dependence of the kinetic parameters for folding and unfolding of the sequences G_3T-T_4-T (circles) and G_3T_4-T-T_4 (inverted triangles; left hand panel) and G_3T_4-T-T (squares) and G_3T_4-T_4-T (triangles; right hand panel). Black symbols were derived from the hysteresis between the melting and annealing profiles; k_{-1}, open circles; k_1, closed circles. Open red circles represent the apparent rate constants ($k_1 + k_{-1}$) derived from the temperature-jump experiments.

Sequence	k_1		k_{-1}				ΔH (kJ.mol^{-1})
	E_a (kJ.mol^{-1})	ln (A) s^{-1}	E_a (kJ.mol^{-1})	ln (A) s^{-1}	k^{37} (s^{-1})	$t_{1/2}{}^{37}$ / s	
G_3T	-	-	-	-	-	-	
G_3T-T_4-T	-93 ± 2	-36 ± 1	182 ± 3	60 ± 1	1.80 x 10^{-5}	38485	-275 (-267)
G_3T_4-T-T	-88 ± 3	-34 ± 1	178 ± 2	58 ± 1	2.36 x 10^{-5}	30885	-266 (-262)
G_3T_4-T-T_4	-48 ± 4	-24 ± 1	173 ± 2	58 ± 1	1.45 x 10^{-4}	4787	-221 (-207)
G_3T_4-T_4-T	-59 ± 2	-27 ± 1	174 ± 2	59 ± 1	1.47 x 10^{-4}	4712	-233 (-227)
G_3T_4	-	-	-	-	-	-	

Table 4.5 Thermodynamic and kinetic parameters for association (k_1) and dissociation (k_{-1}) of the series of loop substituted quadruplex forming sequences in 10 mM lithium phosphate pH 7.4, 20 mM potassium chloride. E_a is the activation energy and A is the pre-exponential factor in the equation $k = Ae^{(-Ea/RT)}$. Dissociation half lives are calculated for each structure at 37 °C, with ΔH values calculated from E_a (k_1) – E_a (k_{-1}). Values in parentheses indicate the corresponding ΔH values derived from van't Hoff analysis of the melting profiles. No values are presented for G_3T as it is too stable in 20 mM potassium, or G_3T_4 as more than one folded conformation is thought to exist.

temperature increase promotes quadruplex dissociation, which can then be monitored over successive jumps along the melting curve (James *et al.*, 2003; Merkina and Fox, 2005; Bourdoncle *et al.*, 2006). Successive temperature jumps were carried out on the same sample in 5 °C increments (30-35 °C, 35-40 °C etc.) until the complex had fully dissociated. This process was repeated with a second sample, with temperature jumps between 31-36 °C, 36-41 °C and so on until relaxation curves at all temperatures had been collected.

Representative temperature-jump relaxation profiles for these complexes are shown in Figures 4.8 and 4.9 (right hand panels), and reveal a slow, time-dependent relaxation to the new equilibrium, which is clearly faster for the complexes with two short loops than the sequences with two long loops, irrespective of the loop positions within the quadruplex. The kinetic curves at different temperatures were fitted with single exponential functions and the rate constants obtained are presented as Arrhenius plots in Figure 4.10. For this unimolecular reaction the apparent rate constant is equal to the sum of the folding and unfolding rate constants ($k_1 + k_{-1}$). Although it is not possible to resolve these individual components, at low temperatures the sum is dominated by k_1, while the sum approximates to k_{-1} at high temperatures. Comparing these data to the individual rate constants derived from analysis of the non-equilibrium melting curves (Figure 4.10) reveals there is excellent agreement between the two datasets. The temperature-jump data therefore confirm that sequences with longer loops display slower rates of folding, with little effect on the rate of unfolding.

4.3.6 Effect of G_4 tracts

The sequences used in this study each contained four runs of three guanines, separated by either short or long loops. In addition to these, a second set of oligonucleotides was designed to investigate whether runs of four guanines would alter the trends observed for the quadruplexes containing G_3 tracts (Table 4.6). Increasing the number of guanines to four in each run could potentially result in structures containing four G-quartets, or alternatively three G-quartets with guanines participating in the loop regions. The effect of increasing the length of the G-tracts on quadruplex properties will be investigated in much greater depth in Chapter 6. In this second series of oligonucleotides, only the number of short or long loops was investigated, as the loop positioning within the

quadruplex appeared less significant with regards to the properties of the quadruplex formed.

Name	Oligonucleotide Sequence						
		Loop 1		Loop 2		Loop 3	
G_4T	d-F-TGGGG	T	GGGG	T	GGGG	T	GGGGT-Q
G_4T-T_4-T	d-F-TGGGG	T	GGGG	TTTT	GGGG	T	GGGGT-Q
G_4T_4-T-T_4	d-F-TGGGG	TTTT	GGGG	T	GGGG	TTTT	GGGGT-Q
G_4T_4	d-F-TGGGG	TTTT	GGGG	TTTT	GGGG	TTTT	GGGGT-Q

Table 4.6 Oligonucleotide sequences containing G_4 tracts. F = FAM, Q = Dabcyl

4.3.6.1 Topology

As performed for the first series of sequences, the topology of the structures adopted by the G_4 series of oligonucleotides was assessed by circular dichroism. The CD spectra for these sequences in the presence of sodium or potassium are shown in Figure 4.11. Samples were heated to 95 °C before being slowly cooled to room temperature over a period of 10 hours in a thermal controller. In the presence of sodium, G_4T and G_4T-T_4-T both exhibit peaks indicative of parallel stranded quadruplex formation. G_4T_4-T-T_4 displays peaks at both 265 and 295 nm whereas G_4T_4 has a typical antiparallel spectrum. This trend is similar to the spectra observed for the sequences containing G_3-tracts with equivalent length linkers. In the presence of potassium (Figure 4.11b), each of the sequences containing at least one single T loop display spectra consistent with parallel-stranded structures. G_4T_4 displays peaks at both 265 nm and 295 nm, which would suggest that unlike G_3T_4 which displays an antiparallel fold (although a number of structures may co-exist), G_4T_4 exhibits a mixed spectrum; this will be considered further in the Discussion. The structure(s) formed by G_4T_4 also appear to be independent of ionic strength (inset to Figure 4.11b), in contrast to G_3T_4 where the folding topology was shown to be dependent on the potassium ion concentration. The proton NMR spectra of each of the sequences (in the presence of 200 mM potassium) showed resonances in the imino region, as would be expected for structures containing G-quartets, but they were poorly defined, indicating multiple structures are likely to exist for all of the G_4 sequences (data not shown).

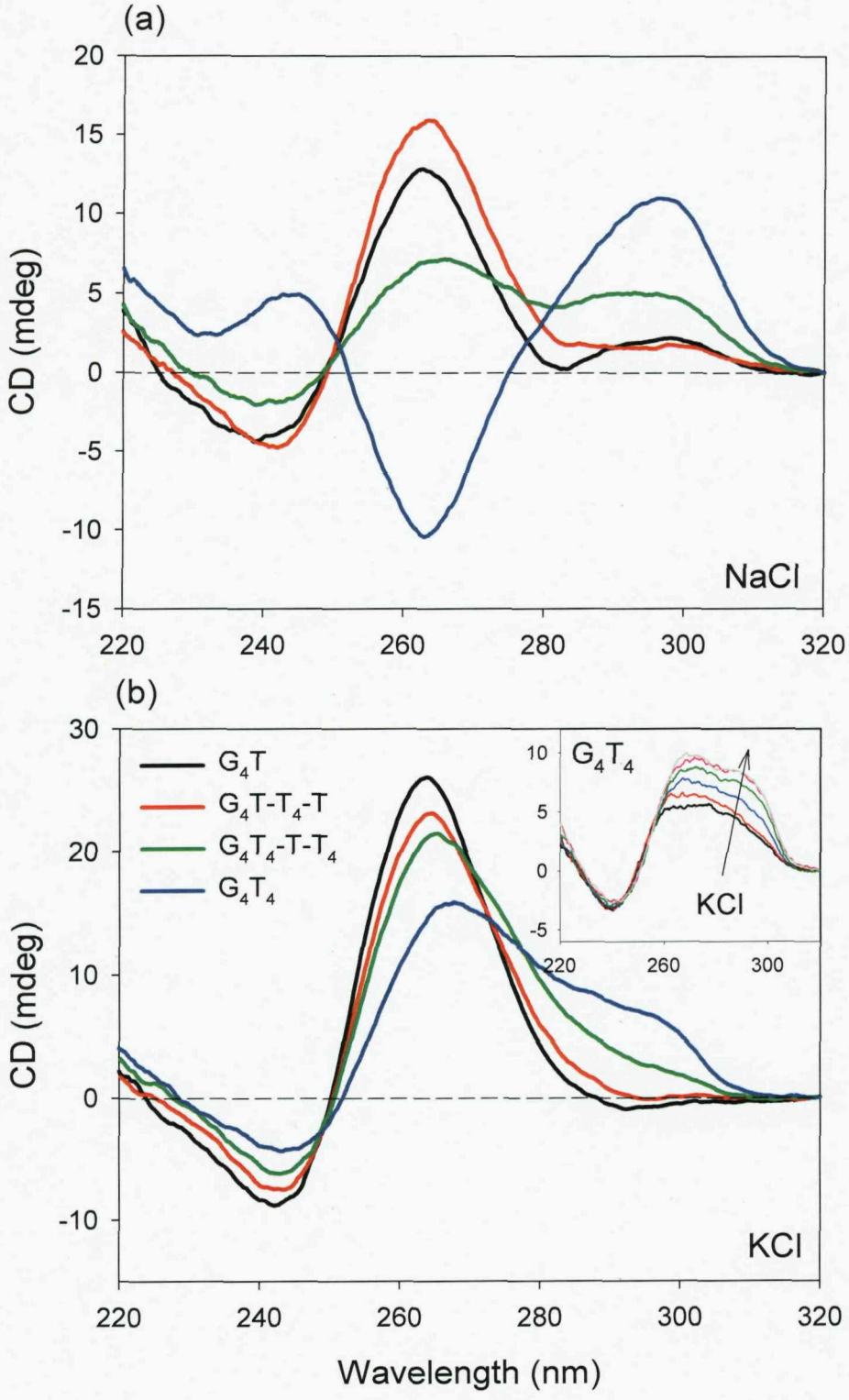


Figure 4.11 (a) CD spectra of the fluorescently-labelled quadruplex-forming oligonucleotides (5 μM) in the presence of 10 mM lithium phosphate pH 7.4 containing either (a) 200 mM NaCl, or (b) 200 mM KCl. Spectra are buffer normalised and a zero correction applied at 320 nm. G_4T (black); G_4T-T_4-T (red); G_3-T_4-T-T_4 (green); G_4T_4 (blue). The inset to the lower panel shows the CD spectrum of G_4T_4 in the presence of increasing concentrations of KCl; black 1 mM; red, 5 mM; blue 10 mM; green 20 mM; pink, 50 mM; grey, 200 mM.

4.3.6.2 Fluorescence Melting:

Table 4.7 shows the T_m values for the G_4 series of oligonucleotides, together with the equivalent loop isomers with G_3 tracts in the presence of potassium or sodium. Profiles were obtained at a heating and cooling rate of 0.2 °C.min^{-1}. In the presence of 100 mM potassium, G_4T is too stable and therefore does not display a melting transition. G_4T_4-T-T_4 is 10 °C less stable than G_4T-T_4-T, however G_4T_4 displays a large hysteresis, thereby making it difficult to accurately determine the T_m value. It is however noticeable that the values for both T_m (melt) and T_m (anneal) are greater than that of G_4-T_4-T-T_4. This is also the case in sodium, where G_4T_4 has a similar stability to G_4-T-T_4-T. This is unusual since the sequences with all long loops would be expected to have the lowest thermal stability. Interpretation of the melting data for the G_4 sequences must be made with caution however, as none of the four oligonucleotides fold to form a unique structure in the presence of potassium. Rather than forming a stack of four G-quartets with either T or T_4 loops, it is possible they could form three G-quartets, with guanines slipping into the loop regions.

A structure containing an additional G-quartet would be expected to have an increased thermal stability (Smirnov and Shafer, 2000). However a structure containing three quartets with longer loops (containing guanines) would likely cause a reduction in stability (Hazel *et al.*, 2005). Comparing the T_m values of the G_4 sequences with their G_3 equivalent loop isomers (Table 4.7) reveals clear differences as the short loops are substituted for longer loops, and these trends are similar in both sodium and potassium. G_3T is more stable than G_4T, but as the short loops are replaced, the G_4 sequences become increasingly more stable relative to their G_3 equivalents, with G_4T_4 over 20 °C more stable than G_3T_4.

(a)

Loop type	T_m (KCl)		ΔT_m
	G_3-tracts	G_4-tracts	
T-T-T	-	-	-
T-T_4-T	87.5	85.5	- 2.0
T_4-T-T_4	65.3	74.5	+ 9.2
T_4-T_4-T_4	56.3	76.4/84.9	+ 20.1/28.6

(b)

Loop type	T_m (NaCl)		ΔT_m
	G_3-tracts	G_4-tracts	
T-T-T	73.3	66.4	- 6.9
T-T_4-T	58.6	62.6	+ 4.0
T_4-T-T_4	43.4	52.5	+ 9.1
T_4-T_4-T_4	41.5	62.3	+ 20.8

Table 4.7 T_m values (°C) for the fluorescently labelled G_4 oligonucleotides and their equivalent G_3 loop isomers, determined in the presence of 10 mM lithium phosphate pH 7.4, containing **(a)** 100 mM KCl, or **(b)** 200 mM NaCl. Samples were heated and cooled at a rate of 0.2 °C.min^{-1}. Where two numbers are present, the first represents the T_m value from the annealing curve and the second the T_m value from the melting curve.

4.4 Discussion

In this study, the effects of loop length and distribution have been examined with respect to quadruplex stability, folding topology and kinetics. The results demonstrate that each of these aspects shows little sensitivity to the position of each loop within the quadruplex, but are significantly affected by the number of short or long loops within the structure.

4.4.1 Topology

Circular dichroism is often used to indicate the folding topology of DNA quadruplexes (Hardin *et al.*, 1991; Balagurumoorthy and Brahmachari, 1994; Hazel *et al.*, 2005; Paramasivan *et al.*, 2007). Characteristic signals arise due to the stacking geometry of the G-quartets, with the sequential alignment of guanine bases around the glycosilic bond (either *syn* or *anti*) suggested to be a key determinant of the strength and position of the peaks. Quadruplexes in which the strands are parallel contain guanines which all exhibit *anti* glycosilic angles (Phillips *et al.*, 1997) and display a positive peak at 265 nm. Antiparallel structures contain both *syn* and *anti* glycosilic bonds in varying ratios, depending on the specific nature of the fold. Several recent studies have provided evidence linking CD spectral signatures with G-quartet polarity and glycosilic angle through the use of modified oligonucleotides (Esposito *et al.*, 2004; Esposito *et al.*, 2005; Virgilio *et al.*, 2005). The substitution of a guanine for an $8^{Me}G$ (which favours the *syn* conformation) at the 5' end of the sequence d(TGGGT) resulted in a parallel-stranded structure which contained both *syn* and *anti* G-quartets, and a CD spectrum with peaks at both 265 nm and 295 nm (Virgilio *et al.*, 2005). A peak at 265 nm therefore can only be generated by an all-*anti* conformation, which in turn must be formed by a parallel-stranded structure. Interpretation of CD spectra for defining quadruplex topology have generally proven to be reliable, with one notable exception (Jing *et al.*, 1997; Dapic *et al.*, 2003). Nonetheless, CD spectra are still useful indicators of changes in global quadruplex folds for series of related oligonucleotides.

In the presence of potassium, all the oligonucleotides that contain at least one single T loop display similar CD spectra which are indicative of a parallel-stranded topology. It therefore appears that in potassium the presence of only one single T loop, in any position, is sufficient to promote all the other loops to form double-chain reversals. In

principle, these oligonucleotides could adopt several different folded conformations, yet the NMR and electrophoresis experiments suggest that only one predominates. When all three loops contain T_4, there is a dramatic change in the CD spectrum to a form which is indicative of an antiparallel conformation. The topology also appears to be dependent on the ionic strength, suggesting G_3T_4 can adopt multiple conformations. This again is consistent with the NMR and electrophoresis experiments which indicate the presence of multiple folded forms. Previous studies have shown that the quadruplex formed by the sequence $d(G_3T_4G_3)_2$ adopts an antiparallel hairpin dimer in the presence of both sodium and potassium (Keniry *et al.*, 1995; Strahan *et al.*, 1998). The formation of an antiparallel topology, once all the loops are longer, suggests that single nucleotide double-chain reversals may constrain longer loops into a similar orientation.

A similar effect is seen for G_3T, G_3T-T_4-T and G_3T_4-T-T in the presence of sodium ions. These sequences also display CD spectra that are consistent with parallel-stranded topologies. However the greater propensity to form antiparallel structures in the presence of sodium is seen with T_4-T-T_4 and T_4-T_4-T loops, which have spectral peaks at both 265 nm and 295 nm. This may indicate the presence of multiple structural forms, but is more likely due to the formation of a structure that contains both edgewise (T_4) and double-chain reversal (T) loops, an increasingly observed topology in biologically related quadruplex-forming sequences (Wang and Patel, 1994; Dai *et al.*, 2005; Ambrus *et al.*, 2006; Luu *et al.*, 2006).

Increasing the length of the G-tracts from three to four guanines resulted in the formation of multiple structures in the presence of potassium. G_4T displays a CD spectrum indicative of a parallel stranded fold, however the NMR spectrum revealed multiple ill-defined peaks. Geometrically, a single thymine loop is able to span between the top and bottom of a stack of four G-quartets (as a double-chain reversal). Despite this, it appears that a three-quartet core may be more favourable, with guanines slipping into the loop regions. This has been observed recently in the promoter region of the human RET oncogene where a sequence containing four G_4 tracts, separated by single nucleotide loops, folded to form three G-quartets rather than four (Guo *et al.*, 2007). A similar effect has also been seen with the *Tetrahymena* telomeric repeat sequence, which folds to form three G-quartets despite having four G_4-tracts (Wang and Patel, 1994). The appearance of multiple imino peaks, particularly for sequences which contain longer loops, could either be a result of several loop conformers existing in

equilibrium with the same number of stacked tetrads, or a combination of structures containing three or four G-quartets.

Substituting either one or two of the short loops in G_4T for longer T_4 loops (G_4T-T_4-T and G_4T_4-T-T_4), resulted in CD spectra which also indicated the formation of a parallel stranded structure. G_4T_4 displayed a mixed spectrum, unlike G_3T_4 which is antiparallel, and the topology appeared to be independent of ionic strength. Trends from the G_3 series of sequences suggested that quadruplexes with all long loops fold to form antiparallel structures, although these are restrained into parallel stranded folds when at least one short loop is present. G_4T_4 seems not to follow this trend as it is not antiparallel, although its CD spectrum is different from the other sequences. Previous studies have shown that a similar *Oxytricha* telomeric sequence can fold to form an antiparallel topology (Dapic *et al.*, 2003). However this sequence lacked any flanking bases at the 5' and 3'-ends, which may influence the final folded topology. G_4T_4 has also been shown to have unusual melting properties, with biphasic melting profiles and exceptionally large hysteresis (Brown *et al.*, 2005). The slow kinetics of folding may therefore be a further influence on the topology assumed.

In the presence of sodium, the CD spectra match those of their G_3 loop sequence equivalents. Two short loops are able to enforce a parallel-stranded structure, but the greater propensity for sodium-stabilised quadruplexes to form antiparallel structures is evident in the sequences containing one short loop (resulting in a mixed spectrum) and all longer loops (antiparallel).

4.4.2 Stability

The fluorescence melting experiments show that the number of short loops, rather than their position, has the greatest effect on quadruplex stability. In the presence of potassium ions, G_3T is most stable and in concentrations above 5 mM it does not display a melting transition. Substituting a T_4 loop into either the first or second loop decreases the T_m by about 20 °C, with a further 20 °C decrease on introducing a second T_4 substitution. Replacing all the T loops with T_4 results in the least stable structure, which as discussed previously adopts a different folded conformation. A similar trend is seen in the presence of sodium ions, though there is only a small decrease in stability on changing the third loop to T_4, consistent with the CD spectra which show that G_4T_4-T-

T_4, G_3T_4-T_4-T and G_3T_4 display some antiparallel characteristics in contrast to the other oligonucleotides. Although the number of short loops is the dominant factor in determining stability, the position of the loops does have a small effect. Interestingly, the sequences containing a central single T loop are 2 - 3 °C less stable and have lower gel mobilities than their sequence isomers containing T_4 in this position (i.e. comparing G_3T_4-T-T with G_3T-T_4-T, or G_3T_4-T-T_4 with G_3T_4-T_4-T). It appears that sequences with a central T_4 loop are more compact and have higher thermal stability.

Based on electrophoresis data, it has previously been suggested that G_3T (as well as G_3A and G_3C) folds to form tetramolecular, parallel-stranded quadruplexes rather than unimolecular complexes (Vorlickova *et al.*, 2007). The minimal variation in thermal stability on increasing the oligonucleotide concentration (between 0.1 and 10 μM) suggests that under these experimental conditions, G_3T and each of the other sequences studied fold to form an intramolecular quadruplex.

The order of stability of the G_4 sequences is slightly different in that the structures with the longest loops are not the least stable. It appears that only the sequences with longer loops are able to form four G-quartets. Comparing the stability of the G_3 sequences with their G_4-tract loop equivalents, gives an indication of the number of G-quartets formed; G_4T is less stable than G_3T, suggesting it contains three G-quartets with longer loops. As the short loops are replaced with longer ones, the G_4 sequences become increasingly more stable relative to their G_3 equivalents. An additional G-quartet has been shown to add around 20 °C to the thermal stability of a G-quadruplex with the same loop lengths (Smirnov and Shafer, 2000). G_4T_4 is therefore likely to contain four G-quartets in both sodium and potassium, but each of the other sequences only three. Single nucleotide loops can span both two (Kettani *et al.*, 2000) and three (Phan *et al.*, 2004) stacked G-quartets, however it seems that a quadruplex with four G-quartets cannot tolerate even one single nucleotide loop, rather it is more favourable to form a structure containing three G-quartets with longer loops.

4.4.3 Potassium ion binding

The variation of ΔG with ionic strength allows the estimation of differences between the number of potassium ions bound in the folded and unfolded structures (Jing *et al.*, 1997). As expected, this value is close to two for G_3T, since two potassium ions can

bind between the three stacked quartets. Although the precise values of Δn should be interpreted with caution, it is noticeable that there is a steady increase in this value as the number of longer loops is increased. The value of Δn is similar for both G_3T_4-T-T and G_3T-T_4-T, and is lower than for both G_3T_4-T-T_4 with G_3T_4-T_4-T. These results suggest that the longer loops may be involved in cation binding. The Δn value for G_3T_4 is significantly larger and may be a result of the antiparallel folding topology, as edgewise or diagonal loops are more likely to interact with the terminal quartets than double-chain reversal loops.

4.4.4 Kinetics

The kinetics of quadruplex formation appear to be strongly influenced by the length of the loop regions, with structures that contain short loops faster to fold and display longer dissociation half-lives. Initial observations when varying the rate of heating and cooling revealed that G_3T_4 was the slowest to fold and/or unfold, whereas G_3T has much faster folding and unfolding parameters. No hysteresis was observed for G_3T and temperature-jump experiments showed very fast re-equilibration times. Comparing the kinetic data for the sequences with one or two T_4 loops (Table 4.5) reveals that the unfolding parameters are very similar, while there are clear differences in the folding reactions. Complexes with longer loops have higher (less negative) activation energies and larger values for the pre-exponential factor (which is related to the entropy of the transition state). It is clear that the folding of a quadruplex with single nucleotide loops is fast. It could be imagined that if one loop is composed of a single nucleotide, the two flanking G-tracts could rapidly associate in an initial folding step. This could then provide a platform for the other two G-tracts to associate into the final folded quadruplex. The results with the oligonucleotides containing one or two T_4 loops imply that the position of the single base loop has little effect on the kinetics. G_3T-T_4-T and G_4T-T_4-T have very similar k_1 values, as do G_3T_4-T_4-T and G_4T_4-T-T_4, suggesting that the most important factor is the number of longer loops. It is interesting to note that none of these sequences show any hysteresis in the presence of sodium ions, even though G_3T and the sequences with two single base loops appear to adopt a similar global structure to those formed in potassium. The higher stability and slower kinetics in the presence of potassium may therefore reflect conformational changes subsequent to the initial folding event (Jing *et al.*, 1997b).

4.4.5 Biological Significance

Quadruplex structures that are formed from telomeric DNA commonly contain loops of equal length due to the highly repetitive nature of these sequences. Sequences located elsewhere in the genome, such as the many G-rich regions found in gene promoters and enhancers, have a much greater variation in loop length. It has been noted recently that despite the huge variety in potential quadruplex folding topologies, many of the structures determined in these promoter regions have a parallel-stranded folding pattern (Burge *et al.*, 2006). Within the sequences used in these experiments, the presence of only one single nucleotide loop (in the presence of potassium) was sufficient to induce an all parallel-stranded structure. Only once longer loops had replaced all the short loops did the structure become unrestrained in terms of folding pattern. The results from this study suggest a possible explanation for this observation. Table 4.8 shows a list of quadruplex forming motifs found in the promoter regions of a number of genes. It is clear that G_3NG_3 is a motif common to almost all of these sequences. The hypothesis that single nucleotide loops can enforce a parallel quadruplex fold is supported in many of these cases. Myc-1245 has a 1-6-1 loop composition with a parallel topology; the central six-nucleotide loop forming a double-chain reversal loop. RET, VEGF and c-kit21T all contain two single nucleotide loops, with a third loop of at least three nucleotides that also forms a double-chain reversal. There is however less data concerning sequences with one single nucleotide loop. The promoter region of the human bcl-2 gene contains a G-rich strand which has seven runs of two or more contiguous guanines (Dexheimer *et al.*, 2006). The predominant quadruplex species within this region contains one single nucleotide loop, although the overall topology is a hybrid parallel/antiparallel fold (Dai *et al.*, 2006). The *c-kit87-up* sequence was initially thought to have a 1-4-4 loop arrangement with a parallel-stranded topology (Rankin *et al.*, 2005). This arrangement would also have been predicted by the rules suggested in the present study. However a recent NMR structure revealed it adopts an unprecedented quadruplex scaffold (Phan *et al.*, 2007) in which a guanine within the central loop folds back to participate in G-quartet formation. This finding adds an additional layer of complexity to quadruplex folding topologies. The ability of one single nucleotide loop to enforce an all-parallel quadruplex fold may not therefore be as universal as when two short loops are present.

Name	Quadruplex-forming sequence							Topology	Reference
		Loop		*Loop*		*Loop*			
myc-2345*	5'-GGG	T	GGG	GA	GGG	T	GGG	Parallel	Phan *et al.*,2004
myc-1245*	5'-GGG	A	GGG	TTTTTA	GGG	T	GGG	Parallel	Phan *et al.*, 2004
myc22-G14T/G23T*	5'-GGG	A	GGG	TA	GGG	T	GGG	Parallel	Ambrus *et al.*, 2005
c-kit21T	5'-GGG	C	GGG	CGCGA	GGG	A	GGG	Parallel	Fernando *et al.*, 2006
c-kit87up*	5'-GGG	A	GGG	C (G^1) CT	GGG	AGGAG	GG	Parallel	Phan *et al.*, 2007
RET	5'-GGG	C	GGG	GCG	GGG	C	GGG	Parallel	Guo *et al.*, 2007
VEGF	5'-GGG	C	GGG	CCGG	GGG	C	GGG	Parallel	Sun *et al.*, 2005
bcl-2 MidG4Pu23 -G15T/G16T*	5'-GGG	CGC	GGG	AGGAATT	GGG	C	GGG	Mixed	Dai *et al.*, 2006
HIF-1α	5'-GGG	A	GGG	GAGAGG	GGG	C	GGG	Parallel	De Armond *et al.*, 2005
KRAS 28Rm i)	5'-GGG	A	GGG	AAGGA	GGG	A	GGG	Parallel	Cogoi and Xodo, 2006
ii)	5'-GGG	A	GGG	AAGGAGGGA	GGG	A	GGG	Parallel	Cogoi and Xodo, 2006
wtTel26*	5'-GGG	TTA	GGG	TTA	GGG	TTA	GGG	Mixed	Dai *et al.*, 2007

Table 4.8 Sequence and folding topology of a number of biologically relevant quadruplexes. Sequences shown are often mutated from wild-type in order to provide a unique species for high resolution structural analysis. Where appropriate these mutations are indicated within the sequence name. * indicates structures confirmed by NMR spectroscopy. In other cases the topology has been determined by chemical footprinting or other spectroscopic techniques.

[1] Guanine participates in G-quartet formation

4.4.6 Recent loop studies

Following the completion of this work, a paper by Bugaut and Balasubramanian was published in which the effect of loop-length on quadruplex topology was examined in a sequence-independent context (Bugaut and Balasubramanian, 2008). This work, based around 21 libraries of quadruplex-forming sequences of the type $G_3X_{1-3}G_3X_{1-3}G_3X_{1-3}G_3$, with partially randomized loops, supported our data concerning the importance of single nucleotide loops on quadruplex properties. The presence of two single nucleotide loops, as in the present study, enforced an all-parallel folded conformation as predicted by CD measurements. Sequences containing one single base loop and two three-nucleotide loops in any combination resulted in structures that were predominantly parallel, but also displayed a CD peak at 295 nm, suggesting some alternative structures may be present. These spectra represent an average of all the structures that may co-exist within each sequence library. Sequences with oligo-dT 1-4-4 and 4-1-4 loop combinations in the present experiments both appeared to be exclusively parallel. It is therefore likely that loop sequence can contribute to, or even determine the final folded conformation when longer loops are involved, either through loop-loop interactions, or even loop-G-stack interactions as has been recently observed (Phan *et al.*, 2007). Resolving which combinations of loop sequences result in alternative structures would require a much more in depth and extensive study. Nonetheless, it is clear that short loops still appear to be the dominant influence on quadruplex topology, forming the most stable quadruplexes with the fastest folding kinetics.

CHAPTER 5

Sequence Effects of Single Nucleotide Loops in Intramolecular Quadruplexes

5.1 Introduction

Single nucleotide loops are a common structural element observed in genomic quadruplex-forming sequences. Early quadruplex structural studies were largely based on telomeric DNA regions, in which linker lengths of two or more nucleotides spanned the G-tracts. But as the search for potential genomic quadruplex-forming sequences (PQS) has widened, so too has the prevalence of single base loops. The first bioinformatic searches for PQS revealed as many as 376,000 genome-wide (Huppert and Balasubramanian, 2005; Todd *et al.*, 2005) and amongst these the most common set of loop lengths was 1-1-1. These sequences accounted for 8% of the total number of hits, over four times more than the second most common loop-length combination (Huppert and Balasubramanian, 2005). Single nucleotide loops are not only common, but as shown in the previous chapter, have a strong influence on the final folded topology of the quadruplex. The presence of two (and possibly one) single nucleotide loops can enforce an all-parallel stranded structure (Rachwal *et al.*, 2007; Bugaut and Balasubramanian, 2008). Whilst loop length is known to be a key determinant of quadruplex structural and thermodynamic properties, sequence is also an important factor (Smirnov and Shafer, 2000; Miyoshi *et al.*, 2005; Vorlickova *et al.*, 2007). Since single nucleotide loops are so common amongst PQS, the first part of this chapter focuses on investigating the effect of base sequence on the properties of quadruplexes that contain the loop lengths 1-1-1.

5.1.1 C-kit

The second part of the chapter extends this loop study to focus on one particular biologically relevant quadruplex-forming sequence, *c-kit87up*. This sequence, which contains two single nucleotide loops, was identified as one of two PQS located within the promoter region of the oncogene *c-kit* (Rankin *et al.*, 2005; Fernando *et al.*, 2006).

The c-kit protein, a tyrosine kinase receptor (Yarden *et al.*, 1987), plays a key role in cellular signalling, mediating processes including cell survival, proliferation and differentiation (Vliagoftis *et al.*, 1997). Mutations in the *c-kit* gene have been implicated in a number of human malignancies (Heinrich *et al.*, 2002) and clinically it represents an important target for the treatment of gastrointestinal stromal tumours (Hirota *et al.*, 1998; Lux *et al.*, 2000). Gene function (particularly proto-oncogene function) has recently been shown to correlate with the potential for quadruplex formation (Eddy and Maizels, 2006). The presence of two PQS within the *c-kit* promoter and their proximity to the transcription start site of this gene has highlighted its potential as a target for regulating expression at the level of transcription. Biophysical studies of the two PQS within the *c-kit* promoter suggest they can both form quadruplexes under close to physiological conditions (Rankin *et al.*, 2005; Fernando *et al.*, 2006). *C-kit87up*, the PQS located closer to the transcription start site, is the better characterised of the two. It contains a unique parallel-stranded topology with four loop regions, including two single nucleotide loops (Phan *et al.*, 2007) (Figure 5.1).

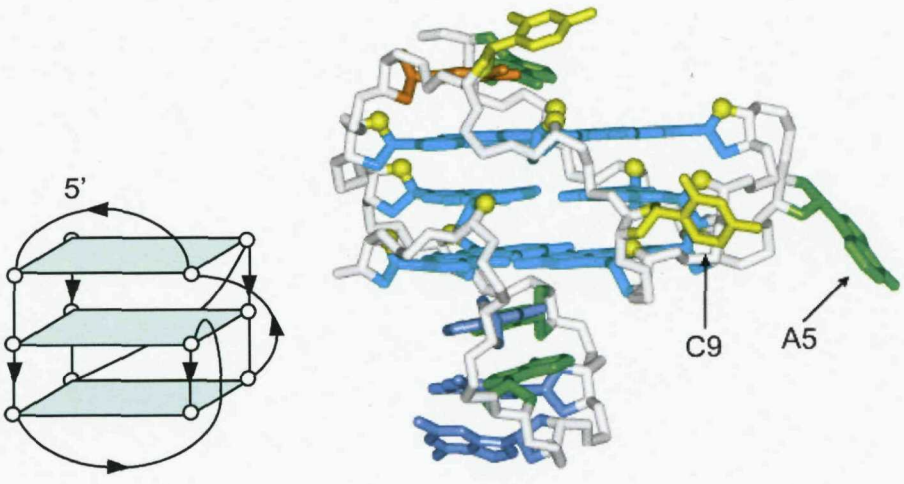


Figure 5.1 Schematic representation of the *c-kit87up* folding topology (left). Solution structure of the *c-kit87up* sequence determined in potassium, with single nucleotide loops indicated (Phan *et al.*, 2007). Figure modified from Patel *et al.* (2007).

Current studies regarding quadruplex formation within the *c-kit* promoter have been limited to biophysical or structural characterisation. As yet there is no biological evidence for their potential role within the promoter. A similar G-rich sequence within the nuclease hypersensitivity element (NHE) III_1 of the *c-myc* promoter (Simonsson *et al.*, 1998) has been shown to repress gene expression through quadruplex formation (Siddiqui-Jain *et al.*, 2002). Of the two possible intramolecular quadruplex folds within

the sequence only one was biologically active as a repressor, suggesting there may be a secondary level of quadruplex structural discrimination with regards to gene activity.

The second part of this chapter first examines how single base loop substitutions within a biologically relevant quadruplex-forming sequence, *c-kit87up* can alter its stability. Based on these findings, both the native *c-kit* promoter sequence and a number of modified sequences were evaluated for promoter activity. In this way, the effect of subtle sequence changes within the quadruplex in the *c-kit* promoter can be assessed.

5.2 Experimental Design

The oligonucleotide sequences chosen for the initial part of the study are based on the sequence d(TGGGNGGGNGGGNGGGT) (G_3N), where N represents each base in turn (except guanine) or 1',2'-dideoxyribose (Table 5.1). Guanine was not considered due to the likelihood of multimeric aggregation. The stability of each complex was examined by fluorescence melting and the topology assessed by CD. The study was then extended to evaluate whether similar trends in stability occur with loop sequence modifications in the *c-kit87up* sequence. This sequence was chosen as it represented one of the few characterised quadruplex-forming sequences that forms a single folded structure and contains two single nucleotide loops. Finally, a number of these sequences were incorporated within a longer *c-kit* promoter sequence. The wild-type promoter and modified sequences were then assessed for their ability to form quadruplex structures using a Taq polymerase inhibition assay, before finally being incorporated as part of a larger construct into an expression vector to evaluate promoter activity.

Name	Oligonucleotide Sequence
G_3T	d-F-TGGG T GGG T GGG T GGGT-Q
G_3C	d-F-TGGG C GGG C GGG C GGGT-Q
G_3A	d-F-TGGG A GGG A GGG A GGGT-Q
$G_3\Phi$	d-F-TGGG Φ GGG Φ GGG Φ GGGT-Q
G_3A-T-A	d-F-TGGG A GGG T GGG A GGGT-Q
G_3T-A-T	d-F-TGGG T GGG A GGG T GGGT-Q

Table 5.1 Oligonucleotide sequences used in this work. F = FAM, Q = Dabcyl, Φ = 1',2'-dideoxyribose.

5.3 Results

5.3.1 Melting Studies

Fluorescence melting and circular dichroism studies have been carried out to investigate the effect of sequence on the stability of quadruplexes that contain single nucleotide loops using oligonucleotides of the type d(TGGGNGGGNGGGNGGGT), where N represents each base in turn or 1',2'-dideoxyribose (Table 5.1). Whilst it would be preferable to use ionic conditions that are physiologically relevant (ie ~ 100-150 mM potassium), quadruplexes containing single base loops are known to be extremely stable (Risitano and Fox, 2003a; Rachwal *et al.*, 2007c). In the presence of 100 mM KCl all these sequences formed structures that melted above 90 °C. However by reducing the potassium concentration to 1 mM the melting temperatures decrease to a measurable range. Although this concentration is lower than is used in most quadruplex studies, it allows comparisons to be made between the sequences. Measurements were also taken in the presence of 100 mM NaCl. Each of the sequences display melting and annealing profiles that are fully reversible at a rate of temperature change of 0.2 °C.min^{-1}, and are independent of oligonucleotide concentration (between 0.1 and 10 μM), suggesting that they all form intramolecular folded complexes (Figure 5.2). Melting profiles for the four oligonucleotides that contain the same base in each loop (TTT, CCC, AAA or ΦΦΦ) are shown in Figure 5.2 in the presence of 1 mM KCl or 100 mM NaCl and the melting temperatures determined at a range of ionic strengths are summarised in Table 5.2.

It can be seen that loops consisting of T, C and 1',2'-dideoxyribose produce the most stable structures (Φ > T > C), while the T_m of the complex with As in the loops is about 25 °C lower. The structures are less stable in sodium-containing buffers, but the rank order of stabilities is the same. The quadruplex that contains all adenine loops has a similar stability to a structure containing propanediol linkers (Risitano and Fox, 2004). Thermodynamic parameters for the folding of each complex in the presence of potassium were estimated from van't Hoff analysis of the melting profiles and these are presented in Table 5.3. The values are consistent with those previously determined for intramolecular quadruplexes, which are typically between 65 and 100 kJ.mol^{-1} (Smirnov and Shafer, 2000; Risitano and Fox, 2003a). These show that the complexes with the lower stability are characterised by a smaller enthalpy and higher entropy.

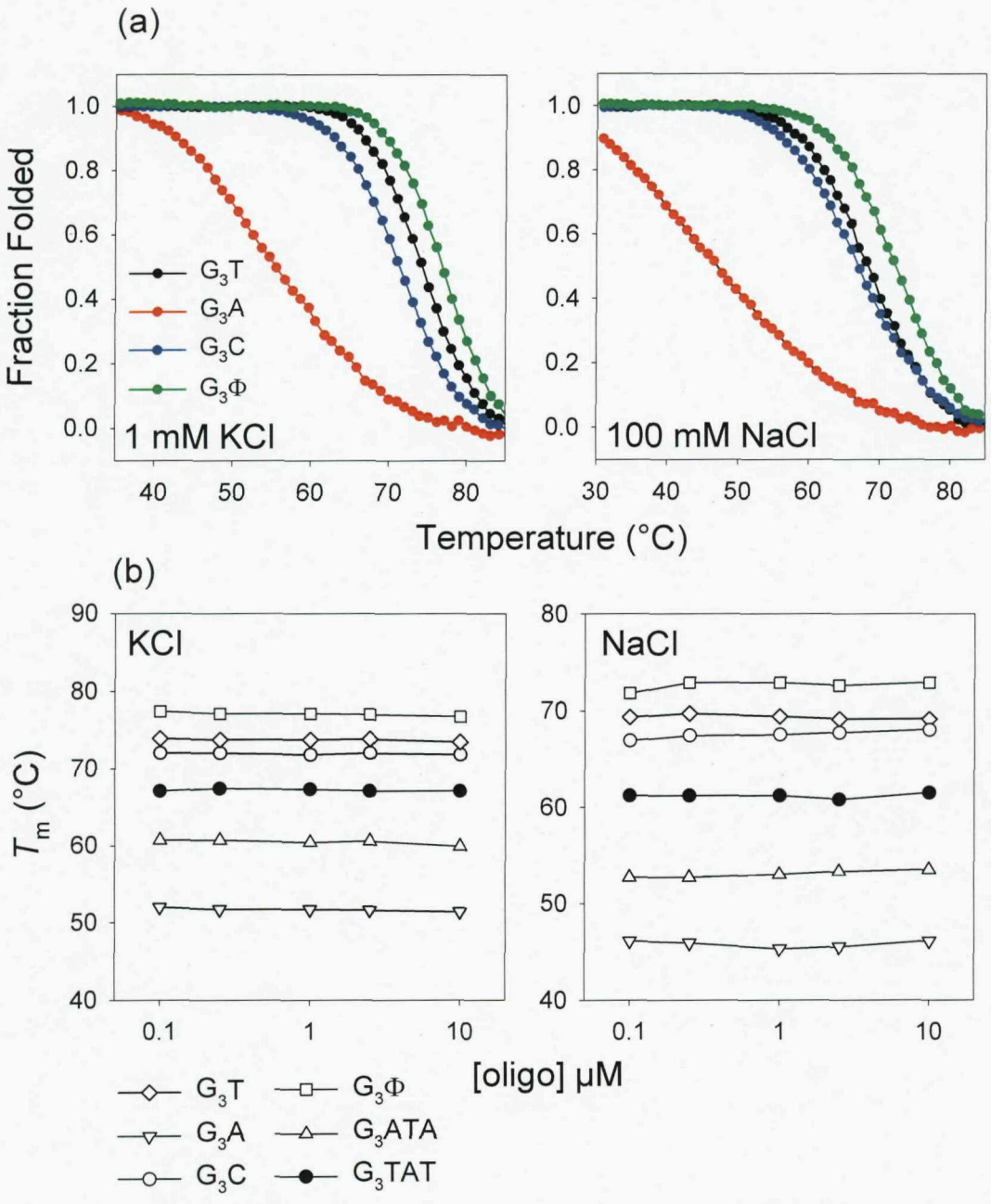


Figure 5.2 a) Representative fluorescence melting profiles for the quadruplex-forming oligonucleotides. The reactions were performed in 10 mM lithium phosphate containing either 1 mM KCl (left) or 100 mM NaCl (right). The curves show the fraction folded (α) as a function of time. b) Concentration dependence of the T_m values between 0.1 and 10 μM oligonucleotide.

	T_m (°C)					
	G_3T	G_3A	G_3C	$G_3\Phi$	G_3A-T-A	G_3T-A-T
KCl (mM)						
0	46.0	< 30	44.4	49.5	36.9	37.9
1	73.9	50.0	71.8	77.0	59.6	67.3
5	85.6	64.4	83.2	88.1	72.1	79.3
10	> 90	70.0	> 90	> 90	78.0	84.2
20	> 90	82.2	> 90	> 90	> 90	> 90
NaCl (mM)						
1	48.3	< 30	47.3	53.2	< 30	40.0
10	53.4	< 30	53.6	60.5	39.6	46.3
50	61.7	< 30	62.3	68.0	47.0	55.6
100	68.3	43.7	67.1	72.5	52.2	60.5
200	73.3	50.2	72.1	77.0	57.8	65.6

Table 5.2 T_m values (°C) for the various oligonucleotides, determined in 10 mM lithium phosphate pH 7.4 containing different concentrations of KCl or NaCl. The oligonucleotide concentration was 0.25 μM and the samples were heated and cooled at a rate of 0.2 °C.min^{-1}. All T_m values ate ± 0.5 °C.

To further study the effect of loop sequence on stability, the properties of oligonucleotides with two Ts in the loop and one A (G_3T-A-T), or one T and two As (G_3A-T-A) were examined. The melting profiles of these sequences are shown in Figure 5.3, alongside those for G_3A and G_3T and the thermodynamic parameters are presented in Table 5.3. Each T to A substitution reduces the melting temperature by about 8 °C in both potassium and sodium-containing buffers, with a change in enthalpy of 46 kJ.mol^{-1} per T to A substitution in 1 mM KCl and 33 kJ.mol^{-1} in 100 mM NaCl.

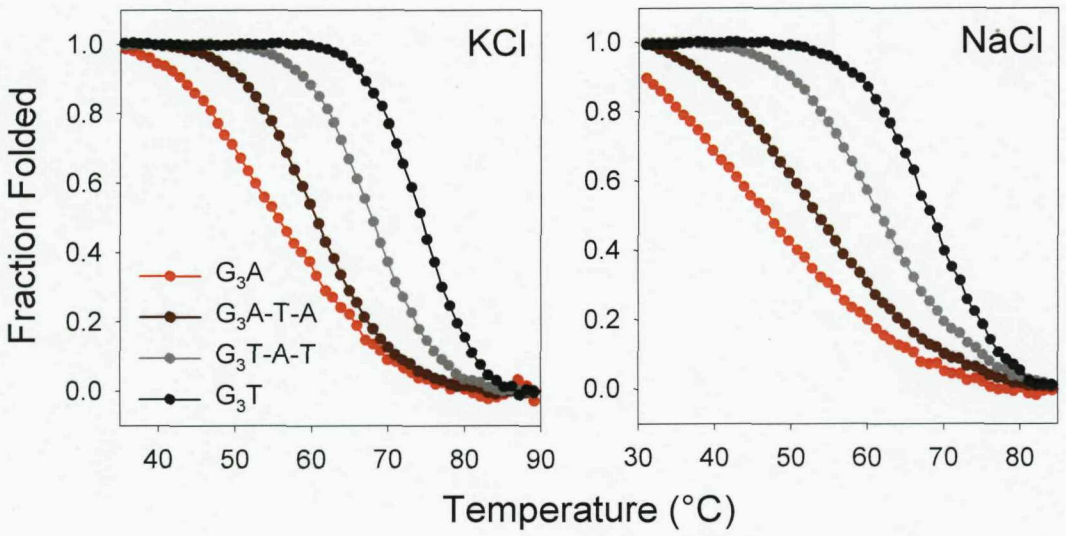


Figure 5.3 Fraction folded of the series of oligonucleotides (0.25 μM) at different temperatures in 10 mM lithium phosphate buffer, pH 7.4 containing a) 1 mM KCl or b) 100 mM NaCl.

5.3.2 Circular Dichroism

The oligonucleotide sequence G_3T with all single base loops has previously been shown to exhibit a CD spectrum indicative of a parallel-stranded structure (Dapic *et al.*, 2003; Hazel *et al.*, 2005; Rachwal *et al.*, 2007c). The CD spectra of these sequences in the presence of either 100 mM KCl or 100 mM NaCl are shown in Figure 5.4 and show that the complexes produce very similar CD spectra with a clear maximum at 265 nm. These spectra therefore strongly suggest that all the complexes adopt the same topology, which is likely to be the parallel form containing all double-chain reversal loops.

Sequence	1 mM KCl				100 mM NaCl			
	T_m	ΔH (kJ.mol^{-1})	ΔS (kJ.mol^{-1}.K^{-1})	ΔG^{37} (kJ.mol^{-1})	T_m	ΔH (kJ.mol^{-1})	ΔS (kJ.mol^{-1}.K^{-1})	ΔG^{37} (kJ.mol^{-1})
G_3T	73.9	-278	-0.80	-30.0	68.3	-206	-0.60	-20.0
G_3C	71.8	-257	-0.75	-24.5	67.1	-181	-0.53	-16.7
G_3A	50.0	-142	-0.44	-5.6	43.7	-104	-0.33	-1.7
$G_3\Phi$	77.0	-296	-0.85	-32.5	72.5	-229	-0.66	-24.4
G_3T-A-T	67.3	-247	-0.73	-20.7	60.5	-158	-0.47	-12.3
G_3A-T-A	59.6	-198	-0.60	-12.0	52.2	-131	-0.40	-7.0

Table 5.3 Thermodynamic parameters for folding of the quadruplex-forming oligonucleotides. All T_m values are ± 0.5 °C, ΔH values varied by around 5%. Experiments were performed in 10 mM lithium phosphate pH 7.4 containing 1 mM KCl or 100 mM NaCl. ΔH was calculated from van't Hoff analysis of the melting profiles, assuming a two-state equilibrium and ΔS was estimated as $\Delta H/T_m$. ΔG was calculated (at 37 °C) as ΔH – T.ΔS.

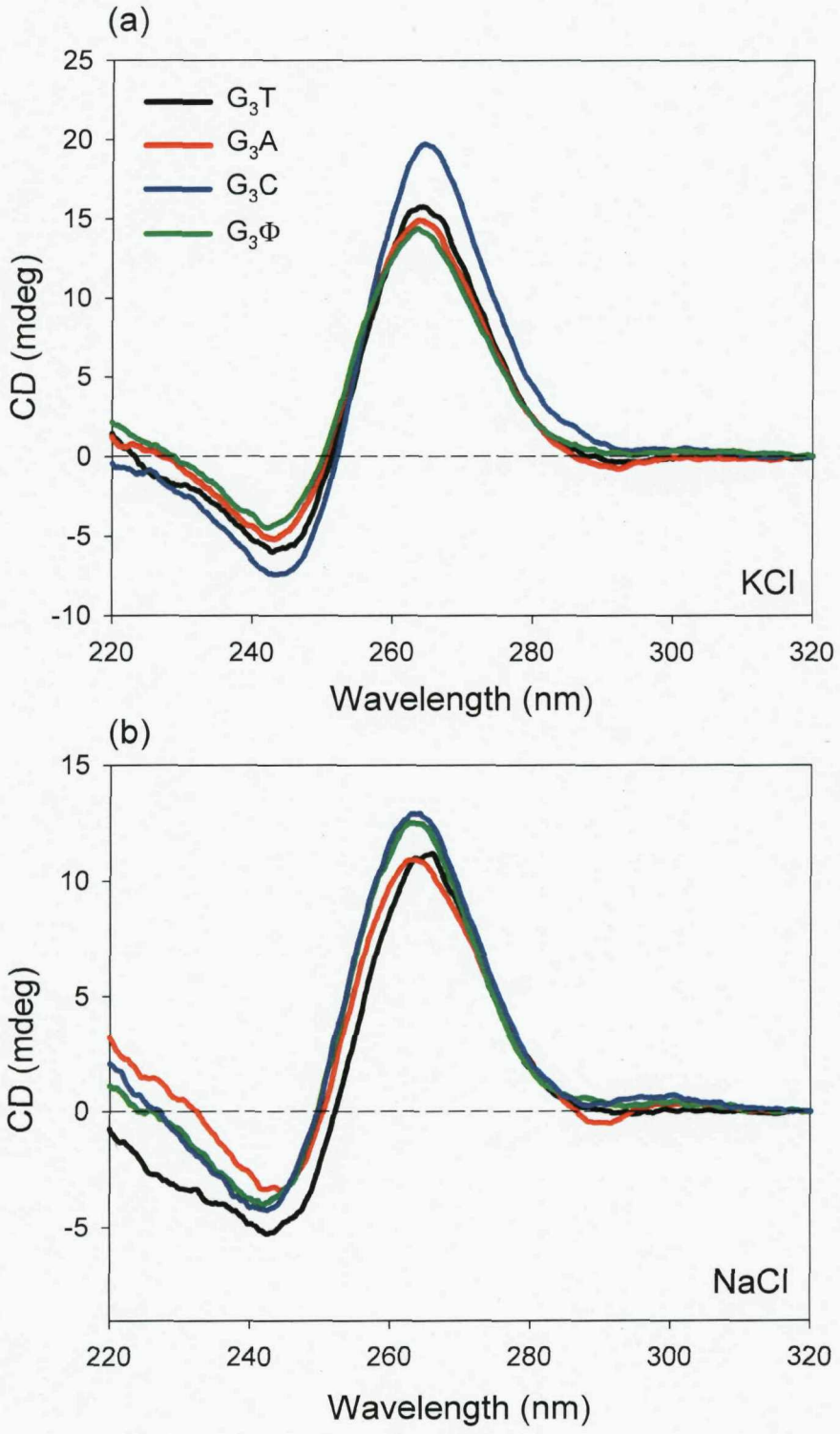


Figure 5.4 CD spectra of the fluorescently-labelled quadruplex-forming oligonucleotides in the presence of 10 mM lithium phosphate pH 7.4 containing either **(a)** 100 mM KCl or **(b)** 100 mM NaCl. Spectra are buffer normalised and a zero correction applied at 320 nm. G_3T (black); G_3A (red); G_3C (blue); $G_3\Phi$ (green)

5.3.3 Single base loops in the *c-kit87up* sequence

Having established that single base loop sequence can moderate quadruplex stability within the model G_3N series, the study was extended to examine whether modifying the sequence of single base loops could alter the properties of a known biologically-relevant quadruplex-forming sequence. The 22 nt *c-kit87up* sequence, found within the promoter region of the *c-kit* gene (Rankin *et al.*, 2005), was chosen for this study as it contains two single nucleotide loops; an A at loop position 1 and a C at loop position 2, although due to the unusual folding topology of this sequence, there are four loop regions in total (Phan *et al.*, 2007). The single base loops within *c-kit87up* have been shown not to be integral to the final conformation of the quadruplex structure (Phan *et al.*, 2007; Todd *et al.*, 2007). Modifying these bases should therefore alter the stability without changing the topology.

5.3.4 *c-kit* melting studies

Three modifications of the wild type *c-kit87up* sequence were made (Table 5.4); Mod 1 contains an A to T mutation at position 5, Mod 2 contains a C to A mutation at position 9 and Mod 3 contains a C to T mutation at position 9. Each sequence was melted and annealed at a rate of 0.2 °C.min^{-1} and no hysteresis was observed (Figure 5.5). The T_m values for these sequences in the presence of sodium and potassium are shown in Table 5.4. Mod 1 with a thymine at position 5 is around 8 °C more stable than the wild-type *c-kit* sequence (which has an A in this position) in the presence of 100 mM potassium. Substituting the C at position 9 for an A (Mod 2) is destabilising, whereas replacing the C9 with a T (Mod 3) results in a complex which is only slightly more stable than the wild-type. These results are consistent with those obtained with the G_3N model sequences and confirm the order of single base loop stability T $\geq$ C >> A. The CD spectra of each of these sequences revealed no change in global topology, as each oligonucleotide displayed the same spectral signature (data not shown).

Previous studies have shown that in the presence of sodium ions, this sequence does not fold to form the same topology (Phan *et al.*, 2007). This sequence could in theory fold to form with a 1-4-4 loop distribution with one single nucleotide loop at position 1, in a similar manner to the model sequences in Chapter 4. Melting profiles in the presence of sodium (with the exception of Mod 3), were shallow with broad melting and annealing

transitions. The single base modifications have little effect on stability with the exception of C9-A (Mod 2) which is destabilising (Table 5.4).

Name	Sequence	T_m (°C) KCl	NaCl
	1 5 10 15 20		
c-kit87up (WT)	AGGGAGGGCGCTGGGAGGAGGG	60.9	44.7
Mod 1 (A5-T)	AGGGTGGGCGCTGGGAGGAGGG	68.3	45.5
Mod 2 (C9-A)	AGGGAGGGAGCTGGGAGGAGGG	56.2	< 30
Mod 3 (C9-T)	AGGGAGGGTGCTGGGAGGAGGG	61.8	41.9

Table 5.4 Sequence and thermal stability of the wild-type *c-kit87up* quadruplex-forming region and three mutant sequences as determined by fluorescence melting experiments. Oligonucleotides (0.25 μM) were dissolved in 10 mM lithium phosphate pH 7.4 containing either 100 mM KCl or 100 mM NaCl. Bases in red indicate the single base loops, underlined bases represent those participating in G-quartet formation.

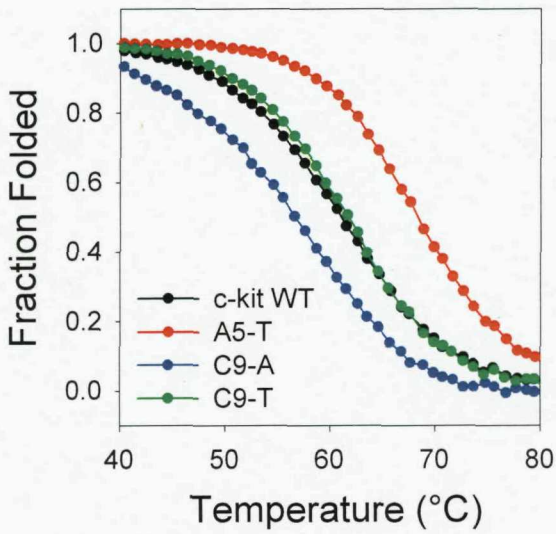


Figure 5.5 Fraction folded plots of the *c-kit87up* sequence and single base substituted sequences. Oligonucleotides were dissolved in 10 mM lithium phosphate pH 7.4, containing 100 mM KCl.

The results of these melting studies showed that the stability of the potassium form of the quadruplex can be moderated by altering the sequence of the single base loops. To evaluate the biological significance of quadruplex formation within the promoter region of *c-kit*, a new series of sequences was designed with two further aims; firstly to assess whether quadruplex formation in the *c-kit* promoter region could affect gene expression, and secondly whether increasing the stability of the quadruplex by modifying the loop sequence could affect its biological properties. The quadruplex-forming ability of these sequences was first assessed by polymerase stop assays and each sequence then

evaluated for promoter activity by assembling a series of reporter constructs which could be tested in luciferase expression assays.

5.3.5 Polymerase stop assays

Three further sequences were designed to test for the ability of *c-kit87up* to form a quadruplex when flanked by the extended promoter sequence, using a Taq polymerase stop assay. The *c-kit21* sequence, which is 20 °C more stable than *c-kit87up* (Fernando *et al.*, 2006), has been shown to form non-duplex states within the natural extended DNA duplex (Shirude *et al.*, 2007). However studies with *c-kit87up* have been limited to the target 22 nt sequence (Rankin *et al.*, 2005; Phan *et al.*, 2007). The addition of longer flanking regions to either side of a quadruplex-forming sequence has previously been shown to be destabilising (Guo *et al.*, 1993), and may therefore hinder the ability of this sequence to form a quadruplex, particularly in light of the complex folding topology. The sequences chosen for the polymerase stop experiments are shown in Figure 5.6. In addition to the wild-type sequence, a second sequence with the single base loop A5 replaced with a thymine (as Mod 1) was introduced. In this way, the potential stabilising effect of the A-to-T substitution could also be assessed. As a negative control, a third scrambled sequence was made in which a number of the guanines and cytosines were switched to eliminate the possibility of G-quadruplex formation, while maintaining the G-C content of the sequence. The sequences flanking the 22 nt quadruplex region were those found in the native promoter.

Wild-type *c-kit* (WT)

CGCGCAGAGGGAGGGCGCTGGGAGGAGGGGCTGCTGCTCGCCGCTCGCGGCTCTGGGGGC

A5-T Point mutation (Point)

CGCGCAGAGGG**T**GGGCGCTGGGAGGAGGGGCTGCTGCTCGCCGCTCGCGGCTCTGGGGGC

Scrambled (Scram)

CGCGCAGAG**C**GA**C**GC**C**GCT**C**GCA**C**CA**C**GC**C**GCTGCTGCTCGCCGCTCGCGGCTCTGGGGGC

Figure 5.6 Oligonucleotide sequences used for the polymerase stop assays. Underlined sequence represents the 22 nt quadruplex-forming region, nucleotides in red represent modifications from the native sequence.

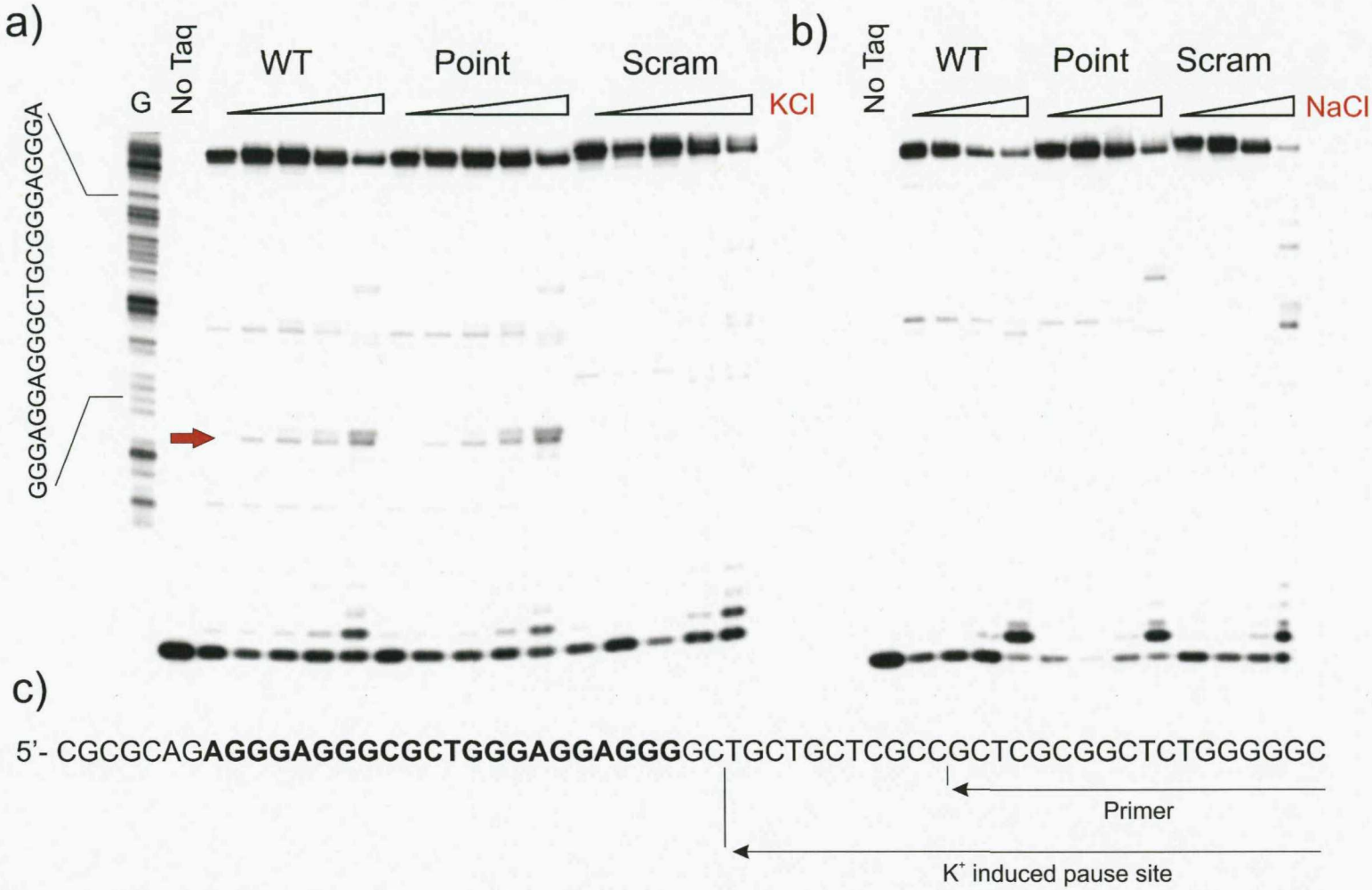


Figure 5.7 Inhibition of Taq polymerase DNA synthesis by quadruplex formation in the *c-kit87up* sequence and two variant sequences in the presence of increasing concentrations of **a)** KCl (0, 25, 50, 75, 100 mM), and **b)** NaCl (0, 25, 50, 100 mM). G represents a guanine marker lane. The red arrow indicates the quadruplex-induced Taq pause site. **c)** *c-kit87up* WT template sequence, showing the primer binding site and potassium-induced Taq pause site.

The results of the Taq polymerase stop assays are shown in Figure 5.7. A ^{32}P-radiolabelled primer was first annealed to the 3' end of each template sequence and extended with Taq polymerase. The folding of a quadruplex structure within the template should inhibit the primer extension and result in a band at the quadruplex-induced pause site. Experiments were carried out in the presence of increasing concentrations of either potassium (0, 25, 50, 75, 100 mM) or sodium ions (0, 25, 50, 100 mM) (Figures 5.7a and b respectively). The control lane indicates the position of the guanine bases within the sequence as determined by Sanger sequencing. In the absence of potassium, only the full-length product is observed. As the concentration of potassium is increased, a band of increasing intensity appears in the wild-type and point mutant sequences (indicated by the red arrow), but is absent in the scrambled sequence. It is also noticeable that as the ionic strength increases, a second band of polymerase stop product appears one base above the first, which is of similar intensity at 100 mM KCl. These stop products are located close to the beginning of the potential quadruplex forming sequence and suggest that the formation of a G-quadruplex has prevented the full extension of the primer (Figure 5.7c). Comparing the intensity of the quadruplex-induced pause site bands for the wild-type and point mutant sequences at 100 mM KCl reveals only very minor differences in band intensity. The increase in quadruplex stability caused by the A-to-T point mutation does not appear to prevent primer extension to any greater extent than the wild-type sequence.

In the presence of sodium, no paused products are observed in this region, even at the highest ionic strength, suggesting that a quadruplex does not form in this sequence with sufficient stability as to prevent primer extension. A number of fainter bands are visible above the quadruplex-induced pause site, both in sodium and potassium, although they do not depend on the ionic strength.

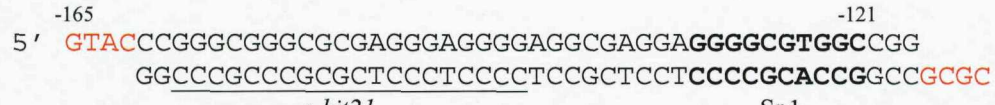
5.3.6 Luciferase promoter assays

To evaluate whether G-quadruplex formation in the *c-kit* promoter region can modulate gene expression, the *c-kit* promoter was ligated in three cassettes into the pGL-3 luciferase expression vector. Reporter assays have previously been used to probe for quadruplex-induced modulation of gene expression for a number of promoter sequences, including *c-myc* (Siddiqui-Jain *et al.*, 2002), HIF-1α (De Armond *et al.*, 2006) and KRAS (Cogoi and Xodo, 2006). The 165 bp of the *c-kit* promoter region was

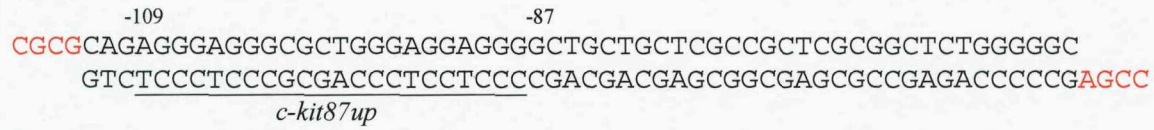
divided into three cassettes (Figure 5.8a), in which the purine-rich strand of cassette 2 (containing the *c-kit87up* sequence) was used for the polymerase stop assays. Cassette 1 contained the second quadruplex-forming region within the promoter sequence (*c-kit21*). Two groups of three reporter plasmids were assembled; one containing all three cassettes, with three variants of cassette 2, and the second group containing each of the variants of cassette 2 plus cassette 3 (i.e. without cassette 1).

a)

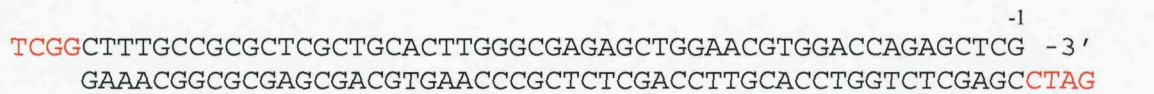
Cassette 1



Cassette 2



Cassette 3



b)

Wild type *c-kit87up*	AGGGAGGGCGCTGGGAGGAGGG TCCCTCCCGCGACCCTCCTCCC
Point mutation	AGGG**T**GGGCGCTGGGAGGAGGG TCCC**A**CCCGCGACCCTCCTCCC
Scrambled	AG**C**GA**C**GC**C**GCT**C**G**C**A**C**C**A**C**G**C TC**G**CT**G**C**G**GCGA**G**C**G**T**G**G**T**G**C**G

Figure 5.8 a) 165 bp of the *c-kit* promoter sequence split into three cassettes, which were cloned into the pGL3 basic luciferase expression vector. Underlined sequence represents the quadruplex-forming regions, bold sequence shows the location of Sp1 binding. **b)** Modifications from the wild-type *c-kit87up* quadruplex-forming region in cassette 2.

The results of the luciferase expression studies are shown in Figure 5.9. As a control for relative activity, the pGL3 basic vector minus a promoter, and the pGL3 control vector, which contains an SV40 promoter were transfected alongside each of the plasmid

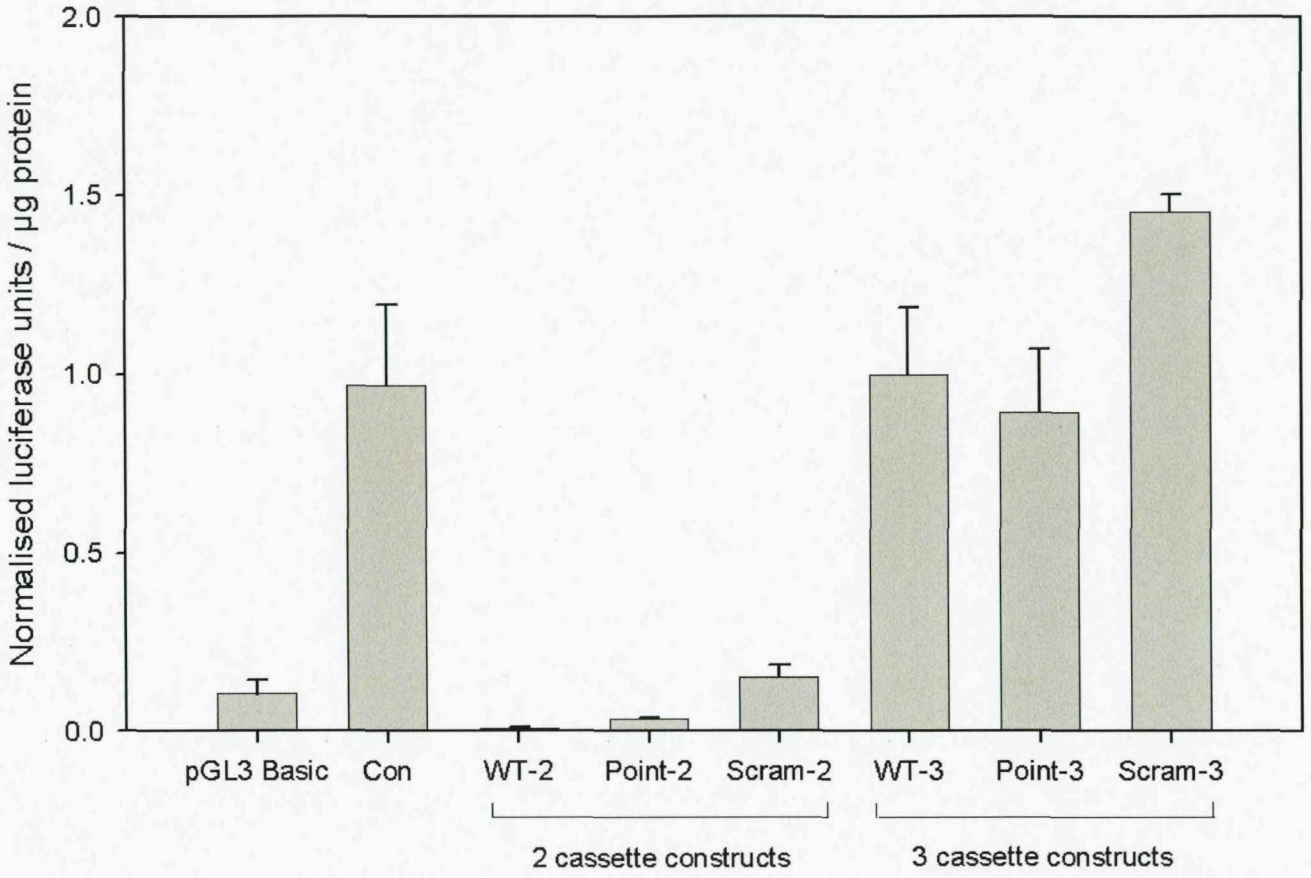


Figure 5.9 Relative activity of the six *c-kit* promoter constructs as determined by luciferase expression. Results are normalised relative to the wild-type 3-cassette construct. pGL3 Basic is the promoter-less construct, Con is the pGL3 Control plasmid containing an SV40 promoter. Error bars indicate standard deviation calculated from three separate experiments.

constructs. Figure 5.9 shows the relative luciferase expression levels from the two-cassette constructs that lack the -118 to -165 bp cassette 1. In accordance with previously observed *c-kit* promoter activity assays (Park *et al.*, 1998), very low gene expression levels were observed from the truncated promoter sequence, most likely because cassette 1 contains an Sp1 binding site at -121/-130 bp upstream of the transcription initiation site (Figure 5.8). Inserting cassette 1 into the plasmid constructs so that the full 165 bp sequence was present resulted in much higher levels of promoter activity (Figure 5.9). Comparing the wild-type construct with the scrambled construct which contains mutations to abolish quadruplex formation, reveals around a 40 % increase in luciferase expression in the scrambled quadruplex construct. Substituting the single base loop A5 in the *c-kit87up* region for a thymine, which increased the T_m of the quadruplex, results in a statistically insignificant reduction in promoter activity relative to the wild-type. These results show that mutating residues within the *c-kit87up* region to eliminate the possibility of quadruplex formation causes an increase in luciferase expression. However mutating the loop sequence so as to increase the stability of the quadruplex has little effect.

5.4 Discussion

The data presented in this chapter demonstrate that loop sequence can have a large effect on quadruplex stability without altering the topology, and that trends observed within a model system also applied in a biologically relevant sequence. The effects of sequence changes and stability moderations of the *c-kit87up* sequence on its ability to prevent Taq polymerisation and ultimately promoter activity are discussed.

5.4.1 Melting studies: G_3N

Fluorescence melting experiments showed that the sequence of single nucleotide loops can have a significant effect on quadruplex stability. Structures with loops containing T, C or 1',2'-dideoxyribose (Φ) are the most stable (Φ > T > C) and these are about 20 °C more stable than those with A in the loops. Each A-containing loop decreases the T_m by about 8 °C. The size of the base may be significant as purine bases (AAA) produce less stable complexes than pyrimidines (TTT and CCC). Comparison of the thermodynamic properties of these sequences reveal that the most stable structures have higher enthalpies which are compensated by less favourable entropic contributions (more

negative ΔS). Surprisingly the most stable complex is produced with 1',2'-dideoxyribose linkers, suggesting that specific interactions with the loop bases are not necessary for forming these intramolecular quadruplexes. Interactions between double-chain reversal loop bases and the G-tetrad edge have previously been observed; single nucleotide adenine loops spanning two tetrads can form hydrogen bonds with the exposed G-quartet edge, resulting in A·(G·G·G·G) pentads (Zhang *et al.*, 2000; Phan *et al.*, 2005) or even A·(G·G·G·G)·A hexad alignments (Kettani *et al.*, 2000). Structural NMR data regarding single nucleotide loops that span three tetrads suggest they are solvent exposed, pointing away from the quadruplex scaffold rather than forming interactions with the external face of the G-quartets (Phan *et al.*, 2004; Ambrus *et al.*, 2005; Phan *et al.*, 2007) (see also Figure 5.1). Based on the finding that 1',2'-dideoxyribose loops form the most stable structures, and adenine loops are destabilising, it appears that for these sequences there is little or no contact between the loop base and the G-quartet edges. The inclusion of the non-nucleosidic linker propanediol (which is the same length as a single nucleotide) in all three loop positions results in a structure of similar stability to G_3A (Risitano and Fox, 2004). It is likely the increased flexibility of this loop when compared to 1',2'-dideoxyribose reduces the stabilising impact of the loop. It should however be noted that the oligonucleotide G_3Prop used by Risitano and Fox lacked a 5' flanking T compared to current sequences which contained both 3' and 5' flanking T's. An alternative explanation is that the structure of the unfolded state may itself be sequence dependent. Modifying the base composition could also affect any structure assumed by the unfolded oligonucleotides via base stacking interactions. It is known that adenine stacks particularly well at the end of duplexes (Bommarito *et al.*, 2000) and any base stacking interactions will need to be reversed in order to form a quadruplex. Studies with longer loops have also observed the stacking of adenine against the terminal quartets (Phan *et al.*, 2004; Ambrus *et al.*, 2005).

Structural studies of various quadruplex-forming sequences have shown several examples of single nucleotide loops that contain T, C or A residues (Table 4.8). Bioinformatic searches have revealed that single nucleotide loops are extremely common in putative quadruplex forming sequences, and short loops appear to be enriched in PQS located within gene promoter regions (Huppert and Balasubramanian, 2007). Amongst these sequences, single base A and T loops appear more frequently than C or G (Todd *et al.*, 2005). It is therefore clear that the lower stability of quadruplexes with A-containing loops does not preclude intramolecular quadruplex

formation, but the inclusion of A instead of T or C may moderate the stability, thereby permitting the interconversion between different structural forms. This may be a quadruplex-duplex equilibrium or conversion between different quadruplex structural types.

5.4.2 Melting studies: *c-kit*

Comparing the results from the model G_3N series with the loop modified *c-kit87up* sequences revealed the same trends. A single A-to-T substitution resulted in quadruplex stabilisation by around 8 °C, typical of a pyrimidine-purine switch, although the C9-A substitution reduced stability by only 4 °C, slightly less than expected from the model sequences. The groove spanned by the single base loop C9 is unusual, as the G-tetrad core is interrupted, which may account for this difference. Sequences containing single nucleotide loops with T or C exhibit similar stabilities and this is reflected in the melting data for Mod 3. In light of the stabilising effect of the A5-T substitution, this modification was maintained for the polymerase inhibition and luciferase expression assays to improve the stability of the quadruplex. It is perhaps noteworthy that informatic searches for structures analogous to the *c-kit87up* quadruplex within the genome, revealed 97 % of the identified sequences (when varying the non-integral residues 5, 9 and 11) also contained thymine at position 5 (Todd *et al.*, 2007).

5.4.3 Polymerase stop and expression studies

When incorporated with the longer promoter sequence, *c-kit87up* still showed evidence of quadruplex formation, as indicated by the potassium induced polymerase pause site, although the effect of increasing quadruplex stability via the A5-T mutation showed a minor increase in the band intensity. In the presence of sodium ions, which produce a structure with a much lower stability, no polymerase pause sites were observed. The results from the luciferase expression studies reiterated the importance of Sp1 binding for *c-kit* promoter activity (Yasuda *et al.*, 1993; Park *et al.*, 1998), as the truncated promoter constructs lacking the Sp1 binding site showed virtually no activity. Nonetheless, the proximity of both *c-kit21* and *c-kit87up* sequences to the important Sp1 binding site raises the possibility that quadruplex formation could contribute to promoter regulation via a blocking mechanism similar to that proposed in the VEGF promoter (Sun *et al.*, 2005). Expression studies with the full 165 bp promoter constructs

revealed far higher promoter activity levels. The scrambled quadruplex-forming sequence showed a clear increase in expression levels relative to the wild-type suggesting the formation of a quadruplex can repress luciferase transcription. Further controls need to be carried out to substantiate the current evidence. Introducing a quadruplex-specific ligand is one way in which quadruplex-formation could be directly linked to the biological response, with the cationic porphyrin $TmPyP_4$ commonly used in quadruplex-related expression assays (Siddiqui-Jain *et al.*, 2002; Cogoi and Xodo, 2006; Qin *et al.*, 2007), although a ligand specific to *c-kit* quadruplex formation would be desirable (Bejugam *et al.*, 2007). Despite an increase in the thermodynamic stability, the single nucleotide loop-modified sequence A5-T (Point) gave very similar expression levels to the wild-type sequence. Within the levels of sensitivity in these experiments there is a negligible change in biological activity.

5.4.4 Future considerations

In addition to the current efforts aimed at investigating the role of quadruplex formation in the *c-kit* promoter, this sequence could also be used in a more generic assay to assess the ability of other G-quadruplex forming sequences to inhibit transcription factor binding. The importance of the Sp1 binding site for *c-kit* gene expression is well documented (Park *et al.*, 1998; Lécuyer et *al.*, 2002), therefore this could become the basis for a quadruplex-related expression assay. For example what would happen if the *c-kit87up* and/or the *c-kit21* sequence were replaced with G_3T, an extremely stable quadruplex? And how would sequence alterations e.g. loop-length / sequence affect the outcome? Not only must the quadruplex be of sufficient stability to show preference over duplex formation, but the final folded topology may also be important (Siddiqui-Jain *et al.*, 2002). Although these would not be biologically relevant sequences, it could provide a proof-of-principle linking quadruplex formation to gene transcription.

CHAPTER 6

Effect of G-tract Length on the Stability and Topology of Intramolecular G-quadruplexes

6.1 Introduction

The simplest requirements for the folding of a unimolecular G-quadruplex can be described as $G_nL_aG_nL_bG_nL_cG_n$ where n represents the length of the G-tracts and L_{a-c} the number and identity of the nucleotides within each loop. The importance of the loop regions in influencing quadruplex topology and stability is well documented, however there has been little investigation on how formation is affected by the length of the guanine-tracts. This chapter examines how G-tract length affects the folding and stability of unimolecular quadruplexes.

Within a G-quadruplex, the core thermodynamic stability is generated by the formation and stacking of the G-quartets, with an enthalpy per quartet estimated to be between -15 and -25 kcal/mol (Pilch *et al.*, 1995; Smirnov and Shafer, 2000). Hence a simple prediction may be that as the length of the G-tract increases, the stability of the structure does likewise. For intermolecular quadruplexes this may be true (Mergny *et al.*, 2005; Petraccone *et al.*, 2005), and extended runs of guanines have been shown to form exceptionally stable structures (Marsh and Henderson, 1994). However for an intramolecular structure, as the number of G-tetrads increases, the distance spanned by the adjoining loops may also increase, which in turn could affect the folding pattern. Work on the thrombin-binding aptamer, which is based on two stacked G-quartets, has shown that the addition of an extra quartet increased stability by 15 kcal/mol, raising the melting temperature by 18 °C (Smirnov and Shafer, 2000; Olsen *et al.*, 2006). In contrast, the *Tetrahymena* telomeric repeat $d(G_4T_2)_n$, which differs from the human telomeric repeat sequence (containing three G-quartets) by exchanging A for G in each repeat, could in principle form four G-quartets linked by two-nucleotide loops. However, although this is observed for the intermolecular dimer (Phan *et al.*, 2004), the intramolecular complex instead folds to form only three quartets with variable loops

consisting of GTTG, TTG and TT (Wang and Patel, 1994). This slipped structure accommodates the additional guanines in the loops, resulting in G·T wobble base pairing. A further consideration with sequences containing extended runs of guanines is the potential for higher order complex formation, as minor changes in base sequence and/or changing the identity of the counterion can cause a conformational switch from intra to intermolecular quadruplex folding (Chen, 1992; Dai *et al.*, 1995; Marotta *et al.*, 1996). Oligomers with the sequence $G_xT_2G_y$ have been shown to form high molecular weight assemblies when G_y contains four or more bases (Marotta *et al.*, 1996).

6.1.1 G-tract length

Reported structures of quadruplexes formed from G-tracts of two guanines are relatively uncommon; the thrombin-binding aptamer is the perhaps the most extensively studied (Bock *et al.*, 1992; Schultze *et al.*, 1994) although trinucleotide repeats of $d(NGG)_n$ have also been shown to form tetrahelical structures (Fry and Loeb, 1994; Matsugami *et al.*, 2001). Repeating units of three or four guanines are more common both in telomeric sequences and wider genomic contexts, although quadruplexes formed from longer stretches of guanine repeats have been observed (Murchie and Lilley, 1992; Qin *et al.*, 2007). This chapter systematically explores the effects of increasing the length of the G-tracts on the folding and stability of a series of model G-quadruplex-forming sequences.

6.2 Experimental Design

In this study, two series of oligonucleotides have been designed containing four G-tracts of equal length between 2 and 7 nucleotides (Table 6.1). Continuing the study of quadruplexes containing single nucleotide linker regions, the first set of sequences contains G-tracts separated by single thymidines (G_nT), with the second series containing two thymidines in the linker regions (G_nT_2). The global structures adopted by these two series of sequences have been probed by circular dichroism and PAGE in the presence of sodium and potassium ions, and the influence of the method of sample preparation (i.e. slow / fast anneal) is assessed. The thermal stability of each of the complexes has been determined by fluorescence melting studies. Quadruplex stability may be an indicator of biological activity, however the formation of quadruplex structures *in vivo* (in non-telomeric contexts) will depend on the relative stability of the

quadruplex and its corresponding duplex. The stability of some of the sequences has therefore been compared with that of the DNA duplexes that are formed on addition of their complementary C-rich oligonucleotide.

6.3 Results

The biophysical properties of the sequences in Table 6.1 have been assessed in order to evaluate their stabilities, global topology (i.e. strand polarities, potential number of G-quartets, molecularity) and the structural equilibrium between duplex and quadruplex.

	Oligonucleotide Sequence
Name	G_nT Series
G_2T	d-F-TGGTGGTGGTGGT-Q
G_3T	d-F-TGGGTGGGTGGGTGGGT-Q
G_4T	d-F-TGGGGTGGGGTGGGGTGGGGT-Q
G_5T	d-F-TGGGGGTGGGGGTGGGGGTGGGGGT-Q
G_6T	d-F-TGGGGGGTGGGGGGTGGGGGGTGGGGGGT-Q
G_7T	d-F-TGGGGGGGTGGGGGGGTGGGGGGGTGGGGGGGT-Q
	G_nT_2 Series
G_2T_2	d-F-TGGTTGGTTGGTTGGT-Q
G_3T_2	d-F-TGGGTTGGGTTGGGTTGGGT-Q
G_4T_2	d-F-TGGGGTTGGGGTTGGGGTTGGGGT-Q
G_5T_2	d-F-TGGGGGTTGGGGGTTGGGGGTTGGGGGT-Q
G_6T_2	d-F-TGGGGGGTTGGGGGGTTGGGGGGTTGGGGGGT-Q
G_7T_2	d-F-TGGGGGGGTTGGGGGGGTTGGGGGGGTTGGGGGGGT-Q
	C_nA Complementary strands
C_3A	d(ACCCACCCACCCACCCA)
C_4A	d(ACCCCACCCCACCCCACCCCA)
C_5A	d(ACCCCCACCCCCACCCCCACCCCCA)
C_6A	d(ACCCCCCACCCCCCACCCCCCACCCCCCA)
C_7A	d(ACCCCCCCACCCCCCCACCCCCCCACCCCCCCA)

Table 6.1 Oligonucleotide sequences used in this work. F = FAM, Q = Dabcyl

6.3.1 Circular Dichroism

6.3.1.1 G_nT series

The CD spectra of the oligonucleotides with the sequences $d(G_nT)_4$ (n = 2-7) in the presence of potassium or sodium are presented in Figure 6.1. Oligonucleotides were prepared by heating to 95 °C before slowly cooling to 5 °C over a period of 10 hours. Heating the samples and annealing again at faster rates yielded similar spectra, outlining the reversibility of complex formation. In the presence of potassium, all the sequences display a major positive peak at 260 nm with a minimum at around 240 nm, suggesting that they adopt a predominantly parallel-stranded configuration. The signal intensity for G_2T is significantly lower than for each of the other oligonucleotides, suggesting it is largely unstructured at room temperature. G_3T, as shown in the previous chapters, exhibits a parallel-stranded conformation and G_4T is also exclusively parallel. G_5T, G_6T and G_7T all display a second positive peak at 295 nm which is greatest for G_6 and G_7. This second peak implies that these sequences may exhibit structural polymorphism with a significant fraction adopting an antiparallel conformation. These spectra may therefore represent the average or collective conformation of the folded species present. Conversely, these spectra are also typical of mixed parallel/antiparallel topologies (Ambrus *et al.*, 2005; Dai *et al.*, 2006) and could also represent the formation of a single mixed strand-polarity structure.

In the presence of sodium, G_2T appears largely unstructured. G_3T is again exclusively parallel with a major peak at 260 nm, however the rest of the sequences show significant ellipticity at 295 nm indicative of some antiparallel nature. These spectral signatures confirm that each of the oligonucleotides folds to form G-quadruplexes (with the exception of G_2T in sodium), but they suggest that the structures may be polymorphic. The dominant peak for all of the sequences is at 260 nm, which indicates they are predominantly parallel, however there are significant peaks at 295 nm which would suggest they may co-exist with the antiparallel form. Comparing these results with previously published data (Hazel *et al.*, 2004), it can be seen that the addition of extra guanines in the stack has a different effect to adding Ts to the loops. $d(G_3T_3)_4$ and $d(G_5T)_4$ both have a six-nucleotide repeat length, yet in 100 mM potassium $d(G_3T_3)_4$ adopts an antiparallel structure whereas $d(G_5T)_4$ is largely parallel.

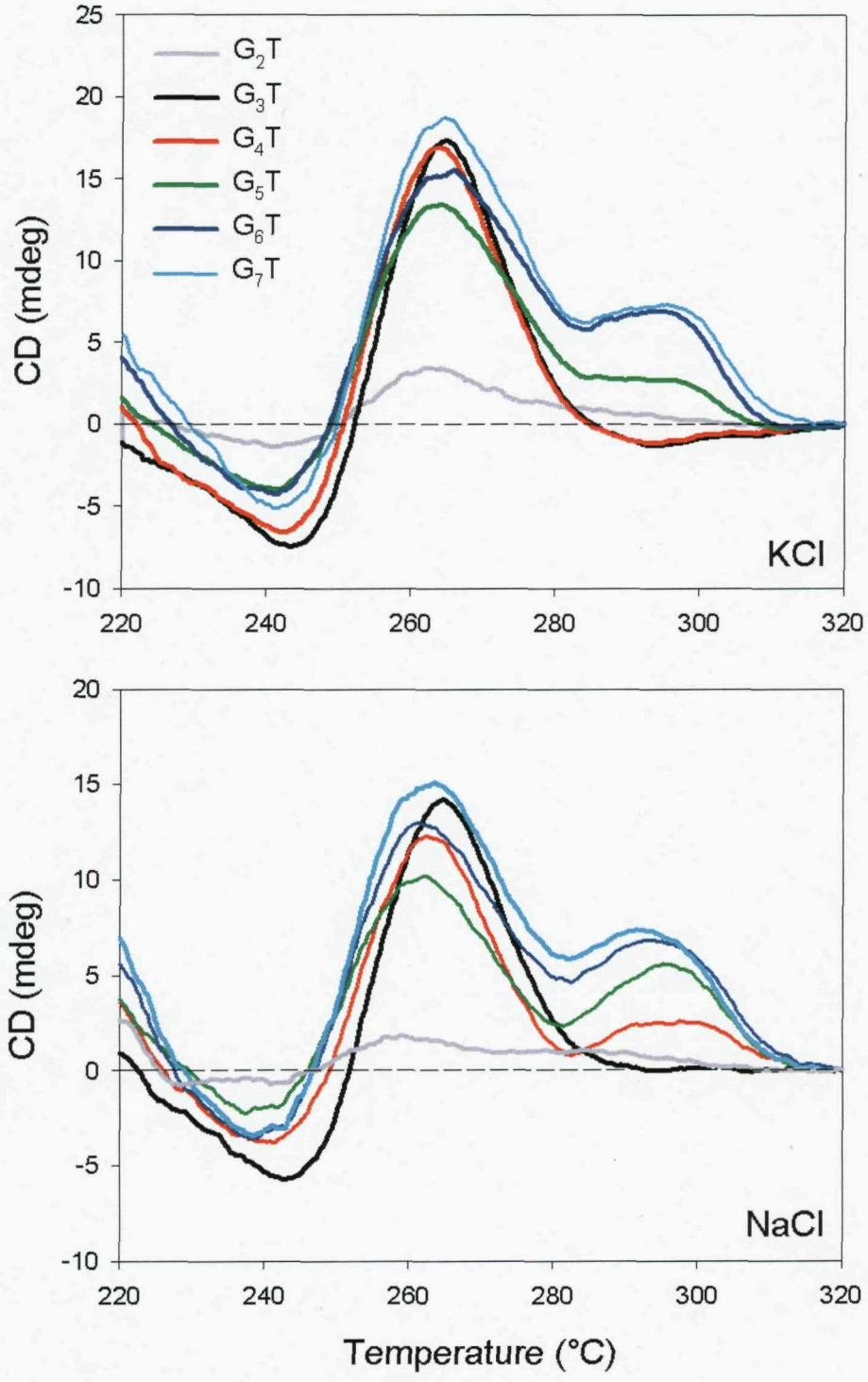


Figure 6.1 CD spectra of the fluorescently labelled oligonucleotides of the series d(G_nT) where n = 2-7. Oligonucleotides (5 μM) were dissolved in 10 mM lithium phosphate (pH 7.4) containing 100 mM potassium chloride (top) or 100 mM sodium chloride (bottom) traces.

6.3.1.2 G_nT_2 series

The CD spectra of the G_nT_2 series of oligonucleotides in the presence of sodium or potassium are presented in Figure 6.2, and these reveal more varied spectra than those observed for the G_nT series. In the presence of potassium, these sequences are more sensitive to the method of annealing, as denaturing the samples and then cooling them rapidly resulted in significantly different spectra (inset panel) to when then the complexes were slowly annealed. Rapidly cooled samples exhibit a greater proportion of the parallel folded form. For the slowly annealed complexes, G_2T_2, like G_2T, displays a much weaker signal than each of the other oligonucleotides, although positive spectral peaks are observed at both 265 nm and 295 nm. In the presence of potassium, G_3T_2 exhibits a major peak at 260 nm indicative of a parallel-stranded structure. G_4T_2, G_6T_2 and G_7T_2 show a similar strength peak at 265 nm, with a second peak at 295 nm which increases in strength with the length of the G-tracts. G_7T_2 displays similar peak intensities at both 260 and 295 nm. The CD spectrum of G_5T_2 is different from each of the other sequences, as it appears to have a predominantly antiparallel fold. The preference for the antiparallel fold in sodium-stabilised quadruplexes is evident for the G_nT_2 series (Figure 6.2 bottom panel), as the dominant spectral peak in each case is at 295 nm with the exception of G_3T_2. The spectra for G_4T_2, G_6T_2 and G_7T_2 are again all very similar but G_5T_2 displays a typical antiparallel spectrum with a distinct negative peak at 265 nm which is absent in the other spectra. These spectra were also independent of the rate of cooling.

6.3.1.3 Comparing the CD spectra of fluorescently and non-fluorescently labelled oligonucleotides

In light of the polymorphic nature of a number of these sequences, there was a concern that the conformational equilibrium may be influenced by the presence of the fluorophore and quencher attached at the termini of the oligonucleotides. In order to assess the affect of the fluorescent groups on the topology of the quadruplexes formed, three of these sequences were synthesised without the fluorophore and quencher attached. G_3T, G_5T and G_5T_2 were chosen due to their varied spectra, with examples of parallel, antiparallel and mixed quadruplex conformations observed. The top panels of Figure 6.3 compare the CD spectra of the fluorescently labelled sequence G_3T, with the same sequence with the fluorescent groups removed. In both sodium and potassium, the

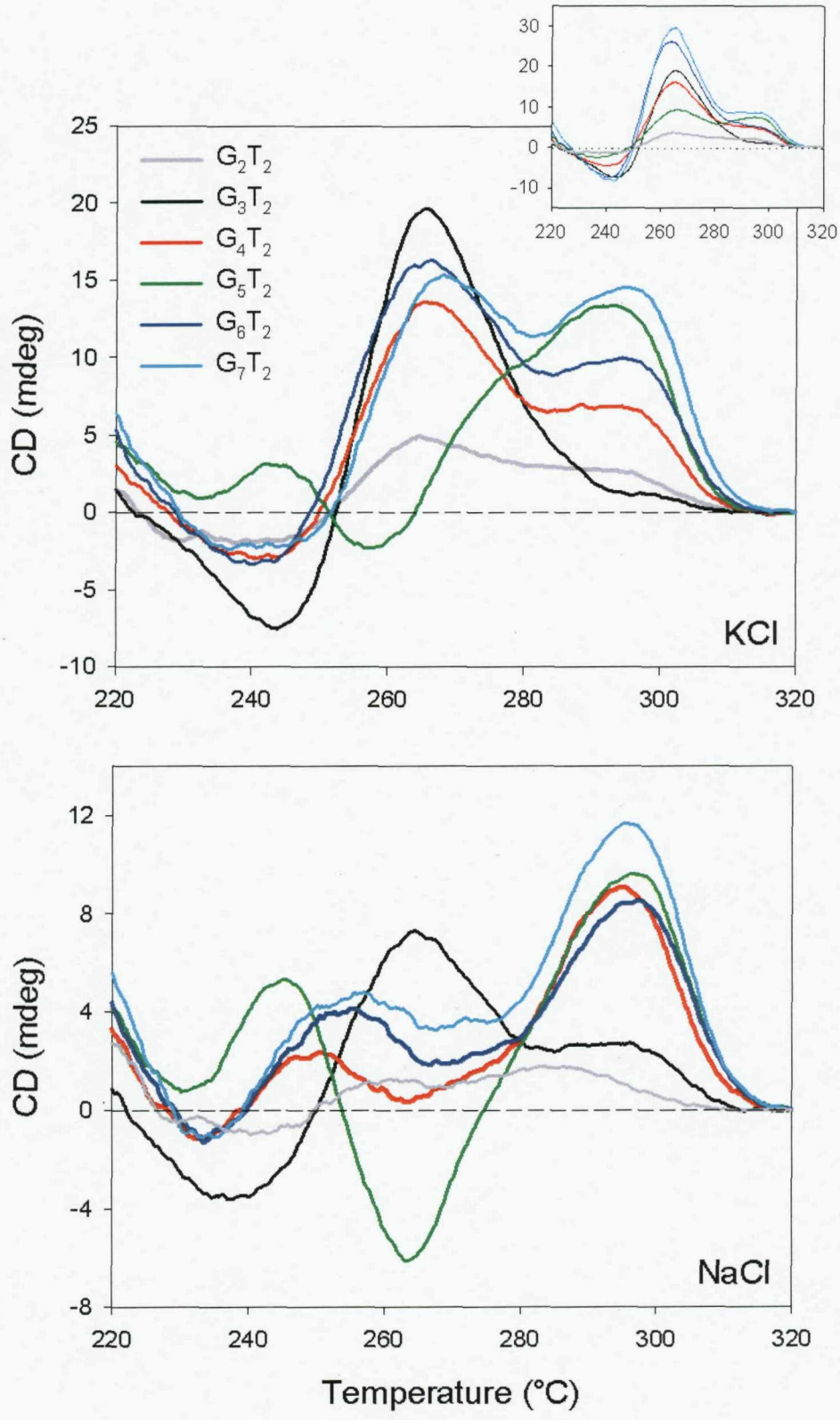


Figure 6.2 CD spectra of the fluorescently labelled oligonucleotides of the series $d(G_nT_2)$. Oligonucleotides (5 μM) were dissolved in 10 mM lithium phosphate (pH 7.4) containing 100 mM potassium chloride (top) or 100 mM sodium chloride (bottom) traces. Inset panel shows the same series of sequences (100 mM KCl) when prepared by rapid cooling.

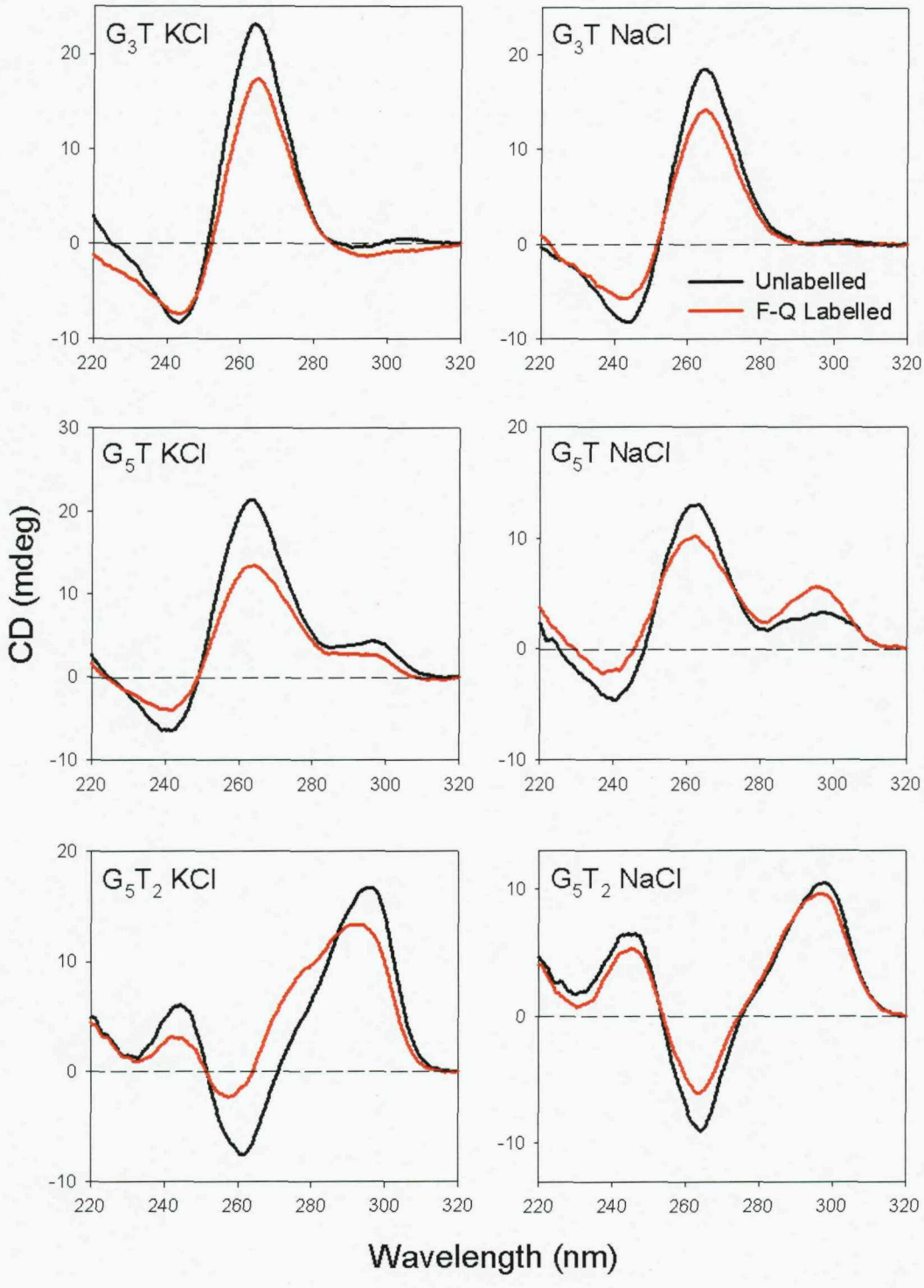


Figure 6.3 CD spectra comparing fluorescently labelled oligonucleotides (red) with the same non-fluorescently labelled sequences (black). Oligonucleotides (5 μM) were dissolved in 10 mM lithium phosphate (pH 7.4) containing 100 mM potassium chloride (left-hand panels) or 100 mM sodium chloride (right-hand panels).

spectral signatures are identical with a positive peak at 265 nm, although the signal intensities are lower for the fluorescently labelled oligonucleotide. This lower signal intensity is a trend observed for all of the fluorescent oligonucleotides, and may be a result of the destabilising influence of the fluorescent groups (Chapter 3; Mergny and Maurizot, 2001). G_5T in potassium is also a close match, however in sodium the fluorescently labelled sequence has an enhanced peak at 295 nm. G_5T_2 in both sodium and potassium show spectra indicative of an antiparallel strand arrangement, although the fluorescently labelled oligonucleotide in potassium contains a shoulder at 275 nm. In general, the spectral peaks of the fluorescently labelled oligonucleotides are a close match to those of their non-labelled counterparts, although the signal intensities are consistently lower. There are however some differences, most notably G_5T in sodium and therefore in some cases, where mixed populations are present, the presence of fluorescent groups may have some influence on the topology.

6.3.2 Fluorescence melting curves

The thermal stability of the two series of oligonucleotides was then assessed by fluorescence melting. Typical fluorescence melting curves for these oligonucleotides are shown in Figure 6.4 and 6.5, and the T_m and ΔH values determined at different ionic concentrations are summarised in Tables 6.2 and 6.3. Each sample was annealed and then melted at a rate of 0.1 °C.min^{-1}. This rate of temperature change was slower than had been used in previous chapters, as preliminary melting analyses revealed evidence of hysteresis in a number of the longer oligonucleotides, and therefore slower rates were employed.

With the exception of G_2T, all the sequences show fluorescence changes that are consistent with quadruplex formation and they are all much more stable in the presence of potassium than sodium. Sequences that contain G_2-tracts, i.e. G_2T and G_2T_2, did not show melting profiles consistent with quadruplex formation in sodium or potassium concentrations below 100 mM. G_2T displayed no melting transition even in ionic strengths in excess of 0.5 M. However, it is interesting to note that the addition of the quadruplex-stabilising ligand BRACO-19 (Read *et al.*, 1999; Gowan *et al.*, 2002) induced quadruplex formation at concentrations around 5 μM (Figure 6.6). G_2T_2 did show a fluorescence profile consistent with quadruplex formation in potassium concentrations of 100 mM and above, though the T_m values were relatively low (35 °C

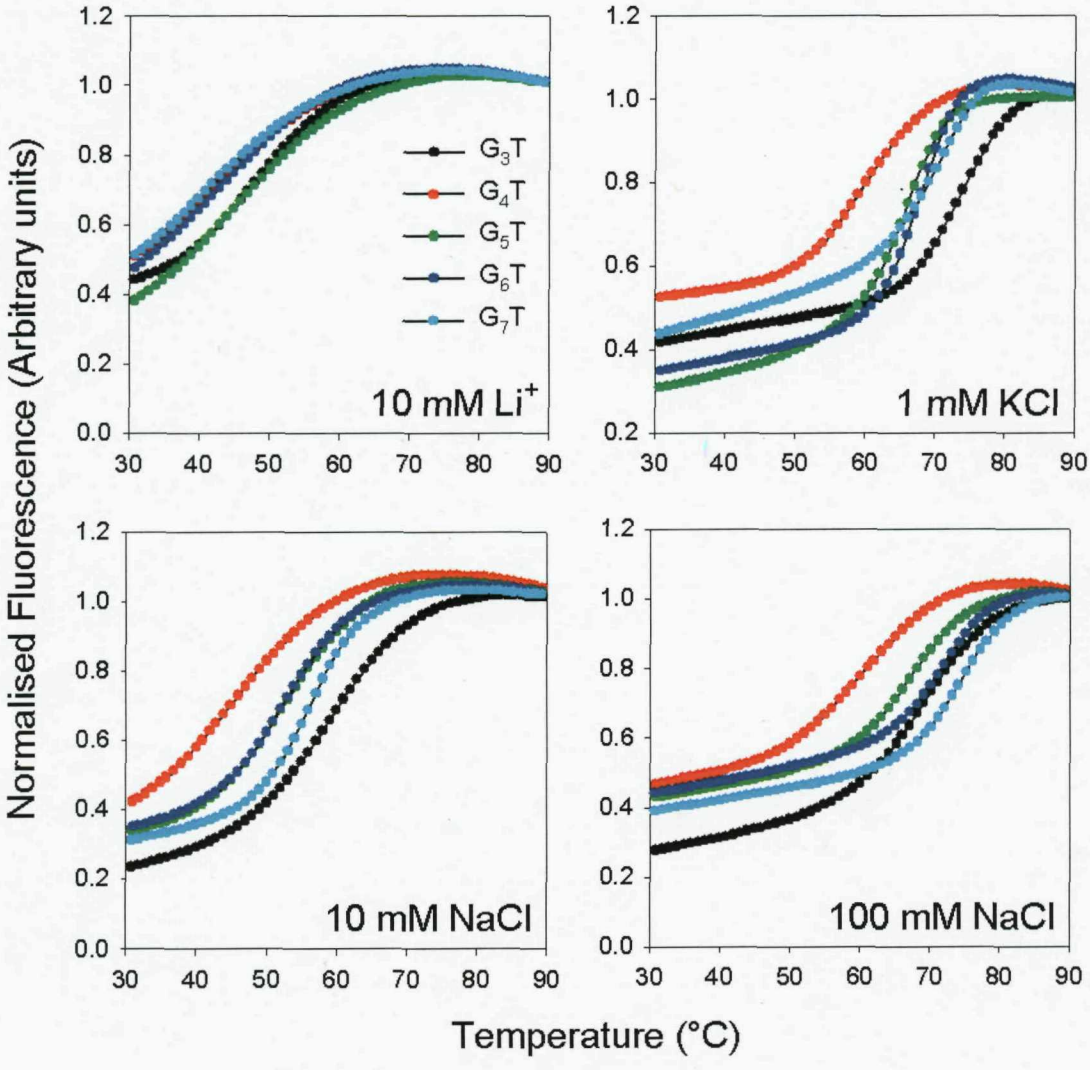


Figure 6.4 Fluorescence melting curves for the oligonucleotides of the series $d(G_nT)_4$. The melting profiles were determined in 10 mM lithium phosphate (pH 7.4) containing different concentrations of potassium or sodium chloride as indicated. The curves have been normalised to the same final fluorescence value; black, G_3T; red, G_4T; green, G_5T; blue, G_6T; cyan, G_7T.

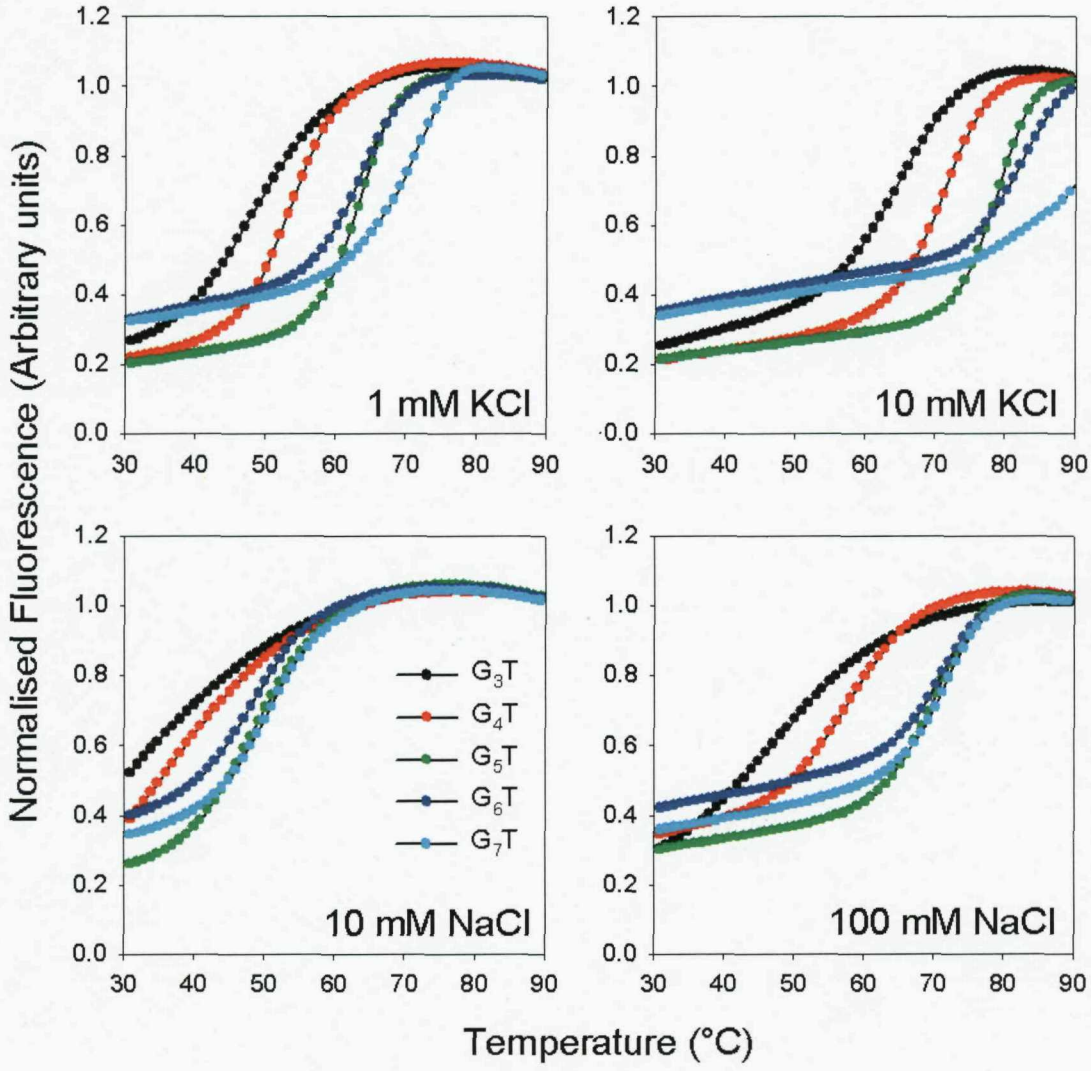


Figure 6.5 Fluorescence melting curves for the oligonucleotides of the series $d(G_nT_2)_4$. The melting profiles were determined in 10 mM lithium phosphate (pH 7.4) containing different concentrations of potassium or sodium chloride as indicated. The curves have been normalised to the same final fluorescence value; black, G_3T_2; red, G_4T_2; green, G_5T_2; blue, G_6T_2; cyan, G_7T_2.

	G_3T		G_4T		G_5T		G_6T		G_7T	
	T_m	ΔH	T_m	ΔH	T_m	ΔH	T_m	ΔH	T_m	ΔH
10 mM Li^+	48.3	122	41.8	102	43.7	104	43.4	111	43.1	100
[KCl] mM										
0.1	58.4	210	46.1	119	50.0	175	51.8	227	54.3	236
1	73.8	271	58.6	200	65.5	272	67.6	322	71.3	326
5	85.4	299	71.3	273	76.4	282	79.1	391	86.5	381
10			76.0	291	80.8	324	83.3	374	88.4	351
Δn	2.35 ± 0.17		2.45 ± 0.32		2.52 ± 0.33		3.34 ± 0.25		3.41 ± 0.48	
[NaCl] mM										
1	52.7	141	44.3	110	46.5	117	47.2	120	47.4	118
10	59.8	177	46.5	118	52.3	179	52.3	177	54.0	213
50	66.0	197	55.5	164	62.8	221	65.4	253	68.7	291
100	69.7	246	60.8	186	67.7	239	71.9	294	75.1	314
200	75.3	198	68.1	202	75.6	214	81.4	340	82.6	346
Δn	1.42 ± 0.41		1.85 ± 0.25		2.02 ± 0.13		4.23 ± 0.50		4.19 ± 0.25	

Table 6.2 Melting temperatures (°C), ΔH values ($kJ.mol^{-1}$) and Δn values derived from fluorescence melting curves for oligonucleotides of the type $(G_nT)_4$ (n = 3-7). The values were determined in 10 mM lithium phosphate pH 7.4 containing different concentrations of sodium or potassium. Each value is the average of four determinations (two melting and two annealing profiles). T_m values are accurate to within 0.5 °C, while ΔH values varied by about 5%. Missing values indicate the complex was too stable to determine the T_m.

	G_3T_2		G_4T_2		G_5T_2		G_6T_2		G_7T_2	
	T_m	ΔH	T_m	ΔH	T_m	ΔH	T_m	ΔH	T_m	ΔH
[KCl] mM										
0.1			36.3	99	47.0	188	46.7	186	51.1*	213*
1	48.1	136	53.3	224	62.9	294	63.0	283	68.3*	289*
5	60.1	199	66.1	263	74.1	359	74.7	345	81.9*	322*
10	65.0	241	71.4	286	78.9	376	80.1	371		
50	76.2	267	82.4	308	89.7	336				
Δn	2.50 ± 0.12		2.48 ± 0.13		2.84 ± 0.30		3.29 ± 0.06		3.07 ± 0.15	
[NaCl] mM										
1							42.5	100	41.4	101
5					45.1	134	47.2	123	47.1	132
10			37.3	97	46.5	157	52.3	155	50.7	168
50	40.3	99	49.9	152	62.3	248	65.4	260	63.7	278
100	46.8	113	57.7	188	70.2	290	71.9	314	71.9	318
200	54.4	144	66.2	225	78.3	323	81.4	354	81.3	334
Δn	1.83 ± 0.28		2.37 ± 0.46		4.21 ± 0.40		4.63 ± 0.54		4.41 ± 0.33	

Table 6.3 Melting temperatures (°C), ΔH values (kJ.mol^{-1}) and Δn values derived from fluorescence melting curves for oligonucleotides of the type $(G_nT_2)_4$ (n = 3-7). The values were determined in 10 mM lithium phosphate pH 7.4 containing different concentrations of sodium or potassium. * indicates hysteresis was observed between the melting and annealing profiles. Each value is the average of four determinations (two melting and two annealing profiles). T_m values are accurate to within 0.5 °C, while ΔH values varied by about 5%. Missing values indicate the complex was either too stable (high ionic strengths) or unstable (low ionic strength) to measure the T_m.

at 100 mM KCl) and the transitions broad. Both G_2T and G_2T_2 were therefore omitted from further melting curve analysis.

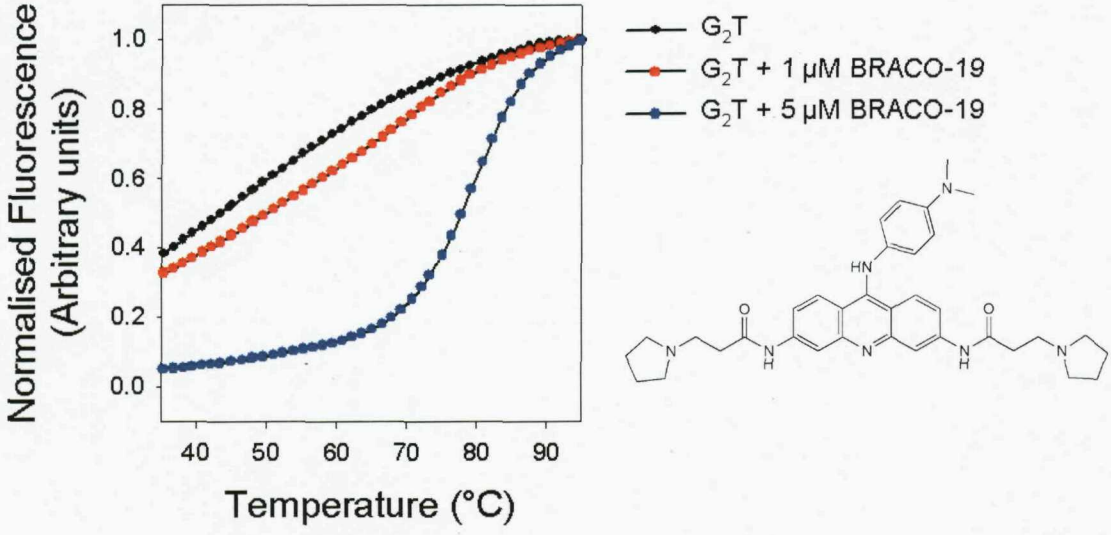


Figure 6.6 Fluorescence melting profiles for the oligonucleotide G_2T in the absence and presence of the 3,6,9-trisubstituted acridine, BRACO-19 (shown right). The oligonucleotide (0.25 µM) was dissolved in 10 mM lithium phosphate (pH 7.4) containing 50 mM KCl and either 0, 1 or 5 µM BRACO-19.

6.3.2.1 G_nT series

Figure 6.4 (top-left panel) shows the melting transitions of each of the remaining G_nT (n = 3-7) series of oligonucleotides in the presence of 10 mM lithium alone. Surprisingly, these sequences still display a melting profile with a T_m of around 45 °C even in the absence of sodium or potassium. A similar effect has been observed with sequences containing G_3-tracts separated by non-nucleosidic loops (Risitano and Fox, 2004). Although this could indicate some quadruplex formation, it may be significant that all of the sequences have very similar T_m and ΔH values, in contrast to their behaviour in the presence of sodium and potassium. The addition of even low concentrations of potassium causes a dramatic increase in quadruplex stability and all the complexes become more stable as the potassium ion concentration is increased. At concentrations above 10 mM KCl, the complexes are too stable to measure. In the presence of 1 mM KCl (Figure 6.4 top right panel), G_3T is the most stable and G_4T the least stable. The stability of the sequences increases with the length of the G-tract ($4 < 5 < 6 < 7$) with the exception of G_3T which is more stable than even G_7T. ΔH values for the transitions were estimated from van't Hoff analysis of the melting profiles (assuming a two-state

equilibrium) and these values are summarised in Table 6.2. The presence of polymorphic quadruplex structures will lead to shallower melting profiles and therefore smaller apparent values for ΔH (Rachwal and Fox, 2007d). This may explain why there appears to be no simple correlation between ΔH and the number of guanine nucleotides in each G-tract, though in general the values are higher for the longer G-tracts (with the exception of G_3T).

Melting profiles in the presence of 10 and 100 mM sodium are presented in the bottom panels of Figure 6.4 and the data summarised in Table 6.2. In contrast to potassium, the addition of 1 mM sodium has only a small effect on the stability of the complexes, and the T_m values increased by only around 3-4 °C relative to that in 10 mM lithium. In the presence of 10 mM sodium, G_3T is again the most stable, with G_4T the least stable. G_5T, G_6T and G_7T have similar T_m values. On increasing the ionic strength to 100 mM sodium, G_7T becomes the most stable, although G_3T is still more stable than each of the other complexes. In general, ΔH values are again higher for longer G-tracts although G_3T is still anomalously high.

6.3.2.2 G_nT_2 series

The thermal melting profiles of the oligonucleotides in which the G-tracts are separated by two Ts are shown in Figure 6.5 and the T_m and ΔH values are summarised in Table 6.3. G_2T_2 is omitted, as it exhibited no melting transitions in the presence of sodium, and only in potassium concentrations in excess of 100 mM. In contrast to the G_nT series, no melting transitions were observed for any of the sequences in the presence of lithium alone. In the presence of either potassium or sodium, the least stable complex is formed by G_3T_2, while G_7T_2 is the most stable. In potassium, the order of stability is $G_3T_2 < G_4T_2 < G_5T_2 = G_6T_2 < G_7T_2$. It should be noted that for G_7T_2 there is some hysteresis between the melting and annealing profiles, indicating the folding of this complex is extremely slow. In sodium, these structures are much less stable with the order of stability $G_3T_2 < G_4T_2 < G_5T_2 = G_6T_2 = G_7T_2$.

6.3.3 Inter or intramolecular?

The oligonucleotides used in this study were designed to form intramolecular quadruplexes, however the presence of extended runs of guanines, particularly in the longer sequences meant it was essential to demonstrate they formed intra and not intermolecular complexes. This was carried out in a number of ways; examining the melting and annealing profiles of each of the oligonucleotides, the traces were fully reversible and reproducible over a number of heating and cooling cycles (with the exception of G_7T_2 in potassium). Higher order quadruplex structures are extremely slow to assemble once dissociated, therefore these profiles are unlikely to be the result of intermolecular quadruplex formation. Secondly, for an intramolecular quadruplex the T_m value should be independent of the oligonucleotide concentration. Figure 6.7 shows the T_m values of each complex determined at concentrations between 0.02 and 5 μM. Within this range, the values are similar to within 0.5 °C. Thirdly, the gel mobility of each of folded complex was examined. The results of gel electrophoresis experiments for each of the oligonucleotides in the presence of sodium or potassium are presented in Figure 6.8. The bands were visualised under UV light as the oligonucleotides were fluorescently tagged. As expected under denaturing conditions, each of the unfolded oligonucleotides ran as a single band with the mobilities dependent on the length of the sequence (Figure 6.8a). G_3T appears to have anomalously fast gel mobility, although this may be due to the complex being partially folded, even under denaturing conditions. Figures 6.8b-e show the gel mobility of the folded structures after they were slowly annealed. The G_nT series in sodium all run as a single band, with G_3T, G_4T and G_5T having similar gel mobility, whereas G_6T and G_7T ran more slowly. In potassium, G_4T and G_5T display two bands, the faster migrating of which appears to be the intramolecular folded species as they co-migrate with G_6T and G_7T. For G_5T a significant fraction appears to be either bi- or tetramolecular. It should be noted that these complexes were prepared with a 20 μM oligonucleotide concentration as compared to 0.25 μM for the melting studies and 5 μM for the CD. In the G_nT_2 series, each complex runs as a single band in both sodium and potassium (although G_7T_2 in potassium is poorly resolved), with no indication of higher order complex formation. The patterns of mobility are however different from those observed in the G_nT series.

These electrophoresis experiments are able to discriminate between structures of different molecularity, but do not seem to be able to distinguish between different

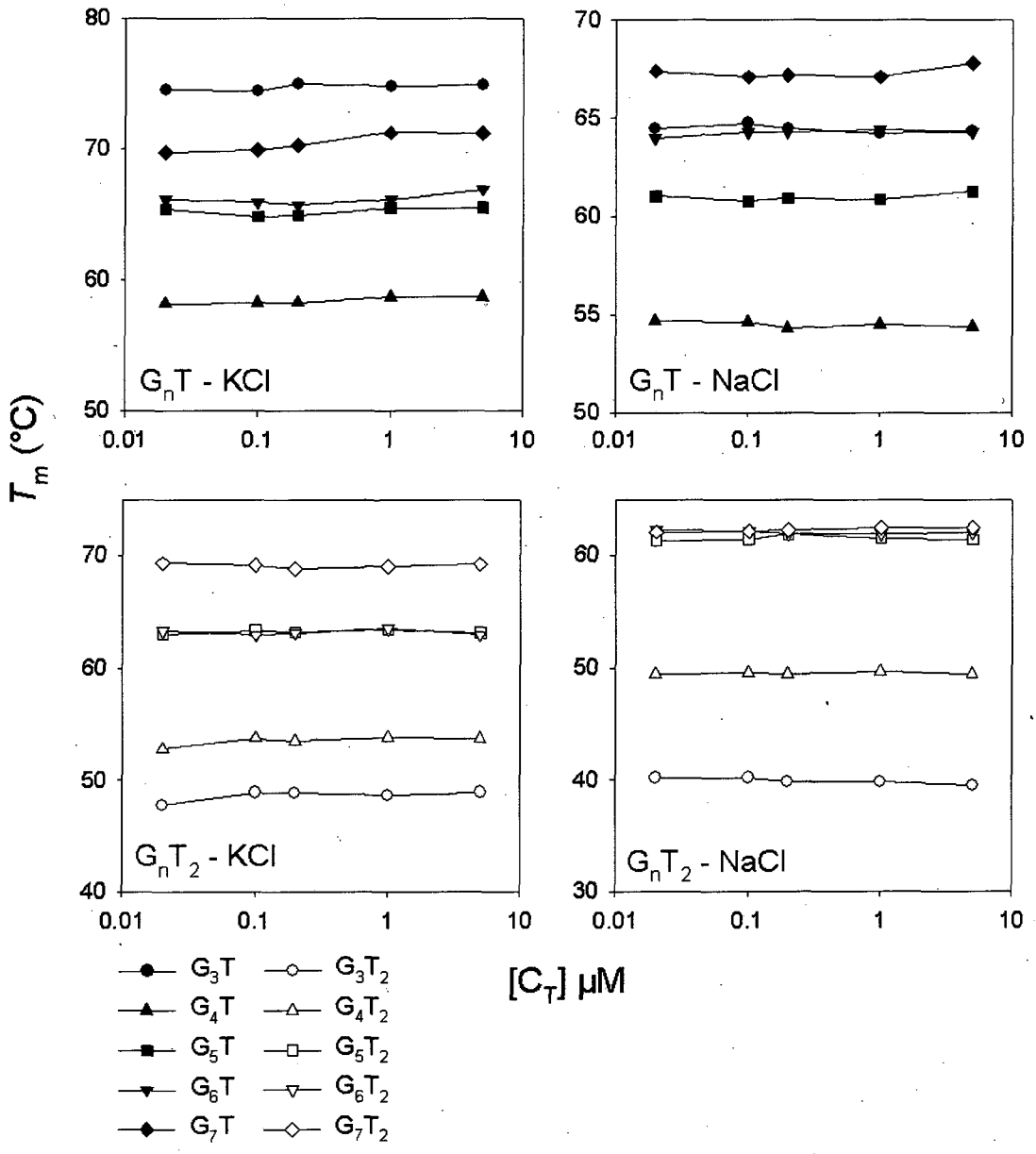


Figure 6.7 Plots of the T_m value dependence on strand concentration for each of the two series of oligonucleotides in the presence of 50 mM sodium (right-hand panels) or 1 mM potassium (left-hand panels).

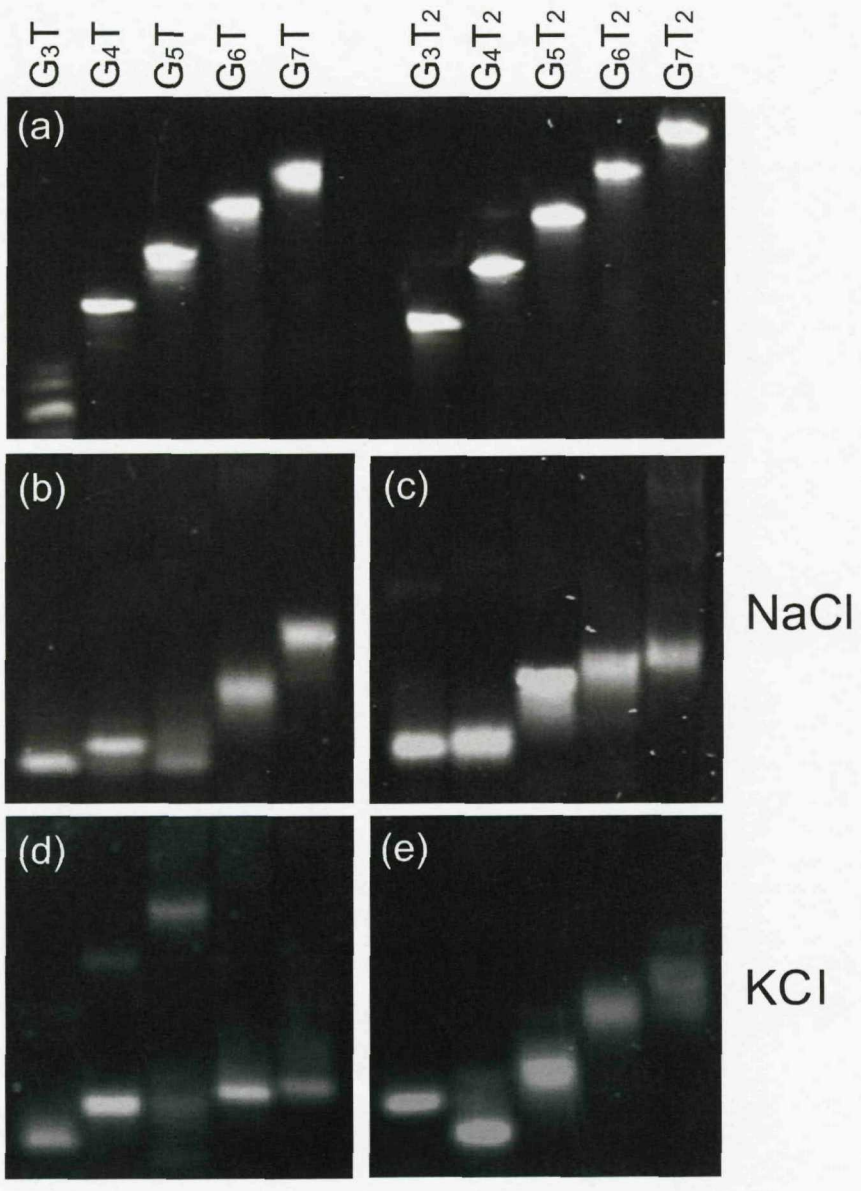


Figure 6.8 Mobility of the fluorescently labelled oligonucleotides in polyacrylamide gels: **(a)** 14% denaturing gel containing 8 M urea. **(b-e)** 16% non-denaturing gels supplemented with 50 mM NaCl (b and c) or 50 mM KCl (d and e).

folded topologies. Results from the CD experiments suggest that for several of the sequences there are a number of different folded conformers in equilibrium. These are not observed in the mobility studies, although if the structures had similar mobilities it would be difficult to discriminate between them.

6.3.4 Ionic strength dependence

The ΔH values for these transitions exhibit a strong dependence on the ionic strength (Tables 6.2 and 6.3), which is consistent with the presence of specific ion binding sites within the quadruplex (Jing *et al.*, 1997). The Δn values for each of the sequences in the presence of sodium and potassium were calculated in order to estimate the specific number of ions bound within each complex (i.e. the difference between the number of ions bound in the folded and unfolded states) and these values are presented in Tables 6.2 and 6.3. Δn values were calculated by plotting the slope of ΔG, derived form the melting curves, verses $\log[M^+]$ as previously described (Cantor and Schimmel, 1980; Jing *et al.*, 1997) (Figure 6.9). Figure 6.10 shows the variation in Δn with the number of guanines in each stack for both series of oligonucleotides. In a structure containing three stacked G-quartets, it could be predicted that the Δn value should be either two (the number of ions located between the stacked quartets) or four (including two additional ions coordinated between the loops and the terminal G-quartets). For a parallel structure a value of two seems more likely, since the loops run parallel to the G-stack and therefore do not interact with the terminal quartets. It would therefore be expected that the addition of a further quartet would increase the value of Δn by one. It may therefore be significant that for the G_nT series in the presence of potassium, G_3T, G_4T and G_5T all have a Δn value of around two, suggesting that they each have three stacked G-quartets. For G_6T and G_7T, Δn increases by about one, suggesting that these complexes contain an additional quartet. The values of Δn exhibit even less variation for the G_nT_2 series in the presence of potassium, again suggesting that there are only three stacked G-quartets. The variations in Δn are more pronounced in the presence of sodium; G_3T, G_4T and G_5T have similar values between 1.5 and 2, which rises to ~4 for G_6T and G_7T. The trend is similar for the G_nT_2 series in sodium, although the transition from a Δn value of 2 to 4 occurs at G_5T_2. It is interesting to note the pattern of Δn with G_n (Figure 6.10) is similar to the pattern of mobility observed for these sequences in the electrophoresis experiments (Figure 6.8), particularly in sodium. These results are considered further in the Discussion.

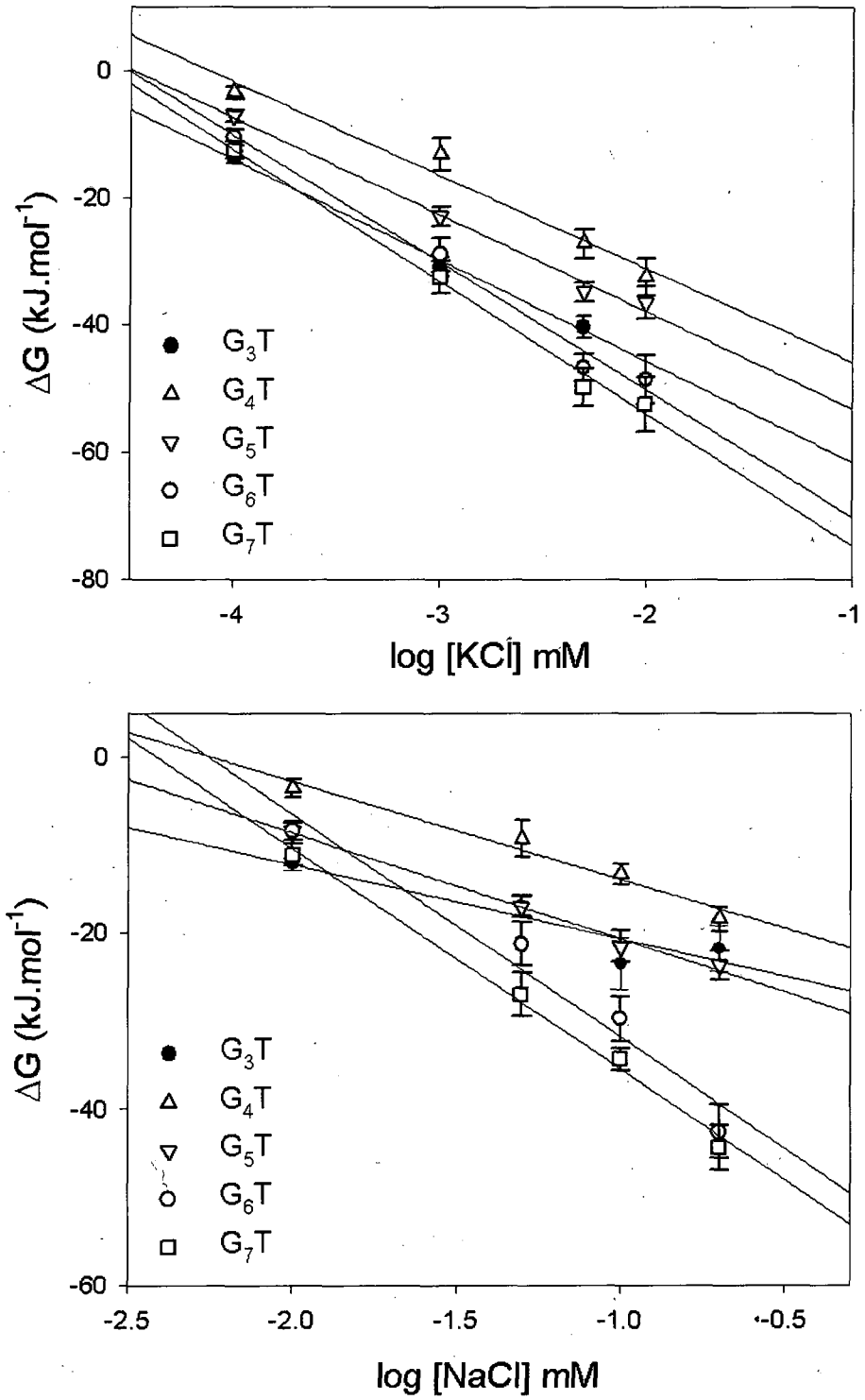


Figure 6.9 Calculated ΔG values for the G_nT series of oligonucleotides plotted against log[KCl] (top panel) or log[NaCl] (bottom panel). These values were fitted to a straight line, yielding a slope of ΔΔG / Δlog[KCl], or Δn. ΔG values were calculated as $\Delta H^*(1\text{-}310/T_m)$, with T_m and ΔH values derived from van't Hoff analysis of the fluorescence melting profiles.

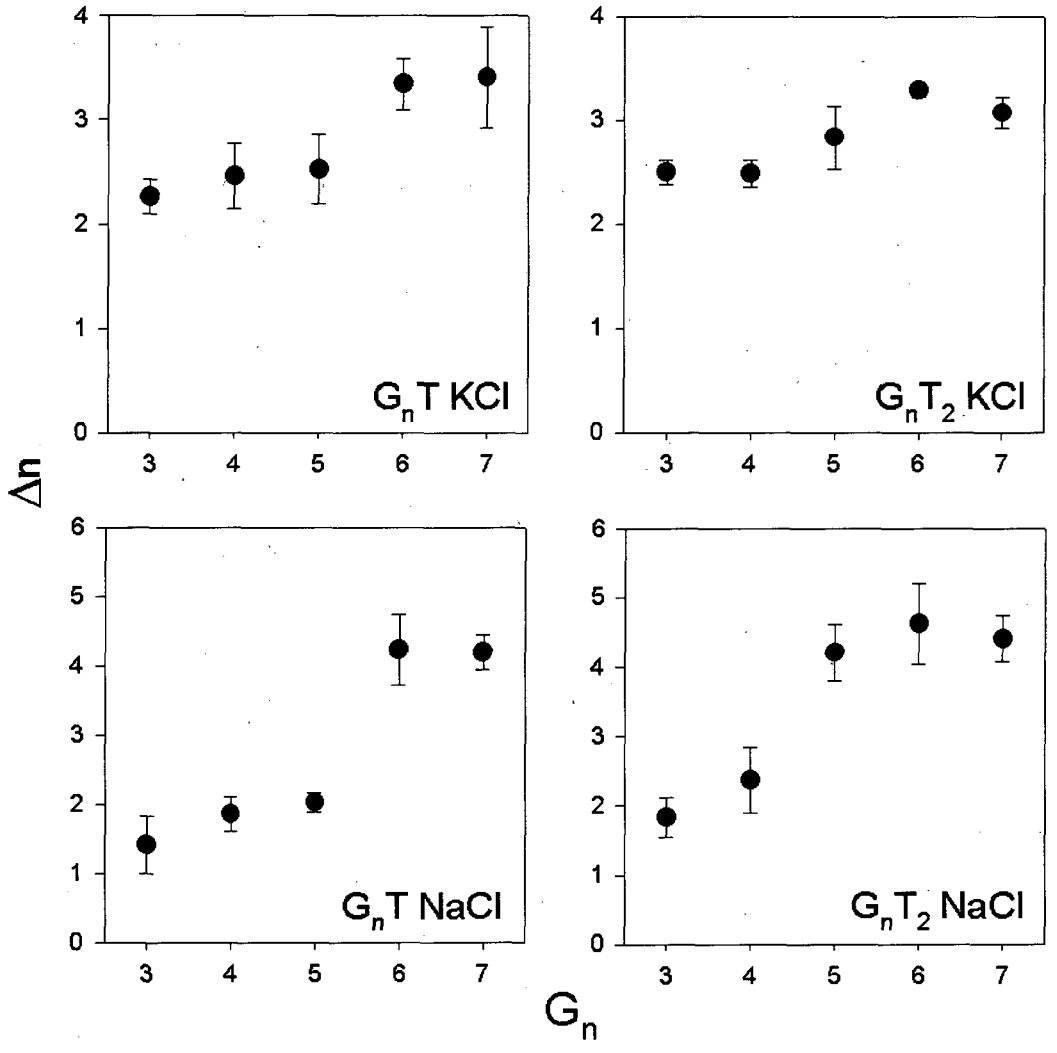


Figure 6.10 Variation in the number of sodium or potassium ions specifically bound within each quadruplex (Δn, y-axis) as a function of the length of the G-tracts (G_n, x-axis). The values for Δn were determined from the dependence of ΔG on ionic strength, as described in the text. Error bars represent the standard error from four determinations, two melting and two annealing.

6.3.5 Comparison of quadruplex stability with duplex stability

Within a genomic context, G-rich DNA sequences with the potential to form quadruplexes will be present alongside their C-rich complementary strand, generating a competition between the Watson-Crick duplex and the Hoogsteen-bonded quadruplex, which will be dependent on their relative stabilities. Fluorescence melting experiments were performed to assess the relative stability of the quadruplex and duplex structures formed by the G_nT series of oligonucleotides. The principle of this assay is outlined in the Methods section (Figure 2.1). Briefly, when the G-rich strand hybridizes with its (unlabelled) complementary strand, the fluorophore and quencher are separated by a large distance producing a large fluorescence signal. When folded as a quadruplex, the fluorescence is quenched, and the fluorescence signal of the random coil is intermediate between the two structural forms. The relative order of fluorescence signal is therefore duplex > quadruplex > random coil.

Fluorescence annealing experiments were performed for of each of the G_nT series of oligonucleotides in the presence of equimolar, 20-fold and 50-fold excess of their complementary strand. Each sample was annealed and then melted at a rate of 0.2 $°C.min^{-1}$ and all profiles were fully reversible and displayed no hysteresis, suggesting the reactions are in thermodynamic equilibrium. Representative annealing profiles for the G_nT series of oligonucleotides are presented in Figure 6.11. Figure 6.11a shows the annealing profiles for G_3T, the most stable of the quadruplex structures studied. In a 1:1 ratio with its complementary strand, G_3T is predominantly in the quadruplex form, as the fluorescence profile is similar to that of the G-rich strand alone. As the concentration of C_3A is increased to a 20-fold excess, an unusual fluorescence profile is observed, which is explained by the competition between the duplex and quadruplex forms. As the samples are annealed, initially the fluorescence decreases as G_3T folds to form a quadruplex between 70 and 80 °C. At lower temperatures (< 50 °C) there is a small increase in fluorescence as C_3A sequesters the dissociated G-rich strand. In a 50-fold excess of the complementary strand the duplex predominates at low temperatures, as the fluorescence level is greater than that of the fully unfolded oligonucleotide (random coil).

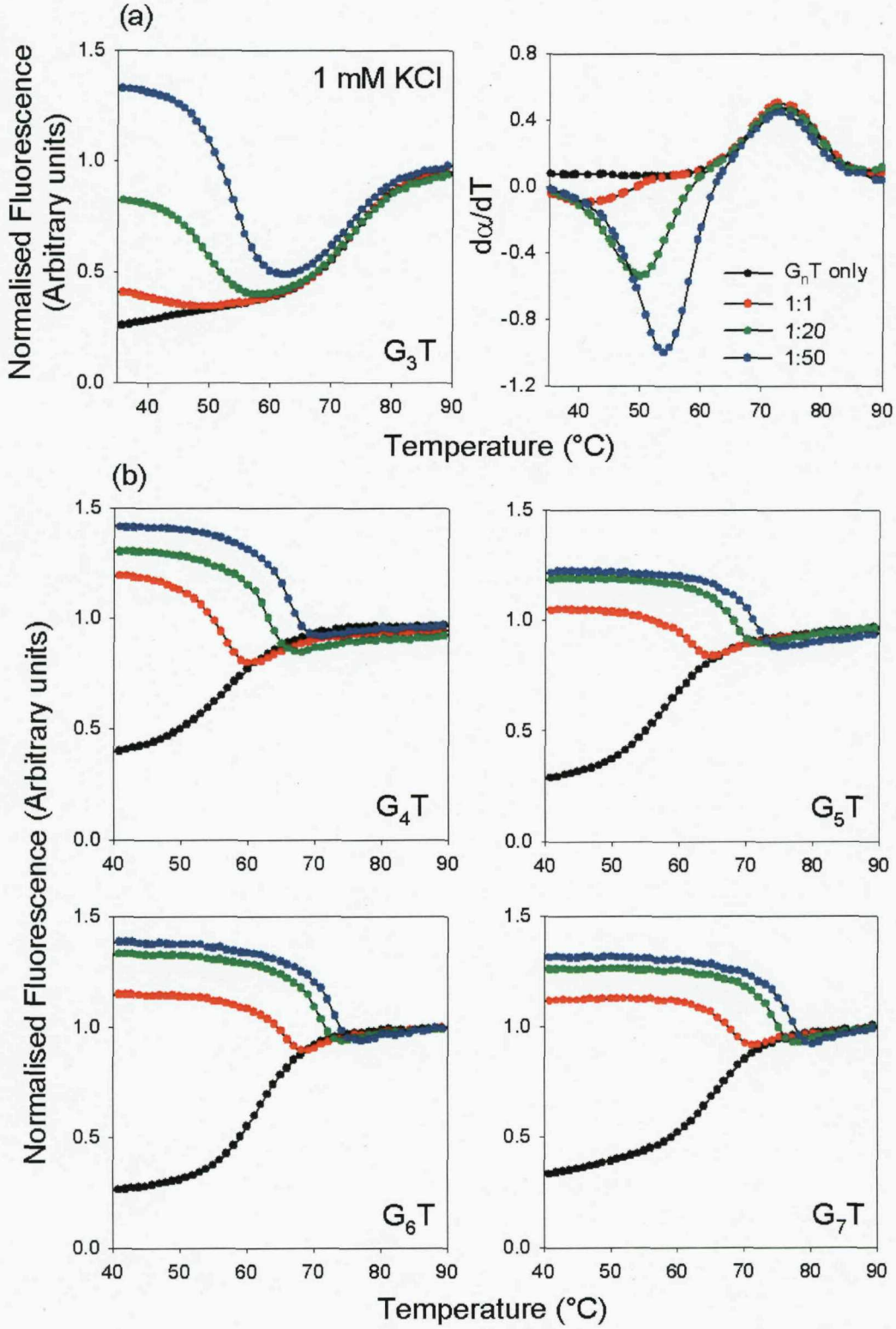


Figure 6.11 (a) Effect of various concentrations of the complementary C-rich oligonucleotide on the fluorescence melting curves of G_3T in the presence of 1 mM KCl. G_3T only (black); 1:1 (G_3T:C_3A) (red); 1:20 (green); 1:50 (blue). Melting and annealing measurements were performed at a rate of temperature change of 0.2 °C.min^{-1}. The first differential of these melting profiles are shown in the right-hand panel. **(b)** Fluorescence melting profiles of each of the other G_nT series of oligonucleotides in the presence of 1 mM KCl and varying concentrations of their complementary strand. G_nT only (black); 1:1 (G_nT:C_nA) (red); 1:20 (green); 1:50 (blue). All melting profiles are normalised to the final fluorescence value.

The panels in Figure 6.11b show that all other sequences in the G_nT series preferentially form duplex over quadruplex in the presence of 1 mM potassium. At equimolar ratios there is a small amount of fluorescence quenching at low temperatures suggesting there is still a small amount of quadruplex formation. The fluorescence values at low temperatures are all greater than that of the final unfolded complex. It is noticeable that as the concentration of C-rich strand is increased, the transition to fully unfolded moves to higher temperatures, as would be expected for an *inter*molecular reaction. This is in contrast to quadruplex formation, which is *intra*molecular and therefore concentration independent.

On increasing the concentration of potassium to 10 mM (Figure 6.12), the equilibrium is shifted in favour of quadruplex formation. Figure 6.12a shows the fluorescence annealing profiles of G_3T on addition of various concentrations of its complement in the presence of 10 mM potassium. The low temperature fluorescence levels are now lower than the unfolded oligonucleotide, even in a 50-fold excess of the complement, suggesting a greater proportion of the DNA has formed quadruplex in favour of duplex. At this raised ionic strength G_3T does not fully melt, as it is too stable, but the fluorescence profiles are still reversible. The preference for quadruplex formation is particularly noticeable in the longer sequences (G_5T, G_6T and G_7T) (Figure 6.12b) where the low temperature fluorescence values are all less than those at high temperatures. In a 1:1 ratio of G-rich to C-rich strand, the quadruplex form appears to predominate at low temperatures for all sequences, with the exception of G_4T, which forms the least stable quadruplex structure in the G_nT series.

In the presence of sodium, all the sequences show a preference for duplex over quadruplex. For G_3T (Figure 6.13a) the quadruplex has a higher T_m than the duplex, though the duplex is still the predominant structural form at low temperatures. Increasing the ionic strength from 10 to 100 mM NaCl has no effect on the structural equilibrium. For each of the other sequences (Figure 6.13b), there is a simple decrease in fluorescence as the temperature is increased, corresponding to a duplex-random coil transition. The quadruplexes formed by these sequences are of lower stability than their duplex equivalents, even at higher (100 mM) ionic strengths.

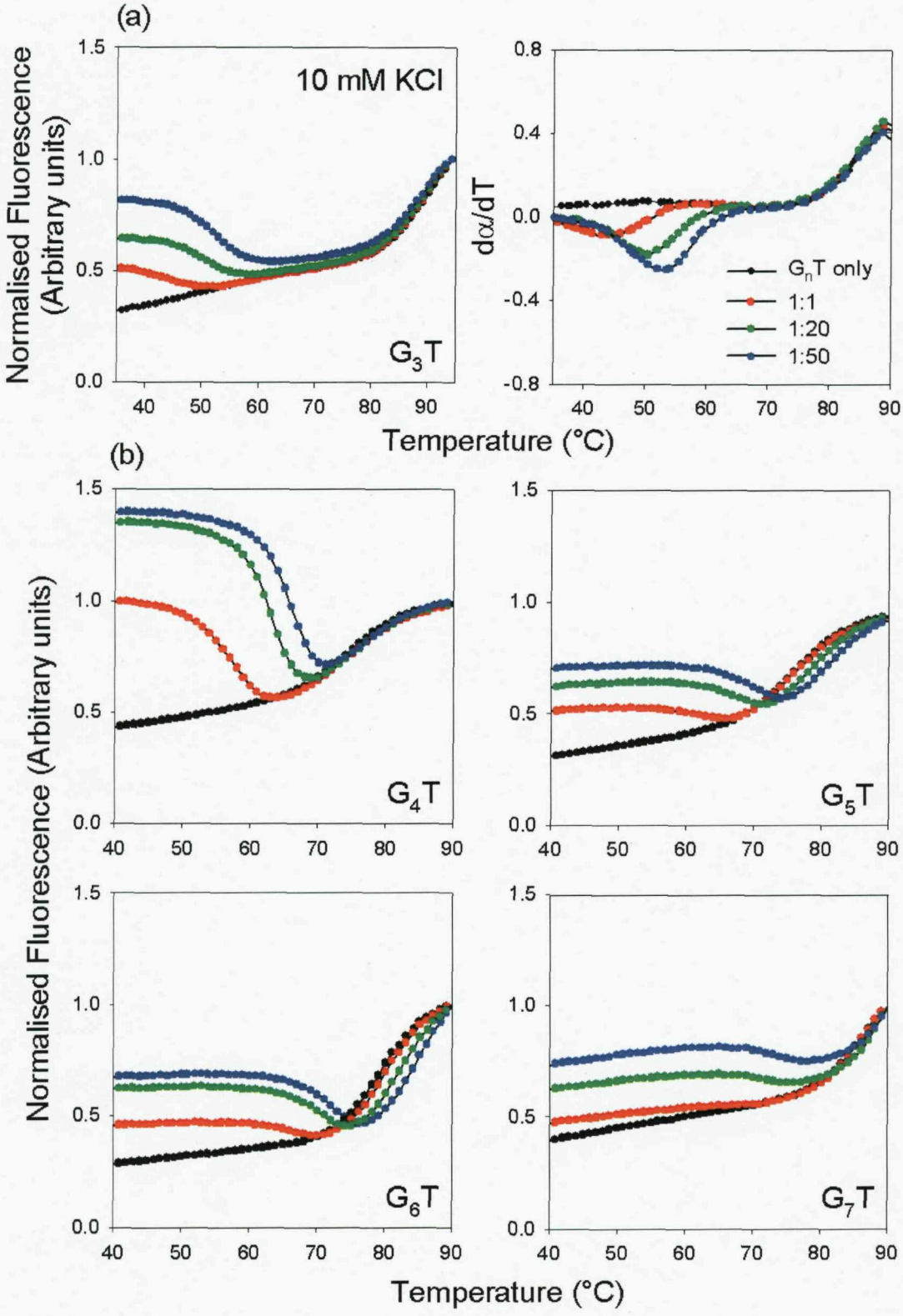


Figure 6.12 (a) Effect of various concentrations of the complementary C-rich oligonucleotide on the fluorescence melting curves of G_3T in the presence of 10 mM KCl. G_3T only (black); 1:1 (G_3T:C_3A) (red); 1:20 (green); 1:50 (blue). Annealing measurements were performed at a rate of temperature change of 0.2 °C.min^{-1}. The first differential of these melting profiles are shown in the right-hand panel. **(b)** Fluorescence melting profiles of the other G_nT series of oligonucleotides in the presence of 10 mM KCl and varying concentrations of their complement. G_nT only (black); 1:1 (G_nT:C_nA) (red); 1:20 (green); 1:50 (blue). Melting profiles are normalised to the final fluorescence value.

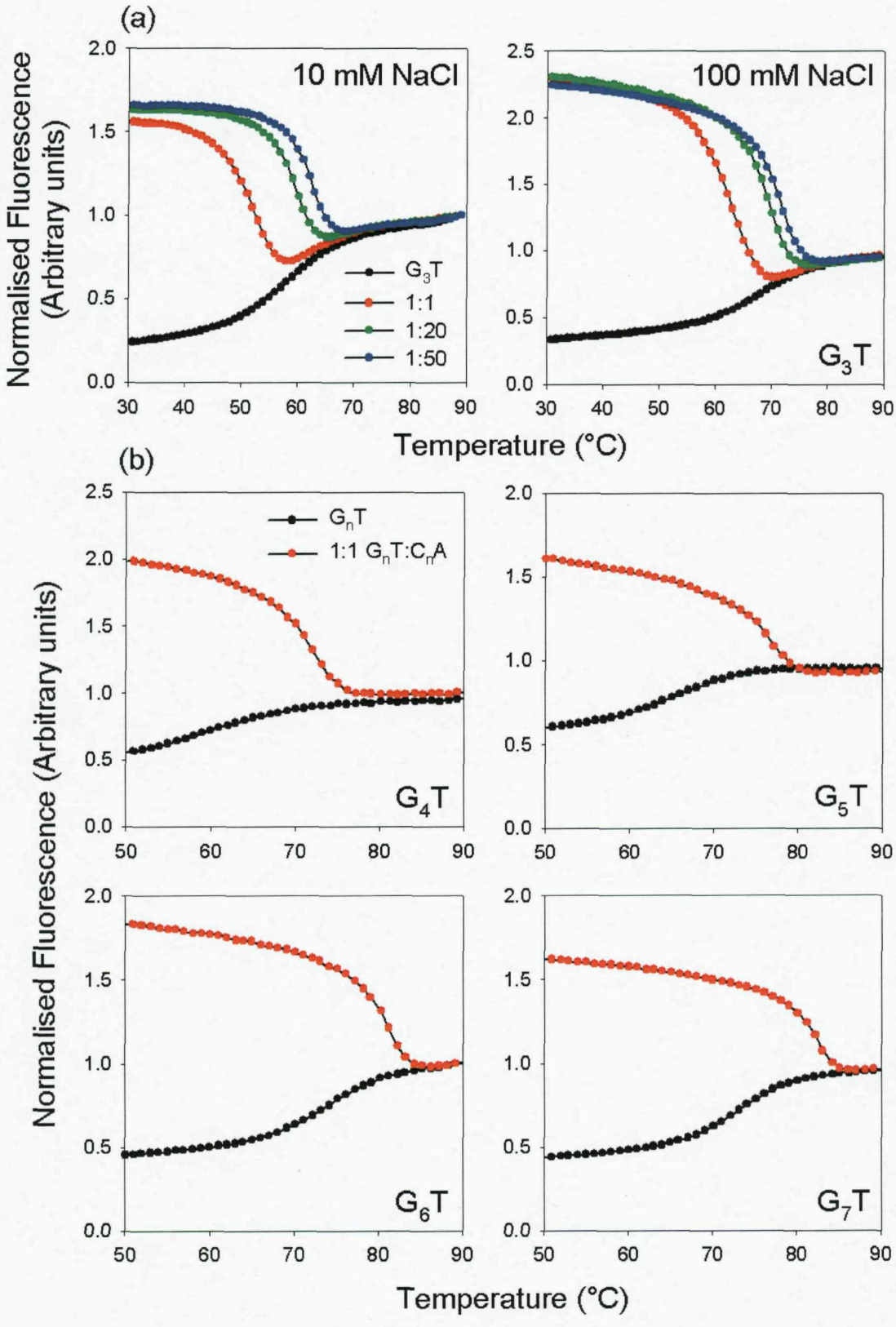


Figure 6.13 (a) Fluorescence annealing profiles for G_3T in the presence of 10 or 100 mM NaCl, and various concentrations of its complementary strand. **(b)** Annealing profiles for each of the other G_nT sequences in the presence of 100 mM NaCl and equimolar concentrations of their complementary strand (0.25 μM). Melting and annealing measurements were performed at a rate of temperature change of 0.2 °C.min^{-1}. All melting curves are normalised to the final fluorescence value.

6.4 Discussion

These results demonstrate that with the exception of G_2T, each of the G-rich sequences studied fold to form quadruplex structures; the CD spectra exhibit positive peaks at either 260 or 295 nm and fluorescence melting experiments show folded structures where the fluorescence is quenched at low temperatures. The folded structures are stabilised by the addition of monovalent cations, with potassium consistently generating more stable complexes than sodium. Each of the structures formed appears to be intramolecular rather than intermolecular, as their T_m values are independent of oligonucleotide concentration. In the presence of the complementary strand, several of these sequences preferentially fold to form a quadruplex in favour of a duplex.

These results are therefore able to address four main questions concerning these quadruplexes; what are their relative stabilities? How many G-quartets are present within each stack? (i.e. do any G residues slip into the loop regions?) Are the structures parallel or antiparallel-stranded, and which (if any) of the structures form a quadruplex even in the presence of their complementary strand?

6.4.1 Stability

Although the fluorescence melting technique cannot distinguish between different quadruplex structures, it provides a way in which to estimate the relative stability of different complexes. In both sodium and potassium, the stability of the quadruplex formed increases with the length of the G-tract, with the exception of G_3T which is anomalously high. Sequences that contained clusters of two contiguous guanines were relatively unstable. G_2T displayed no melting transition, presumably because it has a T_m value less than 25 °C, below the measurable range of the fluorescence melting technique. G_2T could however be induced to form a quadruplex on addition of the quadruplex-stabilising ligand BRACO-19. The additional end-stacking interactions provided by the acridine may promote the stable formation of a two-quartet stack. This current study is limited to four-repeat units of $d(G_2T)$, however there are several reports of trinucleotide repeats and microsatellites of $d(NGG)_n$ forming quadruplex structures (Fry and Loeb, 1994; Matsugami *et al.*, 2001; Zemanek *et al.*, 2005). It is noticeable that these repeating units form higher order structures rather than simple unimolecular folds. G_2T_2 is also unstable in potassium, and does not form at all sodium ions. This

sequence is a close homologue of the thrombin-binding aptamer which forms an antiparallel 'chair-type' fold (Macaya *et al*, 1993; Wang *et al*, 1993; Padmanabhan *et al*, 1993) with a three base central loop. Substitution of this central loop, which spans the wide groove, for a shorter loop has been shown to result in a significantly less favourable enthalpy, due to the limited span of the dinucleotide linker (Smirnov and Shafer, 2000; Marathias and Bolton, 1999).

For the remaining sequences of the G_nT series, the order of stability in potassium is $G_3T > G_7T > G_6T > G_5T > G_4T$. These sequences surprisingly all form quadruplex structures in the presence of lithium alone. A similar observation was made for the sequence T30695, the HIV integrase-binding aptamer (Jing, *et al.*, 1997). It was suggested that an initial folded form was assumed in the presence of lithium, which then switched to its correct folded conformation on addition of potassium. A similar explanation could be imagined for the G_nT series of oligonucleotides, where the addition of even low concentrations of potassium caused a dramatic increase in stability, and at concentrations above 10 mM the complexes were too stable to measure. In sodium, as the ionic strength is increased G_3T moves down the rank order of stability to between G_5T and G_6T. The rank order of stability for the second series, in which the G-tracts are separated by -TT-, shows that longer G-tracts produce more stable structures. G_5T_2 and G_6T_2 have similar T_m values in potassium, while G_7T_2, G_6T_2 and G_5T_2 are all similar in the presence of sodium. These results demonstrate that there is no simple relationship between quadruplex stability and the length of the G-tracts when the linkers are restricted to -T- or -TT-, though the behaviours in sodium and potassium are similar

6.4.2 Topology

The low structural resolution of CD spectra does not allow the determination of detailed tertiary structure of quadruplexes, but can provide an indication of the strand alignment. Spectra can be compared to the spectral signatures of structurally characterised quadruplexes, in order to assign particular oligomer conformations and the type of quadruplex formed. However the structural heterogeneity of these sequences make assigning a topology based on the CD spectrum extremely difficult. In the presence of potassium, the oligonucleotides with single T residues between the G-tracts all adopt a structure which the CD signature suggests is predominantly parallel-stranded. The peak at 295 nm, which increases in intensity with G-tract length, may be a result of the

formation of mixed 3 + 1 strand alignments, but is more likely a result of a combination of parallel and antiparallel stranded structures in equilibrium. This is consistent with the 1D NMR spectra (data not shown), which revealed a broad envelope of ill-defined peaks in the imino region for each of the G_nT sequences except G_3T.

The sequences with -TT- linking the G-tracts show a much stronger propensity to adopt the antiparallel conformation (except G_3T_2, for which the potassium and sodium forms are exclusively parallel). In potassium G_4T_2 is also largely parallel. This sequence is similar to that which was used to solve the NMR structure of the *Tetrahymena* telomeric repeat $d(TTGGGG)_4$ (Wang and Patel, 1994). The 3 + 1 strand alignment of this structure may be favoured over an antiparallel fold due to the restrictive span of the central loop. Addition of a third nucleotide (guanine) to the central position of the central loop i.e. d(GGGGTTGGGGTGTGGGGTTGGGG) has been shown to revert the topology to an antiparallel chair-type topology (Randazzo *et al.*, 2002), similar to that of the thrombin-binding aptamer. G_5T_2 is almost exclusively antiparallel in both sodium and potassium. The requirement for at least three nucleotides in the central loop of an antiparallel fold, suggest G_5T_2 contains either GTTG or GTT/TTG in the central loop (with a four G-quartet stack as predicted by the Δn value (see below)). Several different combinations of peripheral loop lengths could then be imagined depending on the central loop. Extending each G-tract by one or two guanines (G_6T_2 and G_7T_2) reverts the CD spectrum back to mixed parallel/antiparallel in potassium, but remains predominantly antiparallel in sodium. The strong signal at 295 nm for the longer sequences is further evidence against the formation of intermolecular aggregation, which would be expected to be parallel.

6.4.3 How many G-quartets?

For the longer sequences, it is clear that a folded structure must contain some guanine residues in the loops, as for instance a single T will be insufficient to span between the top and bottom of a stack of seven potential G-quartets in G_7T. Moreover, data presented in Chapter 4, and recently observed in a quadruplex formed within the human RET promoter, indicate that single nucleotide double-chain reversals will not span from the top-to-bottom of a stack of four G-quartets (Guo *et al.*, 2007). Molecular modelling studies have also suggested that two nucleotides are required to bridge a four-quartet stack (Qin *et al.*, 2007), although NMR data has shown a single phosphate is capable of

spanning a stack of three tetrads (Crnugelj *et al.*, 2003). The values of Δn, derived from the thermodynamic parameters should be interpreted with caution, as there may be multiple forms in equilibrium, which may vary according to ionic strength. However they do indicate some trends, which could be related to the number of stacked G-quartets in each complex. These trends also show some similarity with the patterns of gel mobility. In potassium, G_3T, G_4T and G_5T all appear to contain only three stacked G-quartets, suggesting their loops contain a single T, GT, and GGT respectively. However for G_4T and G_5T, several structures could be envisaged depending on which guanines participate in quartet formation and which are in the loops; several different forms may co-exist in solution in which the G-strands slip relative to each other. The values of Δn for G_6T and G_7T increase by one, suggesting that these may contain four stacked G-quartets (Figure 6.14). A similar transition is evident in the presence of sodium for G_nT between n = 5 and 6, though in this case the value of Δn is only ~2 for G_3T, G_4T and G_5T and increases to 4 for G_6T and G_7T. The lower value might indicate the presence of one fewer G-quartet, or more likely indicates that sodium ions are less tightly bound between the terminal tetrads and the loops.

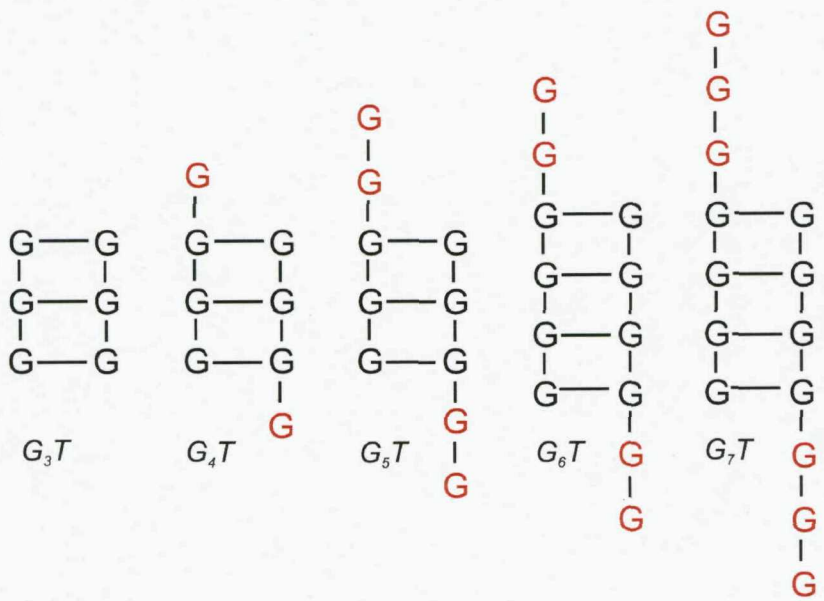


Figure 6.14 Predicted number of G-quartets formed by the G_nT series of oligonucleotides in the presence of potassium. Strand slipping results in some guanines participating in loop formation.

In the G_nT_2 series, the values for Δn are similar for all the complexes in the presence of potassium. The similarity in G_3T_2 and G_4T_2 is consistent with the NMR structure of the intramolecular *Tetrahymena* repeat which shows a three-quartet structure with a 3 + 1 strand alignment (Wang and Patel, 1994). In this case it appears that increasing the length of the G-tracts does not increase the number of stacked G-quartets and the

additional guanines must reside in the loops. In sodium, G_3T_2 and G_4T_2 have the same number of quartets, which increase for G_5T_2, G_6T_2 and G_7T_2.

6.4.4 Quadruplex - duplex equilibria

If quadruplex formation is to occur *in vivo*, in some instances the quadruplex-duplex equilibrium must favour the formation of the quadruplex. Principally, this will be due to the primary sequence, although the structural transition can be modulated by a variety of other factors, including temperature, ionic strength, molecular crowding and pH (Li *et al.*, 2002; Phan and Mergny, 2002; Kumar and Maiti, 2005) as well as ligand or protein binding (Rangan *et al.*, 2001)

Fluorescence annealing profiles were used to generate competition between quadruplex and duplex forms for each of the G_nT series, forcing the G-rich sequence to choose between either fold. Previous studies have shown that sequences with short loops, such as G_3T and two sequences related to the c-myc promoter, preferentially fold to form a quadruplex whereas for sequences with longer loops, including the human telomere repeat $d[G_3(TTAGGG)_3]$, there is a mixture of the two species (Risitano and Fox, 2003). Risitano and Fox suggested that the quadruplex formed by G_3T persists, even in the presence of a 50-fold excess of C_3A. In contrast, results in the present study show for the same sequence the duplex predominates when the complementary strand is in large excess. It should be noted that the previous experiments were performed at a fast rate of annealing (0.1 $^{\circ}C.sec^{-1}$) and it was noted the melting and annealing profiles might not be at thermodynamic equilibrium (Risitano and Fox, 2003). Figure 6.15 shows the effect of increasing the rate of annealing on the fluorescence profile for G_3T in the presence of a 50-fold excess of its complementary strand. When the sample is cooled rapidly, the reaction appears to favour quadruplex formation, while the duplex is favoured at slower rates.

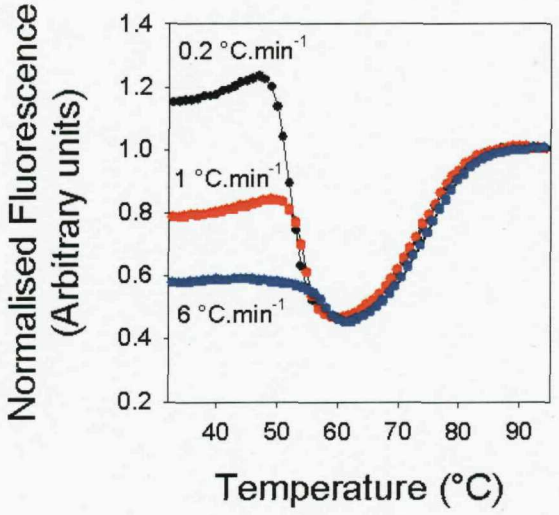


Figure 6.15 Fluorescence annealing profiles of G_3T containing a 50-fold excess of its complement at *different rates of temperature change. Profiles were determined in 10 mM lithium phosphate buffer* containing 1 mM KCl.

In the present study, slower temperature transition rates were employed (0.2 °C.min^{-1}) to ensure equilibrium conditions. The results presented show that in agreement with Risitano and Fox, G_3T preferentially folds to form quadruplex in the presence of an equimolar ratio of C- and G-rich strands. However, there are some differences when the complementary strand is in excess, which can be attributed to both the slower rate of heating, and also the oligonucleotide sequence. The G_3T sequence used previously contained no flanking nucleotides between the G-stack and the fluorescent groups, resulting in a more stable quadruplex structure.

These experiments were carried out in 1 or 10 mM potassium, as under the constraints of the technique it was not possible to increase the concentration of potassium to near physiological levels (100 – 150 mM KCl) and still observe a melting transition. However, the T_m values of the duplexes at 100 mM KCl (with the strands in a 1:1 ratio) are still well below those of the quadruplex structures. Of the other G_nT sequences, G_5T, G_6T and G_7T all favour quadruplex formation in 10 mM potassium, although some duplex is present, whereas for G_4T, which is the least stable quadruplex of the series, both species are present. In sodium, the significantly lower stability of the quadruplex structures results in almost exclusive duplex formation.

These results show that several of the G_nT series of oligonucleotides are able to form quadruplex structures in the presence of their complementary strands. In contrast to

sequences with longer loop lengths, sequences with extended G-tracts still showed preference for quadruplex formation, even in vast excess of their complement. This is significant given longer quadruplex-forming sequences must also compete against increasingly stable G-C rich duplexes. *In vivo* the situation is likely to be more complicated, as the two stands will be kept in close proximity, although strands can be separated during a number of key cellular events, which may provide an opportunity for quadruplex folding. The propensity of the complementary strand to form an i-motif may also contribute to the quadruplex-duplex equilibrium; longer G-tracts increase quadruplex stability, and longer C-tracts do likewise for the i-motif (Mergny *et al.*, 1995) although at physiological pH its influence is likely to be negligible.

CHAPTER 7

General Conclusions

7.1 Summary and Conclusions

G-quadruplexes are currently the subject of considerable interest, as increasing evidence suggests that they can form, or be induced to form in a number of biologically significant genomic locations. Their unique three-dimensional geometry has also highlighted these structures as potential targets for drug design. Bioinformatic searches have unveiled a wealth of data regarding the prevalence of quadruplex-forming motifs throughout the genome, however these searches reveal little in regards to the topology and stability of the structures that may form. Many of these quadruplexes will be unique in terms of their thermodynamic stability and folded conformation, but at present, there is a limited understanding of the rules that govern the formation of these structures and their stability. The work described in this thesis studied the effect of sequence on the formation and stability of intramolecular DNA G-quadruplexes, using series of model sequences. Through extensive use of a number of biophysical techniques, predominantly fluorescence melting and circular dichroism, this work has examined three main aspects in relation to quadruplex folding: (i) The length of the loop regions, (ii) sequence effects of single nucleotide loops and (iii) the length of the G-tracts.

7.1.1 Fluorescence melting

Chapter 3 introduced the use of a fluorescence melting technique that has been applied to the study of intramolecular G-quadruplex formation. This technique is higher throughput than conventional absorbance methods and also requires lower oligonucleotide concentrations (and volumes), thereby promoting the formation of intramolecular structures. The effects of conjugating fluorescent groups (FAM and dabcyl) to the ends of a single-stranded G-rich oligonucleotide are discussed in relation to the stability and topology of the quadruplex formed. In agreement with previous reports, the attachment of fluorescent probes is slightly detrimental to quadruplex stability (Mergny and Maurizot, 2002; Green *et al.*, 2003), however the destabilising

influence is consistent in both sodium- and potassium-containing buffers, and is therefore non-discriminatory between different quadruplex folds. This work also highlights the importance of using a temperature transition rate slow enough to allow equilibrium T_m measurements. Previous studies using this technique have revealed that for some quadruplex-forming sequences there is significant hysteresis between the melting and annealing profiles (Risitano and Fox, 2003a; Risitano and Fox, 2004). Temperature gradient profiles have therefore been designed so as to eliminate hysteresis (in most cases) allowing full thermodynamic analysis of the melting profiles to be performed. Fluorescence melting was used extensively in the later chapters to compare the thermodynamic stability of quadruplex structures, with all labelled oligonucleotides containing a 5'-fluorescein and a 3'-dabcyl quencher.

7.1.2 Loop length

Chapter 4 examined the effect of loop length on the properties of intramolecular quadruplexes, using sequences containing short loops, longer loops or combinations of the two. The results show that quadruplex stability increases with the number of short loops that are present, though the arrangement of short and long loops within the sequence has little effect. The kinetic parameters are also strongly influenced by the length of the loop regions; Analysis of non-equilibrium melting curves has allowed access to the kinetic parameters of folding and unfolding, and these data show that structures containing short loops are faster to fold and display longer dissociation half-lives. Like the trends in stability, the number and not the position of these loops is the more important factor. In the presence sodium ions, this method could not be applied due to the superimposable melting and annealing profiles, even at rapid temperature transition rates (12 $^{\circ}C.min^{-1}$). The structures stabilised by potassium ions displayed greater thermodynamic stabilities and slower kinetics than those stabilised by sodium ions.

Previous studies have demonstrated that single nucleotide loops are restricted to form double-chain reversals due to their short span (Hazel *et al.*, 2005), however the results presented here highlight the ability of short loops to influence the conformation of the remaining loop(s) in the structure; In the presence of potassium ions, G_3T forms a parallel-stranded structure containing all double-chain reversals. When one or two of the short loops are replaced with the longer T_4 loop, in any position, the topology

remains parallel despite the potential for numerous folded topologies. Replacing all of the single-T loops within the sequence results in a conformational switch to antiparallel. The single-thymine loops are therefore able to impose an all-parallel strand arrangement when only one is present within the structure. The trends are slightly different in the presence of sodium ions, reflecting the greater preference for the antiparallel conformation.

7.1.2.1 Recent loop studies

Following the completion of this work, two independent studies have since been published that further explore the effect of loop length and loop sequence on quadruplex folding and stability. In the work described here, each of the loop regions were restricted to thymine residues. Bugaut and Balasubramanian have analysed the influence of loop length in a sequence-independent context, by randomising the loop nucleotides (Bugaut and Balasubramanian 2008). This study supports our data concerning the importance of single nucleotide loops, with sequences containing two such loops enforcing a parallel-stranded topology. Equally, Guédin and co-workers also found sequences of the type d(GGGHGGGN$_{3-9}$GGGHGGG) form a parallel-stranded arrangement (Guédin *et al.*, 2008). It appears however that the topologies of quadruplexes containing one single-base loop are more strongly influenced by the nucleotide composition of the longer loops. In the present study, quadruplexes containing oligo-dT 4-4-1 and 4-1-4 loop lengths are parallel, however randomised loops of 1-3-3, 3-1-3 and 3-3-1 all generate a mixed parallel/antiparallel CD spectrum. These spectra represent an average of all the structures that are present within each sequence library, and therefore defining which sequences adopt different conformations will require more in depth study.

In sodium, the folding rules are more complicated still. Sequences containing two single-nucleotide thymine loops i.e. 1-4-1 and 4-1-1 display a parallel CD signature, as they do in potassium. Guédin *et al.*, have now shown than even this is not universal, as the topology is dependent upon the identity of the two single nucleotide loops (Guédin *et al.*, 2008). Systematic substitution of the single-nucleotide bases has shown that while quadruplexes with loop lengths 1-3-1 are parallel when the peripheral loops are T, in agreement with the present study, T to C substitutions result in a switch to an

antiparallel CD spectrum. Furthermore, combinations of A, C and T loops result in examples of parallel, antiparallel and intermediate spectra (Guédin *et al.*, 2008).

7.1.3 Single nucleotide loop sequence

The data presented in Chapter 5 demonstrate that when *all* the loops are a single nucleotide, modifications to loop sequence can have a significant influence on quadruplex stability, without altering the topology in both sodium and potassium. Amongst the potential quadruplex forming sequences identified in the human genome, single-nucleotide adenines are the most commonly observed loop, ahead of both T and C (Todd *et al.*, 2006). However in terms of stability, adenines are significantly disfavoured over pyrimidine loops with (A-A-A) more than 20 °C more stable than (T-T-T). It is also noteworthy that removing the base altogether resulted in the formation of the most stable structure, but removing the sugar is detrimental to stability.

7.1.4 G-tract length

G_3T forms the most stable structure amongst the series of nucleobase loop isomers, however increasing the length of the G-tracts does not necessarily result in the formation of additional G-quartets. Chapter 6 examined the structure and stability of intramolecular G-quadruplexes that are formed by sequences of the type $d(G_nT)_4$ and $d(G_nT_2)_4$ (where n = 2-7). The results of this study have shown that there is no clear relationship between the length of the G-tracts and the stability and topology of the quadruplex structures formed. It was initially thought that increasing the length of the G-tracts would allow an increased number of stacked G-quartets, and hence more stable structures. In general the T_m values do increase with the length of the G-tracts, although not by as much as would be expected for the stacking of an additional quartet. The inability of these sequences to form more G-quartets result in the formation of slipped structures, in which guanines can participate in either G-quartet formation or as part of the loop. Defining a precise topology for these sequences is therefore virtually impossible, due to the plurality of potential structural combinations.

Competition assays between quadruplex and duplex show that even longer sequences with extended runs of guanines can preferentially fold to form quadruplex. However G_3T again forms the most stable structure, in either series, and predominates even in

vast excess of its complementary strand. This may be significant given the frequency of the G_3NG_3 motif amongst promoter sequences, where the duplex strands may be separated during transcription.

7.2 Future considerations

Predicting the conformation a DNA quadruplex will adopt based on nucleotide sequence is clearly extremely complex, and establishing further folding rules will require substantial systematic investigation. The complexity of quadruplex folding is partly due to the ability of guanines to adopt either *syn* or *anti* conformations about the glycosidic bond when participating in a G-quartet, as this allows the formation of fold-back loops and antiparallel-stranded topologies. RNA quadruplexes have received far less attention than their DNA counterparts, despite the absence of a complementary strand. RNA quadruplex folding rules may differ from those of DNA, due to their preference for the *anti* conformation; the C3'-endo (N-type) conformation of the ribose prevents the base from switching to *syn*. If RNA nucleotides were locked only in *anti*, in theory, RNA oligonucleotides would only be able to adopt parallel-stranded quadruplex topologies, irrespective of the length of the loops. Of the small number of RNA quadruplexes studied to date, many appear to adopt a parallel-stranded arrangement (Wieland and Hartig, 2007; Kumari *et al.*, 2007) or following selective substitution of dG for rG in DNA quadruplexes, drive a change in topology from antiparallel to parallel (Tang and Shafer, 2006). Systematic studies similar to those performed in this work using RNA oligonucleotides could be carried out to examine the folding and stability of such sequences.

References

Aboul-ela, F., Murchie, A.I. and Lilley, D.M. (1992) NMR study of parallel-stranded tetraplex formation by the hexadeoxynucleotide d(TG4T). *Nature*. **360**, 280-282

Alberti, P., Arimondo, P.B., Mergny, J.L., Garestier, T., Hélène, C. and Sun, J.S. (2002) A directional nucleation-zipping mechanism for triple helix formation. *Nucleic Acids Res.* **30,** 5407-5415

Alberti, P. and Mergny, J.L. (2003) DNA duplex-quadruplex exchange as the basis for a nanomolecular machine. *Proc. Natl. Acad. Sci. U.S.A.* **100**, 1569-1573

Alberti, P., Bourdoncle, A., Saccà, B., Lacroix, L. and Mergny, J.L. (2006) DNA nanomachines and nanostructures involving quadruplexes. *Org. Biomol. Chem.* **4**, 3383-3391

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

Ambrus, A., Chen, D., Dai, J., Bialis, T., Jones, R.A. and Yang, D. (2006) Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res.* **34**, 2727-2735

Arimondo, P.B., Riou, J.F., Mergny, J.L., Tazi, J., Sun, J.S., Garestier, T. and Hélène, C. (2000) Interaction of human DNA topoisomerase I with G-quartet structures. *Nucleic Acids Res.* **28**, 4832-4838

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

Balagurumoorthy, P., Brahmachari, S. K., Mohanty, D., Bansal, M. and Sasisekharen, V. (1992) Hairpin and parallel quartet structures for telomeric sequences. *Nucleic Acids Res.* **20**, 4061-4067

Balagurumoorthy, P. and Brahmachari, S.K. (1994) Structure and stability of human telomeric sequence. *J. Biol. Chem.* **269**, 21858-21869

Bang, I. (1910) Untersuchungen über die Guanylsare. *Biochem. Z.* **26**, 293–311

Bardin, C. and Leroy, J.L. (2007) The formation pathway of tetramolecular G-quadruplexes. *Nucleic Acids Res.* **36**, 477-488

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

Bejugam, M., Sewitz, S., Shirude, P.S., Rodriguez, R., Shahid, R. and Balsubramanian, S. (2007) Trisubstituted isoalloxazines as a new class of G-quadruplex binding ligands: Small molecule recognition of c-kit oncogenes expression. *J. Am. Chem. Soc.* **129**, 12926-12927

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

Blume, S.W., Guarcello, V., Zacharias, W. and Miller, D.M. (1997) Divalent transition metal cations counteract potassium-induced quadruplex assembly of oligo(dG) sequences. *Nucleic Acids Res.* **25**, 617-625

Bock, L.C., Griffin, L.C., Latham, J.A., Vermaas, E.H. and Toole, J.J. (1992) Selection of single-stranded DNA molecules that bind and inhibit human thrombin. *Nature*. **355**, 564-566

Bommarito, S., Peyret, N. and SantaLucia, J. (2000) Thermodynamic parameters for DNA sequences with dangling ends. *Nucleic Acids Res.* **28**, 1929-1934

Bouaziz, S., Kettani, A., Patel, D.J. (1998) A K cation-induced conformational switch within a loop spanning segment of a DNA quadruplex containing G-G-G-C repeats. *J. Mol. Biol.* **282**, 637-652

Bourdoncle, A., Estévez-Torres, A., Gosse, C., Lacroix, L., Vekhoff, P., Le Saux, T., Jullien, L. and Mergny, J.L. (2006) Quadruplex-based molecular beacons as tunable DNA probes. *J. Am. Chem. Soc.* **128**, 11094-11105

Brown, N.M., Rachwal, P.A., Brown, T. and Fox, K.R. (2005) Exceptionally slow kinetics of the intramolecular quadruplex formed by the *Oxytricha* telomeric repeat. *Org. Biomol. Chem.* **22**, 4153-4157

Bugaut, A. and Balasubramanian, S. (2008) A sequence-independent study of the influence of short loop lengths on the stability and topology of intramolecular DNA G-quadruplexes. *Biochemistry.* **47**, 689-697

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

Cáceres, C., Wright, G., Gouyette, C., Parkinson, G. and Subirana, J.A. (2004) A thymine tetrad in d(TGGGGT) quadruplexes stabilized with Tl^+/Na^+ ions. *Nucleic Acids Res.* **32**, 1097-1102

Callender, R. and Dyer, R.B. (2002) Probing protein dynamics using temperature jump relaxation spectroscopy. *Curr. Opin. Struct. Biol.* **12**, 628-633

Cantor, C.R. and Schimmel, P.R. (1980) Biophysical chemistry. W.H.Freeman & Co ltd, New York

Catasti, P., Chen, X., Moyzis, R.K., Bradbury, E.M. and Gupta, G. (1996) Structure-function correlations of the insulin-linked polymorphic region. *J. Mol. Biol.* **264**, 534-545

Cech, T.R. (2000) Life at the end of the chromosome: telomeres and telomerase. *Angew. Chem. Int. Ed. Engl.* **39,** 34-43

Cevec, M. and Plavec, J. (2005) Role of loop residues and cations on the formation and stability of dimeric DNA G-quadruplexes. *Biochemistry.* **44**, 15238-15246

Chang, C.C., Kuo, I.C., Ling, I.F., Chen, C.T., Chen, H.C., Lou, P.J., Lin, J.J. and Chang, T.C. (2004) Detection of quadruplex DNA structures in human telomeres by a fluorescent carbazole derivative. *Anal. Chem.* **76**, 4490-4494

Chao, D.T. and Korsmeyer, S.J. (1998) BCL-2 family: regulators of cell death. *Annu Rev Immunol.* **16**, 395-419

Chaput, J.C. and Switzer, C. (1999) A DNA pentaplex incorporating nucleobase quintets. *Proc. Natl. Acad. Sci. U.S.A.* **96**, 10614-10619

Chen, F.M (1992) Sr^{2+} facilitates intermolecular G-quadruplex formation of telomeric sequences. *Biochemistry.* **31**, 3769-3776

Chen, L., Cai, L., Zhang, X. and Rich, A. (1994) Crystal structure of a four-stranded intercalated DNA: d(C4). *Biochemistry.* **33**, 13540-13546

Cheong, C. and Moore, P.B. (1992) Solution structure of an unusually stable RNA tetraplex containing G- and U-quartet structures. *Biochemistry.* **31**, 8406-8414

Cogoi, S. and Xodo, L.E. (2005) G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res.* **34**, 2536-2549

Creze, C., Rinaldi, B., Haser, R., Bouvet, P. and Gouet, P. (2007) Structure of a d(TGGGGT) quadruplex crystallized in the presence of Li^{+} ions. *Acta Crystallogr. D, Biol. Crystallogr.* **63**, 682-688

Crnugelj, M., Hud, N.V. Plavec, J. (2002) The solution structure of d(G(4)T(4)G(3))(2): a bimolecular G-quadruplex with a novel fold. *J. Mol. Biol.* **320**, 911-924

Crnugelj, M., Sket, P. and Plavec, J. (2003) Small change in a G-rich sequence, a dramatic change in topology: New dimeric G-quadruplex folding motif with unique loop orientations. *J. Am. Chem. Soc.* **125,** 7866-7871

Czerlinski, G.H. (1966) Chemical relaxation. New York: Marcel Dekker Inc.

Dai, T-Y., Marotta, S.P. and Sheardy, R.D. (1995) Self-assembly of DNA into high molecular weight species. *Biochemistry.* **34**, 673-682

Dai, J., Dexheimer, T.S., Chen, D., Carver, M., Ambrus, A., Jones, R.A. and Yang, D. (2006) An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution. *J. Am. Chem. Soc.* **128**, 1096-1098

Dai, J., Chen, D., Jones, R.A., Hurley, L.H. and Yang, D. (2006b) NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region. *Nucleic Acids Res.* **34**, 5133-5144

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

Dai, J., Carver, M., Punchihewa, C., Jones, R.A. and Yang, D. (2007b) Structure of the hybrid-2 type intramolecular human telomeric G-quadruplex in K+ solution: insights into structure polymorphism of the human telomeric sequence. *Nucleic Acids Res.* **35**, 4927-4940

Dapic, V., Abdomerovic, V., Marrington, R., Peberdy, J., Rodger, A., Trent, J.O. and Bates, P.J. (2003) Biophysical and biological properties of quadruplex oligodeoxyribonucleotides. *Nucleic Acids Res.* **31**, 2097-2107

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

Darnell, J.C., Jensen, K.B., Jin, P., Brown, V., Warren, S.T. and Darnell, R.B. (2001) Fragile X mental retardation protein targets G quartet mRNAs important for neuronal function. *Cell.* **107**, 489-499

Datta, B. and Armitage, B.A. (2001) Hybridization of PNA to structured DNA targets: quadruplex invasion and the overhang effect. *J. Am. Chem. Soc.* **123**, 9612-9619

De Armond, R., Wood, S., Sun, S., Hurley, L.H. and Ebbinghaus, S.W. (2005) Evidence for the presence of a guanine quadruplex forming region within a polypurine tract of the hypoxia inducible factor 1alpha promoter. *Biochemistry.* **44**, 16341-16350

De Cian, A., Guittat, L., Kaiser, M., Sacca, B., Amrane, S., Bourdoncle, A., Alberti, P., Telaude-Fichou, M-P., Lacroix, L. and Mergny, J.L. (2007) Fluorescence-based melting assays for studying quadruplex ligands. *Methods.* **41**, 183-195

De Cian, A., Lacroix, L., Douarre, C., Temime-Smaali, N., Trentesaux, C., Riou, J.F. and Mergny, J.L. (2008) Targeting telomeres and telomerase. *Biochimie.* **90**, 131-155

Dexheimer, T.S., Sun, D. and Hurley, L.H. (2005) Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the bcl-2 P1 promoter. *J. Am. Chem. Soc.* **128**, 5404-5415

Du, Z., Kong, P., Gao, Y. and Li, N. (2007) Enrichment of G4 DNA motif in transcriptional regulatory region of chicken genome. *Biochem. Biophys. Res. Commun.* **354**, 1067-1070

Duquette, M.L., Handa, P., Vincent, J.A., Taylor, A.F. and Maizels, N. (2004) Intracellular transcription of G-rich DNAs induces formation of G-loops, novel structures containing G4 DNA. *Genes Dev.* **18**, 1618-1629

Eddy, J., Maizels, N. (2006) Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acids Res.* **34**, 3887-3896

Eddy, J. and Maizels, N. (2008) Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes. *Nucleic Acids Res.* **36**, 1321-1333

Esposito, V., Randazzo, A., Piccialli, G., Petraccone, L., Giancola, C. and Mayol, L. (2004) Effects of an 8-bromodeoxyguanosine incorporation on the parallel quadruplex structure $[d(TGGGT)]_4$ *Org. Biomol. Chem.* **2**, 313-318

Esposito, V., Virgilio, A., Randazzo, A., Galeone, A. and Mayol, L. (2005) A new class of DNA quadruplexes formed by oligodeoxyribonucleotides containing a 3'-3' or 5'-5' inversion of polarity site. *Chem. Comm.* **31**, 3953-3955

Fang, G. and Cech, T.R. (1993a) Characterization of a G-quartet formation reaction promoted by the beta-subunit of the Oxytricha telomere-binding protein. *Biochemistry.* **32**, 11646-11657

Fang, G. and Cech, T.R. (1993b) The beta subunit of Oxytricha telomere-binding protein promotes G-quartet formation by telomeric DNA. *Cell.* **74**, 875-885

Feigon, J., Koshlap, K.M. and Smith, F.W. (1995) 1H NMR spectroscopy of DNA triplexes and quadruplexes. *Methods Enzymol.* **261**, 225-255

Fernando, H., Reszka, A.P., Huppert, J., Ladame, S., Rankin, S., Venkitaraman, A.R., Neidle, S. and Balasubrmanian, S. (2006) A conserved quadruplex motif located in a transcription activation site of the human c-kit oncogene. *Biochemistry.* **45**, 7854-7860

Fersht, A. (1977) Enzyme structure and mechanism. W.H.Freeman & Co Ltd, New York

Fry, M. (2007) Tetraplex DNA and its interacting proteins. *Front. Biosci.* **12**, 4336-4351

Fry, M. and Loeb, L.A. (1994) The fragile X syndrome $d(CGG)_n$ nucleotide repeats form a stable tetrahelical structure. *Proc. Natl. Acad. Sci. USA.* **91**, 4950-4954

Galezowska, E., Gluszynska, A. and Juskowiak, B. (2007) Luminescence study of G-quadruplex formation in the presence of Tb^{3+} ions. *J. Inorg. Biochem.* **101**, 678-685

Gehring, K., Leroy, J.L. and Guéron, M. (1993) A tetrameric DNA structure with protonated cytosine.cytosine base pairs. *Nature.* **363**, 561-565

Gellert, M., Lipsett, M.N. and Davies, D.R. (1962) Helix formation by guanylic acid. *Proc. Natl. Acad. Sci. U.S.A.* **48**, 2013-2018

Ghosal, G. and Muniyappa, K. (2005) Saccharomyces cerevisiae Mre11 is a high-affinity G4 DNA-binding protein and a G-rich DNA-specific endonuclease: implications for replication of telomeric DNA. *Nucleic Acids Res.* **33**, 4692-4703

Giraldo, R. and Rhodes, D. (1994a) The yeast telomere-binding protein RAP1 binds to and promotes the formation of DNA quadruplexes in telomeric DNA. *EMBO J.* **13**, 2411-2420

Giraldo, R., Suzuki, M., Chapman, L. and Rhodes, D. (1994b) Promotion of parallel DNA quadruplexes by a yeast telomere binding protein: a circular dichroism study. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 7658-7662

Gomez, D., Lemarteleur, T., Lacroix, L., Mailliet, P., Mergny, J.L. and Riou, J.F. (2004) Telomerase downregulation induced by the G-quadruplex ligand 12459 in A549 cells is mediated by hTERT RNA alternative splicing. *Nucleic Acids Res.* **32**, 371-379

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

Grand, C.L., Han, H., Muñoz, R.M., Weitman, S., Von Hoff, D.D., Hurley, L.H. and Bearss, D.J. (2002) The cationic porphyrin TMPyP4 down-regulates c-MYC and human telomerase reverse transcriptase expression and inhibits tumor growth in vivo. *Mol. Cancer Ther.* **1**, 565-573

Granotier, C., Pennarun, G., Riou, L., Hoffschir, F., Gauthier, L.R., De Cian, A., Gomez, D., Mandine, E., Riou, J.F., Mergny, J.L., Mailliet, P., Dutrillaux, B., Boussin, F.D. (2005) Preferential binding of a G-quadruplex ligand to human chromosome ends. *Nucleic Acids Res.* **33**, 4182-4190

Griffith, J.D., Comeau, L., Rosenfield, S., Stansel, R.M., Bianchi, A., Moss, H. and de Lange, T. (1999) Mammalian telomeres end in a large duplex loop. *Cell.* **97**, 503-514

Gros, J., Rosu, F., Amrane, S., De Cian, A., Gabelica, V., Lacroix, L. and Mergny, J.L. (2007) Guanines are a quartet's best friend: impact of base substitutions on the kinetics and stability of tetramolecular quadruplexes. *Nucleic Acids Res.* **35**, 3064-3075

Guédin, A., De Cian, A., Gros, J., Lacroix, L., Mergny, J.L. (2008) Sequence effects in single-base loops for quadruplexes. *Biochimie.* **90**, 686-696

Guo, Q., Lu, M. and Kallenbach, N.R. (1993) Effect of thymine tract length on the structure and stability of model telomeric sequences. *Biochemistry.* **32**, 3596-3603

Guo, K., Pourpak, A., Beetz-Rogers, K., Gokhale, V., Sun, D. and Hurley, L.H. (2007) Formation of a pseudosymmetrical G-quadruplex and i-motif structures in the proximal promoter region of the RET oncogene. *J. Am. Chem. Soc.* **129**, 10220-10228

Haider, S., Parkinson, G.N. and Neidle. S. (2002) Crystal structure of the potassium form of an Oxytricha nova G-quadruplex. *J. Mol. Biol.* **320**, 189-200

Halder, K. and Chowdhury, S. (2005) Kinetic resolution of bimolecular hybridization versus intramolecular folding in nucleic acids by surface plasmon resonance:

application to G-quadruplex/duplex competition in human c-myc promoter. *Nucleic Acids Res.* **33**, 4466-4474

Han, H., Bennett, R.J. and Hurley, L.H. (2000) Inhibition of unwinding of G-quadruplex structures by Sgs1 helicase in the presence of N,N'-bis[2-(1-piperidino)ethyl]-3,4,9,10-perylenetetracarboxylic diimide, a G-quadruplex-interactive ligand. *Biochemistry.* **39**, 9311-9316

Hanakahi, L.A., Sun, H. and Maizels, N. (1999) High affinity interactions of nucleolin with G-G-paired rDNA. *J. Biol. Chem.* **274**, 15908-15912

Hardin, C.C., Henderson, E., Watson, T. and Prosser, J.K. (1991) Monovalent cation induced structural transitions in telomeric DNAs: G-DNA folding intermediates. *Biochemistry.* **30**, 4460-4472

Hardin, C.C., Corregan, M.J., Lieberman, D.V. and Brown II, B.A. (1997) Allosteric interactions between DNA strands and monovalent cations in DNA quadruplex assembly: Thermodynamic evidence for three linked association pathways. *Biochemistry.* **36**, 15428-15450

Harley, C.B., Futcher, A.B. and Greider, C.W. (1990) Telomeres shorten during ageing of human fibroblasts. *Nature.* **345**, 458-460

Hazel, P., Huppert, J., Balasubramanian, S. and Neidle, S. (2004) Loop-length-dependent folding of G-quadruplexes. *J. Am. Chem. Soc.* **126**, 16405-16415

Hazel, P., Parkinson, G.N. and Neidle, S. (2006) Topology variation and loop structural homology in crystal and simulated structures of a bimolecular DNA quadruplex. *J. Am. Chem. Soc.* **126,** 5480-5487

He, Y.J., Neumann, R.D. and Panyutin, I.G. (2004) Intramolecular quadruplex conformation of human telemore DNA assessed with I-125-radioprobing. *Nucleic Acids Res.* **32**, 5359-5367

Heinrich, M.C., Blanke, C.D., Druker, B.J. and Corless, C.L. (2002) Inhibition of KIT tyrosine kinase activity: A novel molecular approach to the treatment of KIT-positive malignancies. *J. Clin. Oncol.* **20**, 1692-1703

Henriksson, M. and Lüscher, B. (1996) Proteins of the Myc network: essential regulators of cell growth and differentiation. *Adv. Cancer Res.* **68**, 109-182

Hershman, S.G., Chen, Q., Lee, J.Y., Kozak, M.L., Yue, P., Wang, L.S. and Johnson, F.B. (2008) Genomic distribution and functional analyses of potential G-quadruplex-forming sequences in Saccharomyces cerevisiae. *Nucleic Acids Res.* **36**, 144-156

Hirota, S., Isozaki, K., Moriyama, Y., Hashimoto, K., Nishida, T., Ishiguro, S., Kawano, K., Hanada, M., Kurata, A., Takeda, M., Muhammad Tunio, G., Matsuzawa, Y., Kanakura, Y., Shinomura, Y. and Kitamura, Y. (1998) Gain-of-function mutations of c-kit in human gastrointestinal stromal tumors. *Science.* **279**, 577–580

Hovarth, M.P. and Schultze, S.C. (2001) DNA G-quartets in a 1.86 Á resolution structure of an Oxytricha nova telomeric protein-DNA complex. *J. Mol. Biol.* **310**, 367-377

Howard, F.B. and Miles, H.T. (1982) Poly(inosinic acid) helices: essential chelation of alkali metal ions in the axial channel. *Biochemistry*. **21**, 6736-6745

Howell, R.M., Woodford, K.J., Weitzmann, M.N. and Usdin, K. (1996) The Chicken B-globin gene promoter forms a novel "cinched" tetrahelical structure. *J. Biol. Chem.* **271**, 5208-5214

Huber, M.D., Lee, D.C. and Maizels, N. (2002) G4 DNA unwinding by BLM and Sgs1p: substrate specificity and substrate-specific inhibition. *Nucleic Acids Res.* **30**, 3954-3961

Hud, N.V., Smith, F.W., Anet, F.A. and Feigon, J. (1996) The selectivity for K+ versus Na+ in DNA quadruplexes is dominated by relative free energies of hydration: a thermodynamic analysis by 1H NMR. *Biochemistry*. **35**, 15383-15390

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

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

Hurley, L.H. (2002) DNA and its associated processes as targets for cancer therapy. *Nat. Rev. Cancer*. **2**, 188-200

Hurley, L.H., Von Hoff, D.D., Siddiqui-Jain, A. and Yang, D. (2006) Drug targeting of the c-MYC promoter to repress gene expression via a G-quadruplex silencer element. *Semin. Oncol.* **33**, 498-512

Jahn, C.L. and Klobutcher, L.A. (2002) Genome remodeling in ciliated protozoa. *Annu Rev Microbiol.* **56**, 489-520

James, P.L., Brown, T. and Fox, K.R. (2003) Thermodynamic and kinetic stability of intermolecular triple helices containing different proportions of C^+.GC and T.AT triplets. *Nucleic Acids Res.* **31,** 5598-5606

Jing, N., Rando, R.F., Pommier, Y. and Hogan, M.E. (1997) Ion selective folding of loop domains in a potent anti-HIV oligonucleotide. *Biochemistry*. **36**, 12498-12505

Jing, N., Gao, X., Rando, R.F. and Hogan, M.E. (1997b) Potassium-induced loop conformational transition of a potent anti-HIV oligonucleotide. *J. Biomol. Struct. Dyn.* **15**, 573-585

Kang, C.H., Berger, I., Lockshin, C., Ratliff, R., Moyzis, R. and Rich, A. (1994) Crystal structure of intercalated four-stranded d(C3T) at 1.4 A resolution. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 11636-11640

Keniry, M.A., Strahan, G.D., Owen, E.A. and Shafer, R.H. (1995) Solution structure of the Na^+ form of the dimeric guanine quadruplex $[d(G_3T_4G_3)]_2$. *Eur. J. Biochem.* **233**, 631-643

Keniry, M.A., Owen, E.A. and Shafer, R.H. (1997) The contribution of thymine-thymine interactions to the stability of folded dimeric quadruplexes. *Nucleic Acids Res.* **25**, 4389-4392

Kettani, A., Kumar, R.A. and Patel, D.J. (1995) Solution structure of a DNA quadruplex containing the fragile X syndrome triplet repeat. *J. Mol. Biol.* **254**, 638-656

Kettani, A., Bouaziz, S., Gorin, A., Zhao, H., Jones, R.A. and Patel, D.J. (1998) Solution structure of a Na cation stabilized DNA quadruplex containing G.G.G.G and G.C.G.C tetrads formed by G-G-G-C repeats observed in adeno-associated viral DNA. *J. Mol. Biol.* **282**, 619-636

Kettani, A., Gorin, A., Majumdar, A., Hermann, T., Skripkin, E., Zhao, H., Jones, R. and Patel, D.J. (2000) A dimeric DNA interface stabilized by stacked A.(G.G.G.G).A hexads and coordinated monovalent cations. *J. Mol. Biol.* **297**, 627-644

Kikin, O., D'Antonio, L. and Bagga, P.S. (2006) QGRS Mapper: a web-based server for predicting G-quadruplexes in nucleotide sequences. *Nucleic Acids Res.* **34**, W676-682

Kim, N.W., Piatyszek, M.A., Prowse, K.R., Harley, C.B., West, M.D., Ho, P.L., Coviello, G.M., Wright, W.E., Weinrich, S.L. and Shay, J.W. (1994) Specific association of human telomerase activity with immortal cells and cancer. *Science.* **266**, 2011-2015

Kohwi, Y. (1989) Cationic metal-specific structures adopted by the poly(dG) region and the direct repeats in the chicken adult beta A globin gene promoter. *Nucleic Acids Res.* **17**, 4493-4502

Krishnan-Ghosh, Y., Liu, D. and Balasubramanian, S. (2004) Formation of an interlocked quadruplex dimer by d(GGGT). *J. Am. Chem. Soc.* **126**, 11009-11016

Kumar, N. and Maiti, S. (2005) The effect of osmolytes and small molecule on quadruplex-WC duplex equilibrium: a fluorescence resonance energy transfer study. *Nucleic Acids Res.* **33**, 6723-6732

Kumar, N. and Maiti, S. (2007) Role of locked nucleic acid modified complementary strand in quadruplex/Watson-Crick duplex equilibrium. *J. Phys. Chem. B.* **111**, 12328-12337

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

Kuryavyi, V., Kettani, A., Wang, W., Jones, R. and Patel, D.J. (2000) A diamond-shaped zipper-like DNA architecture containing triads sandwiched between mismatches and tetrads. *J. Mol. Biol.* **295**, 455-469

Kuryavyi, V., Majumdar, A., Shallop, A., Chernichenko, N., Skripkin, E., Jones, R. and Patel, D.J. (2001) A double chain reversal loop and two diagonal loops define the architecture of a unimolecular DNA quadruplex containing a pair of stacked G(syn)-G(syn)-G(anti)-G(anti) tetrads flanked by a G-(T-T) Triad and a T-T-T triple. *J. Mol. Biol.* **310**, 181-194

Laughlan. G., Murchie, A.I., Norman. D.G., Moore, M.H., Moody, P.C., Lilley, D.M. and Luisi, B. (1994) The high-resolution crystal structure of a parallel-stranded guanine tetraplex. *Science.* **265**, 520-524

Lécuyer, E., Herblot, S., Saint-Denis, M., Martin, R., Begley, C.G., Porcher, C., Orkin, S.H. and Hoang, T. (2002) The SCL complex regulates c-kit expression in hematopoietic cells through functional interaction with Sp1. *Blood.* **100**, 2430-2440

Lee, M.P, Parkinson, G.N., Hazel, P. and Neidle, S. (2007) Observation of the coexistence of sodium and calcium ions in a DNA G-quadruplex ion channel. *J. Am. Chem. Soc.* **129**, 10106-10107

Leroy, J.L. and Guéron, M. (1995) Solution structures of the i-motif tetramers of d(TCC), d(5methylCCT) and d(T5methylCC): novel NOE connections between amino protons and sugar protons. *Structure.* **3**, 101-120

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

Li, W., Wu, P., Ohmichi, T. and Sugimoto, N. (2002) Characterization and thermodynamic properties of quadruplex/duplex competition. *FEBS Lett.* **526**, 77-81

Liu, Z. and Gilbert, W. (1994) The yeast KEM1 gene encodes a nuclease specific for G4 tetraplex DNA: implication of in vivo functions for this novel DNA structure. *Cell.* **77**, 1083-1092

Liu, Z., Lee, A. and Gilbert, W. (1995) Gene disruption of a G4-DNA-dependent nuclease in yeast leads to cellular senescence and telomere shortening. *Proc. Natl. Acad. Sci. U.S.A.* **92**, 6002-6006

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

Lux, M.L., Rubin, B.P., Biase, T.L., Chen, C.J., Maclure, T., Demetri, G., Xiao, S., Singer, S., Fletcher, C.D. and Fletcher, J.A. (2000) KIT extracellular and kinase domain mutations in gastrointestinal stromal tumors. *Am. J. Pathol.* **156**, 791-795

Macaya, R.F., Schultze, P., Smith, F.W., Roe, J.A., Feigon, J. (1993) Thrombin-binding DNA aptamer forms a unimolecular quadruplex structure in solution. *Proc. Natl. Acad. Sci. U.S.A.* **90**, 3745-3749

Maizels, N. (2006) Dynamic roles for G4 DNA in the biology of eukaryotic cells. *Nat. Struct. Mol. Biol.* **13**, 1055-1059

Marathias, V.M. and Bolton, P.H. (1999) Determinants of DNA quadruplex structural type: sequence and potassium binding. *Biochemistry*. **38**, 4355-4364

Marayama, T., Shinohara, T., Ichinose, H., Kitaoka, M., Okamura, N., Kamiya, N. and Goto, M. (2005) Mutation detection in DNA oligonucleotides based on a guanine quenching method coupled with enzymatic digestion of single-stranded DNA. *Biotechnol. Lett.* **27**, 1349-1354

Marchand, C., Pourquier, P., Laco, G.S., Jing, N. and Pommier, Y. (2002) Interaction of human nuclear topoisomerase I with guanosine quartet-forming and guanosine-rich single-stranded DNA and RNA oligonucleotides. *J. Biol. Chem.* **277**, 8906-8911

Marotta, S.P., Tumaburri, P.A. and Sheardy, R.D. (1996) Sequence and environmental effects of the self-assembly of DNA oligomers processing $G_xT_2G_y$ segments. *Biochemistry*. **35**, 10484-10492

Marras, S.A., Kramer, F.R. and Tyagi, S. (2002) Efficiencies of fluorescence resonance energy transfer and contact-mediated quenching in oligonucleotide probes. *Nucleic Acids Res.* **30**, e122

Marsh, C.M. and Henderson, E. (1994) G-wires: Self-assembly of a telomeric oligonucleotide d(GGGGTTTTGGGG) into large superstructures. *Biochemistry*. **33**, 10718-10724

Matsugami, A., Ouhashi, K., Kanagawa, M., Liu, H., Kanagawa, S., Uesugi, S. and Katahira, M. (2001) An intramolecular quadruplex of (GGA)(4) triplet repeat DNA with a G:G:G:G tetrad and a G(:A):G(:A):G(:A):G heptad, and its dimeric interaction. *J. Mol. Biol.* **313**, 255-269

Matsugami, A., Xu, Y., Noguchi, Y., Sugiyama, H. and Katahira, M. (2007) Structure of a human telomeric DNA sequence stabilized by 8-bromoguanosine substitutions, as determined by NMR in a K+ solution. *FEBS J.* **274**, 3545-3556

Mazumder, A., Neamati, N., Ojwang, J.O., Sunder, S., Rando, R.F. and Pommier, Y. (1996) Inhibition of the human immunodeficiency virus type 1 integrase by guanosine quartet structures. *Biochemistry*. **35**, 13762-13771

McManus, S.A. and Li, Y. (2008) A deoxyribozyme with a novel guanine quartet-helix pseudoknot structure. *J. Mol. Biol.* **375**, 960-968

Mergny, J.L., Lacroix, L., Han, X., Leroy, J.L. and Helene, C. (1995) Intramolecular folding of pyrimidine. oligodeoxynucleotides into an i-DNA motif. *J. Am. Chem. Soc.* **117**, 8887-8898

Mergny J.L, and Lacroix, L. (1998) Kinetics and thermodynamics of i-DNA formation: phosphodiester verses modified oligodeoxynucleotides. *Nucleic Acids Res.* **26**, 4797-4803

Mergny, J.L., Phan, A.T. and Lacroix, L. (1998) Following G-quartet formation by UV-spectroscopy. *FEBS Lett.* **435**, 74-78

Mergny, J.L. (1999) Fluorescence energy transfer as a probe for tetraplex formation: the i-motif. *Biochemistry*. **29**, 9261-9268

Mergny, J.L., Lacroix, L., Teulade-Fichou, M-P., Hounsou, C., Guittat, L., Hoarau, M., Arimondo, P.B., Vigneron, J-P., Lehn, J-M., Riou, J-F., Garestier, T. and Helene, C. (2001) Telomerase inhibitors based on quadruplex ligands selected by a fluorescence assay. *Proc. Natl. Acad. Sci. USA*. **98**, 3062-3067

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

Mergny, J.L., De Cian, A., Ghelab, A., Saccà, B. and Lacroix, L. (2005) Kinetics of tetramolecular quadruplexes. *Nucleic Acids Res*. **33**, 81-94

Mergny, J.L., De Cian, A., Amrane, S. and Webba da Silva M. (2006) Kinetics of double-chain reversals bridging contiguous quartets in tetramolecular quadruplexes. *Nucleic Acids Res*. **34**, 2386-2397

Merkina, E.E. and Fox, K.R. (2005) Kinetic stability of intermolecular DNA quadruplexes. *Biophys. J.* **89**, 365-373

Miles, H.T and Frazier, J. (1978) Poly(I) helix formation. Dependence on site-specific complexing to alkali metals. *J. Am. Chem. Soc*. **100**, 8037-8038

Miyoshi, D., Matsumura, S., Nakano, S. and Sugimoto, N. (2004) Duplex dissociation of telomere DNAs induced by molecular crowding. *J. Am. Chem. Soc*. **126**, 165-169

Miyoshi, D., Karimata, H. and Sugimoto, N. (2005) Drastic effect of a single base difference between human and tetrahymena telomere sequences on their structures under molecular crowding conditions. *Angew. Chem. Int. Ed*. **44**, 3740-3744

Mohaghegh, P., Karow, J.K., Brosh Jr, R.M., Bohr, V.A. and Hickson, I.D. (2001) The Bloom's and Werner's syndrome proteins are DNA structure-specific helicases. *Nucleic Acids Res*. **29**, 2843-2849

Monchaud, D. and Teulade-Fichou, M.P. (2008) A hitchhiker's guide to G-quadruplex ligands. *Org. Biomol. Chem*. **6**, 627-636

Moriera, B.G., You, Y., Behlke, R. and Owczarzy, R. (2005) Effects of fluorescent dyes, quenchers, and dangling ends on DNA duplex stability. *Biochem. Biophys. Res. Commun.* **327**, 473-484

Murchie, A.I.H. and Lilley, D.M.J. (1992) Retinoblstoma susceptibility genes contain 5' sequences with a high propensity to form guanine-tetrad structures. *Nucleic Acids. Res*. **20**, 49-53

Nagatoishi, S., Nojima, T., Galezowska, E., Juskowiak, B. and Takenaka, S. (2006) G-quadruplex-based FRET probes with the thrombin-binding aptamer (TBA) sequence designed for the efficient fluorimetric detection of the potassium ion. *Chembiochem.* **7**, 1730-1737

Nagesh, N., Bhargava, P. and Chatterji, D. (1992) Terbium(III)-induced fluorescence of four-stranded G4-DNA. *Biopolymers.* **32**, 1421-1424

Olsen, C.M., Gmeiner, W.H. and Marky, L.A. (2006) Unfolding of G-quadruplexes: energetic, and ion and water contributions of G-quartet stacking. *J. Phys. Chem. B.* **110**, 6962-6969

Opresko, P.L., Mason, P.A., Podell, E.R., Lei, M., Hickson, I.D., Cech, T.R. and Bohr, V.A. (2005) POT1 stimulates RecQ helicases WRN and BLM to unwind telomeric DNA substrates. *J. Biol. Chem.* **280**, 32069-32080

Padmanabhan, K., Padmanabhan, K.P., Ferrera, J.D., Sadler. J.E. and Tulinsky, A. (1993) The structure of alpha-thrombin inhibited by a 15-mer single-stranded DNA aptamer. *J. Biol. Chem.* **268**, 17651-17654

Paeschke, K., Simonsson, T., Postberg, J., Rhodes, D. and Lipps, H.J. (2005) Telomere end-binding proteins control the formation of G-quadruplex DNA structures in vivo. *Nat. Struct. Mol. Biol.* **12**, 847-854

Pan, B., Xiong, Y., Shi, K., Deng, J. and Sundaralingam, M. (2003) Crystal structure of an RNA purine-rich tetraplex containing adenine tetrads: implications for specific binding in RNA tetraplexes. *Structure.* **11**, 815-823

Pan, B., Shi, K., Sundaralingam, M. (2006) Base-tetrad swapping results in dimerization of RNA quadruplexes: implications for formation of the i-motif RNA octaplex. *Proc. Natl. Acad. Sci. U.S.A.* **103**, 3130-3134

Paramasivan, S., Rujan, I. and Bolton, P. (2007) Circular dichroism of quadruplex DNAs: Applications to structure, cation effects and ligand binding. *Methods.* **43**, 324-331

Park, G.H., Plummer III, H.K. and Krystal, G.W. (1998) Selective Sp1 binding is critical for maximal activity of the human *c-kit* promoter. *Blood.* **92**, 4138-4149

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

Parkinson, G.N., Ghosh, R. and Neidle, S. (2007) Structural basis for binding of porphyrin to human telomeres. *Biochemistry.* **46**, 2390-2397

Patel, D.J., Phan, A.T. and Kuryavyi, V. (2007) Human telomere, oncogenic promoter and 5'-UTR G-quadruplexes: diverse higher order DNA and RNA targets for cancer therapeutics. *Nucleic Acids Res.* **35**, 7429-7455

Petraccone, L., Pagano, B., Esposito, V., Randazzo, A, Piccialli, G., Barone, G., Mattia, C.A. and Giancola, C. (2005) Thermodynamics and kinetics of PNA-DNA quadruplex-forming chimeras. *J. Am. Chem. Soc.* **127**, 16215-16223

Petraccone, L., Erra, E., Duro, I., Esposito, V., Randazzo, A., Mayol, L., Mattia, C.A., Barone, G. and Giancola, C. (2005) Relative stability of quadruplexes containing different number of G-tetrads. *Nucleosides Nucleotides Nucleic Acids.* **24**, 757-760

Phan, A.T. and Mergny, J.L. (2002) Human telomeric DNA: G-quadruplex, i-motif and Watson-Crick double helix. *Nucleic Acids Res.* **30**, 4618-4625

Phan, A.T. and Patel, D.J. (2003) Two-repeat human telomeric d(TAGGGTTAGGGT) sequence forms interconverting parallel and antiparallel G-quadruplexes in solution: Distinct topologies, thermodynamic properties, and folding/unfolding kinetics. *J. Am. Chem. Soc.* **125**, 15021-15027

Phan, A.T., Modi, Y.S. and Patel, D.J. (2004) Two-repeat Tetrahymena telomeric d(TGGGGTTGGGGT) Sequence interconverts between asymmetric dimeric G-quadruplexes in solution. *J. Mol. Biol.* **338**, 93-102

Phan, A.T., Modi, Y.S., Patel, D.J. (2004b) Propeller-type parallel-stranded G-quadruplexes in the human c-myc promoter. *J. Am. Chem. Soc.* **126**, 8710-8716

Phan, A.T., Kuryavyi, V., Gaw, H.Y. and Patel, D.J. (2005) Small-molecule interaction with a five-guanine-tract G-quadruplex structure from the human MYC promoter. *Nat. Chem. Biol.* **1**, 167-173

Phan, A.T., Kuryavyi, V., Ma, J.B., Andreola, M.L. and Patel, D.J. (2005) An interlocked dimeric parallel-stranded DNA quadruplex: A potent inhibitor of HIV-1 integrase. *Proc. Nat. Acad. Sci. USA.* **102**, 634-639

Phan, A.T., Luu, K.N. and Patel, D.J. (2006) Different loop arrangements of intramolecular human telomere (3+1) G-quadruplexes in K^+ solution. *Nucleic Acids Res.* **34**, 5715-5719

Phan, A.T., Kuryavyi, V., Burge, S., Neidle, S. and Patel, D.J. (2007) Structure of an unprecedented G-quadruplex scaffold in the human c-kit promoter. *J. Am. Chem. Soc.* **129**, 4386-4392

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

Phillips, K., Dauter, Z., Murchie, A.I., Lilley, D.M. and Luisi, B. (1997) The crystal structure of a parallel-stranded guanine tetraplex at 0.95 A resolution. *J. Mol. Biol.* **273**, 171-182

Pilch, D.S., Plum, G.E. and Breslauer, K.J. (1995) The thermodynamics of DNA structures that contain lesions or guanine tetrads. *Curr. Opin. Struct. Biol.* **5**, 334-342

Qi, J. and Shafer, R.H. (2007) Human telomere quadruplex: Refolding and selection of individual conformers via RNA/DNA chimeric editing. *Biochemistry.* **46**, 7599-7606

Rachwal, P.A., Brown, T. and Fox, K.R. (2007a) Effect of G-tract length on the topology and stability of intramolecular DNA quadruplexes. *Biochemistry.* **46**, 3036-3044

Rachwal, P.A., Brown, T. and Fox, K.R. (2007b) Sequence effects of single base loops in intramolecular quadruplex DNA. *FEBS. Lett.* **581**, 1657-1660

Rachwal, P.A., Findlow, I.S., Werner, J.M., Brown, T. and Fox, K.R. (2007c) Intramolecular DNA quadruplexes with different arrangements of short and long loops. *Nucleic Acids Res.* **35**, 4214-4222

Rachwal, P.A. and Fox. K.R. (2007d) Quadruplex melting. *Methods.* **43**, 291-301

Raghuraman, M.K. and Cech, T.R. (1990) Effect of monovalent cation-induced telomeric DNA structure on the binding of Oxytricha telomeric protein. *Nucleic Acids Res.* **18**, 4543-4552

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

Rangan, A., Fedoroff, O.Y. and Hurley, L.H. (2001) Induction of duplex to G-quadruplex transition in the c-myc promoter region by a small molecule. *J. Biol. Chem.* **276**, 4640-4646

Rankin, S., Reszka, A.P., Huppert, J., Zloh, M., Parkinson, G.N., Todd, A.K., Ladame, S., Balasubramanian, S. and Neidle, S. (2005) Putative DNA quadruplex formation within the human c-kit oncogene. *J. Am. Chem. Soc.* **127**, 10584-10589

Rawal, P., Kummarasetti, V.B., Ravindran, J., Kumar, N., Halder, K., Sharma, R., Mukerji, M., Das, S.K. and Chowdhury, S. (2006) Genome-wide prediction of G4 DNA as regulatory motifs: role in Escherichia coli global regulation. *Genome Res.* **16**, 644-655

Read, M.A., Harrison, J.R., Romagnoli, B., Tanious, F.A., Gowan, S.M., Reszka, A.P., Wilson, W.D., Kelland, L.R. and Neidle, S. (2001) Structure-based design of selective and potent G-quadruplex telomerase inhibitors. *Proc Natl. Acad. Sci. USA* **98**, 4844-4849

Redon, S., Bombard, S., Elizondo-Riojas, M-A. and Chottard, J-C. (2003) Platinum cross-linking of adenines and guanines on the quadruplex structures of the $AG_3(T_2AG_3)_3$ and $(T_2AG_3)_4$ human telomere sequences in Na^+ and K^+ solutions. *Nucleic Acids Res.* **31,** 1605-1613

Risitano, A. and Fox, K.R. (2003a) Stability of intramolecular DNA quadruplexes: Comparison with DNA duplexes. *Biochemistry.* **42**, 6507-6513

Risitano, A. and Fox, K.R. (2003b) The stability of intramolecular DNA quadruplexes with extended loops forming inter- and intra-loop duplexes. *Org. Biomol. Chem.* **1**, 1852-1855

Risitano, A. and Fox, K.R. (2004) Influence of loop size on the stability of intramolecular DNA quadruplexes. *Nucleic Acids Res.* **32**, 2598-2606

Risitano, A. and Fox, K.R. (2005) Inosine substitutions demonstrate that intramolecular DNA quadruplexes adopt different conformations in the presence of sodium and potassium. *Bioorg. Med. Chem. Lett.* **15**, 2047-2050

Rougée, M., Faucon, B., Mergny, J.L., Barcelo, F., Giovannangeli, C., Garestier, T and Hélène, C. (1992) Kinetics and thermodynamics of triple-helix formation: Effects of ionic strength and mismatches. *Biochemistry*. **31**, 9269-9278

Rujan, I.N., Meleney, J.C. and Bolton, P.H. (2005) Vertebrate telomere repeat DNAs favor external propellar quadruplex structures in the presence of high concentrations of potassium. *Nucleic Acids Res*. **33**, 2022-2031

Saccà, B., Lacroix, L., and Mergny, J.L. (2005) The effect of chemical modifications on the thermal stability of different G-quadruplex-forming oligonucleotides. *Nucleic Acids Res.* **33**, 1182-1192

Saito, Y., Mizuno, E., Bag, S.S., Suzuka, I. and Saito, I. (2007) Design of a novel G-quenched molecular beacon: A simple and efficient strategy for DNA sequence analysis. *Chem Comm*. **43**, 4492-4494

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

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

Schultze, P., Hud, N.V., Smith, F.W. and Feigon, J. (1999) The effect of sodium, potassium and ammonium ions on the conformation of the dimeric quadruplex formed by the Oxytricha nova telomere repeat oligonucleotide d(G(4)T(4)G(4)). *Nucleic Acids Res*. **27**, 3018-3028

Seenisamy, J., Rezler, E.M., Powell, T.J., Tye, D., Gokhale, V., Joshi, C.S., Siddiqui-Jain, A. and Hurley, L.H. (2004) The dynamic character of the G-quadruplex element in the c-MYC promoter and modification by TMPyP4. *J. Am. Chem. Soc.* **126**, 8702-8709

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

Sen, D. and Gilbert, W. (1990) A sodium-potassium switch in the formation of four-stranded G4-DNA. *Nature*. **344**, 410-414

Shirude, P.S., Okumus, B., Ying, L., Ha, T. and Balasubramanian, S. (2007) Single-molecule conformational analysis of G-quadruplex formation in the promoter DNA duplex of the proto-oncogene c-kit. *J. Am. Chem. Soc*. **129**, 7484-7485

Siebenlist, U., Hennighausen, L., Battey, J. and Leder, P. (1984) Chromatin structure and protein binding in the putative regulatory region of the c-myc gene in Burkitt lymphoma. *Cell*. **37**, 381-391

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

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

Simonsson, T. and Sjöback, R. (1999) DNA tetraplex formation studied with fluorescence resonance energy transfer. *J. Biol. Chem.* **274**, 17379-17383

Sket, P. and Plavec, J. Not all G-quadruplexes exhibit ion-channel-like properties: NMR study of ammonium ion (non)movement within the d(G(3)T(4)G(4))(2) quadruplex. *J. Am. Chem. Soc.* **129**, 8794-8800

Smirnov, I. and Shafer, R.H. (2000) Effect of loop sequence and size on DNA aptamer stability. *Biochemistry.* **39**, 1462-1468

Smirnov, I. and Shafer, R.H. (2000b) Lead is unusually effective in sequence-specific folding of DNA. *J. Mol. Biol.* **296**, 1-5

Smith, F.W. and Feigon, J. (1992) Quadruplex structure of Oxytricha telomeric DNA oligonucleotides. *Nature*. **356**, 164-168

Smith, F.W., Lau, F.W. and Feigon, J. (1994) d(G3T4G3) forms an asymmetric diagonally looped dimeric quadruplex with guanosine 5'-syn-syn-anti and 5'-syn-anti-anti N-glycosidic conformations. *Proc. Natl. Acad. Sci. U.S.A.* **91**, 10546-10550

Smith, F.W., Schultze, P., Feigon, J. (1995) Solution structures of unimolecular quadruplexes formed by oligonucleotides containing Oxytricha telomere repeats. *Structure*. **3**, 997-1008

Spackova, N., Berger, I. and Sponer, J. (1999) Nanosecond molecular dynamics simulations of parallel and antiparallel guanine quadruplex DNA molecules. *J. Am. Chem. Soc.* **121**, 5519-5534

Stefl, R., Spackova, N., Berger, I., Koca, J. and Sponer, J. (2001) Molecular dynamics of DNA quadruplex molecules containing inosine, 6-thioquanine and 6-thiopurine. *Biophys. J.* **80**, 455-468

Strahan, G.D., Keniry, M.A. and Shafer, R.H. (1998) NMR structure refinement and dynamics of the K^+ -$[d(G_3T_4G_3)]_2$ quadruplex via particle mesh Ewald molecular dynamics simulations. *Biophys. J.* **75**, 968-981

Sun, H., Yabuki, A. and Maizels, N. (2001) A human nuclease specific for G4 DNA. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 12444-12449

Sun, Y., Guo, K.X., Rusche, J.J. and Hurley,L.H. (2005) Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents. *Nucleic Acids Res.* **33**, 6070-6080

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

Tang, C.F. and Shafer, R.H. (2006) Engineering the quadruplex fold: nucleoside conformation determines both folding topology and molecularity in guanine quadruplexes. *J. Am. Chem. Soc.* **128**, 5966-5973

Todd, A.K., Johnston, M. and Neidle, S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.* **33**, 2901-2907

Todd, A.K., Haider, S.M., Parkinson, G.N. and Neidle, S. (2007) Sequence occurance and structural uniqueness of a G-quadruplex in the human c-kit promoter. *Nucleic Acids Res.* **35**, 5799-5808

Todd, A.K., Neidle, S. (2008) The relationship of potential G-quadruplex sequences in cis-upstream regions of the human genome to SP1-binding elements. *Nucleic Acids Res. In press*

Tsujimoto, Y. and Croce, C.M. (1986) Analysis of the structure, transcripts, and protein products of bcl-2, the gene involved in human follicular lymphoma. *Proc. Natl. Acad. Sci. U.S.A.* **83**, 5214-5218

Tuveson, D.A., Willis, N.A., Jacks, T., Griffin, J.D., Singer, S., Fletcher, C.D., Fletcher, J.A. and Demetri, G.D. (2001) STI571 inactivation of the gastrointestinal stromal tumor c-KIT oncoprotein: biological and clinical implications. *Oncogene.* **20**, 5054-5058

Venczel, E.A. and Sen, D. (1993) Parallel and antiparallel G-DNA structures from a complex telomeric sequence. *Biochemistry.* **32**, 6220-6228

Virgilio, A., Esposito, V., Randazzo, A., Mayol, L. and Galeone, A. (2005) 8-Methyl-2'-deoxyguanosine incorporation into parallel DNA quadruplex structures. *Nucleic Acids Res.* **33**, 6188-6195

Vliagoftis, H., Worobec, A.S. Metcalfe, D.D. (1997) The protooncogene c-kit and c-kit ligand in human disease. *J Allergy Clin Immunol.* **100**, 435–440

Völker, J., Klump, H.H. and Breslauer, K.J. (2001) Communication between noncontacting macromolecules. *Proc. Natl. Acad. Sci. U.S.A.* **98**, 7694-7699

Vorlickova, M., Chladkova, J., Kejnovska, I., Fialova, M. and Kypr, J. (2005) Guanine tetraplex topology of human telomere DNA is governed by the number of (TTAGGG) repeats. *Nucleic Acids Res.* **33**, 5851-5860

Vorlickova, M., Bednarova, K., Kejnovska, I. and Kypr, J. (2007) Intramolecular and intermolecular guanine quadruplexes of DNA in aqueous salt and ethanol solutions. *Biopolymers.* **86**, 1-10

Wallace, M.I., Ying, L., Balasubramanian, S. and Klenerman, D. (2001) Non-Arrhenius kinetics for the loop closure of a DNA hairpin. *Proc. Natl. Acad. Sci. USA.* **98,** 5584-5589

Wang, K.Y., McCurdy, S., Shea, R.G., Swaminathan, S. and Bolton, P.H. (1993) A DNA aptamer which binds to and inhibits thrombin exhibits a new structural motif for DNA. *Biochemistry.* **32**, 1899-1904

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

Wang, Y., Patel, D.J. (1995) Solution structure of the Oxytricha telomeric repeat d[G4(T4G4)3] G-tetraplex. *J. Mol. Biol.* **251**, 76-94

Wang, Y., Rusling, D.A., Powers, V.E.C., Lack, O., Osbourne, S., Fox, K.R. and Brown, T. (2005) Stable recognition of TA interruptions by triplex forming oligonucleotides containing a novel nucleoside. *Biochemistry.* **44**, 5884-5892

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

Webba da Silva, M. (2005) Experimental demonstration of T:(G:G:G:G):T hexad and T:A:A:T tetrad alignments within a DNA quadruplex stem. *Biochemistry.* **44**, 3754-3764

Webba da Silva, M. (2007) Geometric formalism for DNA quadruplex folding. *Chemistry.* **13**, 9738-9745

Weitzmann, M.N., Woodford, K.J. and Usdin, K. (1997) DNA secondary structures and the evolution of hypervariable tandem arrays. *J. Biol. Chem.* **272**, 9517-9523

Wieland, M. and Hartig, J.S. (2007) RNA quadruplex-based modulation of gene expression. *Chem. Biol.* **14**, 757-763

Wlodarczyk, A., Grzybowski, P., Patkowski, A., Dobek, A. (2005) Effect of ions on the polymorphism, effective charge, and stability of human telomeric DNA. Photon correlation spectroscopy and circular dichroism studies. *J. Phys. Chem. B.* **109**, 3594-3605

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

Wu, L., Multani, A.S., He, H., Cosme-Blanco, W., Deng, Y., Deng, J.M., Bachilo, O., Pathak, S., Tahara, H., Bailey, S.M., Deng, Y., Behringer, R.R. and Chang, S. (2006) Pot1 deficiency initiates DNA damage checkpoint activation and aberrant homologous recombination at telomeres. *Cell.* **126**, 49-62

Wyatt, J.R., Davis, P.W. and Freier, S.M. (1996) Kinetics of G-quartet-mediated tetramer formation. *Biochemistry.* **35**, 8002-8008

Xu, Y., Noguchi, Y. and Sugiyama, H. (2006) The new models of the human telomere d[AGGG(TTAGGG)$_3$] in K^+ solution. *Bioorg. Med. Chem.* **14**, 5584-5591

Yadav, V.K., Abraham, J.K., Mani, P., Kulshrestha, R. and Chowdhury, S. (2007) QuadBase: genome-wide database of G4 DNA--occurrence and conservation in human, chimpanzee, mouse and rat promoters and 146 microbes. *Nucleic Acids Res.* **36**, D381-385

Yarden, Y., Kuang, W.J., Yang-Feng, T., Coussens, L., Munemitsu, S., Dull, T.J., Chen, E., Schlessinger, J., Francke, U. and Ullrich, A. (1987) Human proto-oncogene c-kit: a new cell surface receptor tyrosine kinase for an unidentified ligand. *EMBO J.* **6**, 3341-3351

Yasuda, H., Galli, S.J. and Geissler, E.N. (1993) Cloning and functional analysis of the mouse *c-kit* promoter. *Biochem. Biophys. Res. Commun.* **191**, 893-901

Ying, L.M., Green, J.J., Li, H.T., Klenerman, D. and Balasubramanian, S. (2003) Studies on the structure and dynamics of the human telomeric G-quadruplex by single molecule fluorescence resonance energy transfer. *Proc. Natl. Acad. Sci. USA.* **100,** 14629-14634

Young, R.L. and Korsmeyer, S.J. (1993) A negative regulatory element in the bcl-2 5'-untranslated region inhibits expression from an upstream promoter. *Mol. Cell. Biol.* **13**, 3686-3697

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

Zahler, A.M., Williamson, J.R., Cech, T.R. and Prescott, D.M. (1991) Inhibition of telomerase by G-quartet DNA structures. *Nature*. **350**, 718-720

Zemanek, M., Kypr, K. and Vorlickova, M. (2005) Conformational properties of DNA containing (CCA)n and (TGG)n trinucleotide repeats. *Int. J. Biol. Macromol.* **36**, 23-32

Zhang, N., Gorin, A., Majumdar, A., Kettani, A., Chernichenko, N., Skripkin, E. and Patel, D.J. (2001) Dimeric DNA quadruplex containing major groove-aligned A·T·A·T and G·C·G·C tetrads stabilised by intersubunit Watson-Crick A·T and G·C base pairs. *J. Mol. Biol.* **312**, 1073-1088

Zhang, N., Gorin, A., Majumdar, A., Kettani, A., Chernichenko, N., Skripkin, E. and Patel, D.J. (2001b) V-shaped scaffold: A new architectural motif identified in an A·(G·G·G·G) pentad containing dimeric DNA quadruplex involving stacked G(anti)·G(anti)·G(anti)·G(syn) tetrads. *J. Mol. Biol.* **311**, 1063-1079

Zhang, N., Phan, A.T. and Patel, D.J. (2005) (3 + 1) Assembly of three human telomeric repeats into an asymmetric dimeric G-quadruplex. *J. Am. Chem. Soc.* **127**, 17277-17285

Zhao, Y., Kan, Z.Y., Zeng, Z.X., Hao, Y.H., Chen, H. and Tan, Z. (2004) Determining the folding and unfolding rate constants of nucleic acids ny biosensor. Application to telomere G-quadruplex. *J. Am. Chem. Soc.* **126**, 13255-13264

Zimmerman, S.B., Cohen, G.H. and Davies, D.R. (1975) X-ray fiber diffraction and model-building study of polyguanylic acid and polyinosinic acid. *J. Mol. Biol.* **92**, 181-192

PUBLICATIONS

Available online at www.sciencedirect.com
ScienceDirect
Methods 43 (2007) 291–301
METHODS
www.elsevier.com/locate/ymeth

Quadruplex melting

Phillip A. Rachwal, Keith R. Fox *

School of Biological Sciences, University of Southampton, Bassett Crescent East, Southampton SO16 7PX, UK

Accepted 11 May 2007

Abstract

Melting curves are commonly used to determine the stability of folded nucleic acid structures and their interaction with ligands. This paper describes how the technique can be applied to study the properties of four-stranded nucleic acid structures that are formed by G-rich oligonucleotides. Changes in the absorbance (at 295 nm), circular dichroism (at 260 or 295 nm) or fluorescence of appropriately labelled oligonucleotides, can be used to measure the stability and kinetics of folding. This paper focuses on a fluorescence melting technique, and explains how this can be used to determine the T_m ($T_{1/2}$) of intramolecular quadruplexes and the effects of quadruplex-binding ligands. Quantitative analysis of these melting curves can be used to determine the thermodynamic (ΔH, ΔG, and ΔS) and kinetic (k_1, k_{-1}) parameters. The method can also be adapted to investigate the equilibrium between quadruplex and duplex DNA and to explore the selectivity of ligands for one or other structure.

Keywords: Quadruplex; Melting; Thermal stability; Fluorescence melting

1. Introduction

G-rich nucleic acid sequences can fold into four-stranded DNA structures that contain stacks of G-quartets [1–4]. These quadruplexes (tetraplexes) can be formed by the intermolecular association of four DNA molecules, dimerization of sequences that contain two G-tracts, or by the intramolecular folding of a single strand containing four blocks of guanines.

DNA melting studies have been widely employed to investigate the stability of duplex DNA and its interaction with ligands [5–7]. GC-rich sequences melt at higher temperatures than AT-containing ones and compounds that bind to double-stranded DNA selectively stabilise this form over the single-stranded random coil, thereby increasing the melting temperature. These duplex melting experiments are usually (though not exclusively) performed with high molecular weight fragments (natural DNAs or synthetic polynucleotides) and the DNA melts in a highly cooperative fashion. The melting transitions are usually detected by measuring the change in absorbance at 260 nm, which increases by about 25% on denaturation. Similar studies can be performed with higher order DNA structures (*e.g.*, triplexes or quadruplexes), though the results for triplexes may be difficult to interpret as two or more transitions are present (triplex to duplex and duplex to single strands). DNA quadruplexes show only small changes in absorbance at their UV maximum (260 nm), and so this technique has been less widely employed. However, a greater signal is obtained at 295 nm, at which there is a large *decrease* in absorbance on melting [8]. Other means of monitoring the melting behaviour have therefore been employed including circular dichroism [9,10], NMR [8,11] and fluorescence melting [12–16] studies. Two different CD spectral signatures have been described for quadruplexes, which are usually attributed to the presence of distinct conformations; a peak around 265 nm is thought to correspond to the parallel-stranded form, in which the nucleotides are arranged all *anti*, while a positive maximum around 295 nm corresponds to one of several antiparallel forms that contain nucleotides in both the *syn* and *anti* conformation [17–19]. The melting of both these forms can be

* Corresponding author. Fax +44 23 8059 4459.
E-mail address: K.R.Fox@soton.ac.uk (K.R. Fox).

1046-2023/$ - see front matter © 2007 Elsevier Inc. All rights reserved.
doi:10.1016/j.ymeth.2007.05.004

followed by the changes in the CD spectra at appropriate wavelengths [9,10,20–22]. In contrast to duplex DNA, most studies with quadruplex DNA have employed synthetic oligonucleotides which form intermolecular (tetramolecular and dimeric) or intramolecular structures. Since it is straightforward to add fluorescence reporter groups during oligonucleotide synthesis, the use of fluorescence melting profiles has become a popular method for measuring quadruplex stability and its interaction with ligands.

1.1. Absorbance

There is only a small (4%) change in absorbance at 260 nm (the wavelength used for studying duplex melting) on G-quadruplex formation. However, although the absolute absorbance is lower at higher wavelengths the change is greatest around 295 nm, for which there is a 50–80% *increase* in absorbance on quadruplex formation (*i.e.*, a *decrease* on melting) [8]. This is therefore the preferred wavelength for obtaining high quality melting data that can be used for determining thermodynamic properties of quadruplexes. Since this transition is inverted relative to that of duplex melting profiles, it provides a good means for comparing quadruplex and duplex formation. The values measured at this wavelength compare well with those determined by other methods, such as CD or NMR [8]. A typical absorbance melting profile is shown in Fig. 1b and c. The difference between the absorbance spectra of the folded and single-stranded forms (thermal difference spectrum) is also characteristic of quadruplex DNA, and is distinct from that of other DNA structures [23].

1.2. Circular dichroism

Since quadruplex DNA structures have distinctive circular dichroism spectra, temperature dependent changes in CD have often been used to determine quadruplex stability [9,10,20–22]. The wavelength used depends on the specific sequence (260 nm for intermolecular parallel complexes), but 295 nm for intramolecular antiparallel quadruplexes. These experiments require sophisticated equipment, which is not available in many laboratories, and use relatively high oligonucleotide concentrations (typically 5 μM or higher). In addition these experiments often do not have good temperature resolution.

1.3. Fluorescence

In this technique [12,16], synthetic oligonucleotides are prepared that contain a fluorophore (typically fluorescein) at one end and a fluorescence quencher (dabcyl or methyl red) at the other [12,24,25]. The principle of this method is illustrated in Fig. 1a. When the oligonucleotide adopts a folded configuration the reporter groups are close together and the fluorescence is quenched. When the structure melts these groups are separated and there is a large increase in the fluorescence signal. We routinely use this technique to compare the stability of different DNA quadruplexes [26–32]. A typical fluorescence melting profile is shown in Fig. 1b and c. Some laboratories use fluorescence donor and acceptor pairs (such as fluorescein and TAMRA) so that there is FRET (fluorescence energy transfer) between the donor and acceptor in a distance related fashion [13,14,33–38]. However, it should be noted that, although such FRET oligonucleotides give excellent results, the distance between the two groups in the folded complex is usually too short for FRET and the decrease in donor emission on quadruplex formation is probably the result of collisional quenching rather than FRET. This quenching can however be reduced by adding suitable spacer nucleotides between the quadruplex and the fluorescent probes [39,40]. Most experiments that use FRET pairs for fluorescence melting only report on changes in the fluorescence donor rather than the acceptor. For intermolecular complexes it is also possible to use oligonucleotides that possess a single fluorescent label which undergoes self-quenching on quadruplex formation [27]. If the quadruplex-forming region is located within a longer sequence the fluorophores and/or quenchers can be placed within the oligonucleotide (instead of at the 5′- and 3′-ends) usually by attachment to the 5-position of T, or by incorporation of dR-FAM as an internal unpaired base [41]. However, these internal modifications can have a greater effect on stability, which needs to be checked. Note that the fluorescent reporter groups are attached to the oligonucleotides by relatively long chemical linkers, and there is no significant difference in quenching efficiency between different folded forms. This technique therefore cannot be used to distinguish between the various quadruplex topologies (parallel and antiparallel).

1.3.1. Advantages

The fluorescence melting technique has several advantages over conventional UV absorbance studies. First, absorbance changes are not large (typically only 25%), while the fluorescence signal can change by 10-fold or greater. Second, UV absorbance melting is usually a low throughput technique, as most spectrophotometers handle no more than six samples at once. Third, the typical format used for fluorescence melting experiments (using real-time PCR machines as described below) only requires small amounts of material (typically 20 μl of a 0.25 μM solution) in contrast to UV studies that require relatively large volumes (1–3 ml) of a solution with an OD_{260} of at least 0.2 (*i.e.*, a total of about 20 nmole of bases). The lower concentrations will also favour the formation of intramolecular (rather than intermolecular) quadruplexes.

1.3.2. Comparison between methods

Although fluorescence melting is a convenient and simple technique, it requires the attachment of reporter molecules to the oligonucleotide which may influence quadruplex structure and/or stability. The results should therefore be compared with UV melting studies to assess

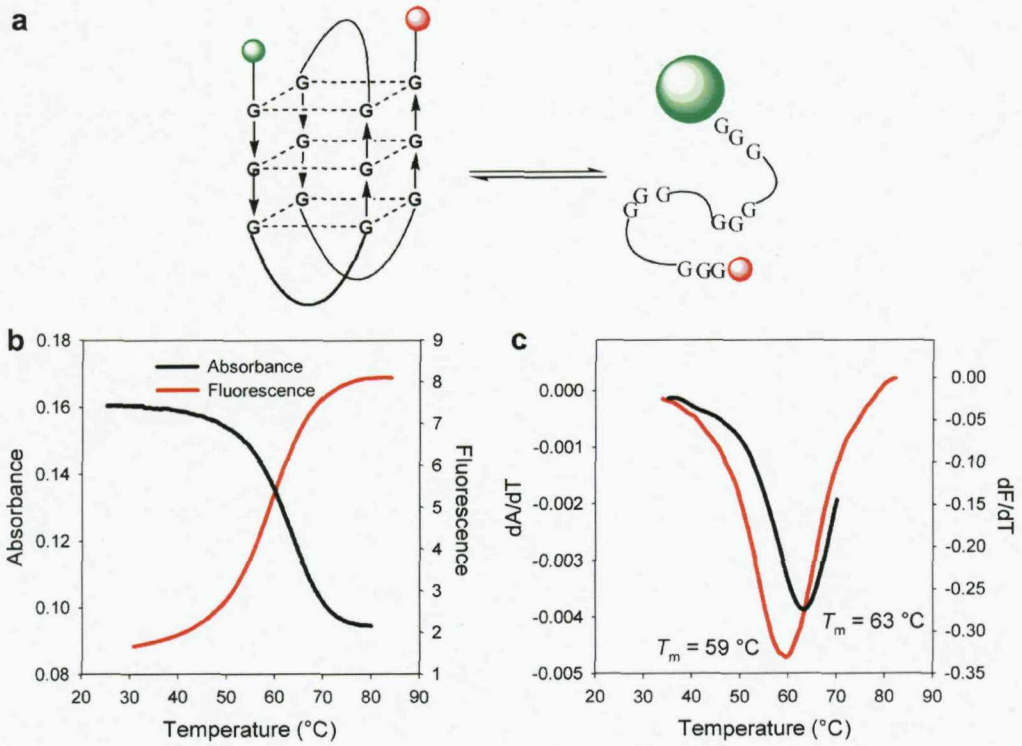


Fig. 1. (a) Schematic representation of the melting of a quadruplex-forming oligonucleotide that has been labelled with a fluorophore (green) and quencher (red). The four G_3-tracts are separated by loop sequences and the folded structure is drawn in an antiparallel topology for the sake of illustrative purposes only. (b) Comparison of melting curves obtained for GGGTTAGGGTTAGGGTTAGGG, determined by changes in absorbance at 295 nm (black line) and F-GGGTTAGGGTTAGGGTTAGGG-Q (F = 6-amidohexylfluorescein (FAM), and Q = C7 dabcyl) determined by changes in the fluorescence (red line) measured in10 mM lithium phosphate pH 7.4 containing 50 mM KCl. (c) First derivative plots of the data shown in (b). (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

the effects of the added fluorophores. Fig. 1b and c shows a comparison of fluorescence melting profiles with UV melting data with for labelled and unlabelled oligonucleotides based on the human telomeric repeat sequence. It can be seen that the two methods give very similar profiles, though addition of the fluorophores decreases the T_m by a few degrees. While fluorescent dyes have been shown to have relatively little effect on duplex [42] and triplex [43] stability, studies with fluorescently labelled quadruplex [14] and i-motif [44] forming oligonucleotides show a more pronounced effect on stability. Generally, the addition of fluorescent groups is destabilising, although the observed thermodynamic effects can be dye specific. Whenever possible, CD spectra should be determined to ensure that the added groups have not radically altered the topology, indeed there is one report of an instance in which FAM/TAMRA tagged oligos show differences in the CD spectra [14].

2. Methods

2.1. Equipment

Fluorescence melting can be performed using various real-time PCR platforms though, since most of these were not designed or marketed specifically for this type of melting curve analysis, there are a number of factors to consider when selecting the appropriate instrument.

2.1.1. Temperature range

The instrument should ideally have a temperature range which spans from ambient (or lower) to 95 °C. The Roche LightCycler® models 1 and 1.5 have temperature ranges between ambient and 95 °C; however, the latest model (LightCycler 2) cannot record below 40 °C, thereby limiting the usefulness of this model for fluorescence melting purposes.

2.1.2. Measuring melting and annealing profiles

In order to assess the reversibility of the transition, fluorescence data must be collected during both melting (heating) and annealing (cooling), since quadruplex melting/annealing profiles can show a considerable hysteresis. Some real-time PCR instruments only record during the melting phase, which does not allow the unambiguous determination of equilibrium melting curves.

2.1.3. Rate of temperature change

Real-time PCR machines typically heat and cool the samples as quickly as possible, with fast rate temperature

gradients (20 °C s^{-1}). However, it is critical that the instrument is capable of slower rates of heating and cooling. The Rotor-Gene instruments (3000 and 6000) allow some temperature gradient manipulation, although we have found precise rates are difficult to ascertain due to temperature equilibration lag times (*i.e.*, 1 °C min^{-1} over a 65 °C temperature range takes longer than 65 min; closer to 70–75 min). The Roche LightCycler allows a more accurate control of temperature gradients. The slowest rate of continuous temperature change on the LightCycler is 0.1 °C s^{-1}, but slower rates can be easily obtained by programming temperature changes in 1 °C steps, holding the samples at each temperature for a specified time before recording the fluorescence. In principle this can produce temperature gradients as slow as 0.01 °C min^{-1} (though we rarely work at slower rates than 0.1 °C min^{-1}). Although there are a number of instruments that can be used for fluorescence melting studies, we have used the LightCycler 1 or 1.5.

2.2. Choice of oligonucleotide sequence

For oligonucleotides with terminal fluorophores and quenchers, there are a number of issues regarding the exact location of these groups. Should they be attached immediately adjacent to the G-tracts, or separated from this by other nucleotides, and if so, which ones? We recommend that the fluorophore (usually fluorescein) should not be placed immediately adjacent to a G, as this is known to quench fluorescence [45–47], but that one or more bases should separate the reporter groups from the quadruplex. For studies with the human telomeric sequence the 5′ and 3′-terminal bases should be A and T, respectively, but for other model quadruplexes the choice is less clear. For any series of related quadruplexes (for instance when varying the nature of the loops) these bases should be kept constant. Table 1 shows the effect of altering the flanking bases on the stability of the human telomeric repeat sequence. The addition of extra bases adjacent to the G-quartets is destabilising. The identity of these bases is also significant; the presence of a 5′-A, which is found in the native telomeric sequence, lowers the stability more than a 5′-T. The addition of longer single-stranded ends further decreases quadruplex stability [14]. The fluorophore is usually placed at the 5′-end, with the quencher at the 3′-end, as this is the simplest and cheapest method for synthesis. The position of these fluorescent groups can be reversed, however we find that this can have an unusually large effect on quadruplex stability; placing methyl red at the 5′-end and fluorescein at the 3′-end increases the stability of the human telomeric repeat sequence by almost 10 °C, compared to the equivalent sequence with the fluorescent reporter groups the other way round (Table 1). For intermolecular complexes we observe that addition of T to the unlabelled end of an oligonucleotide increases quadruplex stability, while placing a T between the fluorophore and the oligonucleotide leads to a decrease in stability. dR-FAM-GGGGT is about 20 °C more stable than GGGGT-dR-FAM and TGGGG-dR-FAM is about 20 °C more stable than dR-FAM-TGGGG [27].

Table 1
Melting temperatures (T_m, °C) for quadruplex-forming oligonucleotides containing different nucleotides between adjacent to the fluorophore and quencher

Oligonucleotide sequence	T_m
FAM- GGGTTAGGGTTAGGGTTAGGG -dabcyl	59.4
FAM-**T** GGGTTAGGGTTAGGGTTAGGG **T**-dabcyl	58.1
FAM-**A** GGGTTAGGGTTAGGGTTAGGG **T**-dabcyl	55.9
FAM -**TT** GGGTTAGGGTTAGGGTTAGGG **TT**-dabcyl	52.8
Methyl red-**A** GGGTTAGGGTTAGGGTTAGGG **T**-FAM	64.8

The experiments were performed in 10 mM lithium phosphate pH 7.4 containing 50 mM KCl, and were heated at a rate of 0.2 °C min^{-1}. The flanking bases are shown in bold.

2.2.1. Purification

HPLC purification of the quadruplex-forming oligonucleotides is recommended, though this can be difficult due to the inherent ability of G-rich oligonucleotides to adopt folded structures. Oligonucleotides can also be purified by denaturing gel electrophoresis. However for terminally labelled oligonucleotides (with 5′-FAM), removal of $n-1$ products may not be a problem as the fluorophore is the last group that is added. As a result, only the full length products will have both the fluorophore and quencher attached and thus show a temperature dependent change in fluorescence.

2.3. Experimental procedure

Although it possible to use oligonucleotide concentrations of 100 nM or lower, for fluorescence melting experiments, we find that 0.25 μM gives a good signal. Higher oligonucleotide concentrations are required when examining the concentration dependency of the process and up to 10 μM can be used without saturating the instrument. Higher concentrations may be required at lower pHs as the fluorescence of fluorescein is pH sensitive (pK 6.4). LightCycler capillaries can use as little as 10 μl of material, though we normally work with a total reaction volume of 20 μl. In a typical experiment, we mix 5 μl each of buffer, oligonucleotide, and ligand (each at four times the final concentration), making the volume up to 20 μl with water. In this way it is simple to change the concentration of any reactants, or exchange one reactant for another (such as adding the complementary strand). In order to avoid any evaporation during the melting transition (which can take over 20 h for a complete melting and annealing profile at a temperature gradient of 0.1 °C min^{-1}), a small amount of mineral oil can be added to each sample.

2.3.1. Buffers

A wide range of buffers can be used for these experiments. Although quadruplex formation is not pH sensitive (unlike triplex formation) good pH control will require a

buffer that does not show a strong temperature dependency and for this reason Tris buffers should be avoided. We usually use phosphate buffers, though cacodylate is also widely used. Since quadruplex stability is dependent on the nature and concentration of the monovalent cations, we avoid buffers that contain sodium or potassium. Since lithium does not support the formation of most quadruplexes, we use 10 mM lithium phosphate pH 7.4 as the base buffer, and add NaCl or KCl to this as required. Although it would be best to work with physiological potassium ion concentrations (~140 mM) this is often not possible as some quadruplexes are too stable under these conditions. KCl (50–100 mM) is typically used, though it is sometimes necessary to lower this to 1 mM KCl.

2.3.2. Annealing and temperature profiles

Since some quadruplexes can adopt a number of metastable forms it essential that oligonucleotides should be properly annealed before use and not simply diluted from a stock solution. This can be achieved by heating the mixture to 95 °C and slowly cooling to 30 °C (as slowly as possible) before recording the melting profiles. There is often hysteresis between the melting and annealing curves at fast rates of temperature change, as a result of the slow kinetics of quadruplex association and dissociation, since the reaction is not at thermodynamic equilibrium during the reaction. An example of such hysteresis is shown in Fig. 2 for the melting profiles of F-$TG_4T_4G_4TG_4T_4G_4$T-Q. It can be seen that at fast rates of heating (6 °C min^{-1}; the slowest continual rate of temperature change on the LightCycler) the melting and annealing curves show a 13 °C difference in T_m (62 °C for annealing, compared with 75 °C for melting). When the rate of temperature change is slowed to 1 °C min^{-1} the hysteresis is reduced to 5 °C (68 °C for annealing, 73 °C for melting). Only on slowing to 0.2 °C min^{-1} do the melting and annealing curves coincide with a T_m of 70 °C. Although this hysteresis can be used to determine kinetic parameters (see below) it should be avoided for most melting experiments. It is therefore essential to confirm the reversibility of the process by comparing melting and annealing curves at different rates of temperature change. We therefore adopt a profile in which the samples are repeatedly melted and annealed at different rates. In a typical profile the samples are heated and cooled at between 1–6 °C min^{-1}, followed by subsequent melts at slower rates (0.05–0.2 °C min^{-1}). The slowest rate of continual sampling on the LightCycler is 0.1 °C s^{-1} though slower rates can be achieved by programming the machine to increase the temperature in steps of 1 °C, leaving the samples to equilibrate for specified times, and recording the fluorescence at the end of this time interval.

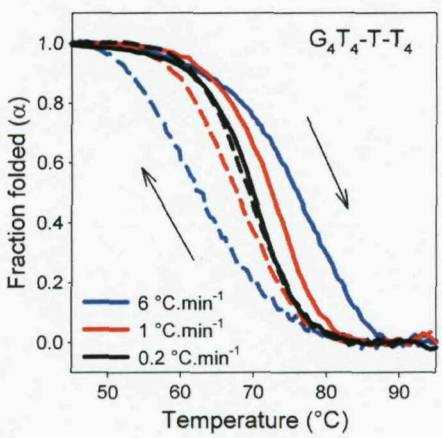


Fig. 2. Effects of different rates of heating and cooling on the fluorescence profiles of F-TGGGGTTTTGGGGTGGGGTTTTGGGGT-Q. Solid lines are melting curves while dashed lines show the annealing. Black, 0.2 °C min^{-1}; red, 1 °C min^{-1}; blue, 6 °C min^{-1}. The graphs show the fraction folded as a function of temperature. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

3. Parameters to be determined

3.1. T_m or $T_{1/2}$

The most common application of melting curves is to compare closely related oligonucleotides or to assess the relative binding strengths of different quadruplex-binding ligands. For a simple comparison the melting temperature (T_m) of each transition is determined and compared between samples. The T_m is the mid-point of a melting curve at which the complex is 50% dissociated. This can be estimated either from the temperature at which the fluorescence is midway between the initial and final values or, more accurately, from the maximum in the first derivative of the melting profile. In conditions where the kinetics are known to be slow, and the reaction is not at thermodynamic equilibrium, this point is often referred to as the $T_{1/2}$ [48,49]. If the unfolding is a single step then there will only be two species in equilibrium (folded and unfolded) and the relative amounts of each species can be estimated by assuming that the florescence at high temperatures corresponds to the unfolded random coil, while that at low temperatures corresponds to the folded form. The fraction folded (α) can then be calculated at each temperature. A complicating factor is that sloping baselines are often observed at high and low temperatures, and it is therefore necessary to fit straight lines to these, which are extrapolated to higher and lower temperatures. These sloping baselines are generally assumed to reflect the temperature dependence of the fluorescence, but they may also indicate that some other molecular process is occurring. Fig. 3 illustrates how the fraction folded is determined from the melting profiles. Comparison of the T_m values of different sequences enables a simple comparison of their relative stabilities. This is illustrated in Fig. 4 which compares the fluorescence melting profiles (converted into fraction folded plots) for a series of quadruplex-forming oligonucleotides with repeats of G_5T, G_4T_2, and G_3T_3, from which it

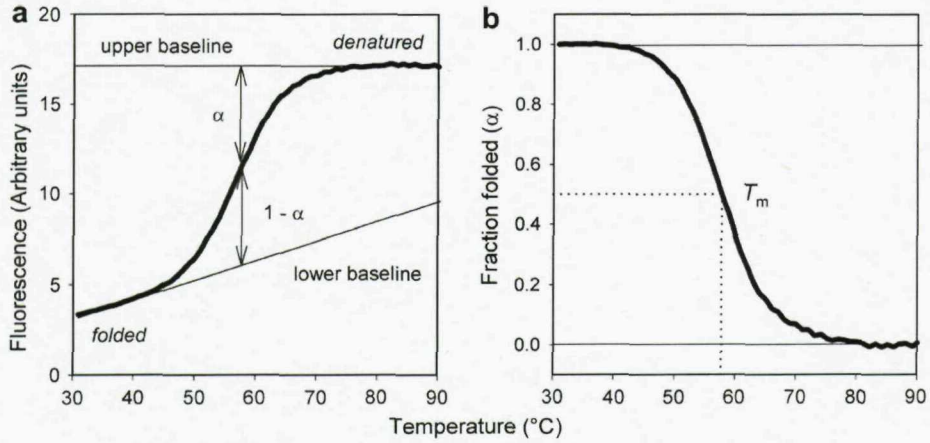


Fig. 3. Determining fraction folded plots from melting profiles. (a) Original fluorescence (or absorbance data). The fraction folded at each temperature is determined by the distance between the unfolded (high temperature) and folded (low temperature) states. The sloping baselines at high and low temperatures are fitted by straight lines, which are extrapolated to higher and lower temperatures. (b) Fraction folded (α) derived from the melting data shown in (a). The T_m (or $T_{1/2}$) corresponds to the temperature of the mid-point of the transition (when $\alpha = 0.5$).

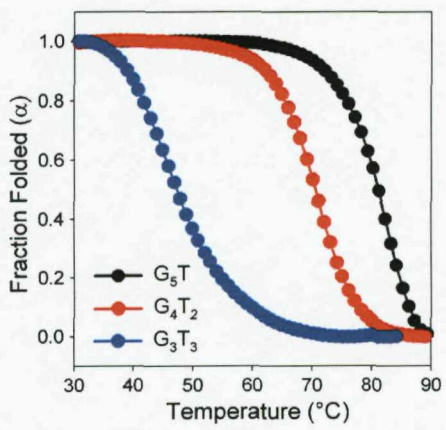


Fig. 4. Examples of fluorescence melting curves. These were determined for the oligonucleotides F-T(G_5T)$_4$-Q, black; F-T(G_4T_2)$_3G_4$T-Q, red; F-T(G_3T_3)$_3G_3$T-Q, blue. The data have been converted into fraction folded (α) plots. The samples were prepared in 10 mM lithium phosphate containing 10 mM KCl and were heated at a rate of 0.2 °C min^{-1}. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

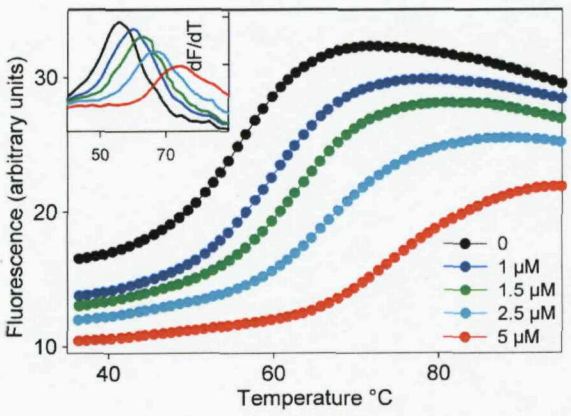


Fig. 5. Effect of various concentrations of BRACO20 on the fluorescence melting curves of F-A(GGGTTA)$_3$GGGT-Q. 0 μM, black; 1 μM, blue; 1.5 μM, green; 2.5 μM, cyan; 5 μM, red. The oligonucleotide concentration was 0.25 μM and the reactions were performed in 10 mM lithium phosphate pH 7.4 containing 50 mM KCl. The inset shows the first differential of these melting profiles. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

can be seen that the sequences with longer G-tracts are more stable.

One of the most important uses of melting curves is to assess the interaction with quadruplex-binding ligands [13,36,50–53] Fig. 5 shows examples of the effects of the tri-substituted acridine BRACO-20 [54] on the melting of an oligonucleotide containing repeats of the human telomeric sequence. As expected the ligand increases the melting temperature, though it should be noted that it quenches the fluorescence and the signal is reduced at higher ligand concentrations. There are several different ways to compare the relative affinities of quadruplex-binding ligands. The simplest means is to determine the changes in melting temperature (ΔT_m) produced by different ligands at the same at concentration (often 1 or 10 μM) [36,55]. This may also be expressed as the ligand concentration required to produce a given ΔT_m [15,34,36]. Alternatively pseudo-binding curves are fitted to plots of ΔT_m against concentration, producing values of $\Delta T_{m,max}$, the (theoretical) maximum increase in melting temperature produced at high ligand concentrations and C_{50}, the concentration required to produce half of this maximum change [56]. It should noted that these parameters have no physical basis (for example C_{50}s should not be confused with binding constants as each of the values on the plot will be determined at a different temperature) but they provide a useful means for comparing related ligands. The maximum increase in T_m is often the same for a series of related compounds and the ligand concentration that produces half this change is a good

indicator of relative binding affinity. It should be noted that C_{50} values that are similar to or less than the oligonucleotide concentration are of questionable quantitative significance. The observation of simple melting curves in the presence of a quadruplex-binding ligand indicates that the free and bound ligands are in relatively fast exchange. Slow exchange rates would generate two different oligonucleotide populations (free and bound) which would melt at different temperatures; the relative amounts of each would vary according to the ligand concentration. We are not aware of any published examples of this phenomenon for quadruplex-binding ligands, but it has been observed for a number of minor groove binding ligands [56].

3.2. Thermodynamic parameters (ΔH, ΔG, ΔS)

A more quantitative analysis of melting data for different quadruplexes involves fitting equations to the profiles. This is relatively straightforward for intramolecular quadruplexes as the folding and unfolding process is a simple unimolecular reaction. If the unfolding is a single step then there will only be two species in equilibrium (folded and unfolded) and the relative proportions at each temperature will be determined by the equilibrium constant ($K = C_{\text{folded}}/C_{\text{unfolded}}$). If the fraction folded (α) is calculated at each temperature then $K = \alpha/(1-\alpha)$. Since $K_T = K_0\exp(-\Delta H/\text{RT})$, ΔH can be obtained by plotting $\ln(K)$ against $1/T$, or by fitting the data with suitable programmes. Note that this is only appropriate for unimolecular processes and a more complex analysis is required to describe the melting of bimolecular or tetramolecular complexes, which accounts for the absolute concentration of the species. This analysis assumes that ΔH is independent of temperature (*i.e.* $\Delta C_p = 0$), that the reaction is only a two step process (*i.e.* that there are no significant reaction intermediates) and that there is only one folded form of the quadruplex.

Since $\Delta G = \Delta H - T\Delta S$ and $\Delta G = 0$ at the T_m (*i.e.*, $K = 1$) ΔS is simply estimated as $\Delta H/T_m$. Reported ΔH values are typically between -65 and 100 kJ mol^{-1} per quartet for an intramolecular quadruplex [28,57,58], while ΔS is usually negative. Since quadruplexes contain specifically bound cations, ΔH values show a strong dependence on ionic strength [28,29,59]. The slope of plots of ΔG against $\log[M^+]$ can be used to determine Δn, the difference between the number of ions bound in the folded and unfolded states [28,29,59,60].

3.3. Kinetic parameters

Melting conditions are usually chosen so as to minimise any hysteresis between the melting and annealing curves. However, kinetic parameters can be derived by analysing the difference between the heating and cooling profiles [61–63]. If α_c and α_h represent the fraction folded in the cooling (α_c) and heating (α_h) curves at any temperature (T), then $d(\alpha_c)/dT = d(\alpha_c)/dt \times (dT/dt)^{-1}$ and $d(\alpha_h)/dT = d(\alpha_h)/dt \times (dT/dt)^{-1}$. If k_1 and k_{-1} are the association and dissociation rate constants for quadruplex folding then $d(\alpha_c)/dt = k_1(1-\alpha_c) - k_{-1}\alpha_c$ and $d(\alpha_h)/dt = k_1(1-\alpha_h) - k_{-1}\alpha_h$. By measuring $d(\alpha_c)/dT$, $d(\alpha_h)/dT$, α_c and α_h, the individual rate constants can be estimated at each temperature. This analysis requires that there is sufficient difference between the melting and cooling curves, but also that there is some overlap between the two profiles (*i.e.*, that there are temperatures for which both $0 < \alpha < 1$ for both heating and cooling). The analysis can be checked by confirming that the estimated rates constants are independent of the rate of heating and cooling. Since $k_{-1} = k_1$ at the T_m, this can be used to estimate T_m values for complexes with very slow kinetics for which some hysteresis is still evident even at slow rates of heating and cooling.

The kinetics of quadruplex unfolding can also be determined by measuring the rate of change of fluorescence after rapidly increasing the temperature as the reaction re-equilibrates [26,56,64]. These experiments can only be conducted at temperatures around the T_m (typically $T_m \pm 10$ °C) where there is significant change in fluorescence with temperature. The samples are first equilibrated at the lower temperature, followed by increasing the temperature (5–10 °C) at the fastest rate (20 °C s^{-1} on the LightCycler). This temperature change causes the quadruplex to partially unfold, moving along the melting curve. Although the theoretical dead-time under these conditions is only 0.25 s, we usually ignore any fluorescence changes that occur in the first 2 s. Successive temperature jumps can then be performed on the same sample. The time dependent changes in fluorescence can then be fitted with an exponential function, providing the relaxation rate constant (k) at each temperature. For a unimolecular reaction the relaxation rate constant is equal to the sum of the folding (k_1) and unfolding (k_{-1}) rate constants. Although it is not possible to determine the individual rate constants, the reaction is dominated by k_1 at low temperatures, while k_{-1} dominates at high temperatures [64].

Note that these kinetic analyses only provide rate constants at temperatures around the T_m. Arrhenius plots can be used to extrapolate to physiological temperatures, but the absolute values of the parameters obtained should treated with caution as this can involve a long extrapolation. This is further complicated by the observation that k_1 often shows a negative activation energy (*i.e.*, the association appears to be faster at lower temperatures) [26,62]. This is explained by suggesting that the reaction proceeds via a nucleation-zipper mechanism, in which transient complexes in the folding pathway are stabilised at lower temperatures.

4. Duplex–quadruplex equilibria

Fluorescence melting curves can also be used to compare the relative stability of duplexes and quadruplexes and the effect of ligands on these transitions. The principle of this melting assay is shown in Fig. 6. When the G-rich oligonucleotide binds to its complementary C-rich sequence, the fluorophore and quencher are separated by

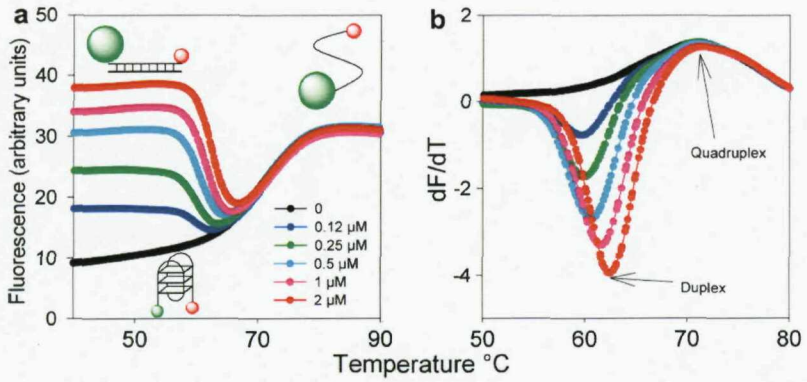


Fig. 6. (a) Effect of various concentrations of the complementary C-rich oligonucleotide on fluorescence melting curves of F-A(GGGTTA)$_3$GGGT-Q. 0 μM, black; 0.12 μM, blue; 0.25 μM, green; 0.5 μM, cyan; 1 μM, pink; 2 μM red. The concentration of the G-rich oligonucleotide was 0.25 μM and the reactions were performed in 50 mM potassium phosphate pH 7.4. (b) The first differential of these melting profiles shown in (a); the peaks corresponding to the duplex and quadruplex melts are indicated. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

a large distance, at opposite ends of the DNA duplex. This will produce a large fluorescence signal. Since the duplex is more rigid than the random coil, the time-averaged distance (affecting the collisional frequency) between the fluorophore and quencher will be greater in the duplex form, resulting in a larger fluorescence signal. The relative order of the fluorescence signals for the different species will therefore be duplex > single strand >> quadruplex. Addition of the complementary C-rich strand can have a number of different effects on quadruplex melting profiles. For the most stable quadruplexes, which form in preference to interacting with the complementary C-rich strand, the usual quadruplex melting profiles are observed [29]. In contrast, less stable quadruplexes do not form in the presence of the complement and a simple duplex melt is observed, with a corresponding decrease in fluorescence [29]. This situation is often observed in the presence of sodium. Quadruplexes with intermediate stability produce unusual melting profiles on addition of the C-rich complement (Fig. 6). At low temperatures the duplex form, which has a high fluorescence, predominates. This melts (accompanied by a small decrease in fluorescence) at a temperature below the quadruplex T_m, and the single-stranded G-rich strand folds to form a quadruplex, with the lowest fluorescence signal. The quadruplex then melts at higher temperatures in the usual fashion. The first (duplex) transition moves to higher temperatures in the presence of increasing concentrations of the C-rich oligonucleotide (see Fig. 6), as expected for an intermolecular reaction, in contrast to the T_m of the quadruplex which is a concentration-independent intramolecular process.

4.1. Effect of ligands on quadruplex–duplex equilibria

As well as comparing the relative stability of quadruplexes and duplexes, this equilibrium can be used to compare the interaction of ligands with each form, by examining their effects on the two transitions. Quadruplex-specific ligands increase the T_m of the higher transition, but have little or no effect on the melting of the duplex. A similar effect can be tested by adding unlabelled duplex DNA to a quadruplex–ligand complex. If the compound selectively binds to the quadruplex then the addition of duplex DNA will have little or no effect on ΔT_m, while compounds that are less selective will be sequestered by the added duplex, producing a reduction in ΔT_m. Examples showing the effect of different ligands on quadruplex–duplex equilibria are shown in Fig. 7. BRACO-20 (Fig. 7a) binds better to quadruplex than duplexes; it decreases the intensity of the duplex-single strand transition, having little effect on the T_m of this transition, while shifting the quadruplex-single strand transition to higher temperatures. In contrast distamycin and echinomycin, which are duplex-binding ligands, merely shift the duplex-single strand transition to higher temperatures (Fig. 7b), without generating a quadruplex melt. Ethidium (Fig. 7c), which can bind to both quadruplexes and duplexes, also appears to decrease the intensity of the duplex transition, but has only a small effect on the quadruplex melting temperature.

These experiments with complementary C-rich oligonucleotides should normally be used only to provide a qualitative indication of the selectivity for duplexes and quadruplex; a full quantitative analysis would require (i) dynamic equilibrium between the quadruplex and duplex forms, which will be very slow and (ii) a comparison of the competition between three bimolecular reactions (the interaction of the ligand with the duplex or the quadruplex and the annealing of the two duplex strands), with the unimolecular folding of the intramolecular quadruplex.

5. Potential problems

5.1. Reversibility

As noted above it is important to ensure that the melting temperature is independent of concentration, to

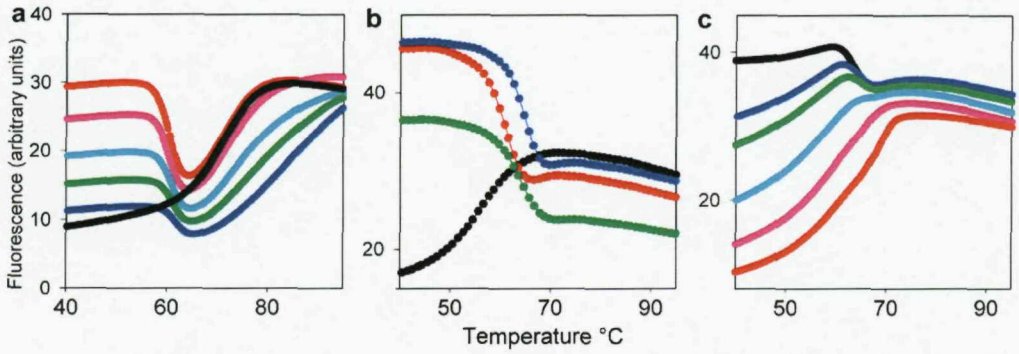


Fig. 7. Effects of different ligands on the fluorescence melting curves of 0.25 μM F-A(GGGTTA)$_3$GGGT-Q in the presence of the complementary C-rich oligonucleotide. (a) BRACO-20. G-rich oligonucleotide alone, black; in the presence of 0.5 μM C-rich complement and the following concentrations of BRACO-20: 0 μM, red; 0.4 μM, pink; 0.6 μM, cyan; 0.8 μM green; 1 μM, blue. (b) G-rich oligonucleotide alone, black; in the presence of 1 μM C-rich complementary oligonucleotide, red; 20 μM distamycin, blue or 20 μM echinomycin, green. (c) G-rich oligonucleotide plus 0.25 μM C-rich complementary oligonucleotide and various concentrations of ethidium: 0 μM, black; 0.6 μM , blue; 1 μM, green; 2 μM, cyan; 4 μM, pink; 10 μM red. All the reactions were performed in 10 mM lithium phosphate pH 7.4 containing 50 mM KCl. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

confirm that it is a unimolecular process. However, concentration-independent melting curves are also observed for some intermolecular complexes, for which re-association is exceptionally slow and dissociation is effectively irreversible. However the irreversible nature of these melting profiles means that the kinetic parameters of association and dissociation can be monitored independently [49]. This emphasises the importance of comparing melting and annealing curves. An additional problem is that some ligands are chemically unstable at elevated temperatures. Their effectiveness therefore decreases on repeated melting and annealing as the concentration of the active ligand decreases.

5.2. Fluorophores

The addition of fluorophores and quenchers may affect the quadruplex stability and this should be checked by comparing the results with other techniques, such as UV absorbance studies. Other studies with duplex DNA have shown that addition of different fluorescent reporters has only small effect on stability [42].

5.3. Ligands

We find that some ligands (such as the porphyrin TmPyP4) cannot be studied by this technique as they quench the fluorescence. Some ligands are insoluble in aqueous buffers and stock solutions are therefore prepared in organic solvents such as DMSO. It is important to check the effect of these solvents on quadruplex stability; we find that even low concentrations of DMSO can increase the T_m of intramolecular quadruplexes as shown in Table 2. Others have also shown that methanol and ethanol stabilise DNA quadruplexes [65].

Table 2
Effect of DMSO concentration (v/v) on the T_m (°C) of the quadruplex-forming oligonucleotides F-A(GGGTTA)3GGGT-Q (F, fluorescein; Q, methyl red)

DMSO % (v/v)	T_m (°C) (ΔT_m)
0	63.8 (–)
1	64.2 (0.4)
2	64.4 (0.5)
5	64.9 (1.1)
10	67.7 (3.9)
15	68.8 (6.0)
20	72.2 (8.4)
30	78.8 (14.5)
40	84.2 (20.7)

The experiments were performed in 10 mM lithium phosphate pH 7.4 containing 50 mM KCl, and were heated at a rate of 0.2 °C min^{-1}.

5.4. Multiple structures

Melting profiles reveal nothing about quadruplex topology and structure, though parallel topologies are usually more stable than antiparallel ones, and potassium ions produces more stable complexes than sodium. Analysis of the melting profiles assumes that there is only one folded conformation present in solution. This needs to be confirmed by other techniques, such as gel electrophoresis or NMR. NMR studies suggest that even the simple human telomeric repeat adopts more than one stable conformation in solution. In principle a mixture of non-interconverting stable conformers would generate biphasic melting profiles. Although this is rarely observed, biphasic melting profiles have been noted for the melting of some quadruplex-forming oligonucleotides [26,29] and mixtures of intermolecular complexes [27]. If two conformers in a mixture have similar melting temperatures then their profiles will overlap and generate what appears to be a single transition, though

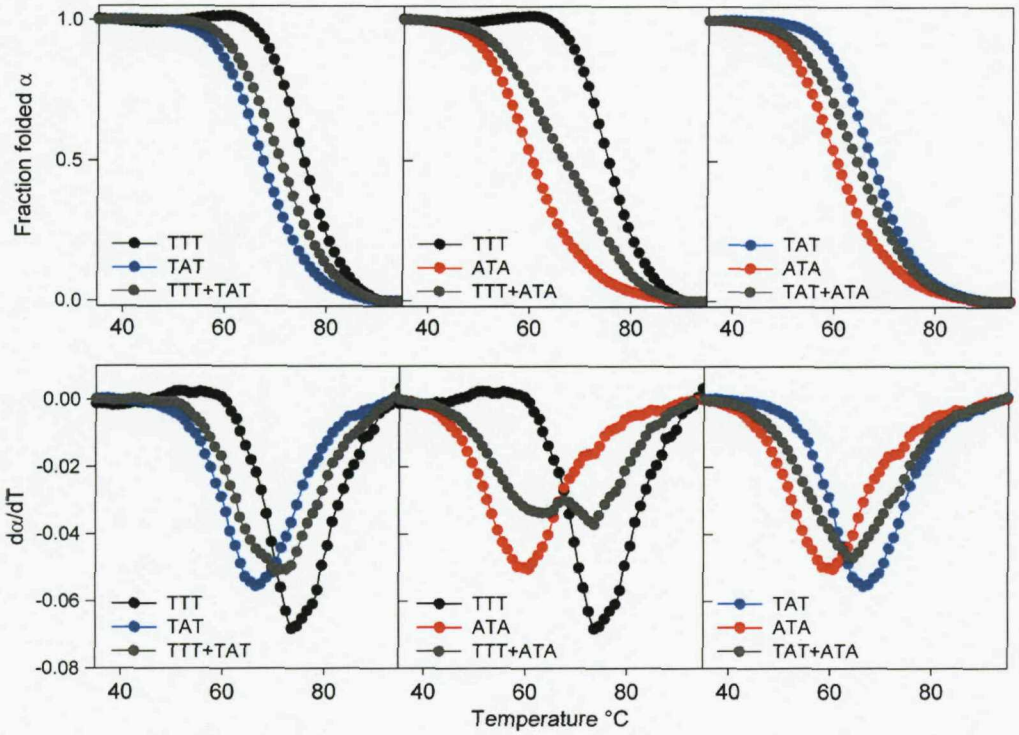


Fig. 8. Fluorescence melting profiles for mixtures of G-quadruplex-forming oligonucleotides. TTT = F-TGGGTGGGTGGGTGGGT-Q, black; TAT = F-TGGGTGGGAGGGTGGGT-Q, blue; ATA = F-TGGGAGGGTGGGAGGGT-Q, red. Mixtures of equimolar concentrations of pairs of oligonucleotides are shown in grey. In each case the total oligonucleotide concentration was 0.25 μM and the reactions were performed in 10 mM lithium phosphate pH 7.4, containing 1 mM KCl. The upper panels show plots of the fraction folded (α) as a function of temperature, while the lower panels show the first derivatives of these profiles. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

with a lower slope. Since the slope is related to ΔH the presence of multiple species will appear as a single transition with lower ΔH. This is illustrated in Fig. 8, showing melting profiles of 1:1 mixtures of F-$TG_3TG_3TG_3TG_3T$-Q (TTT), F-$TG_3TG_3AG_3TG_3T$-Q (TAT) and F-$TG_3AG_3TG_3AG_3T$-Q (ATA). Each of these generates simple melting transition with T_ms of 73.9, 67.3, and 59.6 °C and ΔH values of −278, −247, and −198 kJ mol^{-1}, respectively [57]. The mixture of TAT and ATA also produces a single transition with an intermediate T_m of 62.8 °C, but an apparent ΔH of −179 kJ mol^{-1}, which is lower than either of the individual species. A similar effect is seen with TTT and TAT in which the mixture has a T_m of 70.8 °C and ΔH of −215 kJ mol^{-1}. In contrast the mixture of TTT and ATA, which have the most diverse T_m values shows a biphasic profile with two clear peaks in the first differentials, which correspond to the T_ms of the individual components. This demonstrates that mixtures of species can give the appearance of anomalously low ΔH values and emphasises that the presence of single conformations should be confirmed by other means.

Acknowledgments

We thank Prof. Stephen Neidle for providing the samples of BRACO-20. Work in the authors' laboratory is funded by the BBSRC and EU (LSHB-CT-2004-005204). P.A.R. is supported by a BBSRC research studentship.

References

[1] S. Burge, G.N. Parkinson, P. Hazel, A.K. Todd, S. Neidle, Nucleic Acids Res. 34 (2006) 5402–5415.
[2] J.T. Davis, Angew. Chem. Int. Ed. Engl. 43 (2004) 668–698.
[3] A.T. Phan, V. Kuryavyi, D.J. Patel, Curr. Opin. Struct. Biol. 16 (2006) 288–298.
[4] T. Simonsson, Biol. Chem. 382 (2001) 621–628.
[5] J. SantaLucia, in: M.G. Gore (Ed.), Spectrometry and Fluorimetry, Oxford University Press, Oxford UK, 2000, pp. 329–356.
[6] W.D. Wilson, F.A. Tanious, Methods Mol. Biol. 90 (1997) 219–240.
[7] A.S. Benight, P. Pancoska, R. Owczarzy, P.M. Vallone, J. Nesetril, P.V. Riccelli, Method Enzymol. 340 (2001) 165–192.
[8] J.L. Mergny, A.T. Phan, L. Lacroix, FEBS Lett. 435 (1998) 74–78.
[9] L. Petraccone, G. Barone, C. Giancola, Curr. Med. Chem. AntiCancer Agents 13 (2005) 463–475.
[10] L. Petraccone, E. Erra, V. Esposito, A. Randazzo, A. Galeone, G. Barone, C. Giancola, Biopolymers 77 (2005) 75–85.
[11] C.C. Hardin, E. Henderson, T. Watson, J.K. Prosser, Biochemistry 30 (1991) 4460–4472.
[12] R.A.J. Darby, M. Sollogoub, C. McKeen, L. Brown, A. Risitano, N. Brown, C. Barton, T. Brown, K.R. Fox, Nucleic Acids Res. 30 (2002) e39.
[13] J.L. Mergny, L. Lacroix, M.P. Teulade-Fichou, C. Hounsou, L. Guittat, M. Hoarau, P.B. Arimondo, J.P. Vigneron, J.M. Lehn, J.F. Riou, T. Garestier, C. Helene, Proc. Natl. Acad. Sci. USA 98 (2001) 3062–3067.

[14] J.L. Mergny, J.C. Maurizot, Chembiochem. 2 (2001) 124–132.
[15] C.M. Schultes, W. Guyen, J. Cuesta, S. Neidle, Bioorg. Med. Chem. Lett. 14 (2004) 4347–4351.
[16] A. De Cian, L. Guittat, M. Kaiser, B. Saccà, S. Amrane, A. Bourdoncle, P. Alberti, M.-P. Teulade-Fichou, L. Lacroix, J.-L. Mergny, Methods 42 (2007) 183–195.
[17] P. Balagurumoorthy, S.K. Brahmachari, J. Biol. Chem. 269 (1994) 21858–21869.
[18] M. Lu, Q. Guo, N.R. Kallenbach, Biochemistry 31 (1992) 2455–2459.
[19] I.N. Rujan, J.C. Meleney, P.H. Bolton, Nucleic Acids Res. 33 (2005) 2022–2031.
[20] V. Esposito, A. Randazzo, G. Piccialli, L. Petraccone, C. Giancola, L. Mayol, Org. Biomol. Chem. 2 (2004) 313–318.
[21] C.M. Olsen, W.H. Gmeiner, L.A. Marky, J. Phys. Chem. B 110 (2006) 6962–6969.
[22] Y. Xu, Y. Noguchi, H. Sugiyama, Bioorg. Med. Chem. 14 (2006) 5584–5591.
[23] J.L. Mergny, J. Li, L. Lacroix, S. Amrane, J.B. Chaires, Nucleic Acids Res. 33 (2005) e138.
[24] L. Guittat, A. De Cian, F. Rosu, V. Gabelica, E. De Pauw, E. Delfourne, J.L. Mergny, Biochim. Biophys. Acta 1724 (2005) 375–384.
[25] B. Juskowiak, Curr. Anal. Chem. 2 (2006) 261–270.
[26] N.M. Brown, P.A. Rachwal, T. Brown, K.R. Fox, Org. Biomol. Chem. 3 (2005) 4153–4157.
[27] E.E. Merkina, K.R. Fox, Biophys. J. 89 (2005) 365–373.
[28] P.A. Rachwal, T. Brown, K.R. Fox, Biochemistry 46 (2007) 3036–3044.
[29] A. Risitano, K.R. Fox, Biochemistry 42 (2003) 6507–6513.
[30] A. Risitano, K.R. Fox, Org. Biomol. Chem. 1 (2003) 1852–1855.
[31] A. Risitano, K.R. Fox, Nucleic. Acids Res. 32 (2004) 2598–2606.
[32] A. Risitano, K.R. Fox, Bioorg. Med. Chem. Lett. 15 (2005) 2047–2050.
[33] J.J. Green, S. Ladame, L.M. Ying, D. Klenerman, S. Balasubramanian, J. Am. Chem. Soc. 128 (2006) 9809–9812.
[34] B. Guyen, C.M. Schultes, P. Hazel, J. Mann, S. Neidle, Org. Biomol. Chem. 2 (2004) 981–988.
[35] K. Jantos, R. Rodriguez, S. Ladame, P.S. Shirude, S. Balasubramanian, J. Am. Chem. Soc. 128 (2006) 13662–13663.
[36] M.J.B. Moore, C.M. Schultes, J. Cuesta, F. Cuenca, M. Gunaratnam, F.A. Tanious, W.D. Wilson, S. Neidle, J. Med. Chem. 49 (2006) 582–599.
[37] T. Simonsson, R. Sjoback, J. Biol. Chem. 274 (1999) 17379–17383.
[38] L.M. Ying, J.J. Green, H.T. Li, D. Klenerman, S. Balasubramanian, Proc. Natl. Acad. Sci. USA 100 (2003) 14629–14634.
[39] J.L. Mergny, J.F. Riou, P. Mailliet, M.P. Teulade-Fichou, E. Gilson, Nucleic Acids Res. 30 (2002) 839–865.
[40] S. Nagatoishi, T. Nojima, E. Galezowska, B. Juskowiak, S. Takenaka, Chembiochem 7 (2006) 1730–1737.
[41] J. Casals, L. Debethune, K. Alvarez, A. Risitano, K.R. Fox, A. Grandas, E. Pedroso, Bioconjug. Chem. 17 (2006) 1351–1359.
[42] B.G. Moreira, Y. You, M.A. Behlke, R. Owczarzy, Biochem. Biophys. Res. Commun. 327 (2005) 473–484.
[43] Y. Wang, D.A. Rusling, V.E.C. Powers, O. Lack, S.D. Osborne, K.R. Fox, T. Brown, Biochemistry 44 (2005) 5884–5892.
[44] J.L. Mergny, Biochemistry 38 (1999) 1573–1581.
[45] J.P. Cooper, P.J. Hagerman, Biochemistry 29 (1990) 9261–9268.
[46] T. Maruyama, T. Shinohara, H. Ichinose, M. Kitaoka, N. Okamura, N. Kamiya, M. Goto, Biotechnol. Lett. 27 (2005) 1349–1354.
[47] S.A.E. Marras, F.R. Kramer, S. Tyagi, Nucleic Acids Res. 30 (2002).
[48] A. De Cian, E. DeLemos, J.L. Mergny, M.P. Teulade-Fichou, D. Monchaud, J. Am. Chem. Soc. 129 (2007) 1856–1857.
[49] J.L. Mergny, A. De Cian, A. Ghelab, B. Sacca, L. Lacroix, Nucleic Acids Res. 33 (2005) 81–94.
[50] F.X.G. Han, R.T. Wheelhouse, L.H. Hurley, J. Am. Chem. Soc. 121 (1999) 3561–3570.
[51] F. Koeppel, J.F. Riou, A. Laoui, P. Mailliet, P.B. Arimondo, D. Labit, O. Petitgenet, C. Helene, J.L. Mergny, Nucleic Acids Res. 29 (2001) 1087–1096.
[52] G.S. Minhas, D.S. Pilch, J.E. Kerrigan, E.J. Lavoie, J.E. Rice, Bioorg. Med. Chem. Lett. 16 (2006) 3891–3895.
[53] M.P. Teulade-Fichou, C. Carrasco, L. Guittat, C. Bailly, P. Alberti, J.L. Mergny, A. David, J.M. Lehn, W.D. Wilson, J. Am. Chem. Soc. 125 (2003) 4732–4740.
[54] M. Read, R.J. Harrison, B. Romagnoli, F.A. Tanious, S.H. Gowan, A.P. Reszka, W.D. Wilson, L.R. Kelland, S. Neidle, Proc. Natl. Acad. Sci. USA 98 (2001) 4844–4849.
[55] A.D. Moorhouse, A.M. Santos, M. Gunaratnam, M. Moore, S. Neidle, J.E. Moses, J. Am. Chem. Soc. 128 (2006) 15972–15973.
[56] P.L. James, T. Brown, K.R. Fox, Nucleic Acids Res. 31 (2003) 5598–5606.
[57] P.A. Rachwal, T. Brown, K.R. Fox, FEBS Letts. 581 (2007) 1657–1660.
[58] F. Pileur, M.L. Andreola, E. Dausse, J. Michel, S. Moreau, H. Yamada, S.A. Gaidamakov, R.J. Crouch, J.J. Toulme, C. Cazenave, Nucleic Acids Res. 31 (2003) 5776–5788.
[59] N.J. Jing, R.F. Rando, Y. Pommier, M.E. Hogan, Biochemistry 36 (1997) 12498–12505.
[60] C.R. Cantor, P.R. Schimmel, in: Biophysical Chemistry, W H Freedman and Company, New York, 1980.
[61] E. Bernal-Mendez, C.J. Leumann, Biochemistry 41 (2002) 12343–12349.
[62] J.L. Mergny, L. Lacroix, Oligonucleotides 13 (2003) 515–537.
[63] M. Rougee, B. Faucon, J.L. Mergny, F. Barcelo, C. Giovannangeli, T. Garestier, C. Helene, Biochemistry 31 (1992) 9269–9278.
[64] P.A. Rachwal, I.S. Findlow, J.M. Werner, T. Brown, K.R. Fox, Nucleic Acids Research 35, in press, doi:10.1093/nar/gkm316.
[65] I.V. Smirnov, R.H. Shafer, Biopolymers 85 (2007) 91–101.

4214–4222 Nucleic Acids Research, 2007, Vol. 35, No. 12
doi:10.1093/nar/gkm316

Published online 18 June 2007

Intramolecular DNA quadruplexes with different arrangements of short and long loops

Phillip A. Rachwal[1], I. Stuart Findlow[1], Joern M. Werner[1], Tom Brown[2] and Keith R. Fox[1,*]

[1]School of Biological Sciences, University of Southampton, Bassett Crescent East, Southampton SO16 7PX, UK and [2]School of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, UK

Received March 10, 2007; Revised and Accepted April 13, 2007

ABSTRACT

We have examined the folding, stability and kinetics of intramolecular quadruplexes formed by DNA sequences containing four G_3 tracts separated by either single T or T_4 loops. All these sequences fold to form intramolecular quadruplexes and 1D-NMR spectra suggest that they each adopt unique structures (with the exception of the sequence with all three loops containing T_4, which is polymorphic). The stability increases with the number of single T loops, though the arrangement of different length loops has little effect. In the presence of potassium ions, the oligonucleotides that contain at least one single T loop exhibit similar CD spectra, which are indicative of a parallel topology. In contrast, when all three loops are substituted with T_4 the CD spectrum is typical of an antiparallel arrangement. In the presence of sodium ions, the sequences with two and three single T loops also adopt a parallel folded structure. Kinetic studies on the complexes with one or two T_4 loops in the presence of potassium ions reveal that sequences with longer loops display slower folding rates.

INTRODUCTION

DNA sequences that contain four or more closely spaced G-tracts can fold to form intramolecular quadruplexes, which consist of stacked G-quartets that are linked by three loops between the four G-strands (1–4). These structures are stabilized by monovalent cations (especially potassium) (5,6) and can adopt a variety of different folding patterns dependent on the relative orientation of the strands and the position of the loops. G-rich sequences with the potential to form quadruplex structures are common in genomic DNA and these have been identified in several biologically important regions (7–9). The most widely studied is telomeric DNA, which in higher eukaryotes is composed of repeats of the sequence GGGTTA (10,11) and for which about 50–100 bases at the 3′-end are single stranded. A number of other non-telomeric G-rich DNA sequences may also form quadruplexes and these have been identified in the promoters of *c-myc* (12–15), Ki-*ras* (16), *bcl*2 (17–19), *c-kit* (20), VEGF gene (21) and HIF 1α (22), as well as in fragile X-syndrome (23) and other trinucleotide repeat sequences (24), the retinoblastoma susceptibility gene (25), the chicken β-globin gene (26) and the insulin gene (27). G-rich sequences are especially abundant in gene promoter regions (8) and there is an overabundance of G-rich sequences in the regulatory regions of muscle-specific genes (28).

For intramolecular quadruplexes, the four G-tracts are separated by loops. These are of various lengths and can be as short as a single nucleotide (29–31). Genomic searches (7,9) have revealed many G-rich sequences which may be able to adopt these structures, the most common of which are successive G-tracts that are separated by single T or A residues. The loops can be arranged in several different ways; double-chain reversal (propeller) loops link two adjacent parallel strands (32), while edgewise or diagonal loops link two antiparallel strands (33). Some structures contain both edge-wise and propeller loops (34–36). In the all-parallel (propeller) structures, the nucleotides are in the *anti* conformation, while the other structures have different combinations of *anti* or *syn* glycosidic bonds (3,4). It is known that loop length and sequence affect quadruplex stability and structure (29,37–40). Sequences with single nucleotide loops between the G_3 tracts only adopt a parallel structure, while longer loops can also adopt an antiparallel arrangement of the strands. Quadruplex stability is also affected by the sequence of the loops (39–41), and the bases that flank the quadruplex (42–44).

There is considerable variation in quadruplex structure, depending on the DNA sequence and the ionic conditions. The biological function of quadruplexes may well depend on the folded conformation that is adopted, especially if this involves interaction with specific proteins. Such an effect has been suggested for the NHE element of the c-myc promoter, which can in principle adopt multiple conformations. Since the loops can have a considerable

*To whom correspondence should be addressed. Tel: +44 23 8059 4374; Fax: +44 23 8059 4459; Email: k.r.fox@soton.ac.uk

effect on quadruplex folding and stability, we have examined how changes in loop length affect quadruplex properties. One very stable intramolecular quadruplex contains four G_3 tracts that are linked by single T residues (30,41,45) and this is known to be an inhibitor of HIV integrase. We have used variations on this sequence to examine the importance of loop length on quadruplex folding and stability. In this study, we have systematically replaced each of the single T loops with T_4 and have used CD, fluorescence melting, 1D-NMR, gel electrophoresis and kinetic studies to examine the effect of loop length and position on quadruplex folding and stability.

MATERIAL AND METHODS

Oligonucleotides

All oligonucleotides were synthesized on an Applied Biosystems ABI 394 automated DNA/RNA synthesiser on the 0.2 μmole scale using the standard cycles of acid-catalysed detritylation, coupling, capping and iodine oxidation procedures. Phosphoramidite monomers and other reagents were purchased from Applied Biosystems, Proligo and Link Technologies. The sequences of the oligonucleotides used in this work are shown in Table 1. Fluorescently labelled oligonucleotiodes were used in all the experiments. These were labelled at the 5′-end with 6-amidohexylfluorescein (FAM), and at the 3′-end with dabcyl using C7 dabcyl cpg (Link Technologies). Oligonucleotides were purified by gel filtration using Nap10 columns (GE Healthcare) and analysed by gel electrophoresis. The bases adjacent to the fluorophore and quencher were the same (T) for all the oligonucleotides to avoid any differences in their effects on quadruplex formation and stability.

Fluorescence melting studies

The thermal melting temperatures of the quadruplexes were determined using the fluorescence melting technique that we have developed (46) and have used previously for assessing the stability of related quadruplexes (39,41,44,47). When the sequence adopts a folded structure the quencher and fluorophore are in close proximity and the fluorescence is quenched. When the structure melts, these groups become separated and there is a large increase in fluorescence. Since the fluorophore and quencher are anchored on relatively long aliphatic tethers the quenching does not depend on the quadruplex topology and the fluorescence is quenched for both parallel and antiparallel complexes. Fluorescence melting experiments were conducted in a Roche LightCycler as previously described (39,41,44,46,47) in a total reaction volume of 20 μl. Oligonucleotides (final concentration 0.25 μM) were prepared in 10 mM lithium phosphate pH 7.4, which was supplemented with various concentrations of potassium chloride or sodium chloride. The LightCycler has one excitation source (488 nm) and the changes in fluorescence were measured at 520 nm. For several of the oligonucleotides initial experiments revealed that there was considerable hysteresis between the heating and annealing profiles when the temperature was changed at $0.2°C.s^{-1}$, indicating that the process was not at thermodynamic equilibrium. Melting experiments were therefore performed at a much slower rate of heating and cooling ($0.2°C.min^{-1}$) by changing the temperature in 1°C steps, leaving the samples to equilibrate for 5 min at each temperature before recording the fluorescence. Under these conditions, no hysteresis was observed (except for some experiments with G_3T_4). In a typical experiment, the oligonucleotides were first denatured by heating to 95°C for 5 min. They were then annealed by cooling to 30°C at $0.2°C.min^{-1}$ and melted by heating to 95°C at the same rate. The fluorescence was recorded during both the annealing and melting steps. In some instances, the formation of intramolecular or intermolecular complexes was examined by determining the melting curves using a range of oligonucleotide concentrations (0.1–10 μM). Melting temperatures (T_m values) were determined from the first derivatives of the melting profiles using the Roche LightCycler software.

Table 1. Sequences of the quadruplex-forming oligonucleotides used in this work

Name	Sequence						
	Loop 1		Loop 2		Loop 3		
G_3T	d-F-TGGG	T	GGG	T	GGG	T	GGGT-Q
G_3T-T_4-T	d-F-TGGG	T	GGG	TTTT	GGG	T	GGGT-Q
G_3T_4-T-T	d-F-TGGG	TTTT	GGG	T	GGG	T	GGGT-Q
G_3T_4-T-T_4	d-F-TGGG	TTTT	GGG	T	GGG	TTTT	GGGT-Q
G_3T_4-T_4-T	d-F-TGGG	TTTT	GGG	TTTT	GGG	T	GGGT-Q
G_3T_4	d-F-TGGG	TTTT	GGG	TTTT	GGG	TTTT	GGGT-Q

F = FAM; Q = dabcyl.

Thermodynamic and kinetic analysis

T_m values were obtained from the maxima of the first derivatives of the melting profiles using the LightCycler software or, together with ΔH, from van't Hoff analysis of the melting profiles using FigP for Windows. The fraction folded was calculated as previously described (48) from the difference between the measured fluorescence and the upper and lower baselines. All reactions were performed at least twice and the calculated T_m values usually differed by <0.5°C with a 5% variation in ΔH. Since $\Delta G = 0$ at the T_m, ΔS was estimated as $\Delta H/T_m$. Values for ΔG at 310 K were then estimated from $\Delta G = \Delta H - T\Delta S$. The van't Hoff analysis assumes that ΔH is independent of temperature (i.e. $\Delta C_p = 0$), that the reaction is only a two-step process (i.e. that there are no significant reaction intermediates) and that there is only one folded form of the quadruplex. The number of specifically bound monovalent cations (Δn), was calculated from the slopes of plots of ΔG against $\log[M^+]$ as previously described (30,49).

Hysteresis between the melting and annealing profiles occurs when the reaction is not at thermodynamic equilibrium as a result of the slow folding and/or unfolding kinetics. Individual folding (k_1) and unfolding

(k_{-1}) rate constants can be derived from this hysteresis as previously described (47,48,50,51).

Temperature jump kinetics

The kinetics of quadruplex unfolding were also determined by measuring the rate of change of fluorescence after rapidly increasing the temperature (47,52). The quadruplexes were equilibrated at a temperature around the T_m, which was then rapidly increased by 5°C at the fastest rate on the LightCycler (20°C.s^{-1}). This temperature change causes the quadruplex to partially unfold, moving along the melting curve. Although the theoretical dead-time under these conditions is only 0.25 s, all fluorescence changes that occurred in the first 2 s were ignored, during equilibration to the new temperature. Successive temperature-jumps were then recorded on the same sample by further increasing the temperature by 5°C. Each experiment was repeated at least twice. The time-dependent changes in fluorescence were fitted by an exponential function $F_t = F_f \times (1 - e^{-kt}) + F_0$, using SigmaPlot 10, where F_t is fluorescence at time t, F_0 is the initial fluorescence and F_f is total change in fluorescence (the final fluorescence is $F_f + F_0$). The relaxation rate constant (k) obtained from this analysis is equal to the sum of the folding (k_1) and unfolding (k_{-1}) rate constants. Arrhenius plots of $\ln(k)$ against $1/T$ were constructed from these data and used to estimate the activation energy E_a and pre-exponential factor A [$k = A \times \exp(-E_a/RT)$].

Gel electrophoresis

Non-denaturing gel electrophoresis was performed using 14% polyacrylamide gels, which were run in TBE buffer that had been supplemented with 20 mM KCl. Bands in the gels were visualised under UV light. The oligonucleotide concentration was 20 μM.

Circular dichroism

CD spectra were measured on a Jasco J-720 spectropolarimeter as previously described (39). Oligonucleotide solutions (5 μM) were prepared in 10 mM lithium phosphate pH 7.4, containing either 200 mM potassium chloride or 200 mM sodium chloride. The samples were heated to 95°C and annealed by slowly cooling to 15°C over a period of 12 h. Spectra were recorded between 220 and 320 nm in 5 mm path length cuvettes. Spectra were averaged over 10 scans, which were recorded at 100 nm.min^{-1} with a response time of 1 s and a bandwidth of 1 nm. A buffer baseline was subtracted from each spectrum and the spectra were normalized to have zero ellipticity at 320 nm.

Proton NMR

One-dimensional ^{1}H NMR experiments were performed on a Varian Inova 600 MHz spectrometer. Oligonucleotides were prepared in 200 mM potassium phosphate pH 7.4 and were annealed by heating to 95°C before slowly cooling to 15°C. 300 μl of the oligonucleotide sample was mixed with 20 μl D_2O and placed in a Shigemi NMR tube. The final strand concentration was 100 μM. 1D proton NMR spectra were recorded at 25°C with a sweep width of 25 p.p.m., WATERGATE water suppression, an acquisition time of 0.5 s and 32 k scans. Data were processed using VNMR software (Varian Inc.) with zero filling and resolution enhancement.

RESULTS

A variety of physical techniques were used to examine the folding, stability and kinetics of the intramolecular quadruplexes that are formed by sequences containing four G_3 tracts separated by either single T or T_4 loops, in different combinations. The sequences of these oligonucleotides are shown in Table 1.

Circular dichroism

Intramolecular quadruplexes can adopt a variety of different topologies in which the strands run in different orientations with lateral, edgewise or diagonal loops. Circular dichroism has frequently been used to indicate whether these fold in a parallel or antiparallel configuration (43,53,54). Parallel quadruplexes, in which the glycosidic bonds are all *anti*, display a positive CD signal around 265 nm, with a negative peak at 240 nm. In contrast, antiparallel topologies, with both *syn* and *anti* bonds, exhibit a positive signal at around 295 nm, with a negative signal or shoulder around 260 nm. CD spectra for these oligonucleotides, in the presence of sodium or potassium ions, are shown in Figure 1 [other studies with related sequences have shown that the fluorophores do not affect the CD spectra (55)].

In the presence of potassium (Figure 1) all the sequences, except G_3T_4 show CD spectra with a positive peak around 265 nm and a minimum around 240 nm, which is typical of the parallel configuration. In contrast, G_3T_4 displays a positive peak at 295 nm, indicative of an antiparallel topology. Quadruplexes with single nucleotide loops are thought to be only able to form 'propeller-type' fold-back loops generating parallel-stranded complexes, while longer T_4 loops can form lateral, edgewise or diagonal loops. These CD spectra suggest that the presence of only one single T loop is sufficient to induce the formation of a parallel-stranded structure and that the complexes only adopt an antiparallel arrangement when all the loops are longer. In general, these CD spectra were independent of the potassium concentration in the range 20–200 mM K^+, though pronounced changes were observed for G_3T_4 (inset to Figure 1a). For this sequence a secondary peak is visible around 260 nm at low potassium ion concentrations, which disappears as the potassium ion concentration is increased; this is accompanied by an increase in the peak at 295 nm. The presence of isoelliptic points in these spectra suggests that this sequence may adopt two distinct structural forms in the presence of low or high potassium ion concentrations.

In the presence of sodium ions, the CD spectra for G_3T, G_3T-T_4-T and G_3T_4-T-T are again typical of a parallel topology, with peaks around 265 nm (Figure 1b). However, the addition of a second T_4 loop results in spectra with

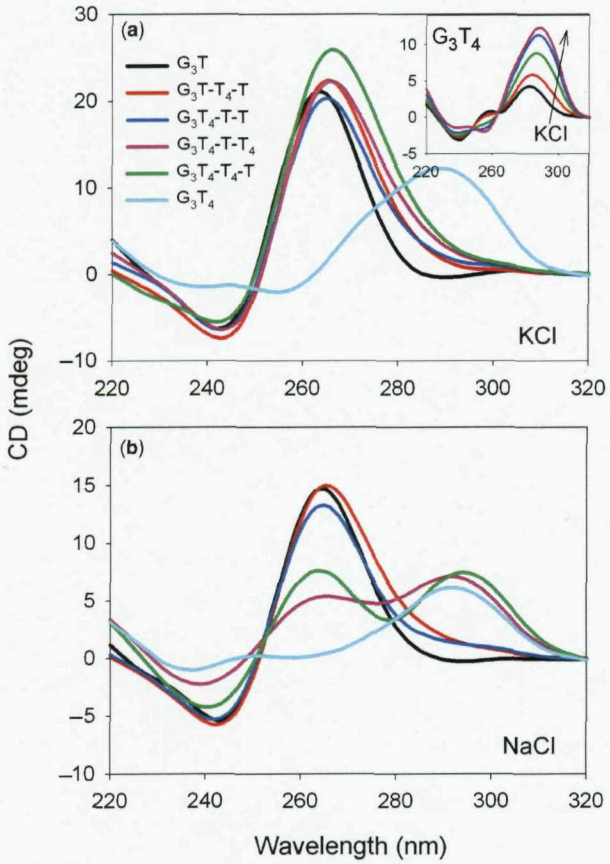


Figure 1. CD spectra of the fluorescently-labelled quadruplex-forming oligonucleotides in the presence of 10 mM lithium phosphate pH 7.4 containing 200 mM KCl (**a**) or 200 mM NaCl (**b**). G_3T, black; G_3T-T_4-T, red; G_3T_4-T-T, blue; G_3T_4-T-T_4, pink; G_3T_4-T_4-T, green; G_3T_4, cyan. The inset to the upper panel shows the CD spectrum of G_3T_4 in the presence of different concentrations of KCl: black 1 mM; red, 5 mM; green 20 mM; blue 50 mM; pink, 200 mM.

equal-sized peaks at 265 nm and 295 nm. It has been suggested that sodium ions promote the formation of antiparallel topologies and it is possible that the two longer T_4 loops are laterally arranged, while the single T loop is in a fold-back arrangement. This mixed spectrum could indicate the co-existence of parallel and antiparallel topologies in solution, but a hybrid structure containing both *syn* and *anti* bonds seems more likely; this will be considered further in the Discussion. The spectrum of G_3T_4 is similar in the presence of sodium and potassium ions, with a peak at 295 nm, suggesting an antiparallel topology.

Gel mobility

We further compared the global structures of these sequences by examining their mobilities in polyacrylamide gels that had been supplemented with 20 mM KCl (Figure 2). Each of the sequences ran as a single band, with the exception of G_3T_4 which is smeared, possibly

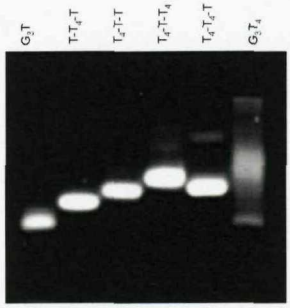


Figure 2. Mobility of the quadruplex-forming oligonucleotides on a 14% polyacrylamide gel supplemented with 20 mM KCl.

because this sequence is less stable under these conditions ($T_m \sim 40°C$, see below). G_3T has the greatest mobility, as expected as it has the lowest molecular weight. Surprisingly we find that G_3T_4-T-T has a lower mobility than G_3T-T_4-T and similarly G_3T_4-T-T_4 is slower than G_3T_4-T_4-T. It appears that the presence of a single T in the central loop reduces the mobility. This will be considered in the Discussion.

Imino proton NMR spectra

One of the defining features of structures that contain G-quartets is the appearance of imino proton resonances between 10.5 and 12.0 p.p.m. in NMR spectra (56). Examination of this region of NMR spectra has often been used to assess whether the sequence adopts a unique structure (20,34,36,56) and the presence of multiple or ill-defined peaks is evidence for the existence of multiple structures. The imino proton spectra for each of these sequences are shown in Figure 3. It can be seen that the spectra of G_3T, G_3T-T_4-T, G_3T_4-T-T and G_3T_4-T-T_4 display between 10 and 12 well-resolved and sharp peaks, indicative of well-defined structures. In the cases where 10 or 11 peaks are resolved, the intensities indicate that one or two imino protons have degenerate chemical shifts. Hence, the number of hydrogen-bonded imino protons is 12, as expected for three stacked G-quartets. G_3T_4-T_4-T also shows 12 major peaks, though the spectrum contains some minor peaks, which might indicate the presence of a small amount of a second structure. In contrast, the imino proton spectrum of G_3T_4 shows multiple peaks confirming that it adopts more than one stable conformation.

Fluorescence melting curves

Representative fluorescent melting curves for these sequences are shown in Figure 4 in the presence of potassium and sodium ions. The T_m values at different ionic strengths, along with the calculated values for ΔH, are shown in Table 2. The samples were melted and annealed at $0.2°C.min^{-1}$; no hysteresis was observed at this rate of temperature change (except for G_3T_4 at low ionic strengths). The melting temperatures were all independent of concentration (between 0.1 and 10 μM; Supplementary material Figure 1) confirming that these oligonucleotide sequences form intramolecular

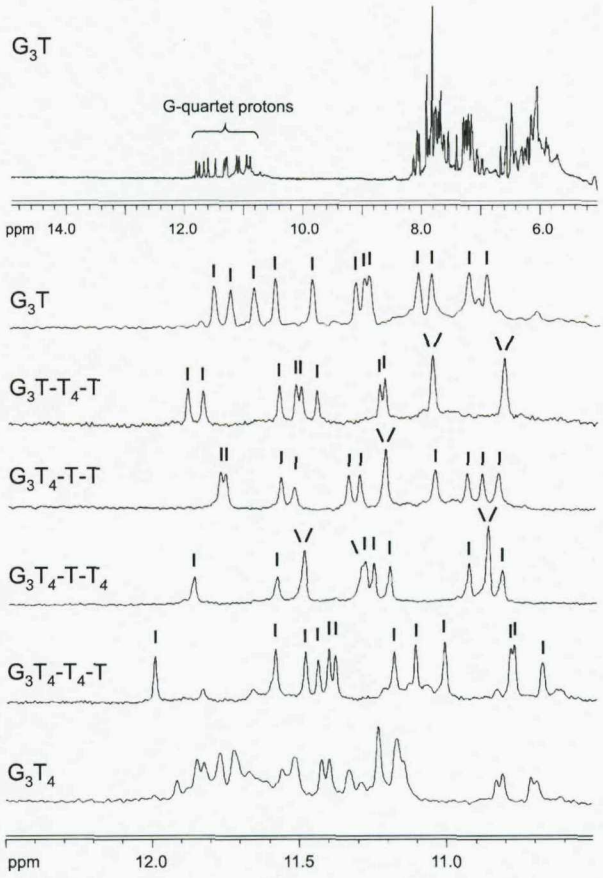


Figure 3. 1D imino proton NMR spectra of the quadruplex-forming oligonucleotides. The samples (100 μM) were prepared in 200 mM potassium phosphate pH 7.4. The top panel shows the 1D-NMR spectrum for G_3T between 5 and 15 p.p.m., while the other panels show the imino proton region for each oligonucleotide. The individual peaks are indicated.

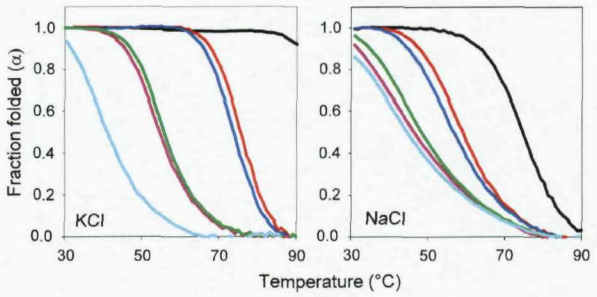


Figure 4. Fluorescence melting profiles for the quadruplex-forming oligonucleotides. The reactions were performed in 10 mM lithium phosphate pH 7.4 containing either 20 mM KCl (left hand panel) or 200 mM NaCl (right hand panel). The temperature was changed at 0.2°C.min^{-1}. The curves show the fraction folded (α) as a function of temperature, calculated as described in the Methods section. G_3T, black; G_3T-T_4-T, red; G_3T_4-T-T, blue; G_3T_4-T-T_4, pink; G_3T_4-T_4-T, green; G_3T_4, cyan.

(not intermolecular) complexes. As expected, all the complexes are more stable in potassium than sodium ions. G_3T is the most stable and, in the presence of potassium, substituting a T_4 loop instead of a loop with a single T decreases the T_m by about 20°C, irrespective of whether the replacement is in a central (G_3T-T_4-T) or peripheral loop (G_3T_4-T-T), though G_3T-T_4-T is about 2–3°C more stable than G_3T_4-T-T. Replacing a second T loop with T_4 causes a further 20°C decrease in T_m and G_3T_4-T_4-T is about 2–4°C more stable than G_3T_4-T-T_4. In each case, the sequence with a single T in the central loop is slightly less stable than the equivalent sequence with T_4 in the same position. Replacing all three single T loops with T_4 decreases the T_m by a further 10°C, though the melting and annealing curves with this sequence show hysteresis at low ionic strengths and the melting (but not the annealing) profiles are biphasic.

The relative order of stability is the same in the presence of sodium ions (Supplementary material Table 1). Replacing one T loop with T_4 decreases the T_m by 15–20°C, and a second substitution causes a further 10°C decrease. Replacing all three loops with T_4 does not affect the stability any further and G_3T_4 has a similar T_m to G_3T_4-T_4-T and G_3T_4-T-T_4. As seen with potassium, sequence isomers with a central T_4 loop are slightly more stable than those with a central T loop (i.e. G_3T_4-T_4-T > G_3T_4-T-T_4 and G_3T-T_4-T > G_3T_4-T-T).

ΔH values for the quadruplex single-strand transition in the presence of potassium were derived from these melting profiles by van't Hoff analysis, assuming that the reaction is a two-state equilibrium, and the values are shown in Table 2. These are typical of those for similar quadruplexes and show a general decrease in ΔH as the overall loop length increases. As previously reported for other quadruplexes, ΔH increases with ionic strength, consistent with the presence of specific cation binding sites within the quadruplex (30). The slopes of plots of ΔG against log[M^+] can be used to determine the stoichiometry of cation binding (30) yielding values of Δn (the difference between the number of ions bound in the folded and unfolded states) and the values of Δn in the presence of potassium are listed in Table 2. For an antiparallel structure containing three stacked G-quartets, Δn would be expected to be either two (the number of potassium ions located between the stacked quartets) or four (including two more ions that may be coordinated between the loops and the terminal quartets). A value of two seems more likely for a parallel topology with single nucleotide loops, in which the loops do not interact with the terminal quartets. The values of Δn are between two and three for all the complexes that contain at least one loop with a single T residue, though there is a steady increase in this value with increased numbers of T_4 loops, which will be considered further in the Discussion. Δn is larger for G_3T_4, consistent with the suggestion that it adopts a different topology, though this value may not be accurate as there is some hysteresis in its melting profiles at low potassium concentrations and the NMR data suggest that it adopts more than one conformation.

Table 2. T_m and ΔH values for the fluorescently labelled quadruplex-forming oligonucleotides, determined in the presence of 10 mM lithium phosphate pH 7.4 containing different concentrations of KCl. The samples were heated and cooled at a rate of 0.2°C.min^{-1}

[KCl] mM	G_3T		G_3T-T_4-T		G_3T_4-T-T		G_3T_4-T-T_4		G_3T_4-T_4-T		G_3T_4	
	T_m°C	ΔH kJ.mol^{-1}	T_m°C	ΔH kJ.mol^{-1}	T_m°C	ΔH kJ.mol^{-1}	T_m°C	ΔH kJ.mol^{-1}	T_m°C	ΔH kJ.mol^{-1}	T_m°C	ΔH kJ.mol^{-1}
0	46.6											
0.1	57.2	−242 ± 4	37.7									
1	73.5	−271 ± 5	53.8	−227 ± 7	51.3	−208 ± 7						
5	85.7	−275 ± 5	65.3	−259 ± 11	63.4	−250 ± 15						
10			70.7	−266 ± 9	67.3	−258 ± 9	46.3	−184 ± 13	48.1	−205 ± 5		
20			75.1	−267 ± 7	73.1	−262 ± 10	52.5	−207 ± 3	54.1	−227 ± 11	44.5/37.6*	
50			81.4	−284 ± 7	79.8	−275 ± 8	59.5	−225 ± 13	60.5	−246 ± 8	47.0/51.0*	−234 ± 11
100			87.5		84.6		65.3	−247 ± 8	67.1	−269 ± 13	56.3	−266 ± 10
200							72.8	−256 ± 9	74.2	−276 ± 10	63.3	−293 ± 9
Δn	2.13 ± 0.10		2.29 ± 0.11		2.31 ± 0.16		2.75 ± 0.11		2.89 ± 0.12		4.01 ± 0.16	

*Indicates a biphasic melting profile. All reactions were performed at least twice and the calculated T_m values usually differed by <0.5°C. ΔH values were typically calculated for melting profiles for which the T_m was between 40°C and 80°C. Missing values at low concentrations of KCl correspond to complexes for which the T_ms were too low to measure (<30°C), while those at high ionic strengths (especially G_3T) were too stable (T_m > 85°C).

Kinetics of quadruplex formation

Hysteresis. The fluorescence melting experiments shown in Figure 4 were performed at a rate of temperature change of 0.2°C.min^{-1} and only G_3T_4 showed hysteresis between the melting and annealing profiles. On increasing the rate to 2°C.min^{-1} there was a 7–10°C hysteresis for the sequences with two T_4 loops in the presence of potassium, though the melting and annealing profiles were identical for the sequences with single T_4 loops. The sequences with single T_4 loops only displayed hysteresis when the rate of heating was increased to 12°C.min^{-1}, while the melting and annealing curves for G_3T were always superimposable. Representative heating and annealing curves at different rates of heating and cooling in the presence of 20 mM potassium are shown in Supplementary material Figure 2 and the different T_m values are summarized in Supplementary Table 1. No hysteresis was observed for any of these sequences in the presence of sodium at even the fastest rate of heating and cooling. Differences between the melting and annealing curves arise because the reaction is not at thermodynamic equilibrium and indicate that either the folding or the unfolding process is slow. The folding (k_1) and unfolding (k_{-1}) rate constants for the unimolecular folding reaction can be obtained by analysis of these data as previously described (47,48). Figure 5a shows the melting and annealing profiles for G_3T_4-T-T and G_3T_4-T-T_4, determined at 12°C.min^{-1} and 2°C.min^{-1}, respectively, while similar plots for G_3T-T_4-T and G_3T_4-T_4-T are included in Supplementary material, Figure 3. Figure 6 shows Arrhenius plots for the folding and unfolding rates constructed from these data for G_3T-T_4-T and G_3T_4-T_4-T (Figure 6a) and G_3T_4-T-T and G_3T_4-T-T_4 (Figure 6b). The kinetic parameters derived from these Arrhenius plots are presented in Table 3. Several factors are apparent from these kinetic data. Firstly, the association reaction shows unusual temperature dependence, with an apparent negative activation energy, i.e. the reaction is faster a lower temperatures. This has been noted by others and is consistent with the reaction occurring by a nucleation-zipper mechanism

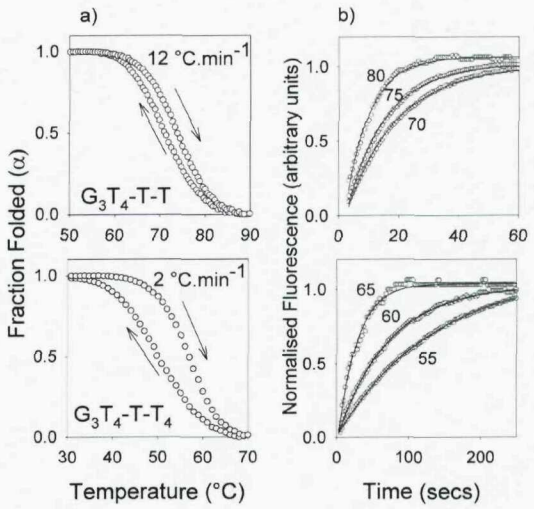


Figure 5. (**a**) Hystersis between the melting and annealing profiles for G_3T_4-T-T (upper panel, with a temperature change of 12°C.min^{-1}) and G_3T_4-T-T_4 (lower panel, with a temperature change of 2°C.min^{-1} in the presence of 10 mM lithium phosphate pH 7.4 containing 20 mM KCl. (**b**) temperature-jump relaxation profiles for G_3T_4-T-T (upper panel) and G_3T_4-T-T_4 (lower panel. The traces show the rate of approach to a new equilibrium following a rapid 5°C increase in temperature to the value shown. The profiles have been normalized to show the fractional change in fluorescence with time.

(47,48). Secondly, the data for G_3T_4-T_4-T are very similar to G_3T_4-T-T_4 and G_3T_4-T-T is similar to G_3T-T_4-T, suggesting that the distribution of the different loops is less important than their length. Thirdly, the unfolding parameters are very similar for all four oligonucleotides, while the folding parameters vary according to the loop lengths. For the association reaction both ln(A) and E_a are less negative for the complexes with longer loops. This kinetic analysis was not performed for G_3T as it showed no hysteresis and for G_3T_4 as the melting and annealing curves were biphasic.

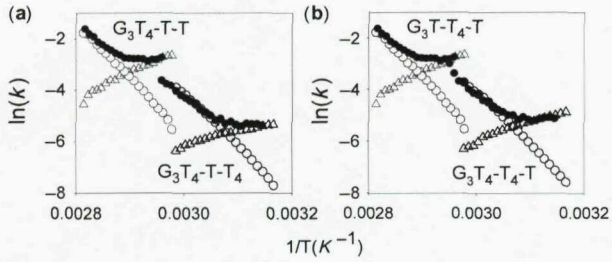


Figure 6. Arrhenius plots showing the temperature dependence of the kinetic parameters for G_3T_4-T-T and G_3T_4-T-T_4 (**a**) and G_3T-T_4-T and G_3T_4-T_4-T (**b**). Open symbols were derived from the hysteresis between the melting and annealing profiles; k_{-1}, open circles; k_1, open triangles. Filled circles show the time constants obtained from the temperature-jump experiments ($k_1 + k_{-1}$).

Table 3. Kinetic parameters for folding (k_1) and unfolding (k_{-1}) of the quadruplex-forming oligonucleotides determined from analysis of the hysteresis between melting and annealing curves

Sequence	k_1		k_{-1}	
	E_a (kJ mol^{-1})	ln (A) s^{-1}	E_a (kJ mol^{-1})	ln (A) s^{-1}
G_3T	-	-	-	-
G_3T-T_4-T	-93 ± 2	-36 ± 1	182 ± 3	60 ± 1
G_3T_4-T-T	-88 ± 3	-34 ± 1	178 ± 2	58 ± 1
G_3T_4-T-T_4	-48 ± 4	-24 ± 1	173 ± 2	58 ± 1
G_3T_4-T_4-T	-59 ± 2	-27 ± 1	174 ± 2	59 ± 1
G_3T_4	-	-	-	-

The experiments were performed in 10 mM lithium phosphate pH 7.4 containing 20 mM potassium chloride. E_a is the activation energy ($kJ.mol^{-1}$) and A is the pre-exponential factor from the equation $k = Ae^{(-Ea/RT)}$. No values are presented for G_3T as it does not show any hysteresis, or G_3T_4 as more than one folded configuration exists in solution.

Temperature-jump kinetics. In order to confirm the kinetic data obtained from the hysteresis experiments we performed temperature-jump relaxation kinetics on these complexes. In this technique, the temperature of the complex (maintained around the T_m) is rapidly increased (by 5°C) and the time-dependent changes in fluorescence are recorded as the reaction relaxes to a new equilibrium. Representative temperature-jump relaxation profiles for these complexes are shown in Figure 5b and reveal a slow time-dependent relaxation to the new equilibrium, which is clearly faster for the complexes with two short loops than the ones with two long loops. The kinetic curves at different temperatures were fitted with single exponential functions and the rate constants obtained are presented as Arrhenius plots in Figure 6. For this unimolecular reaction the apparent rate constant for the relaxation is equal to the sum of the folding and unfolding rate constants ($k_{-1} + k_1$). Although it is not possible to resolve these individual components, at low temperatures the sum is dominated by k_1, while the sum approximates to k_{-1} at high temperatures. It can be seen that there is excellent agreement between the rate constants determined by the two independent methods for each of the sequences, confirming that sequences with longer loops display slower rates of folding with little effect on the rate of unfolding.

DISCUSSION

Topology

Circular dichroism is often used to indicate the folding topology of DNA quadruplexes (20,38,53,54). Antiparallel quadruplexes typically have a positive CD signal at around 295 nm, while parallel quadruplexes display a positive signal around 260 nm. These differences reflect both the arrangements of the strands and the *syn/anti* orientations around the glycosidic bonds. Parallel topologies have all-*anti* glycosidic angles, while antiparallel ones have both *syn* and *anti* in varying ratios. However, it is clear that these spectral signatures are not necessarily an indicator of quadruplex folding as some exceptions have been noted (57,58). Nonetheless, CD spectra are useful indicators of changes in global quadruplex configuration for series of related oligonucleotides. In the presence of potassium, all the oligonucleotides that contain at least one single T loop exhibit a similar CD spectrum that is indicative of a parallel topology. It therefore appears that in potassium the presence of only one single T loop, in any position, is sufficient to promote all the other loops to form a fold-back propeller-like structure. In principle, these oligonucleotides could adopt several different folded configurations, yet the NMR and gel electrophoresis experiments suggest that only one predominates. When all three loops contain T_4, there is a dramatic change in the CD spectrum to a form that is consistent with antiparallel folding, though the details are dependent on the ionic strength suggesting that G_3T_4 can adopt multiple configurations. This again is consistent with the NMR and electrophoresis experiments, which suggest the presence of multiple folded forms. Previous studies have suggested that the quadruplex formed by $d(G_3T_4G_3)_2$ adopts an antiparallel hairpin dimer in the presence of both sodium and potassium (59,60).

A similar effect is seen for G_3T, G_3T_4-T-T and G_3T-T_4-T in the presence of sodium ions and these display CD spectra that are consistent with parallel topologies. However, the greater propensity to form antiparallel structures in the presence of sodium is seen with T_4-T_4-T and T_4-T-T_4 loops, which have CD spectra with peaks at both 260 nm and 295 nm. This may indicate the presence of multiple structural forms, but it is more likely due to the formation of a structure that contains both edgewise (T_4) and fold-back (T) loops, as observed with other sequences (34–36).

Stability

The fluorescence melting experiments show that the number of short loops, rather than their position, has the greatest effect on quadruplex stability. In the presence of potassium ions, G_3T is the most stable and in concentrations above 5 mM it does not display a melting transition. Substituting a T_4 into either the first or second loop decreases the T_m by about 20°C, with a further 20°C decrease on introducing a second T_4 substitution. The same effect is seen in the presence of sodium ions though there is only a small decrease in stability on changing the third loop to T_4, consistent with the CD spectra which

show that G_3T_4-T-T_4, G_3T_4-T_4-T and G_3T_4 display some antiparallel characteristics in contrast to all the other oligonucleotides.

It is noticeable that sequences with a single T loop in the central position are less stable and have lower gel mobilities than their sequence isomers with T_4 in this position (i.e. compare T_4-T-T_4 loops with T_4-T_4-T and T_4-T-T with T-T_4-T). It appears that folded structures with a central T_4 loop are more compact and have higher thermal stability. There is then a further decrease in stability when all three loops are composed of T_4, which as noted above adopts a different configuration.

Potassium ion binding

The variation of ΔG with ionic strength allows us to estimate the difference in the number of potassium ions specifically bound to the folded and unfolded structures. This value is close to two for G_3T as expected, since two potassium ions can bind between the three stacked quartets. Although the precise values of Δn should be interpreted with caution, it is noticeable that there is a steady increase in this value as the number of longer loops is increased. The value of Δn is similar for G_3T-T_4-T and G_3T_4-T-T and is lower than for both G_3T_4-T_4-T and G_3T_4-T-T_4. These results suggest that the longer loops are involved in cation binding. The larger value of Δn seen with G_3T_4 may not be significant, as this sequence adopts multiple configurations.

Kinetic analysis

Comparing the kinetic parameters for the sequences with one or two T_4 loops (Table 3) reveals that the unfolding parameters are very similar, while there are clear differences in the folding reaction. Complexes with longer loops have higher (less negative) activation energies for the association reaction and larger values for the pre-exponential factor (which is related to the entropy of the transition state). In comparison, no hysteresis is observed with G_3T and temperature-jump experiments showed a very fast re-equilibration, while G_3T_4 has slower folding and unfolding parameters, though not as slow as G_4T_4 (47). It is clear that the folding of intramolecular structures with only single T loops is fast, and we imagine that when one loop is composed of a single nucleotide the G-tracts on either side rapidly associate, forming a platform to which the other G-tracts can bind. The results with the oligonucleotides containing one or two T_4 loops suggest that the position of the single-base loop has little effect on the kinetics and that the most important factor is the number of longer loops.

It is interesting to note that none of these sequences show any hysteresis in the presence of sodium ions, even though G_3T and those with two single T loops appear to adopt a similar global structure. The higher stability and slower kinetics in the presence of potassium may therefore reflect conformational changes subsequent to the initial folding events (45).

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

ACKNOWLEDGEMENTS

PAR is supported by a research studentship from BBSRC.

Conflict of interest statement. None declared.

REFERENCES

1. Burge,S., Parkinson,G.N., Hazel,P., Todd,A.K. and Neidle,S. (2006) Quadruplex DNA: sequence, topology and structure. *Nucleic Acids Res.*, **34**, 5402–5415.
2. Davis,J.T. (2004) G-quartets 40 years later: From 5′-GMP to molecular biology and supramolecular chemistry. *Angew. Chem. Int. Ed. Engl.*, **43**, 668–698.
3. Phan,A.T., Kuryavyi,V. and Patel,D.J. (2006) DNA architecture: from G to Z. *Curr. Opin. Struct. Biol.*, **16**, 288–298.
4. Simonsson,T. (2001) G-quadruplex DNA structures - Variations on a theme. *Biol. Chem.*, **382**, 621–628.
5. Sen,D. and Gilbert,W. (1990) A sodium-potassium switch in the formation of 4-stranded G4-DNA. *Nature*, **344**, 410–414.
6. Williamson,J.R., Raghuraman,M.K. and Cech,T.R. (1989) Mono-valent cation induced atructure of telomeric DNA - the G-quartet model. *Cell*, **59**, 871–880.
7. Huppert,J.L. and Balasubramanian,S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2908–2916.
8. Huppert,J.L. and Balasubramanian,S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.*, **35**, 406–413.
9. Todd,A.K., Johnston,M. and Neidle,S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.*, **33**, 2901–2907.
10. Sen,D. and Gilbert,W. (1988) Formation of parallel 4-stranded complexes by guanine-rich motifs in DNA and its implications for meiosis. *Nature*, **334**, 364–366.
11. Sundquist,W.I. and Klug,A. (1989) Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops. *Nature*, **342**, 825–829.
12. Ambrus,A., Chen,D., Dai,J.X., Jones,R.A. and Yang,D.Z. (2005) Solution structure of the biologically relevant G-quadruplex element in the human c-*MYC* promoter. implications for G-quadruplex stabilization. *Biochemistry*, **44**, 2048–2058.
13. Rangan,A., Fedoroff,O.Y. and Hurley,L.H. (2001) Induction of duplex to G-quadruplex transition in the c-myc promoter region by a small molecule. *J. Biol. Chem.*, **276**, 4640–4646.
14. Siddiqui-Jain,A., Grand,C.L., Bearss,D.J. and Hurley,L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-*MYC* transcription. *Proc. Natl Acad. Sci. USA*, **99**, 11593–11598.
15. Simonsson,T., Pecinka,P. and Kubista,M. (1998) DNA tetraplex formation in the control region of c-myc. *Nucleic Acids Res.*, **26**, 1167–1172.
16. Cogoi,S. and Xodo,L.E. (2006) G-quadruplex formation within the promoter of the KRAS proto-oncogene and its effect on transcription. *Nucleic Acids Res.*, **34**, 2536–2549.
17. Dai,J.X., Chen,D., Jones,R.A., Hurley,L.H. and Yang,D.Z. (2006) NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region. *Nucleic Acids Res.*, **34**, 5133–5144.
18. Dai,J.X., Dexheimer,T.S., Chen,D., Carver,M., Ambrus,A., Jones,R.A. and Yang,D.Z. (2006) An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution. *J. Am. Chem. Soc.*, **128**, 1096–1098.
19. Dexheimer,T.S., Sun,D. and Hurley,L.H. (2006) Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the bcl-2 P1 promoter. *J. Am. Chem. Soc.*, **128**, 5404–5415.

20. Rankin,S., Reszka,A.P., Huppert,J., Zloh,M., Parkinson,G.N., Todd,A.K., Ladame,S., Balasubramanian,S. and Neidle,S. (2005) Putative DNA quadruplex formation within the human c-*kit* oncogene. *J. Am. Chem. Soc.*, **127**, 10584–10589.
21. Sun,D.Y., Guo,K.X., Rusche,J.J. and Hurley,L.H. (2005) Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents. *Nucleic Acids Res.*, **33**, 6070–6080.
22. De Armond,R., Wood,S., Sun,D.Y., Hurley,L.H. and Ebbinghaus,S.W. (2005) Evidence for the presence of a guanine quadruplex forming region within a polypurine tract of the hypoxia inducible factor 1 alpha promoter. *Biochemistry*, **44**, 16341–16350.
23. Fry,M. and Loeb,L.A. (1994) The fragile-X syndrome $d(CGG)_n$ nucleotide repeats form a stable tetrahelical structure. *Proc. Natl Acad. Sci. USA*, **91**, 4950–4954.
24. Matsugami,A., Okuizumi,T., Uesugi,S. and Katahira,M. (2003) Intramolecular higher order packing of parallel quadruplexes comprising a G:G:G:G tetrad and a G(:A):G(:A):G(:A):G heptad of GGA triplet repeat DNA. *J. Biol. Chem.*, **278**, 28147–28153.
25. Murchie,A.I.H. and Lilley,D.M.J. (1992) Retinoblastoma susceptibility genes contain 5′ sequences with a high propensity to form guanine-tetrad structures. *Nucleic Acids Res.*, **20**, 49–53.
26. Howell,R.M., Woodford,K.J., Weitzmann,M.N. and Usdin,K. (1996) The chicken beta-globin gene promoter forms a novel "cinched" tetrahelical structure. *J. Biol. Chem.*, **271**, 5208–5214.
27. Lew,A., Rutter,W.J. and Kennedy,G.C. (2000) Unusual DNA structure of the diabetes susceptibility locus IDDM2 and its effect on transcription by the insulin promoter factor Pur-1/MAZ. *Proc. Natl Acad. Sci. USA*, **97**, 12508–12512.
28. Yafe,A., Etzioni,S., Weisman-Shomer,P. and Fry,M. (2005) Formation and properties of hairpin and tetraplex structures of guanine-rich regulatory sequences of muscle-specific genes. *Nucleic Acids Res.*, **33**, 2887–2900.
29. Hazel,P., Parkinson,G.N. and Neidle,S. (2006) Topology variation and loop structural homology in crystal and simulated structures of a bimolecular DNA quadruplex. *J. Am. Chem. Soc.*, **128**, 5480–5487.
30. Jing,N.J., Rando,R.F., Pommier,Y. and Hogan,M.E. (1997) Ion selective folding of loop domains in a potent anti-HIV oligonucleotide. *Biochemistry*, **36**, 12498–12505.
31. Phan,A.T., Modi,Y.S. and Patel,D.J. (2004) Propeller-type parallel-stranded G-quadruplexes in the human c-myc promoter. *J. Am. Chem. Soc.*, **126**, 8710–8716.
32. Parkinson,G.N., Lee,M.P.H. and Neidle,S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. *Nature*, **417**, 876–880.
33. Wang,Y. and Patel,D.J. (1993) Solution structure of the human telomeric repeat $d[AG_3(T_2AG_3)_3]$ G-tetraplex. *Structure*, **1**, 263–282.
34. Ambrus,A., Chen,D., Dai,J.X., Bialis,T., Jones,R.A. and Yang,D.Z. (2006) Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. *Nucleic Acids Res.*, **34**, 2723–2735.
35. Luu,K.N., Phan,A.T., Kuryavyi,V., Lacroix,L. and Patel,D.J. (2006) Structure of the human telomere in K+ solution: An intramolecular (3+1) G-quadruplex scaffold. *J. Am. Chem. Soc.*, **128**, 9963–9970.
36. Phan,A.T., Luu,K.N. and Patel,D.J. (2006) Different loop arrangements of intramolecular human telomeric (3+1) G-quadruplexes in K+ solution. *Nucleic Acids Res.*, **34**, 5715–5719.
37. Cevec,M. and Plavec,J. (2005) Role of loop residues and cations on the formation and stability of dimeric DNA G-quadruplexes. *Biochemistry*, **44**, 15238–15246.
38. Hazel,P., Huppert,J., Balasubramanian,S. and Neidle,S. (2004) Loop-length-dependent folding of G-quadruplexes. *J. Am. Chem. Soc.*, **126**, 16405–16415.
39. Risitano,A. and Fox,K.R. (2004) Influence of loop size on the stability of intramolecular DNA quadruplexes. *Nucleic Acids Res.*, **32**, 2598–2606.
40. Smirnov,I. and Shafer,R.H. (2000) Effect of loop sequence and size on DNA aptamer stability. *Biochemistry*, **39**, 1462–1468.
41. Risitano,A. and Fox,K.R. (2003) Stability of intramolecular DNA quadruplexes: Comparison with DNA duplexes. *Biochemistry*, **42**, 6507–6513.
42. Guo,Q., Lu,M. and Kallenbach,N.R. (1993) Effect of thymine tract length on the structure and stability of model telomeric sequences. *Biochemistry*, **32**, 3596–3603.
43. Lu,M., Guo,Q. and Kallenbach,N.R. (1992) Structure and stability of sodium and potassium complexes of dT_4G_4 and dT_4G_4T. *Biochemistry*, **31**, 2455–2459.
44. Merkina,E.E. and Fox,K.R. (2005) Kinetic stability of intermolecular DNA quadruplexes. *Biophys. J.*, **89**, 365–373.
45. Jing,N.J., Gao,X.L., Rando,R.F. and Hogan,M.E. (1997) Potassium-induced loop conformational transition of a potent anti-HIV oligonucleotide. *J Biomol. Struct. Dyn.*, **15**, 573–585.
46. Darby,R.A.J., Sollogoub,M., McKeen,C., Brown,L., Risitano,A., Brown,N., Barton,C., Brown,T. and Fox,K.R. (2002) High throughput measurement of duplex, triplex and quadruplex melting curves using molecular beacons and a LightCycler. *Nucleic Acids Res.*, **30**, e39.
47. Brown,N.M., Rachwal,P.A., Brown,T. and Fox,K.R. (2005) Exceptionally slow kinetics of the intramolecular quadruplex formed by the Oxytricha telomeric repeat. *Org. Biomol. Chem.*, **3**, 4153–4157.
48. Mergny,J.L. and Lacroix,L. (2003) Analysis of thermal melting curves. *Oligonucleotides*, **13**, 515–537.
49. Cantor,C.R. and Schimmel,P.R. (1980) *Biophysical Chemistry*, W. H. Freedman and Company, New York.
50. Bernal-Mendez,E. and Leumann,C.J. (2002) Stability and kinetics of nucleic acid triplexes with chimaeric DNA/RNA third strands. *Biochemistry*, **41**, 12343–12349.
51. Rougee,M., Faucon,B., Mergny,J.L., Barcelo,F., Giovannangeli,C., Garestier,T. and Helene,C. (1992) Kinetics and thermodynamics of triple-helix formation - effects of ionic-strength and mismatches. *Biochemistry*, **31**, 9269–9278.
52. James,P.L., Brown,T. and Fox,K.R. (2003) Thermodynamic and kinetic stability of intermolecular triple helices containing different proportions of C^+.GC and T.AT triplets. *Nucleic Acids Res.*, **31**, 5598–5606.
53. Balagurumoorthy,P., Brahmachari,S.K., Mohanty,D., Bansal,M. and Sasisekharan,V. (1992) Hairpin and parallel quartet structures for telomeric sequences. *Nucleic Acids Res.*, **20**, 4061–4067.
54. Balagurumoorthy,P. and Brahmachari,S.K. (1994) Structure and stability of human telomeric sequence. *J. Biol. Chem.*, **269**, 21858–21869.
55. Rachwal,P.A., Brown,T and Fox,K.R. (2007) Effect of G-tract length on the structure and stability of intramolecular DNA quadruplexes. *Biochemistry*, **46**, 3036–3044.
56. Feigon,J., Koshlap,K.M. and Smith,F.W. (1995) H-1 NMR spectroscopy of DNA triplexes and quadruplexes. *Methods Enzymol.*, **261**, 225–255.
57. Dapic,V., Abdomerovic,V., Marrington,R., Peberdy,J., Rodger,A., Trent,J.O. and Bates,P.J. (2003) Biophysical and biological properties of quadruplex oligodeoxyribonucleotides. *Nucleic Acids Res.*, **31**, 2097–2107.
58. Esposito,V., Randazzo,A., Piccialli,G., Petraccone,L., Giancola,C. and Mayol,L. (2004) Effects of an 8-bromodeoxyguanosine incorporation on the parallel quadruplex structure $[d(TGGGT)]_4$. *Org. Biomol. Chem.*, **2**, 313–318.
59. Keniry,M.A., Strahan,G.D., Owen,E.A. and Shafer,R.H. (1995) Solution structure of the Na^+ form of the dimeric guanine quadruplex $[d(G_3T_4G_3)]_2$. *Eur. J. Biochem.*, **233**, 631–643.
60. Strahan,G.D., Keniry,M.A. and Shafer,R.H. (1998) NMR structure refinement and dynamics of the K^+-$[d(G_3T_4G_3)]_2$ quadruplex via particle mesh Ewald molecular dynamics simulations. *Biophys. J.*, **75**, 968–981.

FEBS Letters 581 (2007) 1657–1660

Sequence effects of single base loops in intramolecular quadruplex DNA

Phillip A. Rachwal[a], Tom Brown[b], Keith R. Fox[a,*]

[a] School of Biological Sciences, University of Southampton, Bassett Crescent East, Southampton SO16 7PX, UK
[b] School of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, UK

Received 7 February 2007; revised 8 March 2007; accepted 16 March 2007

Available online 28 March 2007

Edited by Hans Eklund

Abstract **We have examined the properties of intramolecular G-quadruplexes in which the G3 tracts are separated by single base loops. The most stable complex contained 1′,2′-dideoxyribose in all three loops, while loops containing T and C were slightly less stable (by about 2 °C). Quadruplexes containing loops with single A residues were less stable by 8 °C for each T to A substitution. These folded sequences display similar CD spectra, which are consistent with the formation of parallel stranded complexes with double-chain reversal loops. These results demonstrate that loop sequence, and not just length, affects quadruplex stability.**

Keywords: DNA quadruplex; DNA structure

1. Introduction

DNA sequences that contain four or more G-tracts can fold to form intramolecular structures that consist of stacks of G-quartets [1–5]. The four G-tracts are separated by different length loops, which can be as short as a single nucleotide [6–8]. The complexes can adopt a range of structures in which the bases are either *anti* or *syn* and in which the strands run parallel or antiparallel [4,5]. The loops can be arranged in several different ways; double chain reversal (propeller) loops link two adjacent parallel strands by a connection between the top and bottom G-tetrads [9], while edgewise or diagonal loops link two antiparallel strands [10]. Some structures contain both edgewise and propeller loops [11–13]. G-rich sequences with the potential to adopt these structures are found in several gene promoters; most notably the c-*myc* [8,14,15], *bcl2* [16,17] and c-*kit* [18,19] oncogenes and several of these contain single base loops. Although it has been shown that loop length affects quadruplex stability and structure [6,20–23] there have been few studies on the effect of loop sequence, though single base changes can have a significant effect (e.g. changing TTA in the human telomeric repeat to TTG [10] in the Tetrahymena repeat [24,25]).

One very stable intramolecular quadruplex contains four G_3 tracts that are linked by single T residues [7,26]. We have investigated the effect of loop sequence on its stability using sequences of the type d(TGGGNGGGNGGGNGGGT), where N is each base in turn (except G) together with 1′,2′-dideoxyribose, Φ.

2. Materials and methods

2.1. Oligonucleotides

All oligonucleotides were synthesized on an Applied Biosystems ABI 394 automated DNA/RNA synthesizer on the 0.2 or 1 μmole scale using the standard cycles of acid-catalysed detritylation, coupling, capping and iodine oxidation procedures. Phosphoramidite monomers and other reagents were purchased from Applied Biosystems or Link Technologies. The sequences of the oligonucleotides used in this work are shown in Table 1. All oligonucleotides were prepared with 5′-fluorescein and 3′-dabcyl (fluorescein C6 phosphoramidite and dabcyl cpg purchased from Link Technologies Ltd.) for use in the fluorescence melting experiments and the same sequences were used for the circular dichroism studies. The bases adjacent to the fluorophore and quencher were the same (T) for all the oligonucleotides to ensure that the terminal base did not affect quadruplex formation and stability. Inclusion of this base also hinders any fluorescence quenching between G and fluorescein.

2.2. Fluorescence melting

Fluorescence melting curves were determined in a Roche LightCycler as previously described [22,26,27] in a total reaction volume of 20 μL. Oligonucleotides (final concentration 0.25 μM) were prepared in 10 mM lithium phosphate pH 7.4, which was supplemented with various concentrations of potassium or sodium chloride. The LightCycler has one excitation source (488 nm) and the changes in fluorescence were measured at 520 nm. In order to avoid hysteresis between heating and cooling curves melting experiments were performed at a slow rate of temperature change (0.2 °C min^{-1}). This was achieved by changing the temperature in 1 °C steps and leaving the samples to equilibrate for 5 min at each temperature before recording the fluorescence. In a typical experiment the oligonucleotides were first denatured by heating to 95 °C for 5 min. They were then annealed by cooling to 30 °C at 0.2 °C min^{-1} and melted by heating to 95 °C at the same rate. The fluorescence was recorded during both the annealing and melting steps.

T_m values were obtained from the maxima of the first derivatives of the melting profiles using the LightCycler software or, together with ΔH, from van't Hoff analysis of the melting profiles [23,26–28]. The fraction folded was calculated as previously described [28] from the difference between the measured fluorescence and the upper and lower baselines. All reactions were performed at least twice and the T_m values differed by <0.5 °C with a 5% variation in ΔH. Since $\Delta G = 0$ at the T_m, ΔS was estimated as $\Delta H/T_m$. It should therefore be noted that ΔS is not determined independently of ΔH and T_m. This analysis assumes a simple two-state equilibrium between the folded and unfolded forms. The presence of polymorphic quadruplex structures will lead to shallower melting profiles and therefore smaller apparent values for ΔH.

2.3. Circular dichroism

CD spectra were measured on a Jasco J-720 spectropolarimeter as previously described [22]. Oligonucleotides solutions (5 μM) were prepared in 10 mM lithium phosphate pH 7.4, containing either 100 mM potassium chloride or 100 mM sodium chloride. The samples were

*Corresponding author. Fax: +44 23 8059 4459.
E-mail address: k.r.fox@soton.ac.uk (K.R. Fox).

0014-5793/$32.00
doi:10.1016/j.febslet.2007.03.040

Table 1
Sequences of the oligonucleotides used in this work

Name	Oligonucleotide sequence
TTT	d(Fam-TGGGTGGGTGGGTGGGT-dabcyl)
CCC	d(Fam-TGGGCGGGCGGGCGGGT-dabcyl)
AAA	d(Fam-TGGGAGGGAGGGAGGGT-dabcyl)
ΦΦΦ	d(Fam-TGGGΦGGGΦGGGΦGGGT-dabcyl)
TAT	d(Fam-TGGGTGGGAGGGTGGGT-dabcyl)
ATA	d(Fam-TGGGAGGGTGGGAGGGT-dabcyl)

Fam = fluorescein; Φ = 1′,2′-dideoxyribose.

heated to 95 °C and annealed by slowly cooling to 15 °C over a period of 12 hours. Spectra were recorded between 220 and 320 nm in 5 mm path length cuvettes. Spectra were averaged over 10 scans, which were recorded at 100 nm min^{-1} with a response time of 1 s and 1 nm bandwidth. A buffer baseline was subtracted from each spectrum and the spectra were normalized to have zero ellipticity at 320 nm.

3. Results

We have investigated the effect of sequence on the stability of quadruplexes that contain single nucleotide loops using oligonucleotides of the type (TGGGNGGGNGGGNGGGT), where N is each base in turn. The sequences of these oligonucleotides are shown in Table 1; they were each labelled at the 5′-end with fluorescein and with dabcyl at the 3′-end. Fluorescence melting curves were used to assess the thermal stability of these complexes. When the quadruplex is folded the fluorophore and quencher are in close proximity and the fluorescence is quenched [22,26–28]. When the structure unfolds the fluorophore and quencher are separated and there is an increase in fluorescence. In the presence of 100 mM KCl all these sequences formed structures that melted above 90 °C. However, by using only 1 mM KCl the melting temperatures decreased to a measurable range. Although this concentration is lower than that used in most quadruplex studies, higher potassium ion concentrations produced melting curves that were too stable to measure. The fluorescence melting profiles (in which the fluorophore and quencher are close together) and the similar patterns of stability in both 1mM KCl and 100 mM NaCl, strongly suggest that these sequences adopt an intramolecular quadruplex. Melting profiles for the four oligonucleotides that contain the same base in each loop (TTT, CCC, AAA or ΦΦΦ)

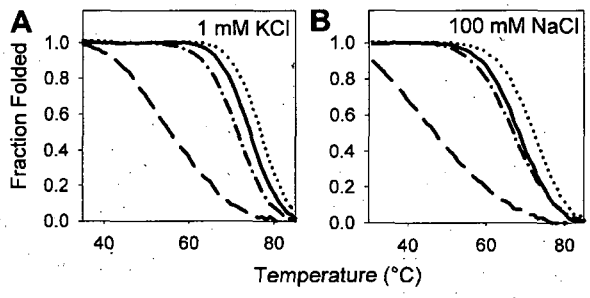


Fig. 1. Fraction of the different oligonucleotides folded at different temperatures (A) in 10 mM lithium phosphate pH 7.4 containing 1 mM KCl. (B) in 10 mM lithium phosphate pH 7.4 containing 100 mM NaCl. Solid line, TTT; dotted line, ΦΦΦ; dot and dash line, CCC; dashed line, AAA. These plots were derived from the melting curves by assuming that the complexes are folded at low temperature and single stranded at elevated temperatures [27].

Table 2
T_m values (°C) for the various oligonucleotides, determined in 10 mM lithium phosphate pH 7.4 containing different concentrations of KCl or NaCl

		TTT	AAA	CCC	ΦΦΦ	ATA	TAT
KCl	0	46.0	<30	44.4	49.5	36.9	37.9
	1	73.9	50.0	71.8	77.0	59.6	67.3
	5	85.6	64.4	83.2	88.1	72.1	79.3
	10	>90	70.0	>90	>90	78.0	84.2
	50	>90	82.2	>90	>90	>90	>90
NaCl	1	48.3	<30	47.3	53.2	<30	40.0
	10	55.4	<30	53.6	60.5	39.6	46.3
	50	63.7	<30	62.3	68.0	47.0	55.6
	100	68.3	43.7	67.1	72.5	52.2	60.5
	200	73.3	50.2	72.1	77.0	57.8	65.6

The oligonucleotide concentration was 0.25 μM and the samples were heated and cooled at 0.2 °C min^{-1}. All T_m values are ±0.5 °C.

are shown in Fig. 1 in the presence of 1 mM KCl or 100 mM NaCl and the melting temperatures determined at a range of ionic strengths are summarized in Table 2. At this rate of heating (0.2 °C min^{-1}) there was no hysteresis between the melting and annealing profiles and the T_m values were independent of oligonucleotide concentration between 0.1 and 10 μM.

It can be seen that loops with T, C and 1′,2′-dideoxyribose produce the most stable complexes (Φ > T > C), while the T_m of the complex with As in the loops is about 25 °C lower. The complexes are less stable in sodium containing buffers, but the rank order of stabilities is the same. The complex with A-containing loops has a similar stability to a complex containing propanediol linkers [22]. Thermodynamic parameters for the folding of each complex in the presence of potassium were estimated from van't Hoff analysis of the melting profiles and these are presented in Table 3. The values are consistent with those previously determined for intramolecular quadruplexes, which are typically between 65 and 100 kJ mol^{-1} per quartet [23,26]. These show that the complexes with the lower stability are characterized by a smaller enthalpy and higher entropy.

To further study the effect of loop sequence on stability we examined the properties of oligonucleotides with two Ts in the loops and one A (TAT), or one T in the loop and two As (ATA). The melting profiles of these sequences are shown in Fig. 2, alongside those for AAA and TTT and the thermodynamic parameters are presented in Table 3. Each T to A modification reduces the melting temperature by about 8 °C in both sodium and potassium-containing buffers, with a change in enthalpy of 46 kJ mol^{-1} per A substitution in 1 mM KCl and 33 kJ mol^{-1} in 100 mM NaCl.

CD spectral signatures are often used as indicators of the folding pattern of intramolecular quadruplexes; parallel structures (with *anti* glycosidic bonds) typically have positive maxima around 260 nm, while antiparallel structures (containing both *anti* and *syn* bonds) show maxima around 295 nm [29–32]. CD spectra of these sequences in the presence of 100 mM KCl are shown in Fig. 3 and show that the complexes produce very similar CD spectra with clear single maxima at 265 nm. These CD spectra were identical in sodium and potassium-containing buffers and strongly suggest that all these complexes adopt the same topology, which is likely to be the

Table 3
Thermodynamic parameters for folding of the quadruplex forming oligonucleotides

Oligo	1 mM KCl			100 mM NaCl		
	T_m (°C)	ΔH (kJ mol^{-1})	ΔS (kJ mol^{-1} K^{-1})	T_m (°C)	ΔH (kJ mol^{-1})	ΔS (kJ mol^{-1} K^{-1})
TTT	73.9	−278 ± 8	−0.80	68.3	−206 ± 3	−0.60
CCC	71.8	−257 ± 3	−0.75	67.1	−181 ± 10	−0.53
AAA	50.0	−142 ± 8	−0.44	43.7	−104 ± 9	−0.33
ΦΦΦ	77.0	−296 ± 10	−0.85	72.5	−229 ± 8	−0.66
TAT	67.3	−247 ± 4	−0.73	60.5	−158 ± 6	−0.47
ATA	59.6	−198 ± 8	−0.60	52.2	−131 ± 11	−0.40

All T_m values are ±0.5 °C; Experiments were performed in 10 mM lithium phosphate pH 7.4 containing 1 mM KCl or 100 mM NaCl. ΔH was calculated from van't Hoff analysis of the melting profiles, assuming a two-state equilibrium and ΔS was estimated as $\Delta H/T_m$.

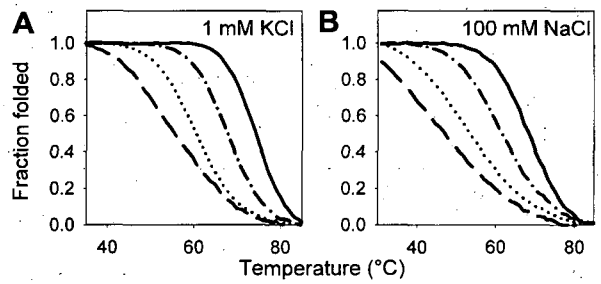


Fig. 2. Fraction of the different oligonucleotides folded at different temperatures (A) in 10 mM lithium phosphate pH 7.4 containing 1 mM KCl. (B) in 10 mM lithium phosphate pH 7.4 containing 100 mM NaCl. Solid line, TTT: dot and dash line, TAT; dotted line, ATA; dashed line, AAA. These plots were derived from the melting curves by assuming that the complexes are folded at low temperature and single stranded at elevated temperatures [27].

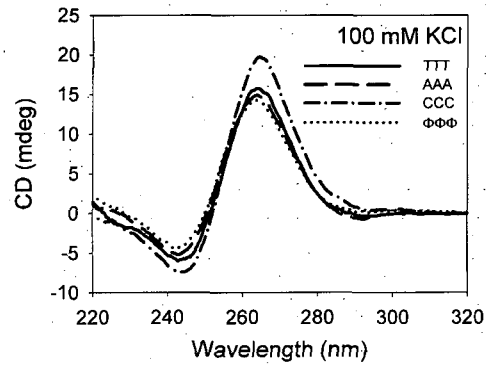


Fig. 3. CD spectra of the different oligonucleotides in 10 mM lithium phosphate, pH 7.4 containing 100 mM KCl. The oligonucleotide concentration was 5 μM. solid line, TTT; dotted line, AAA; dashed line, CCC; dot and dash line, ΦΦΦ.

parallel form containing double-chain reversal, propeller loops.

4. Discussion

These results demonstrate that the loop sequence has a large effect on quadruplex stability. Structures with loops containing T, C or 1′,2′-dideoxyribose (Φ) are the most stable (Φ > T > C) and these are about 20 °C more stable than those with A in the loops. Each A-containing loop decreases the T_m by about 8 °C. The size of the base may be significant as purine bases (AAA) produce less stable complexes than pyrimidines (TTT and CCC). Surprisingly the most stable complex is produced with 1′,2′-dideoxyribose linkers, suggesting that specific interactions with the loop bases are not necessary for forming these intramolecular quadruplexes. However inclusion of the non-nucleosidic propanediol linker (which is the same length as a single nucleotide) has the same effect as A, though in this case the increased flexibility of the loop may limit the stability. An alternative explanation is that the structure of the unfolded state may itself be sequence dependent. It is known that adenine stacks particularly well at the ends of duplexes [33] and any base specific stacking will need to be reversed in order to form a parallel quadruplex. Studies with longer loops have also observed the stacking of adenine against the terminal quartets [8,14], which would not be possible for a parallel quadruplex containing single nucleotide loop. The CD spectra of these sequences are also very similar, suggesting that the differences in stability do not arise from gross differences in the folding topology, for which the most abundant form in each case is probably a parallel orientation of the strands.

Structural studies of various quadruplex forming sequences have shown several examples of single nucleotide loops that contain single T, C or A residues [8,14,17,34]. Searches for potential quadruplex-forming sequences in the human genome reveal that the most common linker between the G-tracts is single base A or T [35,36]. It is therefore clear that the lower stability of quadruplexes with A-containing loops does not preclude intramolecular quadruplex formation, but the inclusion of A instead of T may moderate the stability, thereby permitting the interconversion between different structural forms.

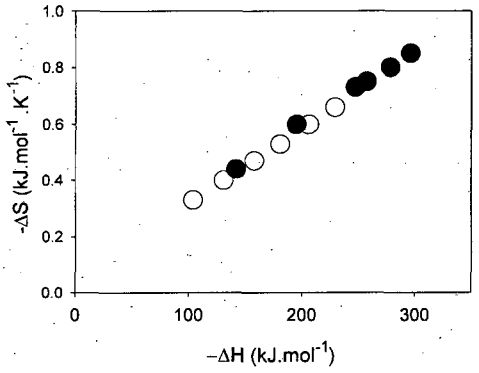


Fig. 4. Relationship between ΔH and ΔS determined for the complexes in the presence of 1 mM KCl (filled circles) and 100 mM NaCl (open circles).

Comparison of the thermodynamic parameters for this closely related series of quadruplexes (Table 3) reveals a correlation between the stability, enthalpy and entropy. The most stable complexes (higher T_m) have higher enthalpies which are compensated by less favourable entropic contributions (more negative ΔS). The relationship between enthalpy and entropy is illustrated in Fig. 4, which reveals that the same effect is seen in the presence of both potassium and sodium. It appears that the larger enthalpic contribution, which may arise from better stacking between the quartets and/or the loops, is accompanied by less flexibility, maybe by forming more ordered loop structures [23,37].

Acknowledgement: PAR is supported by a research studentship from BBSRC.

References

[1] Burge, S., Parkinson, G.N., Hazel, P., Todd, A.K. and Neidle, S. (2006) Quadruplex DNA: sequence, topology and structure. Nucleic Acids Res. 34, 5402–5415.
[2] Davis, J.T. (2004) G-quartets 40 years later: from 5′-GMP to molecular biology and supramolecular chemistry. Angew. Chem., Int. Ed. Engl. 43, 668–698.
[3] Keniry, M.A. (2000) Quadruplex structures in nucleic acids. Biopolymers 56, 123–146.
[4] Phan, A.T., Kuryavyi, V. and Patel, D.J. (2006) DNA architecture: from G to Z. Curr. Opin. Struct. Biol. 16, 288–298.
[5] Simonsson, T. (2001) G-quadruplex DNA structures – variations on a theme. Biol. Chem. 382, 621–628.
[6] Hazel, P., Huppert, J., Balasubramanian, S. and Neidle, S. (2004) Loop-length-dependent folding of G-quadruplexes. J. Am. Chem. Soc. 126, 16405–16415.
[7] Jing, N.J., Rando, R.F., Pommier, Y. and Hogan, M.E. (1997) Ion selective folding of loop domains in a potent anti-HIV oligonucleotide. Biochemistry 36, 12498–12505.
[8] Phan, A.T., Modi, Y.S. and Patel, D.J. (2004) Propeller-type parallel-stranded G-quadruplexes in the human c-*myc* promoter. J. Am. Chem. Soc. 126, 8710–8716.
[9] Parkinson, G.N., Lee, M.P.H. and Neidle, S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA. Nature 417, 876–880.
[10] Wang, Y. and Patel, D.J. (1993) Solution structure of the human telomeric repeat $d[AG_3(T_2AG_3)_3]$ G-tetraplex. Structure 1, 263–282.
[11] Ambrus, A., Chen, D., Dai, J.X., Bialis, T., Jones, R.A. and Yang, D.Z. (2006) Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution. Nucleic Acids Res. 34, 2723–2735.
[12] Luu, K.N., Phan, A.T., Kuryavyi, V., Lacroix, L. and Patel, D.J. (2006) Structure of the human telomere in K^+ solution: an intramolecular (3+1) G-quadruplex scaffold. J. Am. Chem. Soc. 128, 9963–9970.
[13] Phan, A.T., Luu, K.N. and Patel, D.J. (2006) Different loop arrangements of intramolecular human telomeric (3 + 1) G-quadruplexes in K^+ solution. Nucleic Acids Res. 34, 5715–5719.
[14] Ambrus, A., Chen, D., Dai, J.X., Jones, R.A. and Yang, D.Z. (2005) Solution structure of the biologically relevant G-quadruplex element in the human c-*myc* promoter implications for G-quadruplex stabilization. Biochemistry 44, 2048–2058.
[15] Siddiqui-Jain, A., Grand, C.L., Bearss, D.J. and Hurley, L.H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-*myc* transcription. Proc. Natl. Acad. Sci. USA 99, 11593–11598.
[16] Dai, J.X., Chen, D., Jones, R.A., Hurley, L.H. and Yang, D.Z. (2006) NMR solution structure of the major G-quadruplex structure formed in the human BCL2 promoter region. Nucleic Acids Res. 34, 5133–5144.
[17] Dai, J.X., Dexheimer, T.S., Chen, D., Carver, M., Ambrus, A., Jones, R.A. and Yang, D.Z. (2006) An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human BCL-2 promoter region in solution. J. Am. Chem. Soc. 128, 1096–1098.
[18] Fernando, H., Reszka, A.P., Huppert, J., Ladame, S., Rankin, S., Venkitaraman, A.R., Neidle, S. and Balasubramanian, S. (2006) A conserved quadruplex motif located in a transcription activation site of the human c-kit oncogene. Biochemistry 45, 7854–7860.
[19] Rankin, S., Reszka, A.P., Huppert, J., Zloh, M., Parkinson, G.N., Todd, A.K., Ladame, S., Balasubramanian, S. and Neidle, S. (2005) Putative DNA quadruplex formation within the human c-*kit* oncogene. J. Am. Chem. Soc. 127, 10584–10589.
[20] Cevec, M. and Plavec, J. (2005) Role of loop residues and cations on the formation and stability of dimeric DNA G-quadruplexes. Biochemistry 44, 15238–15246.
[21] Hazel, P., Parkinson, G.N. and Neidle, S. (2006) Topology variation and loop structural homology in crystal and simulated structures of a bimolecular DNA quadruplex. J. Am. Chem. Soc. 128, 5480–5487.
[22] Risitano, A. and Fox, K.R. (2004) Influence of loop size on the stability of intramolecular DNA quadruplexes. Nucleic Acids Res. 32, 2598–2606.
[23] Smirnov, I. and Shafer, R.H. (2000) Effect of loop sequence and size on DNA aptamer stability. Biochemistry 39, 1462–1468.
[24] Wang, Y. and Patel, D.J. (1994) Solution structure of the Tetrahymena telomeric repeat $d(T_2G_4)_4$ G-tetraplex. Structure 2, 1141–1156.
[25] Miyoshi, D., Karimata, H. and Sugimoto, N. (2005) Drastic effect of a single base difference between human and Tetrahymena telomere sequences on their structures under molecular crowding conditions. Angew. Chem., Int. Ed. Engl. 44, 3740–3744.
[26] Risitano, A. and Fox, K.R. (2003) Stability of intramolecular DNA quadruplexes: comparison with DNA duplexes. Biochemistry 42, 6507–6513.
[27] Darby, R.A.J., Sollogoub, M., McKeen, C., Brown, L., Risitano, A., Brown, N., Barton, C., Brown, T. and Fox, K.R. (2002) High throughput measurement of duplex, triplex and quadruplex melting curves using molecular beacons and a LightCycler. Nucleic Acids Res. 30, e39.
[28] Mergny, J.L. and Lacroix, L. (2003) Analysis of thermal melting curves. Oligonucleotides 13, 515–537.
[29] Balagurumoorthy, P., Brahmachari, S.K., Mohanty, D., Bansal, M. and Sasisekharan, V. (1992) Hairpin and parallel quartet structures for telomeric sequences. Nucleic Acids Res. 20, 4061–4067.
[30] Balagurumoorthy, P. and Brahmachari, S.K. (1994) Structure and stability of human telomeric sequence. J. Biol. Chem. 269, 21858–21869.
[31] Lu, M., Guo, Q. and Kallenbach, N.R. (1992) Structure and stability of sodium and potassium complexes of dT_4G_4 and dT_4G_4T. Biochemistry 31, 2455–2459.
[32] Rujan, I.N., Meleney, J.C. and Bolton, P.H. (2005) Vertebrate telomere repeat DNAs favor external loop propeller quadruplex structures in the presence of high concentrations of potassium. Nucleic Acids Res. 33, 2022–2031.
[33] Bommarito, S., Peyret, N. and SantaLucia, J. (2000) Thermodynamic parameters for DNA sequences with dangling ends. Nucleic Acids Res. 28, 1929–1934.
[34] Phan, A.T., Kuryavyi, V., Ma, J.B., Faure, A., Andreola, M.L. and Patel, D.J. (2005) An interlocked dimeric parallel-stranded DNA quadruplex: a potent inhibitor of HIV-1 integrase. Proc. Natl. Acad. Sci. USA 102, 634–639.
[35] Huppert, J.L. and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome. Nucleic Acids Res. 33, 2908–2916.
[36] Todd, A.K., Johnston, M. and Neidle, S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. Nucleic Acids Res. 33, 2901–2907.
[37] Olsen, C.M., Gmeiner, W.H. and Marky, L.A. (2006) Unfolding of G-quadruplexes: energetic, and ion and water contributions of G-quartet stacking. J. Phys. Chem. B 110, 6962–6969.

Effect of G-Tract Length on the Topology and Stability of Intramolecular DNA Quadruplexes†

Phillip A. Rachwal,‡ Tom Brown,§ and Keith R. Fox*,‡

School of Biological Sciences, University of Southampton, Bassett Crescent East, Southampton SO16 7PX, U.K., and School of Chemistry, University of Southampton, Highfield, Southampton SO17 1BJ, U.K.

Received October 11, 2006; Revised Manuscript Received January 10, 2007

ABSTRACT: G-Rich sequences are known to form four-stranded structures that are based on stacks of G-quartets, and sequences with the potential to adopt these structures are common in eukaryotic genomes. However, there are few rules for predicting the relative stability of folded complexes that are adopted by sequences with different-length G-tracts or variable-length linkers between them. We have used thermal melting, circular dichroism, and gel electrophoresis to examine the topology and stability of intramolecular G-quadruplexes that are formed by sequences of the type $d(G_nT)_4$ and $d(G_nT_2)_4$ ($n = 3-7$) in the presence of varying concentrations of sodium and potassium. In the presence of potassium or sodium, $d(G_nT)_4$ sequences form intramolecular parallel complexes with the following order of stability: $n = 3 > n = 7 > n = 6 > n = 5 > n = 4$. $d(G_3T)_4$ is anomalously stable. In contrast, the stability of $d(G_nT_2)_4$ increases with the length of the G-tract ($n = 7 > n = 6 > n = 5 > n = 4 > n = 3$). The CD spectra for $d(G_nT)_4$ in the presence of potassium exhibit positive peaks around 260 nm, consistent with the formation of parallel topologies. These peaks are retained in sodium-containing buffers, but when $n = 4$, 5, or 6, CD maxima are observed around 290 nm, suggesting that these sequences [especially $d(G_5T)_4$] have some antiparallel characteristics. $d(G_3T_2)_4$ adopts a parallel conformation in the presence of both sodium and potassium, while all the other $d(G_nT_2)_4$ complexes exhibit predominantly antiparallel features. The properties of these complexes are also affected by the rate of annealing, and faster rates favor parallel complexes.

G-Rich sequences are known to fold into four-stranded structures that contain stacks of G-quartets (*1*−*4*). These structures can be formed by the association of four separate DNA strands (*5*), the dimerization of two strands that each contain two G-tracts (*6*), or the intramolecular folding of a single strand that contains four G-tracts (*7*−*10*). Quadruplex formation requires the presence of monovalent cations (*11*, *12*) (especially potassium), while stable folding is usually inhibited by lithium (*13*). The cations fit within the central core of guanine carbonyls and can lie between or within the plane of each quartet.

There is considerable interest in quadruplexes since G-rich sequences with the potential to adopt these structures are found in several biologically important DNA regions, such as gene promoters and telomeres (*6*, *14*). Telomeres consist of long repeats of G-rich sequences, $(GGGTTA)_n$ in humans and higher eukaryotes, $(G_4T_2)_n$ in *Tetrahymena*, and $(G_4T_4)_n$ in *Oxytricha*. A number of other nontelomeric G-rich DNA sequences may also form quadruplexes, and these have been identified in the NHE element of the *c-myc* promoter (*15*−*18*), the promoters of the Ki-*ras* (*19*), *bcl2* (*20*, *21*), and *c-kit* oncogenes (*22*), the VEGF gene (*23*) and hypoxia inducible factor 1α (*24*), fragile X-syndrome (*25*), and other trinucleotide repeat sequences (*26*), the retinblastoma susceptibility gene (*27*), the chicken β-globin gene (*28*), and the insulin gene (*29*). There is also an overabundance of G-rich sequences in the regulatory regions of muscle-specific genes (*30*). Genome searches reveal that G-rich sequences with the potential to form quadruplexes are abundant in the human genome (*31*−*33*), the most common of which are successive G-tracts that are separated by single T or A residues. We have performed simple BLAST searches on repeated sequences of the type $d(G_nT)_4$ and find that these occur 147, 65, 120, and 2 times and once in the human genome when $n = 3-7$, respectively, and for $(G_nTT)_4$ for which there are 15, 75, and 6 occurrences when $n = 3-5$, respectively [we find no examples of $d(G_6TT)_4$ and $d(G_7TT)_4$].

Quadruplexes are known to adopt several different topologies, depending on their sequence and the ionic conditions. The four strands are all parallel in the intermolecular quadruplex (*5*), and the nucleotides are all in the *anti* conformation. However, multiple forms have been proposed for intramolecular quadruplexes in which the bases are both *syn* and *anti* (*1*). The strands can be arranged antiparallel to each other with the loops at the top and bottom of the stacked quartets (*1*, *9*) or in a parallel orientation with the loops in a lateral arrangement running between the bottom of one stack and the top of the other (*10*). Although there is still debate about which structure is biologically relevant (*34*), NMR (*35*), CD (*36*), and crystallographic studies (*10*) suggest that the human telomeric repeat $d(GGGTTA)_n$ adopts a

†P.A.R. is supported by a research studentship from BBSRC.

* To whom correspondence should be addressed: School of Biological Sciences, University of Southampton, Bassett Crescent East, Southampton SO16 7PX, U.K. Telephone: +44 23 8059 4374. Fax: +44 23 8059 4459. E-mail: k.r.fox@soton.ac.uk.

‡ School of Biological Sciences, University of Southampton, Bassett Crescent East.

§ School of Chemistry, University of Southampton, Highfield.

10.1021/bi062118j CCC: $37.00
Published on Web 02/21/2007

parallel structure in the presence of potassium. Several recent studies have shown that sequences closely related to the human telomeric repeat adopt a mixed topology with one double-chain reversal and two edgewise loops (*37*−*39*).

Some G-rich sequences form exceptionally stable complexes, such as the NHE element of *c-myc* and the inhibitor of HIV integrase $d(G_3T)_4$, which persist even in the presence of the complementary C-rich strand (*40*). However, there are few clear rules for predicting which G-rich sequences form the most stable quadruplexes or what topologies they will adopt. Quadruplex stability and topology are affected by the length of the loops between the G-tracts (*41*); G_3 tracts separated by short T-loops generate parallel topologies, while longer T-loops form antiparallel complexes (*41*). The stability is also affected by the sequence of the loops (*40*, *42*, *43*) and the bases that flank the quadruplex (*44*−*46*). However, there have been few studies on how formation is affected by the length of the G-stack. One might simply predict that longer G-stacks will increase quadruplex stability; however, for an intramolecular complex, this will affect the distance spanned by the connecting loops and may alter the folding pattern. Work with the thrombin-binding aptamer, which is based on two stacked quartets, has shown that the melting temperature is increased by ~20 °C upon addition of an extra G-quartet (*43*). In contrast, the *Tetrahymena* sequence $d(G_4T_2)_n$, which differs from the human sequence by exchanging A for G in each repeat, could in principle generate four stacked G-quartets linked by two-base loops. However, although this is observed for the intermolecular dimer (*47*), the intramolecular complex instead folds to form only three quartets with variable loops consisting of TGGT, TTG, and TT (*48*).

We have therefore explored how the length of the G-stack affects quadruplex stability and structure using synthetic oligonucleotides with $d(G_nT)_4$ and $d(G_nT_2)_4$ sequences (n = 3−7). The stability has been assessed by determining the melting temperatures under different ionic conditions, using a fluorescence melting assay as previously described (*40*, *42*, *49*), while structural changes are assessed from the CD spectra, since parallel and antiparallel quadruplexes generate characteristic CD spectra (*44*, *50*−*52*).

MATERIALS AND METHODS

Oligonucleotides. All oligonucleotides were synthesized on an Applied Biosystems ABI 394 automated DNA/RNA synthesizer on the 0.2 or 1 μmole scale. Phosphoramidite monomers and other reagents were purchased from Applied Biosystems or Link Technologies. The sequences of the oligonucleotides used in this work are given in Table 1. All oligonucleotides were prepared with 5′-fluorescein and 3′-dabcyl (fluorescein C6 phosphoramidite and dabcyl cpg purchased from Link Technologies Ltd.) for use in the fluorescence melting experiments, and the same sequences were used for the circular dichroism studies. The bases adjacent to the fluorophore and quencher were the same for all the oligonucleotides to prevent any differences in their effects on quadruplex formation and stability.

Gel Electrophoresis. Nondenaturing gel electrophoresis was performed using 16% polyacrylamide gels, which were run in TBE buffer that had been supplemented with 50 mM NaCl or KCl. Bands in the gels were visualized under UV light. The oligonucleotide concentration was 20 μM, and the samples were slowly annealed in the appropriate buffer. Denaturing electrophoresis was performed using 14% polyacrylamide gels containing 8 M urea.

Table 1: Sequences of Oligonucleotides Used in This Work

oligonucleotide sequence[a]	name
G_nT series	
5′-F-TGGGTGGGTGGGTGGGT-Q	G_3T
5′-F-TGGGGTGGGGTGGGGTGGGGT-Q	G_4T
5′-F-TGGGGGTGGGGGTGGGGGTGGGGGT-Q	G_5T
5′-F-TGGGGGGTGGGGGGTGGGGGGTGGGGGGT-Q	G_6T
5′-F-TGGGGGGGTGGGGGGGTGGGGGGGTGGGGGGGT-Q	G_7T
G_nT_2 series	
5′-F-TGGGTTGGGTTGGGTTGGGT-Q	G_3T_2
5′-F-TGGGGTTGGGGTTGGGGTTGGGGT-Q	G_4T_2
5′-F-TGGGGGTTGGGGGTTGGGGGTTGGGGGT-Q	G_5T_2
5′-F-TGGGGGGTTGGGGGGTTGGGGGGTTGGGGGGT-Q	G_6T_2
5′-F-TGGGGGGGTTGGGGGGGTTGGGGGGGTTGGGGGGGT-Q	G_7T_2

[a] F represents fluorescein and Q dabcyl.

Circular Dichroism. CD spectra were recorded on a Jasco J-720 spectropolarimeter as previously described (*42*). Oligonucleotide solutions (5 μM) were prepared in 10 mM lithium phosphate (pH 7.4) containing either 50 mM potassium chloride or 50 mM sodium chloride. The samples were heated to 95 °C and annealed by being slowly cooled to 15 °C over a period of 12 h. Spectra were recorded between 220 and 320 nm in 5 mm path length cuvettes. Spectra were averaged over 16 scans, which were recorded at 100 nm/min with a response time of 1 s and a bandwidth of 1 nm. A buffer baseline was subtracted from each spectrum, and the spectra were normalized to have zero ellipticity at 320 nm.

Fluorescence Melting. Fluorescence melting curves were determined in a Roche LightCycler as previously described (*40*, *42*, *49*) in a total reaction volume of 20 μL. Oligonucleotides (final concentration of 0.25 μM) were prepared in 10 mM lithium phosphate (pH 7.4), which was supplemented with various concentrations of potassium or sodium chloride. The LightCycler has one excitation source (488 nm), and the changes in fluorescence were measured at 520 nm. For several of the oligonucleotides, initial experiments revealed that there was considerable hysteresis between the heating and annealing profiles when the temperature was changed at a rate of 0.2 °C/s, indicating that the process was not at thermodynamic equilibrium. Melting experiments were therefore performed at a much slower rate of heating and cooling (0.1 °C/min) by changing the temperature in 1 °C steps, leaving the samples to equilibrate for 10 min at each temperature before the fluorescence was recorded. Under these conditions, no hysteresis was observed (except for some experiments with G_7T_2). In a typical experiment, the oligonucleotides were first denatured by heating to 95 °C for 5 min. They were then annealed by cooling to 30 °C at a rate of 0.1 °C/min and melted by heating to 95 °C at the same rate. The fluorescence was recorded during both the annealing and melting steps. In some instances, the formation of intramolecular or intermolecular complexes was examined by determining the melting curves using a range of oligonucleotide concentrations (from 20 nM to 2 μM).

Data Analysis. T_m values were obtained from the maxima of the first derivatives of the melting profiles using the LightCycler software or, together with ΔH, from van't Hoff

analysis of the melting profiles. In some instances, the melting curves showed a linear change in fluorescence with temperature in regions outside the melting transition. We accounted for this by fitting a straight line to the first and last portions of the fluorescence curve. The fraction folded was calculated as previously described (*53*) from the difference between the measured fluorescence and the upper and lower baselines. All reactions were performed at least twice, and the calculated T_m values usually differed by <0.5 °C with a 5% variation in ΔH. Since $\Delta G = 0$ at the T_m, ΔS was estimated to be $\Delta H/T_m$. Values for ΔG at 310 K were then estimated from $\Delta G = \Delta H - T\Delta S$ (assuming that $\Delta C_p = 0$). This analysis assumes a simple two-state equilibrium between the folded and unfolded forms. The presence of polymorphic quadruplex structures will lead to shallower melting profiles and therefore smaller apparent values for ΔH. The number of specifically bound monovalent cations (Δn) was calculated from the slopes of plots of ΔG against $\log[M^+]$ as previously described (*54*, *55*).

RESULTS

We have recorded the CD spectra and thermal melting profiles of two series of oligonucleotides that contain G-tracts of varying lengths. Oligonucleotides in the first series contain four identical G-tracts with lengths between three and seven bases, which are separated by single thymines. In the second series, the G-tracts are each separated by two thymines. Both series of oligonucleotides were prepared for use in fluorescence melting experiments and contain fluorescein at the 5′-end and dabcyl at the 3′-end. Previous studies have shown that the addition of fluorescent groups has only small effects on quadruplex stability (*56*).

Circular Dichroism

Several studies have shown that the CD spectra of quadruplexes can be used to indicate whether they fold in a parallel or antiparallel configuration (*44*, *50*−*52*). Antiparallel quadruplexes exhibit a positive CD signal at around 295 nm, with a negative signal or shoulder around 260 nm. In contrast, parallel quadruplexes display a positive signal around 260 nm, with a negative peak at 240 nm. Unfolded oligonucleotides do not display these spectral signatures. Clearly, these signals not only reflect the arrangements of the strands but also are sensitive to the *syn*/*anti* orientations around the glycosidic bonds. Parallel topologies have all-*anti* glycosidic angles, while antiparallel ones have *syn* and *anti* in varying ratios depending on the particular arrangement. Nonetheless, they are a good indicator of changes in the global quadruplex configuration.

d(G$_n$T)$_4$. The CD spectra of oligonucleotides with a d(G$_n$T)$_4$ sequence ($n = 3-7$) in the presence of potassium or sodium are presented in Figure 1. These were obtained after slowly cooling the oligonucleotides from 95 °C, though identical spectra were observed when the complexes were rapidly cooled. It can be seen that, in the presence of potassium, all the sequences display a major positive peak around 260 nm with a minimum around 240 nm, which is typical of a parallel arrangement of the strands. G$_3$T and G$_4$T exhibit no ellipicity around 290 nm, though the other sequences show some signal in this region which is greatest for G$_6$T. This second peak suggests that these sequences may

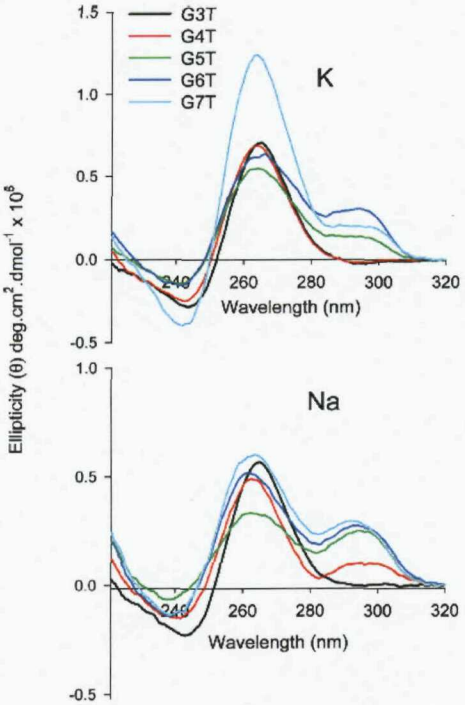


FIGURE 1: CD spectra of the fluorescently labeled oligonucleotides of the series d(G$_n$T)$_4$. The oligonucleotides (5 μM) were dissolved in 10 mM lithium phosphate (pH 7.4) containing 50 mM potassium chloride (top traces) or 50 mM sodium chloride (bottom traces): black for G$_3$T, red for G$_4$T, green for G$_5$T, blue for G$_6$T, and cyan for G$_7$T.

exhibit structural polymorphism with a significant fraction adopting an antiparallel configuration (though it is possible that unpaired guanines in the loops might also affect the CD spectra). However, these spectra are also typical of mixed parallel/antiparallel topologies (*18*, *20*). In the presence of sodium, G$_3$T again shows a single peak at 260 nm, indicative of a parallel complex, while the other oligonucleotides display significant ellipticity at 295 nm. In sodium, the relative intensity of the 295 nm peak compared to that at 260 nm is greatest for G$_5$T.

These results confirm that the oligonucleotides all fold to form G-quadruplexes but suggest that their structures may be polymorphic. They predominantly fold into a parallel configuration (especially G$_3$T), though these coexist with significant amounts of the antiparallel form (especially G$_5$T in the presence of sodium). Via a comparison of these results with those published for sequences of the type (G$_3$T$_n$)$_4$ (*41*), it is clear that adding G bases to the sequence has an effect different from the effect of adding extra T bases to the loops. d(G$_3$T$_3$)$_4$ and d(G$_5$T)$_4$ have the same repeat length; however, the former adopts an antiparallel topology, while the latter is largely parallel.

d(G$_n$T$_2$)$_4$. The CD spectra of sequences containing four repeats of G$_n$T$_2$ ($n = 3-7$) in the presence of sodium or potassium are shown in Figure 2. In the presence of potassium, the spectra are different when the complexes are rapidly cooled on ice (inset) or slowly annealed. After the rapid cooling, the spectra are similar to those of the G$_n$T series, with a strong peak at 260 nm and a weaker one at 295 nm, indicating that the parallel configuration is the major form. In contrast, the peaks at 295 nm are much more

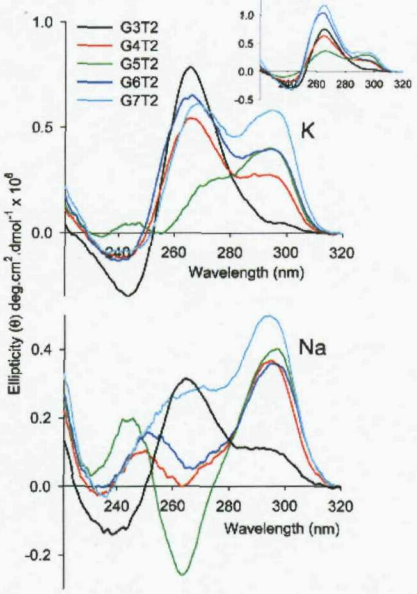


FIGURE 2: CD spectra of the fluorescently labeled oligonucleotides of the series $d(G_nT_2)_4$. The oligonucleotides (5 μM) were dissolved in 10 mM lithium phosphate (pH 7.4) containing 50 mM potassium chloride (top traces) or 50 mM sodium chloride (bottom traces): black for G_3T_2, red for G_4T_2, green for G_5T_2, blue for G_6T_2, and cyan for G_7T_2.

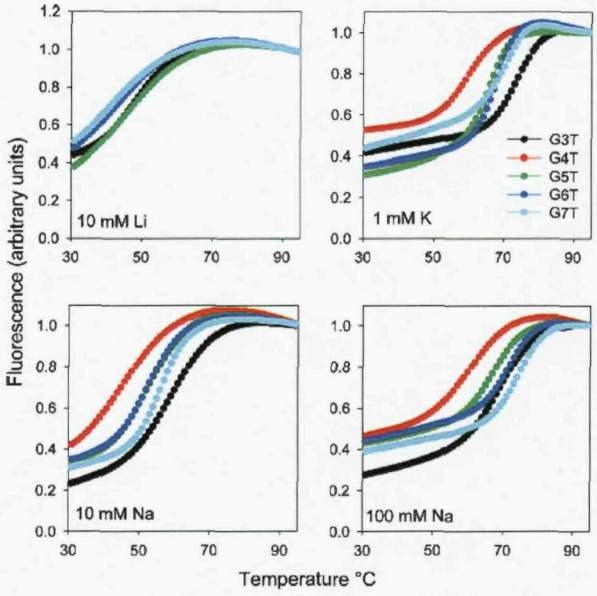


FIGURE 3: Fluorescence melting curves for oligonucleotides of the series $d(G_nT)_4$. The melting profiles were determined in 10 mM lithium phosphate (pH 7.4) containing different concentrations of potassium or sodium chloride as indicated. The curves have been normalized to the same final fluorescence value: black for G_3T, red for G_4T, green for G_5T, blue for G_6T, and cyan for G_7T.

pronounced when the oligonucleotides are slowly annealed. G_3T_2 appears to be mainly parallel, while G_4T_2 has a significant shoulder at 295 nm. G_5T_2 is predominantly in the antiparallel form, though the proportion of the antiparallel species appears to decrease with G_6T_2 and G_7T_2.

In the presence of sodium, the spectra exhibit large peaks at 295 nm, indicative of antiparallel complex formation. Even the shortest sequence, G_3T_2, reveals a significant shoulder in this region, though it is predominantly in the parallel form. In contrast, G_4T_2, G_5T_2, and G_6T_2 are mainly in an antiparallel form with no positive peak at 260 nm. G_7T_2 is mainly antiparallel but has a significant shoulder around 260 nm. These spectra in the presence of sodium are independent of the rate of cooling.

For these CD studies, we were concerned that the fluorophore and quencher might affect the conformational equilibrium. We therefore determined the CD spectra of unlabeled sequences that showed examples of very different spectra (G_3T, G_5T, and G_5T_2). These are shown in Figure 1 of Supporting Information and reveal only subtle changes in the details of the spectra. The positions and relative intensities of the peaks at 260 and 295 nm are largely unaffected by removal of the fluorescent groups.

Fluorescence Melting

$d(G_nT)_4$. Typical fluorescence melting curves for these oligonucleotides are shown in Figure 3, and the T_m and ΔH values determined at different ionic concentrations are summarized in Table 2. These labeled oligonucleotides are designed so that the fluorescence is quenched upon formation of the quadruplex, as the fluorophore (fluorescein) and quencher (dabcyl) are close together. These groups are separated when the complex melts, and there is a large increase in fluorescence. This cannot be used to distinguish

Table 2: Melting Temperatures and ΔH Values Derived from Fluorescence Melting Curves for Oligonucleotides of the Type $(G_nT)_4$ ($n = 3-7$)[a]

	T_m (°C)					ΔH (kJ/mol)				
	$n=3$	$n=4$	$n=5$	$n=6$	$n=7$	$n=3$	$n=4$	$n=5$	$n=6$	$n=7$
10 mM Li	48.3	41.8	43.7	43.4	43.1	122	102	104	111	100
0.1 mM K^+	58.4	46.1	50.0	51.8	54.3	210	119	175	227	236
1 mM K^+	73.8	58.6	65.5	67.6	71.3	291	200	272	322	326
5 mM K^+	85.4	71.3	76.4	79.1	86.5	299	273	324	391	381
10 mM K^+		76.0	80.8	83.3	88.4		*291*	282	374	*351*
1 mM Na^+	52.7	44.3	46.5	47.2	47.4	141	110	117	120	118
10 mM Na^+	59.8	46.5	52.3	52.3	54.0	177	118	179	177	213
50 mM Na^+	66.0	55.5	62.8	65.4	68.7	197	164	221	253	291
100 mM Na^+	69.7	60.8	67.7	71.9	75.1	246	186	239	294	314
200 mM Na^+	75.3	68.1	75.6	81.4	82.6	198	202	214	340	346

[a] The values were determined in 10 mM lithium phosphate (pH 7.4) containing different concentrations of sodium or potassium. Each value is the average of four determinations (two melting and two annealing profiles). T_m values are accurate to within 0.5 °C, while ΔH values varied by ~5%.

between the different folded forms, as the fluorophore and quencher are close together in both the parallel and the antiparallel topologies. All the sequences exhibit fluorescence changes that are consistent with quadruplex formation, and they are all much more stable in the presence of potassium than sodium. We confirmed that these transitions are not intermolecular complexes by examining the concentration dependence of the melting profiles (from 20 nM to 2 μM). These results (shown in Figure 2 of the Supporting Information) show that the T_m values are independent of concentration. We can therefore assume that these are intramolecular complexes over this range of concentrations. All these melting curves were fully reversible at this rate of heating and cooling (0.1 °C/min).

The first panel of Figure 3 shows the transitions in the presence of 10 mM lithium alone. Surprisingly, these complexes display a melting profile (with a T_m of ~45 °C), even in the absence of potassium or sodium. This is similar to the effect seen with sequences containing four G_3 tracts that are connected by non-nucleosidic linkers (*42*). Although this may indicate some quadruplex formation, it may be significant that these complexes all have very similar T_m and ΔH values, in contrast to their behavior in the presence of sodium or potassium. The addition of even low concentrations of potassium causes a dramatic increase in stability, and the complexes are too stable to measure at potassium concentrations of >10 mM.

In the presence of ≤1 mM potassium, G_3T is the most stable while G_4T is the least stable. In general, the stability increases with the length of the G-tract ($n = 4 < n = 5 < n = 6 < n = 7$), though the behavior of G_3T is anomalous since it is more stable than G_7T. All the complexes become more stable when the potassium ion concentration increases. ΔH values for these transitions were estimated from van't Hoff analysis of the melting profiles, assuming a two-state equilibrium, and these are also summarized in Table 2. The presence of polymorphic quadruplex structures will lead to shallower melting profiles and therefore smaller apparent values of ΔH. This may explain why there is no simple correlation between ΔH and the number of G bases in each G-tract, though in general the values are higher for the longer G-tracts (with the exception of G_3T).

The results of similar experiments in the presence of 10 and 100 mM sodium are shown in the bottom panels of Figure 3, and all the data are summarized in Table 2. In contrast to potassium, the addition of 1 mM sodium has an only small effect on the melting profiles, increasing the T_m values by only 3−4 °C relative to that in 10 mM lithium. In the presence of 10 mM sodium, G_3T again forms the most stable structure; G_4T is the least stable, and G_5T, G_6T, and G_7T have similar T_m values. G_7T becomes the most stable when the sodium concentration is increased to 100 mM, though G_3T is still more stable than G_4T, G_5T, or G_6T. In general, the ΔH values are higher for the longer G-tracts, though G_3T is again anomalous.

$d(G_nT_2)_4$. The results of similar studies with oligonucleotides in which the G-tracts are separated by two T bases are shown in Figure 4, and the T_m and ΔH values are summarized in Table 3. No melting transitions were observed in the presence of lithium alone. In the presence of either potassium or sodium, the least stable complex is formed by G_3T_2, while G_7T_2 is the most stable. In potassium, the order

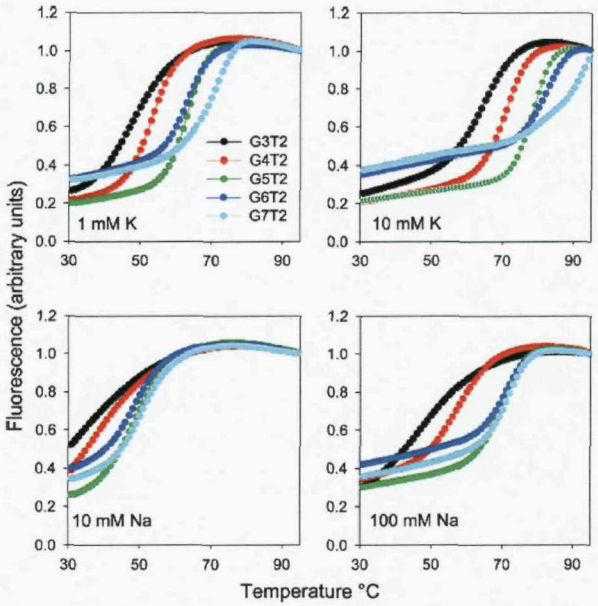


FIGURE 4: Fluorescence melting curves for oligonucleotides of the series $d(G_nT_2)_4$. The melting profiles were determined in 10 mM lithium phosphate (pH 7.4) containing different concentrations of potassium or sodium chloride as indicated. The curves have been normalized to the same final fluorescence value: black for G_3T, red for G_4T, green for G_5T, blue for G_6T, and cyan for G_7T.

of stability is as follows: $G_3T_2 < G_4T_2 < G_5T_2 = G_6T_2 < G_7T_2$ (though there is some hysteresis with G_7T_2 for which the melting temperatures are 3, 8, and 10 °C higher than the annealing temperatures in the presence of 0.1, 1, and 5 mM potassium, respectively). These structures are unstable in the presence of sodium concentrations below 10 mM, and the order of stability is as follows: $G_3T_2 < G_4T_2 < G_5T_2 = G_6T_2 = G_7T_2$.

Ionic Strength Dependence

The ΔH values for these transitions exhibit a strong dependence on the ionic strength, which is consistent with the presence of specific ion binding sites within the quadruplex (*54*). The slopes of plots of ΔG versus $\log[M^+]$ can be used to determine the stoichiometry of ion binding as previously described (*54*, *55*), yielding values of Δn (the difference between the number of ions bound in the folded and unfolded states). Figure 5 shows the variation in Δn with the number of guanines in each stack for both series of oligonucleotides, in the presence of sodium or potassium. We have previously determined a value of 2.6 for the human telomeric repeat sequence in the presence of potassium (*42*), and a value of 3.7 ± 0.2 has been reported for T30695, which has the same repeat sequence as G3T (*54*). For a structure containing three stacked G-quartets, we would predict that Δn should be either 2 (the number of potassium ions located between the stacked quartets) or 4 (including two further ions that may be coordinated between the loops and the terminal quartets). A value of 2 seems more likely for a parallel topology, in which the loops do not interact with the terminal quartets. In either case, we would expect that adding a further G-quartet should increase the value of Δn

Table 3: Melting Temperatures and ΔH Values Derived from Fluorescence Melting Curves for Oligonucleotides of the Type $(G_nT_2)_4$ (n = 3–7)[a]

	T_m (°C)					ΔH (kJ/mol)				
	$n=3$	$n=4$	$n=5$	$n=6$	$n=7$	$n=3$	$n=4$	$n=5$	$n=6$	$n=7$
0.1 mM K^+		36.3	47.0	46.7	51.1[b]		99	188	186	213[b]
1 mM K^+	48.1	53.3	62.9	63.0	68.3[b]	136	224	294	283	289[b]
5 mM K^+	60.1	66.1	74.1	74.7	81.9[b]	199	263	359	345	322[b]
10 mM K^+	65.0	71.4	78.9	80.1		241	286	376	371	
50 mM K^+	76.2	82.4	89.7			267	308	336		
1 mM Na^+				42.5	41.4				100	101
5 mM Na^+			45.1	47.2	47.1			134	123	132
10 mM Na^+		37.3	46.5	52.3	50.7		97	157	155	168
50 mM Na^+	40.3	49.9	62.3	65.4	63.7	99	152	248	260	278
100 mM Na^+	46.8	57.7	70.2	71.9	71.9	113	188	290	314	318
200 mM Na^+	54.4	66.2	78.3	81.4	81.3	144	225	323	354	334

[a] The values were determined in 10 mM lithium phosphate (pH 7.4) containing different concentrations of sodium or potassium. Each value is the average of four determinations (two melting and two annealing profiles). T_m values are accurate to within 0.5 °C, while ΔH values varied by ~5%. [b] Hysteresis was observed between the melting and annealing profiles.

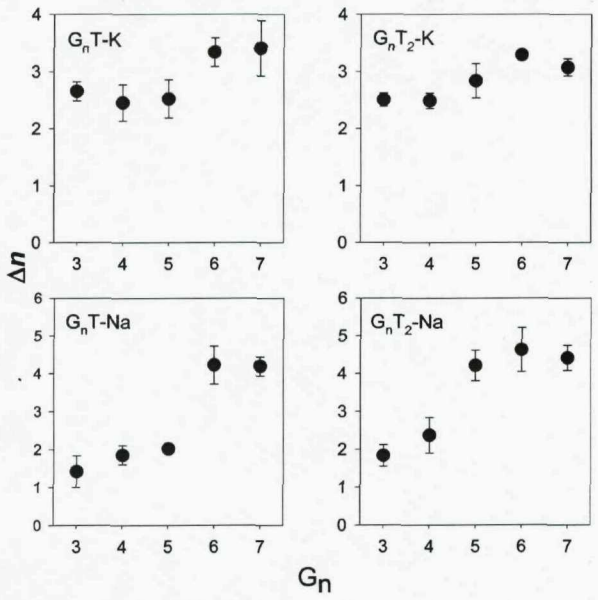


FIGURE 5: Variation in the number of potassium or sodium ions specifically bound to each quadruplex (Δn, y-axis) as a function of the length of the G-tracts (G_n, x-axis). The values for Δn were determined from the dependence of ΔG on ionic strength, as described in the text.

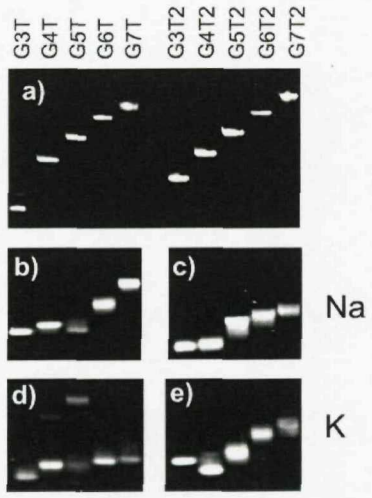


FIGURE 6: Mobility of the fluorescently labeled oligonucleotides in polyacrylamide gels: (a) mobility on a 14% polyacrylamide gel containing 8 M urea and (b–e) mobility on 16% nondenaturing polyacrylamide gels supplemented with 50 mM NaCl (b and c) or 50 mM KCl (d and e).

by 1. It is therefore significant that, for the G_nT series in the presence of potassium, Δn is the same for G_3T, G_4T, and G_5T, suggesting that these each contain only three stacked G-quartets. Δn increases by ~1 for G_6T and G_7T, suggesting that these complexes contain an additional quartet. The values of Δn exhibit even less variation for the G_nT_2 series in the presence of potassium, again suggesting that there are only three stacked G-quartets. The variations in Δn are more pronounced in the presence of sodium; G_3T, G_4T, and G_5T have similar values, between 1.5 and 2, which rises to ~4 for G_6T and G_7T. A similar pattern is seen for the G_nT_2 series in sodium, though the transition between Δn values of 2 and 4 occurs with G_5T_2. These results are considered further in the Discussion.

Gel Mobility

We sought to further explore the topologies of these intramolecular quadruplexes by comparing their mobilities in polyacrylamide gels, and the results are presented in Figure 6. As expected, the unfolded oligonucleotides run as single bands on denaturing polyacrylamide gels (Figure 6a), with mobilities that depend on their length. G_3T appears to have an anomalously fast mobility, probably because it is still partially folded even under these harsh conditions. In contrast, the relative mobility of the various folded species is very different. In the presence of sodium (Figure 6b,c), although single bands are observed for each oligonucleotide, G_3T, G_4T, and G_5T have similar mobilities, while G_6T and G_7T run more slowly (Figure 6b). The pattern is different for the G_nT_2 series (Figure 6c) in which G_3T_2 and G_4T_2 have similar mobilities, while G_5T_2, G_6T_2, and G_7T_2 run more slowly and have similar mobilities. These patterns are similar to the variations in the apparent number of sodium ions released on the melting of each complex (Figure 5, bottom panels), suggesting that they reflect the number of stacked G-quartets in each complex.

The gel mobilities in the presence of 50 mM potassium show a different pattern. For the G_nT series, two bands are evident with G_3T, G_4T, and G_5T, the faster of which

comigrates with G_6T and G_7T, though G_3T has a slightly faster mobility. The intensity of the slowly migrating species was dependent on the method of annealing, and these were more abundant when the complexes were prepared by slow annealing (as shown). We assume that the slow-migrating species (which is especially abundant with G_5T) corresponds to intermolecular complexes, while the faster species are intramolecular. It should be noted that these complexes were prepared with 20 μM oligonucleotide (in contrast to 0.25 μM oligonucleotide required for the fluorescence melting experiments), and this 80-fold higher concentration will favor the formation of bimolecular or tetramolecular complexes. The observation that all the faster species have similar mobilities is consistent with the suggestion that they contain the same number of stacked G-quartets. For the G_nT_2 series, the mobility decreases as the number of G bases increases from G_4T_2 to G_7T_2, though G_3T_2 is unusually slow.

DISCUSSION

These results demonstrate that, as expected, all the G-rich sequences fold to form quadruplexes; the CD spectra exhibit positive peaks at either 260 or 295 nm, and the fluorescence melting profiles show a folded conformation in which the fluorophore and quencher are close together at low temperatures. The folded complexes are stabilized by addition of monovalent cations, and potassium consistently generates more stable complexes than sodium. The complexes appear to be intramolecular (rather than intermolecular) as the T_m values are generally independent of oligonucleotide concentration.

These results address three main questions concerning these quadruplexes: what their relative stabilities are, how many stacked G-quartets are present in each complex (i.e., are any of the G bases placed in the loops), and whether they adopt a parallel or antiparallel topology.

Stabilities. Although the fluorescence melting data cannot distinguish between different conformations, they provide an estimation of the overall stability of the different complexes. Under all the conditions that were investigated, the stability increases with the length of the G-tracts, except for G_3T, which has an anomalously high stability in the presence of both sodium and potassium. In low concentrations of both potassium and sodium, the order of T_m values is as follows: $G_3T > G_7T > G_6T > G_5T > G_4T$. The unusual stability of G_3T has previously been noted (*40*, *41*) and may result from conformational changes after the initial formation of the complex (*54*). This order is retained at higher potassium concentrations, except that G_7T and G_3T display very similar T_m values. In high concentrations of sodium, G_3T is further down the rank order of stabilities and is intermediate between G_5T and G_6T. The rank order of stability for the second series, in which the G-tracts are separated by two T bases, shows that longer G-tracts produce more stable structures. In the presence of potassium, G_5T_2 and G_6T_2 have very similar T_m values, while G_7T_2, G_6T_2, and G_5T_2 are similar in the presence of sodium. These results demonstrate that there is no simple relationship between quadruplex stability and the length of the G-tracts, though the behaviors in sodium and potassium are similar.

Parallel or Antiparallel. The presence of peaks at 260 or 295 nm in the CD spectrum is typically used as an indicator of whether the quadruplex adopts a parallel or antiparallel topology (*44*, *50*−*52*), as a result of the *syn*/*anti* configuration of the glycosidic bonds, which are *anti* for the parallel structures and *syn* and *anti* in varying ratios for antiparallel complexes (*60*, *61*). The results show that the shortest oligonucleotides (G_3T, G_3T_2, and G_4T) adopt parallel topologies in the presence of both sodium and potassium. This is consistent with previous studies on the effect of loop length (*41*), which suggested that short loops are not sufficient to generate antiparallel topologies. The parallel structure proposed for G_3T (T30695, HIV integrase inhibitor) is at variance with the published NMR structure, which suggested an antiparallel structure containing two stacked G-quartets (*54*, *57*, *59*), though it has been noted that this structure requires further investigation (*59*).

In the presence of potassium, the oligonucleotides with single T residues between the G-tracts all adopt a structure for which the CD signature suggests a predominantly parallel topology. The parallel topology also appears to be the main form in the presence of sodium, though the greater CD signal at 295 nm suggests that this may be in equilibrium with the antiparallel form. This is especially noticeable for G_5T, for which the peaks at 260 and 295 nm are closer in intensity. In contrast, the CD spectra of G_6T are very similar in the presence of both sodium and potassium.

The sequences with two T bases between the G-tracts show a much stronger propensity to adopt the antiparallel conformation (except G_3T_2, for which the potassium and sodium forms are exclusively parallel). G_4T_2 is largely parallel in the presence of potassium but switches to the antiparallel form in sodium. G_5T_2 adopts a topology which is mainly antiparallel in both sodium and potassium. In contrast, G_6T_2 and G_7T_2 show mixed conformations in potassium but switch to mainly antiparallel conformations in sodium. The strong signal at 295 nm for G_6T and G_7T argues against aggregation into an intermolecular form, which would be expected to be parallel.

How Many G-Quartets? For the longer sequences, it is clear that a folded structure must contain some G residues in the loops as, for instance, a single T will be insufficient to span between the top and bottom of a stack of seven potential G-quartets in G_7T. The values of Δn, derived from the thermodynamic parameters, should be interpreted with caution, as there may be multiple forms in equilibrium, which may vary according to the ionic strength. However, they do indicate some trends concerning the number of stacked G-quartets in each complex, which are consistent with the gel mobility patterns. In potassium, G_3T, G_4T, and G_5T all appear to contain only three stacked quartets, suggesting that their loops contain a single T, GT, and GGT, respectively. For G_4T and G_5T, several structures could be envisaged depending on which G bases are stacked and which are in the loops; several different forms may coexist in solution in which the G-strands slip relative to each other. The values of Δn for G_6T and G_7T appear to increase by one, suggesting that these may contain four stacked G-quartets, with GGT and GGGT in their (lateral) loops. A similar transition is evident for G_nT between $n = 5$ and 6 in the presence of sodium, though in this case the value of Δn is only $\sim$2 for G_3T, G_4T, and G_5T and increases to 4 for G_6T and G_7T. The lower value might indicate the presence of one fewer stacked quartet or more likely indicates that the sodium ions

are less tightly held between the terminal G bases and the loops.

In the presence of potassium, the values of Δn are similar for all the complexes. The similarity in G_3T_2 and G_4T_2 is consistent with the NMR structure of the intramolecular *Tetrahymena* repeat (G_4T_2) which shows an antiparallel complex containing three stacked G-quartets. In this case, it appears that increasing the length of the G-tracts does not increase the number of stacked G-quartets and the additional G bases must reside in the loops. In sodium, G_3T_2 and G_4T_2 have the same number of G-quartets, which increase for G_5T_2, G_6T_2, and G_7T_2.

Comparison with Other Studies. In our previous publication (*40*), we suggested that G_3T_2 was more stable than G_4T_2, in contrast to our study in which the melting temperatures increase with the longer G-tracts. Several factors may contribute to this difference. First, this study was performed in 10 mM lithium phosphate buffer to which different concentrations of potassium or sodium had been added, in contrast to the earlier work which simply used sodium or potassium phosphate buffers. Second, different fluorophores and quenchers have been used in the two studies, which may affect the relative stabilities. In this study, these are separated from the quadruplex by single thymines at each end, while in the previous work, the fluorophores were positioned immediately adjacent to the terminal G bases. Third, and probably most importantly, we have employed much slower rates of heating and cooling. Risitano and Fox (*40*) measured annealing profiles at a rate of temperature change of 0.1 °C/s and noted that there was hysteresis between the heating and cooling curves, some of which were biphasic. In this study, we have shown that, in the presence of potassium, the CD spectra of the G_nT_2 series are affected by the rate of annealing; rapid cooling produces parallel structures, while slow cooling generates antiparallel complexes. In this study, we therefore used a 60-fold slower rate of temperature change in determining the melting profiles (0.1 °C/min) and observed no hysteresis between the heating and cooling curves, and the profiles were all monophasic. This work has therefore been performed under equilibrium conditions, allowing a proper analysis of relative stabilities for the different folded quadruplexes.

SUPPORTING INFORMATION AVAILABLE

Full details comparing the CD spectra of labeled and unlabeled oligonucleotides and the concentration dependence of the melting profiles. This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Simonsson, T. (2001) G-Quadruplex DNA structures: Variations on a theme, *Biol. Chem. 382*, 621−628.
2. Burge, S., Parkinson, G. N., Hazel, P., Todd, A. K., and Neidle, S. (2006) Quadruplex DNA: Sequence, topology and structure, *Nucleic Acids Res. 34*, 5402−5415.
3. Keniry, M. A. (2000) Quadruplex structures in nucleic acids, *Biopolymers 56*, 123−146.
4. Davies, J. T. (2004) G-Quartets 40 years later: From 5′-GMP to molecular biology and supramolecular chemistry, *Angew. Chem., Int. Ed. 43*, 668−698.
5. Phillips, K., Dauter, Z., Murchie, A. I., Lilley, D. M. J., and Luisi, B. (1997) The crystal structure of a parallel-stranded guanine tetraplex at 0.95 Angstrom resolution, *J. Mol. Biol. 273*, 171−182.
6. Sundquist, W. I., and Klug, A. (1989) Telomeric DNA dimerizes by formation of guanine tetrads between hairpin loops, *Nature 342*, 825−829.
7. Henderson, E., Hardin, C. C., Walk, S. K., Tinoco, I., and Blackburn, E. H. (1987) Telomeric DNA oligonucleotides form novel intramolecular structures containing guanine-guanine base pairs, *Cell 51*, 899−908.
8. Macaya, R. F., Schultze, P., Smith, F. W., Roe, J. A., and Feigon, J. (1993) Thrombin-binding DNA aptamer forms a unimolecular quadruplex structure in solution, *Proc. Natl. Acad. Sci. U.S.A. 90*, 3745−3749.
9. Wang, Y., and Patel, D. J. (1993) Solution structure of the human telomeric repeat d[$AG_3(T_2AG_3)_3$] G-tetraplex, *Structure 1*, 76−94.
10. Parkinson, G. N., Lee, M. P. H., and Neidle, S. (2002) Crystal structure of parallel quadruplexes from human telomeric DNA, *Nature 417*, 876−880.
11. Williamson, J. R., Raghuraman, M. K., and Cech, T. R. (1989) Monovalent cation induced structure of telomeric DNA: The G-quartet model, *Cell 59*, 871−880.
12. Sen, D., and Gilbert, W. (1990) A sodium-potassium switch in the formation of 4-stranded G4-DNA, *Nature 344*, 410−414.
13. Sen, D., and Gilbert, W. (1992) Guanine quartet structures, *Methods Enzymol. 211*, 191−199.
14. Sen, D., and Gilbert, W. (1988) Formation of parallel four-stranded complexes by guanine rich motifs in DNA and its implications for meiosis, *Nature 344*, 410−414.
15. Simonsson, T., Pecinka, P., and Kubista, M. (1998) DNA tetraplex formation in the control regions of c-myc, *Nucleic Acids Res. 26*, 1167−1172.
16. Rangan, A., Fedoroff, O. Y., and Hurley, L. H. (2001) Induction of duplex to G-quadruplex transition in the c-myc promoter region by a small molecule, *J. Biol. Chem. 276*, 4640−4646.
17. Siddiqui-Jain, A., Grand, C. L., Bearss, D. J., and Hurley, L. H. (2002) Direct evidence for a G-quadruplex in a promoter region and its targeting with a small molecule to repress c-MYC transcription, *Proc. Natl. Acad. Sci. U.S.A. 99*, 11593−11598.
18. Ambrus, A., Chen, D., Dai, J. X., Jones, R. A., and Yang, D. Z. (2005) Solution structure of the biologically relevant G-quadruplex element in the human c-MYC promoter. Implications for G-quadruplex stabilization, *Biochemistry 44*, 2048−2058.
19. Cogoi, S., and Xodo, L. E. (2006) G-Quadruplex formation within the promoter of the K-ras proto-oncogene and its effect on transcription, *Nucleic Acids Res. 34*, 2536−3549.
20. Dai, J. X., Dexheimer, T. S., Chen, D., Carver, M., Ambrus, A., Jones, R. A., and Yang, D. Z. (2006) An intramolecular G-quadruplex structure with mixed parallel/antiparallel G-strands formed in the human bcl-2 promoter region in solution, *J. Am. Chem. Soc. 128*, 1096−1098.
21. Dexheimer, T. S., Sun, D., and Hurley, L. H. (2006) Deconvoluting the structural and drug-recognition complexity of the G-quadruplex-forming region upstream of the bcl-2 P1 promoter, *J. Am. Chem. Soc. 128*, 5404−5415.
22. Rankin, S., Reszka, A P., Huppert, J., Zloh, M., Parkinson, G. N., Todd, A. K., Ladame, S., Balasubramanian, S., and Neidle, S. (2005) Putative DNA quadruplex formation within the human c-kit oncogene, *J. Am. Chem. Soc. 127*, 10584−10589.
23. Sun, Y., Guo, K. X., Rusche, J. J., and Hurley, L. H. (2005) Facilitation of a structural transition in the polypurine/polypyrimidine tract within the proximal promoter region of the human VEGF gene by the presence of potassium and G-quadruplex-interactive agents, *Nucleic Acids Res. 33*, 6070−6080.
24. De Armond, R., Wood, S., Sun, D. Y., Hurley, L. H., and Ebbinghaus, S. W. (2005) Evidence for the presence of a guanine quadruplex forming region within a polypurine tract of the hypoxia inducible factor 1α promoter, *Biochemistry 44*, 16341−16350.
25. Fry, M., and Loeb, M. A. (1994) The fragile-X syndrome d$(CGG)_n$ nucleotide repeats form a stable tetrahelical structure, *Proc. Natl. Acad. Sci. U.S.A. 91*, 4950−4954.
26. Matsugami, A., Okuizumi, T., Uesugi, S., and Katahira, M. (2003) Intramolecular higher order packing of parallel quadruplexes comprising a G:G:G:G tetrad and a G(:A):G(:A):G(:A):G heptad of GGA triplet repeat DNA, *J. Biol. Chem. 278*, 28147−28153.
27. Murchie, A. I., and Lilley, D. M. J. (1992) Retinoblastoma susceptibility genes contain 5′ sequences with a high propensity to form guanine-tetrad structures, *Nucleic Acids Res. 20*, 49−53.
28. Howell, R. M., Woodford, K. J., Weitzmann, M. N., and Usdin, K. (1996) The chicken β-globin gene promoter forms a novel "cinched" tetrahelical structure, *J. Biol. Chem. 271*, 5208−5214.

29. Lew, A., Rutter, W. J., and Kennedy, G. C. (2000) Unusual DNA structure of the diabetes susceptibility locus IDDM2 and its effect on transcription by the insulin promoter factor Pur-1/MAZ, *Proc. Natl. Acad. Sci. U.S.A. 97*, 12508−12512.
30. Yafe, A., Etzioni, S., Weisman-Shomer, P., and Fry, M. (2005) Formation and properties of hairpin and tetraplex structures of guanine-rich regulatory sequences of muscle-specific genes, *Nucleic Acids Res. 33*, 2887−2990.
31. Todd, A. K., Johnston, M., and Neidle, S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA, *Nucleic Acids Res. 33*, 2901−2907.
32. Huppert, J. L., and Balasubramanian, S. (2005) Prevalence of quadruplexes in the human genome, *Nucleic Acids Res. 33*, 2908−2916.
33. Huppert, J. L.; and Balasubramanian, S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res. 35*, 406−413.
34. Li, J., Correia, J. J., Wang, L., Trent, J. O., and Chaires, J. B. (2005) Not so crystal clear: The structure of the human telomere G-quadruplex in solution differs from that present in a crystal, *Nucleic Acids Res. 33*, 4649−4659.
35. Phan, A. T., and Patel, D. J. (2003) Two repeat human telomeric d(TAGGGTTAGGGT) sequence forms interconverting parallel and antiparallel G-quadruplexes in solution: Distinct topologies, thermodynamic properties and folding/unfolding kinetics, *J. Am. Chem. Soc. 125*, 15021−15027.
36. Rujan, I. N., Meleney, J. C., and Bolton, P. H. (2005) Vertebrate telomere repeat DNAs favour external propeller quadruplex structures in the presence of high concentrations of potassium, *Nucleic Acids Res. 33*, 2022−2031.
37. Ambrus, A., Chen, D., Dai, J. X., Bialis, T., Jones, R. A., and Yang, D. Z. (2006) Human telomeric sequence forms a hybrid-type intramolecular G-quadruplex structure with mixed parallel/antiparallel strands in potassium solution, *Nucleic Acids Res. 34*, 2723−2735.
38. Luu, K. N., Phan, A. T., Kuryavyi, V., Lacroix, L., and Patel, D. J. (2006) Structure of the human telomere in K^+ solution: An intramolecular (3+1) G-quadruplex, *J. Am. Chem. Soc. 128*, 9963−9970.
39. Xu, Y., Noguchi, Y., and Sugiyama, H. (2006) The new models of the human telomere d[AGGG(TTAGGG)$_3$] in solution, *Bioorg. Med. Chem. 14*, 5584−5591.
40. Risitano, A., and Fox, K. R. (2003) Stability of intramolecular DNA quadruplexes: Comparison with DNA duplexes, *Biochemistry 42*, 6507−6513.
41. Hazel, P., Huppert, J., Balasubramanian, S., and Neidle, S. (2004) Loop-length-dependent folding of G-quadruplexes, *J. Am. Chem. Soc. 126*, 16405−16415.
42. Risitano, A., and Fox, K. R. (2004) Influence of loop size on the stability of intramolecular DNA quadruplexes, *Nucleic Acids Res. 32*, 2598−2606.
43. Smirnov, I., and Shafer, R. H. (2000) Effect of loop sequence and size on DNA aptamer stability, *Biochemistry 39*, 1462−1468.
44. Lu, M., Guo, Q., and Kallenbach, N. R. (1992) Structure and stability of sodium and potassium complexes of dT_4G_4 and dT_4G_4T, *Biochemistry 31*, 2455−2459.
45. Guo, Q., Lu, M., and Kallenbach, N. R. (1993) Effect of thymine tract length on the structure and stability of model telomeric sequences, *Biochemistry 32*, 3596−3603.
46. Merkina, E. E., and Fox, K. R. (2005) Kinetic stability of intermolecular DNA quadruplexes, *Biophys. J. 89*, 365−373.
47. Phan, A. T., Modi, Y. S., and Patel, D. J. (2004) Two-repeat *Tetrahymena* telomeric d(TGGGGTTGGGGT) sequence interconverts between asymmetric dimeric G-quadruplexes in solution, *J. Mol. Biol. 338*, 93−102.
48. Wang, Y., and Patel, D. J. (1994) Solution structure of the *Tetrahymena* telomeric repeat $d(T_2G_4)_4$ G-tetraplex, *Structure 2*, 1141−1156.
49. Darby, R. A. J., Sollogoub, M., McKeen, C., Brown, L., Risitano, A., Brown, N., Barton, C., Brown, T., and Fox, K. R. (2002) High throughput measurement of duplex, triplex and quadruplex melting curves using molecular beacons and a LightCycler, *Nucleic Acids Res. 30*, e39.
50. Balagurumoorthy, P., Brahmachari, S. K., Mohanty, D., Bansal, M., and Sasisekharan, V. (1992) Hairpin and parallel quartet structures for telomeric sequences, *Nucleic Acids Res. 20*, 4061−4067.
51. Balagurumoorthy, P., and Brahmachari, S. K. (1994) Structure and stability of human telomeric sequence, *J. Biol. Chem. 269*, 21858−21869.
52. Lu, M., Guo, Q., and Kallenbach, N. R. (1993) Thermodynamics of G-tetraplex formation by telomeric DNAs, *Biochemistry 32*, 598−601.
53. Mergny, J. L., and Lacroix, L. (2003) Analysis of thermal melting curves, *Oligonucleotides 13*, 515−537.
54. Jing, N., Rando, R. F., Pommier, Y., and Hogan, M. E. (1997) Ion selective folding of loop domains in a potent anti-HIV oligonucleotide, *Biochemistry 36*, 12498−12505.
55. Cantor, C. R., and Schimmel, P. R. (1980) *Biophysical Chemistry*, W. H. Freedman and Co., New York.
56. Mergny, J.-L., and Maurizot, J.-C. (2001) Fluorescence resonance energy transfer as a probe for G-quartet formation by a telomeric repeat, *ChemBioChem 2*, 124−132.
57. Jing, N., and Hogan, M. E. (1998) Structure-activity of tetrad-forming oligonucleotides as a potential anti-HIV therapeutic drug, *J. Biol. Chem. 273*, 34992−34999.
58. Jing, N., Marchand, C., Liu, J., Mitra, R., Hogan, M. E., and Pommier, Y. (2000) Mechanism of inhibition of HIV-1 integrase by G-tetrad forming oligonucleotides in vitro, *J. Biol. Chem. 275*, 21460−21467.
59. Phan, A. T., Kuryavyi, V., Ma, J. B., Faure, A., Andreola, M. L., and Patel, D. J. (2005) An interlocked dimeric parallel-stranded DNA quadruplex: A potent inhibitor of HIV-1 integrase, *Proc. Natl. Acad. Sci. U.S.A. 102*, 634−639.
60. Esposito, V., Randazzo, A., Piccialli, G., Petraccone, L., Giancola, C., and Mayol, L. (2004) Effects of an 8-bromodeoxyguanosine incorporation on the parallel quadruplex structure $[d(TGGGT)]_4$, *Org. Biomol. Chem. 2*, 313−318.
61. Đapić, V., Abdomerović, V., Marrington, R., Peberdy, J., Rodger, A., Trent, J. O., and Bates, P. J. (2003) Biophysical and biological properties of quadruplex oligodeoxyribonucleotides, *Nucleic Acids Res. 31*, 2097−2107.

BI062118J