*5610–5622* *Nucleic Acids Research, 2008, Vol. 36, No. 17* *Published online 30 August 2008*
*doi:10.1093/nar/gkn543*

# A thermodynamic overview of naturally occurring intramolecular DNA quadruplexes

**Niti Kumar and Souvik Maiti***

Proteomics and Structural Biology Unit, Institute of Genomics and Integrative Biology, CSIR, Mall Road, Delhi 110 007, India

Received February 29, 2008; Revised August 7, 2008; Accepted August 8, 2008

## ABSTRACT

**Loop length and its composition are important for the structural and functional versatility of quadruplexes. To date studies on the loops have mainly concerned model sequences compared with naturally occurring quadruplex sequences which have diverse loop lengths and compositions. Herein, we have characterized 36 quadruplex-forming sequences from the promoter regions of various proto-oncogenes using CD, UV and native gel electrophoresis. We examined folding topologies and determined the thermodynamic profile for quadruplexes varying in total loop length (5–18 bases) and composition. We found that naturally occurring quadruplexes have variable thermodynamic stabilities ($\Delta G_{37}$) ranging from −1.7 to −15.6 kcal/mol. Overall, our results suggest that both loop length and its composition affect quadruplex structure and thermodynamics, thus making it difficult to draw generalized correlations between loop length and thermodynamic stability. Additionally, we compared the thermodynamic stability of quadruplexes and their respective duplexes to understand quadruplex–duplex competition. Our findings invoke a discussion on whether biological function is associated with quadruplexes with lower thermodynamic stability which undergo facile formation and disruption, or by quadruplexes with high thermodynamic stability.**

## INTRODUCTION

In nature, guanine-rich sequences are found in important regions such as telomeres, centromeres, immunoglobulin switch regions, mutational hot spots and promoter elements in human genome, and have the potential to form four-stranded structures which are involved in regulatory roles (1–6). These remarkably diverse four stranded structures contain planar arrays of four guanines, paired by intra- or intermolecular Hoogsteen bonds. The lengths of the G-runs and the 'loops' separating them both contribute to overall topology (7–10). Recently a genome wide search has identified 376 000 potential quadruplexes in the functionally important regions of genes (11,12). These findings suggest that the potential for quadruplex formation could contribute to the stability (or instability) of specific classes of genes or reflect mechanisms for global regulation of gene expression (13). These unusual structures are also involved in molecular recognition and may play intricate regulatory roles at the RNA level (14–16). Conserved elements with the potential to form polymorphic G-quadruplex structures in the first intron of human genes suggest that these elements may act as structural targets for regulation of transcription, especially RNA processing and translation (17). The possible existence and roles of G-quadruplexes *in vivo* has been corroborated by the detection of several proteins such as helicases and nucleases that bind specifically to G-DNA with very high affinity (18–24).

The growing awareness of the biological importance of quadruplexes has kindled interest in developing synthetic compounds that can bind selectively to these structures to exploit quadruplexes as a therapeutic target (25,26). Interestingly, the molecular recognition and function of these quadruplexes is influenced by the loop length and its composition. Recent computational approaches have found sequences that can fold into G-quadruplex structures to be widely dispersed in the promoters throughout human genome (27). The prevalence of putative quadruplex-forming sequences depends on the lengths of the guanine tracts and the distribution of the loops. The total loop length and its composition may determine the functionality of the quadruplexes and may thereby modulate their roles. These observations have fueled interest in understanding the role of loop length and composition on the thermodynamic stability, topology and molecular recognition of quadruplexes (28–34). The influence of loop length on the structure adopted by model sequences which form intramolecular quadruplexes has been investigated using a combination of experimental and molecular modeling techniques (28). Both these approaches suggest the preference of antiparallel over

*To whom correspondence should be addressed. Tel: +91 11 2766 6156; Fax: +91 11 2766 7471; Email: souvik@igib.res.in

parallel conformations for longer loops and vice versa for shorter loops. The available literature suggests a strong influence of loop length on quadruplex stability and quadruplexes with shorter loops display higher thermodynamic stability. Additionally loop composition also affects the quadruplex thermodynamic stability; a single T-to-A-loop substitution in an intramolecular quadruplex with G3 tracts separated by single base loops has been shown to lower the stability by 8°C (33). Although the thermodynamic fate of quadruplexes with different loop length and composition has been addressed using model sequences, the data set for this study is limited. Moreover, it is difficult to anticipate whether the results obtained for quadruplexes formed by model sequences can be extrapolated to naturally occurring quadruplex-forming sequences which possess heterogeneous loop length and composition. Hence, thermodynamic characterization of a considerable number of naturally occurring quadruplex sequences is desired. To address these important issues, we have investigated quadruplex-forming sequences in the promoter regions of various proto-oncogenes, examined their folding topologies and determined the thermodynamic profile for structures which have variable loop length and composition. Our data show that it is difficult to find a simple correlation between loop length and the thermodynamic stability of quadruplexes, in contrast to the correlation documented for model sequences. Additionally, we have compared the relative stability of duplex and quadruplex structures to comprehend the predominance of either of these structures at equilibrium (35–41).

## MATERIALS AND METHODS

### Sequence selection

Sequences 2 kb upstream from the Ensembl genes were downloaded via UCSC Genome Browser hg18 using Table browser option. We employed a computational search algorithm Quadfinder (42) to identify potential quadruplex-forming sequences. The Ensembl IDs obtained were subsequently matched with the Ensembl IDs of genes which have been correlated to genomic instability in human malignancies (13). The sequences with three G-quartets and total loop length of 5–18 bases were selected for biophysical investigation.

HPLC purified oligonucleotides were procured from SBS Genetech, China. The concentrations of these oligonucleotides were calculated by extrapolation of tabulated values of the monomer bases and dimers at 25°C using procedures reported earlier (43,44).

CD spectra were recorded in Jasco spectropolarimeter (model 715, Japan) equipped with a thermoelectrically controlled cell holder and a cuvette with a path length of 1 cm. The sequences were heated at 95°C for 5 min followed by programmed cooling (0.15°C/min) in 10 mM sodium cacodylate buffer, pH 7.4, 100 mM KCl. CD spectra for quadruplexes (10 μM) was recorded between 220 and 325 nm at 25°C and the spectra obtained was the average of three scans.

### UV melting experiments

The UV experiments were performed using Cary 100 (Varian) UV-vis spectrophotometer. Both heating and cooling profiles of quadruplex (2 μM) were monitored at 295 nm (45) with a rate of 0.15°C/min in 10 mM sodium cacodylate buffer, pH 7.4, 100 mM KCl. The cooling and heating profile of 80 quadruplex-forming sequences were obtained. The sequences which did not give characteristic quadruplex melting or reversible cooling and heating profile were not considered for further analysis. We performed concentration dependent UV experiments (2–100 μM) and filtered out the sequences which did not display concentration independent melting temperature to differentiate between intra- and intermolecular system (64). Therefore out of 80, we selected 62 sequences which gave characteristic quadruplex melting profile with reversible cooling and heating profile. The excluded 18 sequences are provided in SI Table 2. Out of 62 sequences, we selected 40 sequences which displayed concentration independent $T_m$ for thermodynamic analysis (Tables 2–4). The absorbance profile recorded at 295 nm was analyzed by nonlinear least square curve fitting method. This method involved the contribution from pre- and post-transition baselines and thermodynamic data was obtained using equations described previously (46,47). The analysis was done using Mathematica 5.1 and origin 7.0.

$$A_u = b_u + (m_u^* T) \quad (1)$$

$$A_l = b_l + (m_l^* T) \quad (2)$$

$$K_{eq} = \frac{(1-\alpha)}{\alpha} \quad (3)$$

$$A(T) = \alpha^*(A_u - A_l) + A_l \quad (4)$$

$$K_{eq} = \exp\left(\frac{\Delta G^o}{RT}\right) = \exp\left(\frac{\Delta H^o}{RT} + \frac{\Delta S^o}{R}\right) \quad (5)$$

$A_u$, $A_l$ are linear equations describing the upper and lower baselines, respectively, where $b_u$ and $b_l$ are fitted parameters for the intercepts for the upper and lower baseline with $m_u$ and $m_l$ as respective slopes. $K_{eq}$, is the equilibrium constant for the unstructured–structured transition for an intramolecular system and $\alpha$ is the folded fraction. $A(T)$ is the dependent variable and is the experimentally determined absorbance at each temperature ($T$). Using these equations, van't Hoff enthalpy ($\Delta H_{vH}$) and entropy ($\Delta S_{vH}$) were calculated and $T_m$ was calculated from the peak value of the first derivative of the fitted curve. The absorbance profiles of these sequences were also subjected to Monte Carlo methods for the determination of the error of fitted parameters $\Delta H_{vH}$ and $\Delta S_{vH}$.

The analysis was done using the built in method in Prism 5.0 as described previously (46).

Free energy ($\Delta G_{vH}$) was calculated at 37°C using equation $\Delta G_{vH} = \Delta H_{vH} - T\Delta S_{vH}$ assuming $\Delta H_{vH}$ and $\Delta S_{vH}$ are temperature independent. To obtain the thermodynamic parameters for the duplexes formed by these sequences, we used online tool HYTHER (ozone2.chem.wayne.edu) which employs Nearest Neighbor Method (48,49) for prediction of nucleic acid hybridization thermodynamics using experimental conditions of 2 μM strand concentration and 100 mM monovalent cations.

### Differential scanning calorimetric (DSC) experiments

DSC experiments were performed to measure the heat required for unstructured–structured transition at strand concentration of 50 μM. The experiments were performed in VP-DSC differential scanning calorimeter from Microcal (Northampton, MA). The experiments were performed in 10 mM sodium cacodylate buffer, pH 7.4, 100 mM KCl with a scanning rate of 0.15°C/min. Repeated buffer versus buffer scans were carried out in the temperature range of 110–25°C to obtain an appropriate and reproducible baseline which was subtracted from the sample scan. The subtracted sample scan was normalized for the oligoucleotide concentration. Origin 7.0 analysis package was used to integrate the resulting curve to obtain the molar folding enthalpy ($\Delta H_{cal}$) and $T_m$ was determined from the midpoint of the transition.

### Nondenaturing gel electrophoresis

Out of 62 quadruplex-forming sequences, 40 sequences adopted predominantly one conformation as indicated by CD experiment. These 40 sequences were annealed at 0.15°C/min rate in 10 mM sodium cacodylate buffer, 100 mM KCl, pH 7.4. The samples were electrophoresed on 20% polyacrylamide nondenaturing gel and run in 1 × TBE, pH 7.4, 100 mM KCl buffer. Gel was run at 4°C at constant voltage of 100 V. Oligonucleotide concentration used was 15 μM. After electrophoresis, gel was stained with ethidium bromide and was visualized through BioRad Gel Doc XR.

## RESULTS AND DISCUSSION

This study aims to elucidate the thermodynamic profiles of G-quadruplexes formed by naturally occurring sequences present in the promoter regions of various proto-oncogenes. These sequences were sorted using Quadfinder (42), which employs a consensus unimolecular G-quadruplex sequence motif of the form $G_xL_{y1}G_xL_{y2}G_xL_{y3}G_x$, where x denotes the G-stretch and $L_{y1}$, $L_{y2}$ and $L_{y3}$ denote the length of loops 1, 2 and 3, respectively. The selected sequences belong to different regions of proto-oncogenes, which have been correlated to genomic instability in human malignancies by Eddy and co-workers (13). The sequences investigated in our study have three G-quartets, as they are the shortest G-tracts that form quadruplexes with reasonable stability, for which the influence of total loop length (5–18 bases) on the quadruplex stability can be evaluated. Our dataset includes 62 quadruplex-forming sequences present in the promoter regions of various proto-oncogenes located at different positions with respect to transcription start site (TSS) (SI Table 1). Based on the loop length, we classified the sequences with total loop length ranging from 5 to 18 bases (tabulated in Table 1). We initially examined the folding topologies of these sequences through CD spectroscopy. Though CD does not provide direct evidence for the structural features of quadruplexes, the spectra obtained provide valuable information about the structural characteristics of quadruplexes and agrees with the topology obtained by direct methods such as NMR (50,51). The characteristic CD signals arise from G-G stacking between G-quartets, the strength of which mainly depends on the conformation of guanine bases around the glycosidic bond (*syn* or *anti*). The difference in the orientation of glycosidic torsion conformation gives rise to parallel and antiparallel structures (8–10). The CD signature comprising of a positive peak at 262 nm and a negative peak at 240 nm typically indicates a parallel conformation and signature with a positive peak at 295 nm and a negative peak at 238 nm indicates an antiparallel conformation. However, the presence of both these signatures indicates mixed conformation. The CD signature obtained for all the sequences has been tabulated in SI Table 1 and the spectra are provided in Supplementary Data. The information provided in SI Table 1 shows that out of the 62 sequences investigated in this study, two sequences adopted predominantly antiparallel conformation, 38 sequences adopted predominantly parallel conformation and the remaining 22 sequences adopted both parallel and antiparallel signatures. The characteristic CD spectra of a few sequences are shown in Figure 1. The CD spectra provided in Figure 1 show that c-*MYC* (Q21), c-*KIT* (Q30), *WNT* 3 (Q2), *AKT* 2 (Q34), *MYB* (Q35) have a predominantly parallel population and a small antiparallel population, whereas *VEGF* (Q1), *PDFB* (Q4) and *PIM* 1 (Q9), adopt only the parallel topology. Our results for the CD spectra of c-*MYC* (Q21), c-*KIT* (Q30), *VEGF* (Q1) and *PDFB* (Q4) are in agreement with the topology reported in literature (6,7,50–52). An interesting observation was made from the CD spectra of the quadruplex-forming sequences from the *HCK* protooncogene, Q24 and Q25 (Supplementary Data, SI Table 1), which possess the same total loop length of 10 bases and loop composition, but display variation in base distribution among the three loops. In Q24, the number of bases in loops L1, L2 and L3 is 2, 2 and 6, respectively, while in Q25, these are 2, 6 and 2, respectively. This minor difference in loop distribution results in a remarkable difference in the CD spectra. Q24 adopts a predominant parallel conformation, whereas Q25 adopts a mixed conformation (Supplementary Data). Detailed inspection of the CD spectra of all these sequences shows that sequences with shorter loop lengths such as, *VEGF* (Q1), *WNT 3* (Q2), *WNT5A* (Q3), *PDGFB* (Q4) adopt mainly parallel structures, while longer loops such as, *RALB* (Q59), *EGFR* (Q60), *VAV 1* (Q61), *FGF 6* (Q62) adopt mixed structures comprising both parallel and antiparallel characteristics. A few quadruplexes with longer loops such as, *FLI* (Q48), *THRA* (Q52), *KRAS* (Q57) and *THPO* (Q58) also

**Table 1.** Sequences used in the study with loop length varying from 5 to 18 bases

| Oligo name | Gene | Sequence (5′–3′) | | | | | | | | L 1 | L 2 | L 3 | Total |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Q1 | *VEGF* | GGG | A | GGG | TTG | GGG | T | GGG | | 1 | 3 | 1 | 5 |
| Q2 | *WNT 3* | GGG | CC | GGG | C | GGG | AG | GGG | | 2 | 1 | 2 | 5 |
| Q3 | *WNT5A* | GGG | A | GGG | CA | GGG | AA | GGG | | 1 | 2 | 2 | 5 |
| Q4 | *PDGFB* | GGG | T | GGG | T | GGG | CTCT | GGG | | 1 | 1 | 4 | 6 |
| Q5 | *JUNB* | GGG | TG | GGG | CG | GGG | CC | GGG | | 2 | 2 | 2 | 6 |
| Q6 | *TAL1* | GGG | G | GGG | CGGT | GGG | G | GGG | | 1 | 4 | 1 | 6 |
| Q7 | *THPO* | GGG | TGG | GGG | AG | GGG | A | GGG | | 3 | 2 | 1 | 6 |
| Q8 | *FOS* | GGG | GAGCT | GGG | A | GGG | A | GGG | | 5 | 1 | 1 | 7 |
| Q9 | *PIM1* | GGG | CG | GGG | CGG | GGG | CG | GGG | | 2 | 2 | 3 | 7 |
| Q10 | *FGF 3* | GGG | GGA | GGG | GC | GGG | TA | GGG | | 3 | 2 | 2 | 7 |
| Q11 | *FES* | GGG | A | GGG | CT | GGG | ACCA | GGG | | 1 | 2 | 4 | 7 |
| Q12 | *VAV1* | GGG | CA | GGG | A | GGG | AACT | GGG | | 2 | 1 | 4 | 7 |
| Q13 | *IGF2* | GGG | T | GGG | CC | GGG | AGCTT | GGG | | 1 | 2 | 5 | 8 |
| Q14 | *FOSL1* | GGG | CTG | GGG | CG | GGG | CGC | GGG | | 3 | 2 | 3 | 8 |
| Q15 | *TAL1* | GGG | GCC | GGG | C | GGG | CGCG | GGG | | 3 | 1 | 4 | 8 |
| Q16 | *AKT2* | GGG | GGAG | GGG | GC | GGG | GG | GGG | | 4 | 2 | 2 | 8 |
| Q17 | *THRA* | GGG | CG | GGG | GGC | GGG | GGC | GGG | | 2 | 3 | 3 | 8 |
| Q18 | *FBXW 7* | GGG | CGGC | GGG | GCC | GGG | G | GGG | | 4 | 3 | 1 | 8 |
| Q19 | *HRAS* | GGG | GGGACT | GGG | C | GGG | CA | GGG | | 6 | 1 | 2 | 9 |
| Q20 | *WNT1* | GGG | AAAGA | GGG | CAT | GGG | A | GGG | | 5 | 3 | 1 | 9 |
| Q21 | *c-MYC* | GGG | GA | GGG | TGGGGA | GGG | TG | GGG | | 2 | 6 | 2 | 10 |
| Q22 | *NFKB 2* | GGG | CGC | GGG | GCGC | GGG | CGC | GGG | | 3 | 4 | 3 | 10 |
| Q23 | *GLI 1* | GGG | CCGGC | GGG | A | GGG | CTGG | GGG | | 5 | 1 | 4 | 10 |
| Q24 | *HCK* | GGG | GA | GGG | CC | GGG | CCAGAA | GGG | | 2 | 2 | 6 | 10 |
| Q25 | *HCK* | GGG | CC | GGG | CCAGAA | GGG | AG | GGG | | 2 | 6 | 2 | 10 |
| Q26 | *YES 1* | GGG | T | GGG | AA | GGG | ACGCCGC | GGG | | 1 | 2 | 7 | 11 |
| Q27 | *AKT 1* | GGG | C | GGG | CGGCTCC | GGG | CGC | GGG | | 1 | 7 | 3 | 11 |
| Q28 | *BCL 2* | GGG | CCA | GGG | AGC | GGG | GCGGA | GGG | | 3 | 3 | 5 | 11 |
| Q29 | *NOTCH 1* | GGG | A | GGG | AGCGCAA | GGG | CGC | GGG | | 1 | 7 | 3 | 11 |
| Q30 | *c-KIT* | GGG | TGG | GGG | CCCCGA | GGG | CT | GGG | | 3 | 6 | 2 | 11 |
| Q31 | *JUNB* | GGG | CT | GGG | TT | GGG | CCAGGCC | GGG | | 2 | 2 | 7 | 11 |
| Q32 | *SRC* | GGG | TGAGCT | GGG | CT | GGG | GCT | GGG | | 6 | 2 | 3 | 11 |
| Q33 | *VAV 2* | GGG | CGTG | GGG | ACTGC | GGG | TG | GGG | | 4 | 5 | 2 | 11 |
| Q34 | *AKT 2* | GGG | GGCTG | GGG | GGA | GGG | GGC | GGG | | 5 | 3 | 3 | 11 |
| Q35 | *MYB* | GGG | CGGGGT | GGG | TG | GGG | CCC | GGG | | 6 | 2 | 3 | 11 |
| Q36 | *ROS 1* | GGG | GTGG | GGG | AT | GGG | AGATG | GGG | | 4 | 2 | 5 | 11 |
| Q37 | *TAL 1* | GGG | CGG | GGG | AGC | GGG | GGCCGC | GGG | | 3 | 5 | 4 | 12 |
| Q38 | *WNT1* | GGG | GGCCACC | GGG | CA | GGG | GGC | GGG | | 7 | 2 | 3 | 12 |
| Q39 | *CTTN* | GGG | AGGAGGG | GGG | CA GG | GGG | C | GGG | | 7 | 4 | 1 | 12 |
| Q40 | *AXL* | GGG | CT | GGG | CTGG | GGG | TGGAGGC | GGG | | 2 | 4 | 7 | 12 |
| Q41 | *FGF 3* | GGG | GATTGAT | GGG | GAGC | GGG | AG | GGG | | 7 | 4 | 2 | 12 |
| Q42 | *MAF* | GGG | TGGA | GGG | CAGG CA | GGG | GAG | GGG | | 4 | 6 | 3 | 13 |
| Q43 | *MPL* | GGG | CTGGCCT | GGG | AG | GGG | ATTG | GGG | | 7 | 2 | 4 | 13 |
| Q44 | *MYB* | GGG | AGGCGGC | GGG | ACT | GGG | CGC | GGG | | 7 | 3 | 3 | 13 |
| Q45 | *ABL 1* | GGG | CC | GGG | GGAGG | GGG | TGTCTC | GGG | | 2 | 5 | 6 | 13 |
| Q46 | *MYEOV* | GGG | TAAA | GGG | AGGAGGA | GGG | AA | GGG | | 4 | 7 | 2 | 13 |
| Q47 | *VEGF* | GGG | AG | GGG | ACTGG | GGG | AAGGATA | GGG | | 2 | 5 | 7 | 14 |
| Q48 | *FLI 1* | GGG | AGGAAGA | GGG | GGTGT | GGG | GG | GGG | | 7 | 5 | 2 | 14 |
| Q49 | *FGR* | GGG | CTTTT | GGG | AAGG CC | GGG | CTG | GGG | | 5 | 6 | 3 | 14 |
| Q50 | *WNT 3* | GGG | CT C | GGG | AGGG | GGG | CGGCTGC | GGG | | 3 | 4 | 7 | 14 |
| Q51 | *CSF 1R* | GGG | CA GGA | GGG | TGT | GGG | GCTGCA | GGG | | 5 | 3 | 6 | 14 |
| Q52 | *THRA* | GGG | CGCGGCA | GGG | GGC | GGG | GGGC | GGG | | 7 | 3 | 4 | 14 |
| Q53 | *FES* | GGG | CCCTG | GGG | CCTGCG | GGG | CGC | GGG | | 5 | 6 | 3 | 14 |
| Q54 | *REL* | GGG | ACAGGCA | GGG | GCG | GGG | CGCT | GGG | | 7 | 3 | 4 | 14 |
| Q55 | *FGF 4* | GGG | AGAC | GGG | GAC | GGG | GACAGGCA | GGG | | 4 | 3 | 8 | 15 |
| Q56 | *BCL 3* | GGG | TCCCCA | GGG | GT | GGG | GTTGGAT | GGG | | 6 | 2 | 7 | 15 |
| Q57 | *KRAS* | GGG | CGGTGT | GGG | AAGA | GGG | AAGAG | GGG | | 6 | 4 | 5 | 15 |
| Q58 | *THPO* | GGG | TGG | GGG | AGGGGA | GGG | AGGGGCT | GGG | | 3 | 6 | 7 | 16 |
| Q59 | *RALB* | GGG | TCCTA | GGG | GATGG | GGG | GCAGGT | GGG | | 5 | 5 | 6 | 16 |
| Q60 | *EGFR* | GGG | GACC | GGG | TCCAGA | GGG | GCAGTGCT | GGG | | 4 | 6 | 8 | 18 |
| Q61 | *VAV 1* | GGG | TTTTGT | GGG | CCTT | GGG | GAGGACTT | GGG | | 6 | 4 | 8 | 18 |
| Q62 | *FGF 6* | GGG | CGATGAT | GGG | TGGAG | GGG | AGGGCC | GGG | | 7 | 5 | 6 | 18 |

adopted a parallel conformation. For the given data set, we observed that only two sequences, *FGR* (Q49) and *FES* (Q53), adopted an antiparallel fold. It is worthy to mention that for few sequences the assignment of bases is in the loops is arbitrary, and it is also likely that sequences that have been assigned with longer loops could also fold to adopt topologies with short loop lengths. Examination of the CD spectra of all the naturally occurring sequences

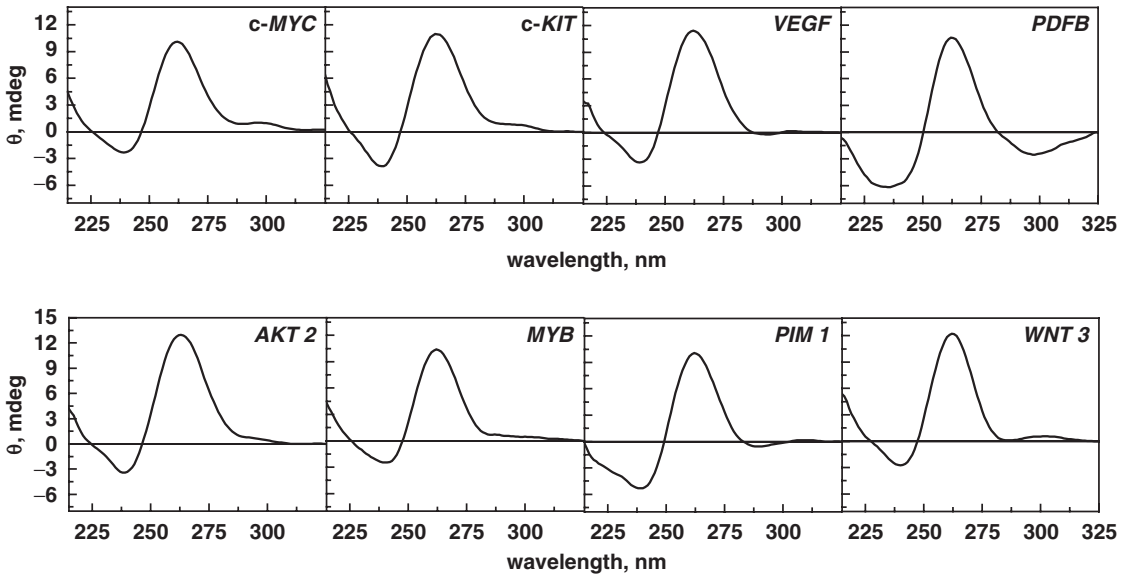


**Figure 1.** CD spectra of quadruplexes (10 μM) obtained in 10 mM sodium cacodylate buffer, 100 mM KCl, pH 7.4.

studied here shows a predominance of the parallel topology, hinting toward a bias for the parallel fold over the antiparallel form as a recognition motif in biology.

Out of the 62 sequences, 40 sequences displayed predominantly one type of conformation, (either parallel or antiparallel) and the remaining 22 sequences displayed mixed conformations. We performed native gel electrophoresis on these 40 sequences to confirm the existence of one type of conformation as observed in CD study. Our gel study showed that these sequences adopt a single conformation with the absence of higher order structures (Supplementary Data). We observed that most of the quadruplexes move faster than the oligo $dT_{25}$ marker. However, a few quadruplex-forming sequences, namely Q6, Q7, Q47 and Q48, have mobility almost similar to the oligo $dT_{25}$ marker. Note that the migration of the oligo $dT_n$ marker does not necessarily correspond to the single strand (53), this marker was chosen to provide an internal migration standard. The anomalous behavior observed for Q6 and Q7 quadruplexes (18 mer) and Q47 and Q48 (26 mer), raises a doubt concerning the molecularity of the structure. However, concentration dependent UV melting study for these sequences shows that these sequences adopt a unimolecular structure. A possible explanation for this anomaly is that quadruplexes with different loop lengths adopt varying topologies which give rise to different electrophoretic mobilities. Since, the loop residues influence the quadruplex structure (compact or an extended), a less compact structure may display a lower mobility.

Further investigations were carried out only with the 40 sequences which displayed predominantly one type of conformation to determine their thermodynamic profile by van't Hoff analysis. We also performed the UV melting of the remaining 22 sequences which showed a mixed population. The data for these sequences are presented in Supplementary Data, and can be used as a resource for future investigations; however, these 22 sequences have been excluded from the present analysis. UV melting of these quadruplexes showed characteristic hypochromic absorbance profile at 295 nm, which showed reversible heating and cooling curves, thus indicating that the system is in thermodynamic equilibrium. We performed concentration-dependent UV melting on all 40 sequences and observed that the melting temperatures were concentration independent (data not shown), which indicates that they form intramolecular structures. Representative UV melting profiles of a few of these quadruplexes is presented in Figure 2; the melting profile of rest of the sequences has been provided in Supplementary Data. In Figure 2, the upper panel displays melting profiles of quadruplexes from the promoters of c-*MYC* (Q21), c-*KIT* (Q30), *VEGF* (Q1) and *PDGFB* (Q4) which have melting temperature, $T_m$ of 75.0, 71.6, 80.8 and 83.0°C, respectively. We identified new quadruplex-forming sequences which demonstrated remarkable stability and are presented in the lower panel of Figure 2. These sequences belong to promoter regions of *AKT* 2 (Q34), *MYB* (Q35), *PIM* 1 (Q9), *WNT* 3 (Q2) which exhibit $T_m$ of 87.0, 88.0, 82.8 and 86.0°C, respectively. We found several examples where sequences with the same loop length display different $T_m$ values. For example, *VEGF* (Q1), *WNT 3* (Q2) and *WNT 5A* (Q3) have total loop lengths of 5 bases and $T_m$s of 80.8, 86.0 and 65.0°C, respectively. It can also be seen that structures with different loop lengths display similar melting temperatures. For example, *JUN B* (Q30), *MAF* (Q42), *FES* (Q53) and *FGF 6* (Q62) have total loop length of 11, 13, 14 and 18 bases, respectively, but all exhibit a $T_m$ of 70.0°C. The melting profiles obtained for sequences Q6, Q7, Q16 and Q17 showed very broad transitions suggestive of multiple quadruplex conformations. However, the UV melting data for these sequences were concentration independent, thereby indicating intramolecular quadruplex formation.

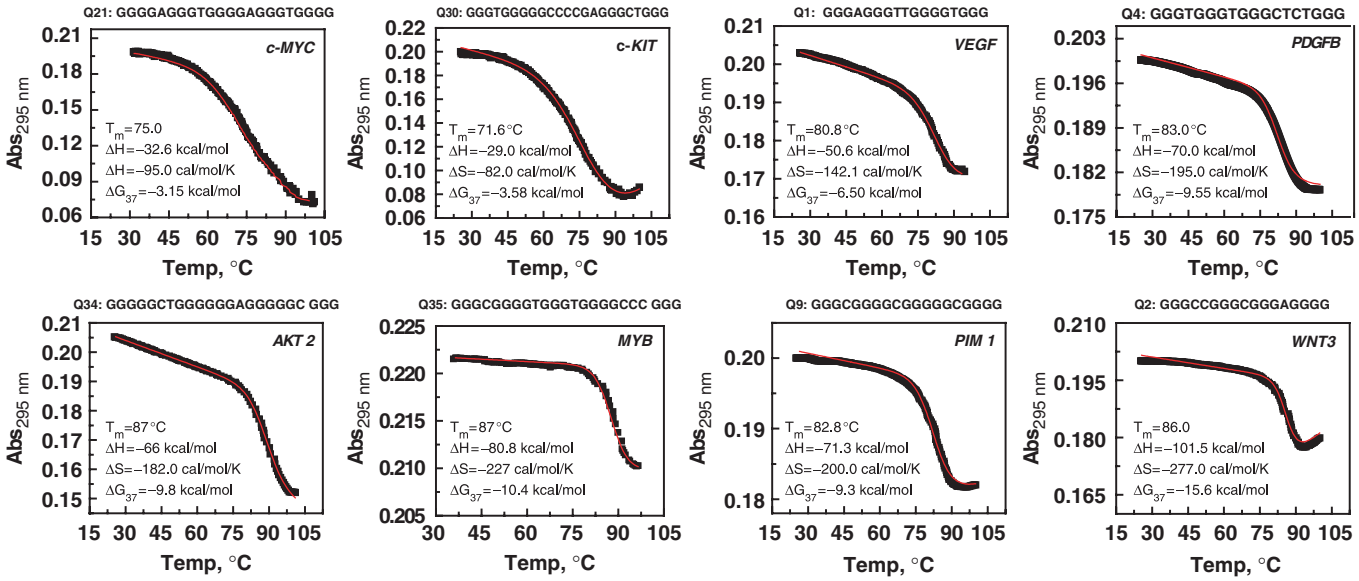


**Figure 2.** UV cooling profile of quadruplexes monitored at 295 nm along with nonlinear least square curve fitting (in red). The experiment was performed in 10 mM sodium cacodylate buffer, pH 7.4, 100 mM KCl. The cooling rate was 0.15°C/min. The thermodynamic parameters obtained by van't Hoff analysis are provided in the inset.

Moreover, the CD and gel studies also showed that these sequences each adopt a single conformation. The multiple transitions in the UV melting profiles of these structures, suggest the presence of conformations which have similar CD spectra and similar mobility but different melting temperatures. These sequences were not considered for further analysis.

The sequences investigated in this study contain additional Gs at the 3′-or 5′-ends, which create ambiguity concerning which Gs are in the loops and which are in the G-stack. This issue is usually avoided in model sequences, where the loops are homogenous in composition and length. Since the promoter regions of proto-oncogenes are enriched in guanines, therefore naturally occurring quadruplex-forming sequences often contain GGG sequences in the loops. For the final analysis using CD, gel and concentration-dependent UV melting studies, we therefore selected only 36 sequences, which predominantly showed one type of conformation and evaluated their thermodynamic parameters using van't Hoff analysis. The thermodynamic parameters obtained from nonlinear least square curve fittings are tabulated in Tables 2–4 The raw data were also subjected to a Monte Carlo method, which involves fitting 1000 simulated data sets by nonlinear least square curve equation, and these results are tabulated in SI Tables 3–5. The thermodynamic parameters obtained by Monte Carlo method are in agreement with the data obtained from the direct fit of the raw data using nonlinear least square equation. Our data show that quadruplexes with same loop length have different enthalpy $\Delta H$ and entropy $\Delta S$. For example, *VEGF* (Q1), *WNT 3* (Q2) and *WNT 5A* (Q3) have total loop length of 5 bases and the enthalpy associated with these structures is −50.6, −101.5 and −44.0 kcal/mol, respectively. The corresponding entropy associated with these structures is −142.1, −277.0 and −130.0 cal/mol/K, respectively. We also observed that quadruplexes with different loop lengths can have similar enthalpy and entropy values. For instance, the quadruplex structure formed by *WNT 3* (Q2) and *MYEOV* (Q46), with total loop lengths of 5 and 13 bases, respectively, possess remarkably high enthalpy (−101.5 and −109.0 kcal/mol, respectively) and unfavorable entropy values (−277.0 and −330.0 cal/mol/K, respectively). We also observed that sequences with different loop lengths and melting temperatures exhibit similar range of enthalpy and entropy values, as seen in case of *FOS* (Q8), *IGF* 2 (Q13), *FBXW* 7 (Q18), c-*MYC* (Q21) and c-*KIT* (Q30). Sequences with similar melting temperatures displayed different enthalpy and entropy values, as observed in case of *VEGF* (Q1) and *JUN B* (Q5). Due to these striking differences, it is difficult to draw generalized reasons for the variable thermal stability observed between these sequences.

Next we performed differential scanning calorimetry (DSC) to obtain model-independent values which were compared with the thermodynamic data obtained from UV study. The DSC experiments were performed for few sequences which were selected randomly to obtain the van't Hoff and calorimetric enthalpy associated with the unstructured–structured transition. The unstructured–structured transition may be associated with nonzero heat capacity changes. These heat capacity changes arise from the change in solvent exposure for nonpolar and polar groups. These heat capacity changes are determined by the difference between pre- and post-transition baseline of the DSC curves. The net effect of solvent exposure during unstructured–structured transition can sometimes be relatively small due to simultaneous exposure or burial of polar and non-olar groups. Precise nonzero heat capacity changes ($\Delta C$p) associated with quadruplex melting

**Table 2.** Thermodynamic profiles obtained for sequences with loop length of 5–9

| Oligo Name | Gene | Total | Quadruplex[a] | | | | Duplex[b] | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $T_m$ (°C) | $\Delta H$ (kcal/mol) | $\Delta S$ (cal/mol/K) | $\Delta G_{37}$ (kcal/mol) | $T_m$ (°C) | $\Delta H$ (kcal/mol) | $\Delta S$ (cal/mol/K) | $\Delta G_{37}$ (kcal/mol) | $\Delta\Delta G_{37}$ (kcal/mol) |
| Q1 | *VEGF* | 5 | 80.8 | −50.6 ± 0.5 | −142.1 ± 1.4 | −6.5 ± 0.1 | 66.0 | −129.5 | −353.8 | −19.8 | −13.3 |
| Q2 | *WNT 3* | 5 | 86.0 | −101.5 ± 2.5 | −277.0 ± 5.0 | −15.6 ± 0.1 | 75.6 | −136.6 | −364.2 | −23.7 | −8.1 |
| Q3 | *WNT5A* | 5 | 65.0 | −44.0 ± 0.8 | −130 ± 2.4 | −3.7 ± 0.1 | 66.2 | −129.8 | −355.1 | −19.7 | −16.0 |
| Q4 | *PDGFB* | 6 | 83.0 | −70.0 ± 1.2 | −195.0 ± 3.4 | −9.6 ± 0.2 | 69.1 | −139.7 | −380.7 | −21.7 | −12.1 |
| Q5 | *JUNB* | 6 | 77.7 | −26.0 ± 0.1 | −73.5 ± 0.3 | −3.3 ± 0.02 | 78.0 | −145.5 | −386.8 | −25.6 | −22.3 |
| Q6 | *TAL1*[c] | 6 | 80.0 | nd | | | 77.4 | −141.1 | −375.0 | −24.8 | nd |
| Q7 | *THPO*[c] | 6 | 82.0 | nd | | | 70.6 | −136.7 | −370.3 | −22.0 | nd |
| Q8 | *FOS* | 7 | 72.4 | −30.7 ± 0.3 | −87.7 ± 0.8 | −3.5 ± 0.1 | 70.1 | −146.0 | −397.6 | −22.6 | −19.1 |
| Q9 | *PIM1* | 7 | 82.8 | −71.3 ± 1.2 | −200.0 ± 3.3 | −9.3 ± 0.2 | 82.5 | −157.0 | −414.0 | −28.7 | −19.4 |
| Q10 | *FGF 3* | 7 | 62.0 | −23.3 ± 0.3 | −69.2 ± 0.8 | −1.8 ± 0.1 | 72.9 | −147.6 | −399.1 | −24.0 | −22.1 |
| Q11 | *FES* | 7 | 58.0 | −53.0 ± 0.5 | −160.0 ± 1.1 | −3.4 ± 0.2 | 70.7 | −146.8 | −399.5 | −23.0 | −19.5 |
| Q12 | *VAV1* | 7 | 62.8 | −17.7 ± 0.2 | −48.0 ± 0.5 | −2.8 ± 0.1 | 68.5 | −146.7 | −402.0 | −22.1 | −19.3 |
| Q13 | *IGF2* | 8 | 59.6 | −30.6 ± 0.3 | −91.0 ± 0.8 | −2.4 ± 0.1 | 74.0 | −159.1 | −431.0 | −25.5 | −23.1 |
| Q14 | *FOSL1* | 8 | 73.7 | −36.0 ± 0.2 | −101 ± 0.5 | −4.7 ± 0.1 | 81.9 | −167.1 | −443.2 | −29.7 | −25.0 |
| Q15 | *TAL1* | 8 | 81.5 | −43.8 ± 0.4 | −123.7 ± 1.0 | −5.4 ± 0.1 | 84.3 | −169.4 | −446.5 | −31.0 | −25.6 |
| Q16 | *AKT2*[c] | 8 | 88.0 | nd | | | 80.3 | −156.2 | −414.5 | −27.7 | nd |
| Q17 | *THRA*[c] | 8 | 81.6 | nd | | | 83.9 | −165.0 | −434.7 | −30.2 | nd |
| Q18 | *FBXW 7* | 8 | 75.8 | −31.0 ± 0.9 | −89.0 ± 2.3 | −3.4 ± 0.2 | 83.9 | −165.0 | −434.7 | −30.2 | −26.8 |
| Q19 | *HRAS* | 9 | 53.7 | −28.1 ± 0.3 | −85.2 ± 0.7 | −1.7 ± 0.1 | 77.6 | −167.2 | −449.2 | −28.0 | −26.2 |
| Q20 | *WNT1* | 9 | 50.0 | −46.0 ± 1.5 | −141.5 ± 4.7 | −2.1 ± 0.1 | 68.1 | −161.6 | −446.0 | −23.3 | −21.2 |

[a]Thermodynamic profiles for quadruplexes (2 μM) obtained from UV experiment performed in 10 mM sodium cacodylate buffer, 100 mM KCl, pH 7.4. $T_m$ values differed by ±1.5°C. Thermodynamic parameters were determined by non-linear least square curve fitting (details in Materials and method section).
[b]Predicted thermodynamic parameters for duplex formed by all sequences obtained through HYTHER (ozone2.chem.wayne.edu) using experimental conditions of 2 μM strand concentration and 100 mM monovalent cations.
[c]nd, not determined. The sequences shown are those which displayed predominantly one conformation in CD and gel studies, and concentration independent $T_m$.

can be obtained at very high oligonucleotide concentration (≈200 μM). But the quadruplex melting experiments performed at high concentrations lead to DNA precipitation at high temperatures. At low concentrations (50 μM) such aggregation problem can be avoided but determination of $\Delta C$p was not possible due to the low signal. Due to this limitation, we performed the analysis with the assumption of zero heat capacity changes. The assumption of negligible heat capacity change associated with unstructured–structured transition is very common in this field, as the process yields heat capacity changes within experimental errors (54,55). The shape of the curve provides the van't Hoff enthalpy and the area under the curve provides calorimetric enthalpy (Figure 3). We observed that ratio of van't Hoff and calorimetric enthalpy is nearly equal to 1, therefore, indicating that structured–unstructured transition follows a two-state process. The thermodynamic parameters obtained from DSC experiments correlated well with the parameters obtained from UV study, further indicating that the model used for the analysis in the UV study is correct (Table 4). Recently using isothermal calorimetry (ITC), heat capacity changes associated with quadruplex formation in a solution containing excess of salt were measured (56). ITC measurements differ from thermal denaturation such as DSC and UV melting. The results obtained from the former technique are valid at temperature selected for the experiment, which is usually lower than $T_m$, while, the results obtained from the latter technique are reliable at or near the $T_m$ of the unstructured–structured transition. In the ITC experiments, the quadruplex formation is measured under a high salt concentration where the oligonucleotide is subjected to changes in salt conditions; while the thermal melting experiments do not involve such changes in salt conditions. Therefore disagreement in the heat capacity changes obtained through ITC and DSC experiments is quite likely.

Examination of Tables 2–4 shows that the enthalpy and entropy associated with quadruplex formation for these 36 sequences varies from −17.7 to −109.0 kcal/mol and −48.0 to −330.0 cal/mol/K, respectively. This indicates that quadruplex melting is associated with favorable enthalpy and unfavorable entropy changes. During quadruplex formation the enthalpic stabilization is compensated at entropic cost due to ordering or stacking of bases in the loop region. This is to be noted that for sequences with G-tracts longer than three guanines can potentially adopt multiple structures, and the thermodynamic parameters obtained for these sequences may contain large errors. The presence of multiple conformations with similar stabilities will give shallow melting profiles with apparently low values for enthalpy. Due to the limited studies on thermodynamic stability of the naturally occurring quadruplex-forming sequences, we can only compare our data with literature values reported for model sequences. It has been reported that quadruplexes formed by model sequences with loop length varying from T2 to T5 have enthalpy values ranging from −44.0 to

**Table 3.** Thermodynamic profile obtained for sequences with loop length of 10–14

| Oligo Name | Gene | Total | Quadruplex[a] | | | | Duplex[b] | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $T_m$ (°C) | $\Delta H$ (kcal/mol) | $\Delta S$ (cal/mol/K) | $\Delta G_{37}$ (kcal/mol) | $T_m$ (°C) | $\Delta H$ (kcal/mol) | $\Delta S$ (cal/mol/K) | $\Delta G_{37}$ (kcal/mol) | $\Delta\Delta G_{37}$ (kcal/mol) |
| Q21 | *c-MYC* | 10 | 75.0 | −32.6 ± 0.4 | −95.0 ± 1.0 | −3.2 ± 0.1 | 75.8 | −169.6 | −458.6 | −27.4 | −24.3 |
| Q22 | *NFKB 2* | 10 | 75.0 | −42.0 ± 0.3 | −119.0 ± 0.7 | −5.1 ± 0.1 | 87.1 | −194.2 | −511.6 | −35.6 | −30.5 |
| Q23 | *GLI 1* | 10 | 73.5 | −44.6 ± 0.3 | −129.0 ± 0.7 | −4.6 ± 0.1 | 81.6 | −178.7 | −476.3 | −31.0 | −26.4 |
| Q24 | *HCK* | 10 | 52.0 | −39.0 ± 0.8 | −119.5 ± 2.5 | −2.0 ± 0.03 | 76.7 | −174.2 | −470.4 | −28.4 | −26.4 |
| Q25 | *HCK* | 10 | 62.4 | nd | | | 76.7 | −174.2 | −470.4 | −28.4 | nd |
| Q26 | *YES 1* | 11 | 66.0 | nd | | | 77.3 | −180.6 | −488.0 | −29.4 | nd |
| Q27 | *AKT 1* | 11 | 57.3 | nd | | | 85.5 | −197.8 | −524.0 | −35.4 | nd |
| Q28 | *BCL 2* | 11 | 66.6 | nd | | | 80.4 | −186.7 | −500.6 | −31.5 | nd |
| Q29 | *NOTCH 1* | 11 | 70.0 | nd | | | 79.1 | −191.0 | −514.8 | −31.4 | nd |
| Q30 | *c − KIT* | 11 | 71.6 | −29.0 ± 0.2 | −82.0 ± 0.4 | −3.6 ± 0.1 | 80.4 | −183.2 | −490.75 | −31.1 | −27.5 |
| Q31 | *JUNB* | 11 | 70.0 | nd | | | 79.3 | −185.2 | −498.0 | −30.8 | nd |
| Q32 | *SRC* | 11 | 57.2 | nd | | | 76.5 | −183.0 | −496.0 | −29.3 | nd |
| Q33 | *VAV 2* | 11 | 80.8 | nd | | | 78.7 | −187.3 | −504.8 | −30.8 | nd |
| Q34 | *AKT 2* | 11 | 87.0 | −66.0 ± 1.8 | −182.0 ± 5.1 | −9.6 ± 0.2 | 82.4 | −182.3 | −485.2 | −32.0 | −22.3 |
| Q35 | *MYB* | 11 | 87.0 | −80.8 ± 2.3 | −227.0 ± 6.2 | −10.0 ± 0.4 | 83.1 | −186.4 | −495.8 | −32.7 | −22.3 |
| Q36 | *ROS 1* | 11 | 77.5 | −70.0 ± 1.0 | −198.0 ± 2.6 | −8.6 ± 0.2 | 73.4 | −176.5 | −481.8 | −27.1 | −18.5 |
| Q37 | *TAL 1* | 12 | 81.2 | nd | | | 86.3 | −201.4 | −532.8 | −36.2 | nd |
| Q38 | *WNT1* | 12 | 75.0 | −36.8 ± 0.3 | −104.7 ± 0.7 | −4.3 ± 0.1 | 84.4 | −195.6 | −519.7 | −34.5 | −30.2 |
| Q39 | *CTTN* | 12 | 78.7 | nd | | | 82.8 | −202.7 | −542.0 | −34.7 | nd |
| Q40 | *AXL* | 12 | 74.0 | −45.4 ± 0.4 | −130.0 ± 1.0 | −5.2 ± 0.1 | 81.2 | −201.3 | −540.6 | −33.7 | −28.6 |
| Q41 | *FGF 3* | 12 | 65.8 | −38.0 ± 0.8 | −111.0 ± 2.2 | −3.6 ± 0.1 | 74.8 | −195.9 | −535.6 | −30.0 | −26.3 |
| Q42 | *MAF* | 13 | 70.0 | nd | | | 78.6 | −196.9 | −532.3 | −32.0 | nd |
| Q43 | *MPL* | 13 | 66.8 | nd | | | 77.2 | −195.8 | −531.5 | −31.0 | nd |
| Q44 | *MYB* | 13 | 67.0 | nd | | | 83.3 | −210.3 | −562.6 | −36.0 | nd |
| Q45 | *ABL 1* | 13 | 75.0 | −74.0 ± 5.0 | −210.5 ± 14 | −8.7 ± 0.7 | 80.2 | −200.3 | −539.4 | −33.1 | −24.4 |
| Q46 | *MYEOV* | 13 | 63.8 | −109.0 ± 9 | −330.0 ± 20 | −6.7 ± 2.0 | 69.8 | −190.9 | −529.3 | −26.8 | −20.1 |
| Q47 | *VEGF* | 13 | 64.0 | −34.6 ± 1.1 | −102.6 ± 3.4 | −2.8 ± 0.1 | 73.0 | −199.0 | −547.4 | −29.2 | −26.4 |
| Q48 | *FLI 1* | 14 | 71.0 | −27.0 ± 0.2 | −78.0 ± 0.5 | −3.0 ± 0.1 | 77.4 | −201.5 | −547.3 | −31.8 | −28.8 |
| Q49 | *FGR* | 14 | 58.5 | −33.6 ± 0.2 | −102.4 ± 0.6 | −1.8 ± 0.01 | 77.7 | −208.0 | −565.4 | −32.7 | −31.0 |
| Q50 | *WNT 3* | 14 | 70.0 | nd | | | 84.2 | −215.1 | −574.5 | −37.0 | nd |
| Q51 | *CSF 1R* | 14 | 65.0 | nd | | | 78.7 | −207.9 | −563.5 | −33.2 | nd |
| Q52 | *THRA* | 14 | 83.0 | −54.1 ± 0.6 | −152.0 ± 1.6 | −6.8 ± 0.1 | 88.6 | −219.5 | −579.3 | −40.0 | −33.2 |
| Q53 | *FES* | 14 | 68.0 | −34.0 ± 0.3 | −98.2 ± 0.7 | −3.5 ± 0.1 | 86.8 | −217.2 | −576.0 | −38.6 | −35.1 |
| Q54 | *REL* | 14 | 63.7 | nd | | | 82.7 | −213.7 | −573.2 | −36.0 | nd |

[a]Thermodynamic profiles for quadruplexes (2 μM) obtained from UV melting performed in 10 mM sodium cacodylate buffer, 100 mM KCl, pH 7.4. $T_m$ values differed by ±1.5°C. Thermodynamic parameters were determined by non-linear least square curve fitting (Details in Materials and method section). nd, not determined. Thermodynamic profile of mixed quadruplex conformation was not calculated.
[b]Predicted thermodynamic parameters for duplex formed by all sequences obtained through HYTHER (ozone2.chem.wayne.edu) using experimental conditions of 2 μM strand concentration and 100 mM monovalent cations.

**Table 4.** Thermodynamic profiles obtained for sequences with loop length of 15–18

| Oligo name | Gene | Total | Quadruplex[a] | | | | Duplex[b] | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | $T_m$ (°C) | $\Delta H$ (kcal/mol) | $\Delta S$ (cal/mol/K) | $\Delta G_{37}$ (kcal/mol) | $T_m$ (°C) | $\Delta H$ (kcal/mol) | $\Delta S$ (cal/mol/K) | $\Delta G_{37}$ (kcal/mol) | $\Delta G_{37}$ (kcal/mol) |
| Q55 | *FGF 4* | 15 | 57.0 | nd | | | 78.7 | −217.2 | −589.8 | −34.4 | nd |
| Q56 | *BCL 3* | 15 | 60.0 | nd | | | 77.6 | −210.3 | −572.2 | −33.0 | nd |
| Q57 | *KRAS* | 15 | 53.3 | −41.6 ± 1.2 | −125.5 ± 3.7 | −2.3 ± 0.1 | 75.5 | −213.8 | −585.8 | −32.2 | −28.8 |
| Q58 | *THPO* | 16 | 87.0 | −69.0 ± 4.0 | −205.5 ± 11.0 | −5.3 ± 0.6 | 81.5 | −218.8 | −589.5 | −36.1 | −26.5 |
| Q59 | *RALB* | 16 | 65.0 | nd | | | 78.2 | −218.1 | −593.4 | −34.2 | nd |
| Q60 | *EGFR* | 18 | 56.4 | nd | | | 80.7 | −241.0 | −653.6 | −38.4 | nd |
| Q61 | *VAV 1* | 18 | 57.4 | nd | | | 75.3 | −236.1 | −650.2 | −34.6 | nd |
| Q62 | *FGF 6* | 18 | 70.0 | nd | | | 80.9 | −241.3 | −654.0 | −38.6 | nd |

[a]Thermodynamic profiles for quadruplexes (2 μM) obtained from UV melting performed in 10 mM Sodium Cacodylate buffer, 100 mM KCl, pH 7.4. $T_m$ values differed by ± 1.5°C. Thermodynamic parameters were determined by non−linear least square curve fitting (details in Materials and method section). nd, not determined. Thermodynamic profiles of mixed quadruplex conformations were not determined.
[b]Predicted thermodynamic parameters for duplex formed by all sequences obtained through HYTHER (ozone2.chem.wayne.edu) using experimental conditions of 2 μM strand concentration and 100 mM monovalent cations.

**Table 5.** Comparison of thermodynamic data obtained from DSC and UV study performed in 10 mM sodium cacodylate buffer, pH 7.4, 100 mM KCl

| Oligo name | Gene | DSC[a] | | | UV[b] | | |
|---|---|---|---|---|---|---|---|
| | | $T_m$ (°C) | $\Delta H_{cal}$ (kcal/mol) | $\Delta S_{cal}$ (cal/mol/K) | $T_m$ (°C) | $\Delta H_{vH}$ (kcal/mol) | $\Delta S_{vH}$ (cal/mol/K) |
| Q1 | *VEGF* | 82.0 | −53.0 ± 0.4 | −149.3 ± 1.1 | 80.8 | −50.6 ± 0.5 | −142.1 ± 1.4 |
| Q2 | *WNT3* | 87.5 | −105.0 ± 0.8 | −291.0 ± 2.2 | 86.0 | −101.5 ± 2.5 | −277.0 ± 5.0 |
| Q4 | *PDGFB* | 84.8 | −73.0 ± 0.6 | −204.0 ± 1.6 | 83.0 | −70.0 ± 1.2 | −195.0 ± 3.4 |
| Q5 | *JUN B* | 80.2 | −28.5 ± 1.7 | −80.7 ± 0.8 | 77.7 | −26.0 ± 0.1 | −73.5 ± 0.3 |
| Q11 | *FES* | 60.5 | −55.5 ± 0.4 | −166.4 ± 1.2 | 58.0 | −53.0 ± 0.5 | −160 ± 1.1 |
| Q14 | *FOSL1* | 75.6 | −39.4 ± 0.5 | −113.0 ± 1.5 | 73.7 | −36.0 ± 0.2 | −101 ± 0.5 |
| Q30 | *c-KIT* | 73.0 | −32.0 ± 0.6 | −92.5 ± 1.7 | 71.6 | −29.0 ± 0.2 | −82.0 ± 0.4 |
| Q34 | *AKT2* | 87.8 | −69.0 ± 0.8 | −191.2 ± 2.2 | 87.0 | −66.0 ± 1.8 | −182.0 ± 5.1 |
| Q58 | *THPO* | 88.0 | −72.0 ± 0.7 | −200.0 ± 2.0 | 87.0 | −69.0 ± 4.0 | −205.5 ± 11.0 |

[a]DSC study was performed at 50 μM oligonucleotide concentration. The parameters were obtained from three experimental replicas through two-state model with zero heat capacity change. $T_m$ values differed by ±1.0°C.
[b]UV melting study was performed at 2 μM oligonucleotide concentration and the thermodynamic parameters were obtained from non-linear curve fitting. $T_m$ values differed by ±1.5°C.

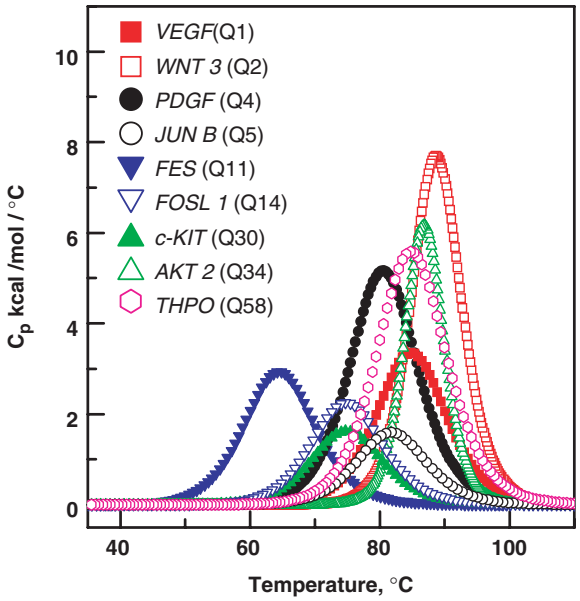


**Figure 3.** Representative DSC curves for different quadruplexes in 10 mM sodium cacodylate buffer, pH 7.4, 100 mM KCl.

−31.0 kcal/mol (35). Another study on stability of intramolecular quadruplexes formed by DNA sequences containing four G3 tracts separated by either single T or T4 loops, showed higher stability with the single T loops and the arrangement of different length loops has little effect on thermodynamic stability of quadruplex (31). Comparing our data with literature reports for model sequences with total loop length of 7 bases, we observe that enthalpy ranges from −17.7 to −53.0 kcal/mol for quadruplexes formed by naturally occurring sequences in contrast to values reported for model sequences which ranges from −26.8 to −33.7 kcal/mol (34). Further for longer loops, we observe that our enthalpy values are slightly lesser than those reported for model sequences, thereby hinting toward multiple conformations adopted by naturally occurring sequences (31,32,34,35).

Both enthalpy and entropy in turn influence $G$, which is an indicator of thermodynamic stability. It is well documented in literature that increases in loop length results in decrease in quadruplex stability; however this observation is reported mainly for model sequences (29–34). A recent study using model sequences has shown that total loop length ≤5 has a major effect on quadruplex stability, and an addition of single nucleotide significantly affects the thermodynamic stability (34). Another study using model sequences with long central loops (6–9 bases) and two side loops composed of single base has shown that pyrimidines provide more stability over adenines and the composition of the central loop can minimize the destabilizing effect of a long central loop on quadruplex stability (29). This is in agreement with another study which demonstrated that, for intramolecular quadruplexes with G3 tracts and single base loops, a T to A loop substitution in reduced stability by 8°C (33). Intriguingly, the thermodynamic data obtained for quadruplexes formed by naturally occurring sequences having different loop lengths show that there is no clear correlation of loop length with the thermodynamic stability. Figure 4 depicts the variability in the thermodynamic stability with respect to the loop length. Quadruplexes with a total loop length of 5 formed by *WNT* 5A (Q3), *VEGF* (Q1) and *WNT* 3 (Q2) have thermodynamic stability ranging from −3.7 to −15.6 kcal/mol, respectively. Similarly, wide variations in the thermodynamic stability (−1.8 to −9.3 kcal/mol) were observed among quadruplex formed by sequences with loop size of 7. Likewise, no significant correlation of loop length and thermodynamic stability was observed for other data sets with longer loops. The lack of correlation between the thermodynamic analysis of quadruplexes formed by model sequences and genomic sequences with varied loop length can be attributed to the loop composition. For the model sequences, the loop length and composition are usually homogeneous. However, the naturally occurring sequences show random occurrences of purines and/or pyrimidines in the loops between the G stretches, which substantially influence the quadruplex thermodynamic stabilities.

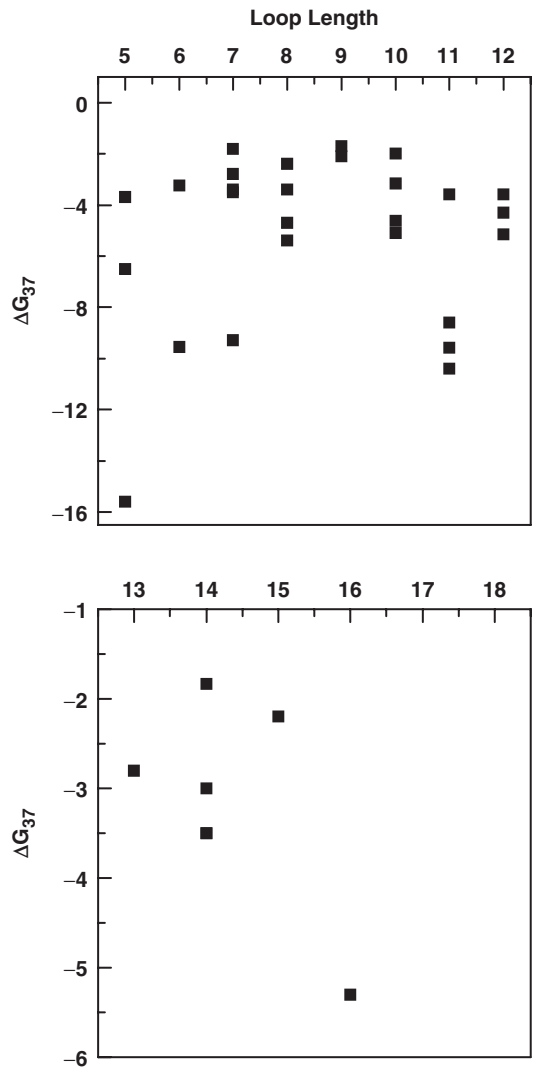


**Figure 4.** Representative graphs showing effect of loop length on thermodynamic stability ($\Delta G_{37}$) of quadruplex.

For instance, *PIM*1 (Q9) and *FGF* 3 (Q10) have total loop lengths of 7 with similar loop composition display remarkably different thermodynamic stabilities of −9.3 and −1.8 kcal/mol, respectively. Apart from total loop length and loop composition, the number of bases in the central loop is critical for quadruplex stability (32–35). We observed that for smaller loops, quadruplexes with a single base central loop are much more stable and the stability decreased upon increasing the bases in the central loop, as observed in case of *WNT 3* (Q2) and *WNT 5A* (Q3) with thermodynamic stability of −15.6 and −3.7 kcal/mol, respectively. Our results shows that the majority (32) of quadruplex-forming sequences of proto-ncogenes have $\Delta G_{37}$ values approximately within the range of −1.8 to −8.6 kcal/mol. Only five genes from our data set have high $\Delta G_{37}$ values ($\geq$−9.0 kcal/mol). These genes include *WNT* 3 (Q2), *MYB* (Q35), *AKT 2* (Q34), *PDGFB* (Q4) and *PIM* 1 (Q9) which form quadruplex structures having 1–3 bases in the central loop and have $\Delta G_{37}$ as −15.6, −10.4, −9.6, −9.6 and −9.3 kcal/mol, respectively. Potential quadruplex-forming sequences of promoter regions of c-*MYC*, c-*KIT*, *VEGF* and *PDGFB* have served as paradigms for quadruplex-mediated gene expression regulation (6,7,51). The structure adopted by promoter regions c-*MYC* (Q21), c-*KIT* (Q30), *VEGF* (Q1) and *PDGFB* (Q4) have thermodynamic stability of −3.2, −3.6, −6.5 and −9.6 kcal/mol, respectively.

Emerging evidence from bioinformatic data analysis indicates a wide distribution of quadruplex-forming sequences in the human genome with an average incidence of 1 quadruplex in 10 000 bases (27). Our study shows that the sequences in the promoter region of proto-oncogenes adopt quadruplex structures which exhibit variable thermodynamic stability ranging from −1.7 to −15.6 kcal/mol. These findings raise a discussion over whether biological function is associated with quadruplex structures having lower thermodynamic stability, which undergoes facile formation and disruption, or is it accomplished, by quadruplex structures with high thermodynamic stability. Our attempt to elucidate the thermodynamic profile of the quadruplexes formed by naturally occurring sequences also invites a systematic effort to integrate the thermodynamics and biological relevance of quadruplexes that would allow better understanding of quadruplex-mediated regulatory mechanism. In the human genome, a number of sites with potential quadruplex-forming structures have been estimated and some of these potential sites for quadruplex formation have been correlated with gene function (13). Quadruplex formation at specific human promoter regions has suggested the notion that quadruplexes might serve as potential regulatory motifs. However, these speculations have only been supported by analysis of either synthetic oligonucleotides or supercoiled DNA. To comprehend how quadruplex formation contributes in gene regulation, it is essential to take into account the contribution of the competing Watson–Crick duplex structure. For quadruplex formation, G-rich regions must be released from the duplex DNA, which occurs under transient denaturation during replication, transcription and recombination. The interconversion of quadruplex and duplex DNA is dependent on the relative stability of these competing secondary structures. In the current study we have examined the role of loop length on thermodynamic stability of quadruplex formed by naturally occurring sequences in human proto-oncogene promoters. However, it is necessary to compare thermodynamic stability of quadruplexes with their respective duplexes to understand duplex–quadruplex interconversion. It is difficult to obtain the thermodynamic parameters involved in duplex formation from the same sequences and their respective complementary strand by UV melting studies, as it includes the contributions from both duplex and quadruplex. We therefore obtained the thermodynamic profiles for the duplexes through a nearest neighbor (NN) method (48,49). This method assumes that the stability of a given base pair depends on the identity and orientation of the neighboring base pairs. During the past decades numerous studies have been performed to calculate thermodynamic parameters of a given duplex under specified experimental conditions by this method.

It is well established that NN method allows evaluation of the thermal stability and thermodynamic parameters of duplexes with precision and data are in agreement with the experimental data. Hyther is a tool that allows calculation of nucleic acid hybridization thermodynamics using NN method. Using Hyther, we calculated the thermodynamic profile of duplexes formed by G-rich sequence with their respective complementary strand with the strand concentration and buffer conditions (Tables 2–4) used in this study. The important parameter that dictates the predominance of either the duplex or quadruplex is the relative free energy difference, the $\Delta\Delta G_{37}$ between the duplex and quadruplex forms. It is noteworthy that in all the cases $\Delta\Delta G_{37}$ values are negative indicating that duplex is predominant structure in the competition (Tables 2–4). We also observed an increase in duplex stability upon increasing the loop length (Tables 2–4). The relative free energy difference, the $\Delta\Delta G_{37}$ between duplex and quadruplex structure increases ($\sim$−8.1 kcal/mol to −35.1 kcal/mol) upon increasing the loop length. The greater the negative magnitude of $\Delta\Delta G_{37}$, the higher is the predominance of duplex at equilibrium. As shown in Figure 5, only the quadruplexes formed by *WNT* 3 (Q2), *PDGFB* (Q4), *VEGF* (Q1) and *WNT 5A* (Q3) demonstrate less negative $\Delta\Delta G_{37}$ values of −8.1 kcal/mol, −12.1 kcal/mol, −13.3 kcal/mol and −16.0 kcal/mol, respectively. These lower $\Delta\Delta G_{37}$ values indicate that although the duplex is more stable than the quadruplex, the contribution from competing quadruplex population is high, which leads to less negative $\Delta\Delta G_{37}$ values.

The genomic DNA predominantly exists in double-stranded conformation; however, during specialized conditions guanine stretches fold spontaneously into quadruplex structures. These structures have the potential to serve as functional elements and identifying their structural and functional characteristics is essential for understanding their gene regulation. Literature cites examples for opposing roles of quadruplexes in biology. Repressor role of quadruplex in *KRAS* and c-*MYC* promoter region has been demonstrated using reporter assay and quadruplex interacting ligand, TMpyP4 (18). In contrast, the stimulatory role of quadruplexes has been shown for chicken beta globin genes and human insulin genes through biochemical and biophysical analysis (18). These contrasting roles observed for quadruplexes can be explained on the basis positional regulomics. Here, the position of the regulatory motif which binds a transcription factor influences its activator or repressor activity in different tissues (57). Recently, a dual role as a transcriptional repressor and activator has been demonstrated for a GGA element in quadruplex-forming region in the c-*MYB* promoter (58). Analysis of a genome-wide study has shown that the presence of potential quadruplex-forming motifs in the transcriptional regulatory region (−500 to +500 bp) is associated with significant enrichment of RNA polymerase II at the transcriptional regulatory region (59). These findings taken together lead to the hypothesis that quadruplex structures may be common regulatory elements involved in transcriptional regulation. The transient separation of duplex DNA during transcription increases the opportunity for the quadruplex formation.

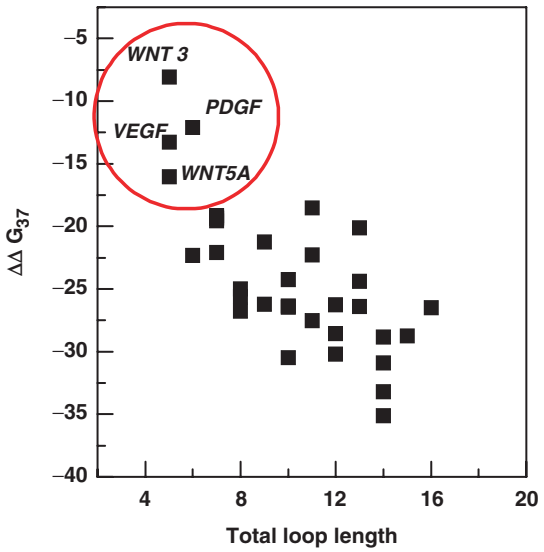


**Figure 5.** Plots showing dependence of relative thermodynamic stability of duplex and quadruplex on total loop length. The encircled points display lower $\Delta\Delta G_{37}$ values suggesting a significant contribution from competing quadruplex structures.

These structural motifs are kinetically trapped species which dissociate slowly and hold the open DNA structure, which thereby render the template available for higher rate of transcription. Apart from the G-rich strand, the C-rich complementary strand also has the potential for forming a secondary structure called i-motif. This structure consists of intercalated hemiprotonated C:$C^+$ base pairs (60,61). Although these structures are stable at acidic pH, they may also form at near physiological pH as in case of c-*MYC* (61). Lately, the evidence of proteins interacting with i-motif structures has aroused interest in biological role of i-motifs (62,63). Therefore, during transcription, ion or pH fluctuations both G- and C-rich strand may adopt competing secondary structures which dissociate slowly and keep the DNA in an open conformation or act as molecular recognition motif for transcription factors.

## CONCLUSION

Though loop length has marked influence on the thermodynamic stability of quadruplexes, the variation in loop composition makes it difficult to find a significant correlation between loop length and quadruplex stability. Apart from loop length and composition, loop symmetry also affects thermodynamic stability of quadruplexes. Furthermore, the results obtained in this study can be discussed in the light of natural scenario, where both quadruplex and duplex structures are likely to coexist and undergo transition to execute their respective biological role. The relative stability of the duplex and quadruplex structure would dictate the predominance of either of these structures at equilibrium. Our results affirm that increase in loop length pushes the equilibrium toward duplex and outcompetes quadruplex formation.

## SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online.

## ACKNOWLEDGEMENT

Authors thank Dr Vinod Scaria for his help in using Quadfinder. The authors also thank Dr Anjan A. Sen from Centre of Theoretical Physics, Jamia Milia Islamia University for the help in use of Mathematica 5.1.

## FUNDING

Council of Scientific and Industrial Research grant (N.K.). Council of Scientific and Industrial Research grant (S.M.). Funding for open access charge: Council of Scientific & Industrial Research (CSIR), India.

*Conflict of interest statement*. None declared.

## REFERENCES

1. Blackburn,E.H. (1994) Telomeres: no end in sight. *Cell*, **77**, 621–623.
2. 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.
3. Simonsson,T., Pecinka,P. and Kubista,M. (1998) DNA tetraplex formation in the control region of c-myc. *Nucleic Acid Res.*, **26**, 1167–1172.
4. Simonsson,T. and Henriksson,M. (2002) c-myc Suppression in Burkitt's lymphoma cells. *Biochem. Biophys. Res. Commun.*, **290**, 11–15.
5. Jain,A.S., 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.
6. Sun,D., Guo,K., 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.
7. Saxena,S., Bansal,A. and Kukreti,S. (2008) Structural polymorphism exhibited by a homopurine.homopyrimidine sequence found at the right end of human c-jun proto-oncogene. *Arch. Biochem. Biophys.*, **471**, 95–108.
8. Williamson,J.R. (1994) G-quartet structures in telomeric DNA. *Annu. Rev. Biophys. Biomol. Struct.*, **23**, 703–730.
9. Phan,A.T., Kuryavyi,V. and Patel,D.J. (2006) DNA architecture: from G to Z. *Curr. Opin. Struct. Biol.*, **16**, 288–298.
10. Simonsson,T. (2001) G-quadruplex DNA structures – variations on a theme. *Biol. Chem.*, **382**, 621–628.
11. Huppert,J.L. and Balasubramanian,S. (2005) Prevalence of quadruplexes in the human genome. *Nucleic Acids Res.*, **33**, 2908–2916.
12. Todd,A.K., Johnston,M. and Neidle,S. (2005) Highly prevalent putative quadruplex sequence motifs in human DNA. *Nucleic Acids Res.*, **33**, 2901–2907.
13. Eddy,J. and Maizels,N. (2006) Gene function correlates with potential for G4 DNA formation in the human genome. *Nucleic Acid Res.*, **34**, 3887–3896.
14. Wieland,M. and Hartig,J.S. (2007) RNA quadruplex-based modulation of gene expression. *Chem. Biol.*, **14**, 757–763.
15. 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.*, **4**, 218–221.
16. Arora,A., Dutkiewicz,M., Scaria,V., Hariharan,M., Maiti,S. and Kurreck,J. (2008) Inhibition of translation in living eukaryotic cells by an RNA G-quadruplex motif. *RNA*, **14**, 1290–1296.
17. Eddy,J. and Maizels,N. (2008) Conserved elements with potential to form polymorphic G-quadruplex structures in the first intron of human genes. *Nucleic Acid Res.*, **36**, 1321–1333.
18. Fry,M. (2007) Tetraplex DNA and its interacting proteins. *Front. Biosci.*, **12**, 4336–4351.
19. Giraldo,R., Suzuki,M., Chapman,L. and Rhodes,D. (1994) Promotion of parallel DNA quadruplexes by a yeast telomere binding protein: a circular dichroism study. *Proc. Natl Acad. Sci. USA*, **91**, 7658–7662.
20. Muniyappa,K., Anuradha,S. and Byers,B. (2000) Yeast meiosis-specific protein Hop1 binds to G4 DNA and promotes its formation. *Mol. Cell. Biol.*, **20**, 1361–1369.
21. 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.
22. 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.
23. Sun,H., Yabuki,A. and Maizels,N. (2001) A human nuclease specific for G4 DNA. *Proc. Natl Acad. Sci. USA*, **98**, 12444–12449.
24. Zaug,A.J., Podell,E.R. and Cech,T.R. (2005) Human POT1 disrupts telomeric G-quadruplexes allowing telomerase extension in vitro. *Proc. Natl Acad. Sci. USA*, **102**, 10864–10869.
25. Hurley,L.H., Wheelhouse,R.T., Sun,D., Kerwin,S.M., Salazar,M., Fedoroff,O.Y., Han,F.X., Han,H., Izbicka,E. and Von Hoff,D.D. (2002) G-quadruplexes as targets for drug design. *Pharmacol. Ther.*, **85**, 141–158.
26. Izbicka,E., Wheelhouse,R.T., Raymond,E., Davidson,K.K., Lawrence,R.A., Sun,D., Windle,B.E., Hurley,L.H. and Von Hoff,D.D. (1999) Effects of cationic porphyrins as G-quadruplex interactive agents in human tumor cells. *Cancer Res.*, **59**, 639–644.
27. Huppert,J.L. and Balasubramanian,S. (2007) G-quadruplexes in promoters throughout the human genome. *Nucleic Acids Res.*, **35**, 406–413.
28. Hazel,P., Huppert,J., Balasubramanian,S. and Neidle,S. (2004) Loop-length-dependent folding of G-quadruplexes. *J. Am. Chem. Soc.*, **126**, 16405–16415.
29. Guédin,A., De Cian,A., Gros,J., Lacroix,L. and Mergny,J.L. (2008) Sequence effects in single-base loops for quadruplexes. *Biochimie*, **90**, 686–696.
30. Smirnov,I. and Shafer,R.H. (2000) Effect of loop sequence and size on DNA aptamer stability. *Biochemistry*, **39**, 1462–1468.
31. 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 Acid Res.*, **35**, 4214–4222.
32. 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.
33. Rachwal,P.A., Brown,T. and Fox,K.R. (2007) Sequence effects of single base loops in intramolecular quadruplex DNA. *FEBS Lett.*, **44**, 1657–1660.
34. 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.
35. Kumar,N., Sahoo,B., Varun,K.A.S., Maiti,S. and Maiti,S. (2008) Effect of loop length variation on quadruplex-Watson Crick duplex competition. *Nucleic Acid Res.*, **36**, 4433–4442.
36. Miura,T. and Thomas,G.J. (1994) Structural polymorphism of telomere DNA: interquadruplex and duplex–quadruplex conversions probed by Raman spectroscopy. *Biochemistry*, **33**, 7848–7856.
37. Risitano,A. and Fox,K.R. (2003) Stability of intramolecular DNA quadruplexes: comparison with DNA duplexes. *Biochemistry*, **42**, 6507–6513.
38. Kumar,N. and Maiti,S. (2004) Quadruplex to Watson–Crick duplex transition of the thrombin binding aptamer: a fluorescence resonance energy transfer study. *Biochem. Biophys. Res. Commun.*, **319**, 759–767.
39. 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 Acid Res.*, **33**, 6723–6732.

40. 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.
41. Li,W., Miyoshi,D., Nakano,S. and Sugimoto,N. (2003) Competition involving G-quadruplex DNA and its complement. *Biochemistry*, **42**, 11736–11744.
42. Scaria,V., Hariharan,M., Arora,A. and Maiti,S. (2006) Quadfinder: server for identification and analysis of quadruplex-forming motifs in nucleotide sequences. *Nucleic Acid Res.*, **34**, W683–W685.
43. Cantor,C.R., Warshaw,M.M. and Shapiro,H. (1970) Oligonucleotide interactions. Circular dichroism studies of the conformation of deoxyoligonucleotides. *Biopolymers*, **9**, 1059–1077.
44. Marky,L.A., Blumenfeld,K.S., Kozlowski,S. and Breslauer,K.J. (1983) Salt-dependent conformational transitions in the self-complementary deoxydodecanucleotide d(CGCAATTCGCG): evidence for hairpin formation. *Biopolymers*, **9**, 1247–1257.
45. Mergny,J.L., Phan,A.T. and Lacroix,L. (1998) Following G-quartet formation by UV spectroscopy,. *FEBS Lett.*, **435**, 74–78.
46. Bishop,G.R., Ren,J., Polander,B.C., Jeanfreau,B.D., Trent,J.O. and Chaires,J.B. (2007) Energetic basis of molecular recognition in a DNA aptamer. *Biophys. Chem.*, **126**, 165–175.
47. McTigue,P.M., Peterson,R.J. and Kahn,J.D. (2004) Sequence-dependent thermodynamic parameters for locked nucleic acid (LNA)-DNA duplex formation. *Biochemistry*, **43**, 5388–5405.
48. SantaLucia,J. Jr. (1998) A unified view of polymer, dumbbell, and oligonucleotide DNA nearest-neighbor thermodynamics. *Proc. Natl Acad. Sci. USA*, **95**, 1460–1465.
49. Peyret,N., Seneviratne,P.A., Allawi,H.T. and SantaLucia,J. Jr. (1999) Nearest-neighbor thermodynamics and NMR of DNA sequences with internal A·A, C·C, G·G, and T·T mismatches. *Biochemistry*, **38**, 3468–3477.
50. 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.
51. Qin,Y., Rezler,E.M., Gokhale,V., Sun,D. and Hurley,L.H. (2007) Characterization of the G-quadruplexes in the duplex nuclease hypersensitive element of the *PDGF-A* promoter and modulation of *PDGF-A* promoter activity by TMPyP4. *Nucleic Acid Res.*, **35**, 7698–7713.
52. 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.
53. Kejnovska,I., Kypr,J. and Vorlickova,M. (2007) Oligo(dT) is not a correct native PAGE marker for single-stranded DNA. *Biochem. Biophys. Res. Commun.*, **353**, 776–779.
54. 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.
55. Lee,H.T., Olsen,C.M., Waters,L., Sukup,H. and Marky,L.A. (2008) Thermodynamic contributions of the reactions of DNA intramolecular structures with their complementary strands. *Biochimie*, **90**, 1052–1063.
56. Majhi,P.R., Qi,J., Tang,C.-F. and Shafer,R.H. (2008) Heat capacity associated with guanine quadruplex formation: an isothermal titration calorimetry study. *Biopolymers*, **89**, 302–309.
57. Tharakaraman,K., Bodenreider,O., Landsman,D., Spouge,J.L. and Mariño-Ramírez,L. (2008) The biological function of some human transcription factor binding motifs varies with position relative to the transcription start site. *Nucleic Acid Res.*, **36**, 2777–2786.
58. Palumbo,S.L., Memmott,R.M., Uribe,D.J., Krotova-Khan,Y., Hurley,L.H. and Ebbinghaus,S.W. (2008) A novel G-quadruplex-forming GGA repeat region in the c-myb promoter is a critical regulator of promoter activity. *Nucleic Acid Res.*, **36**, 1755–1769.
59. Du,Z., Zhao,Y. and Li,N. (2008) Genome-wide analysis reveals regulatory role of G4 DNA in gene transcription. *Genome Res.*, **18**, 233–241.
60. Gehring,K., Leroy,J.L. and Guéron,M. (1993) A tetrameric DNA structure with protonated cytosine.cytosine base pairs. *Nature*, **363**, 561–565.
61. Simonsson,T., Pribylova,M. and Vorlickova,M. (2000) A nuclease hypersensitive element in the human c-*myc* promoter adopts several distinct *i*-tetraplex structures, *Biochem. Biophys. Res. Commun.*, **319**, 158–166.
62. Cortés,A., Huertas,D., Fanti,L., Pimpinelli,S., Marsellach,F.X., Piña,B. and Azorín,F. (1999) DDP1, a single-stranded nucleic acid-binding protein of Drosophila, associates with pericentric heterochromatin and is functionally homologous to the yeast Scp160p, which is involved in the control of cell ploidy. *EMBO J.*, **18**, 3820–3833.
63. Lacroix,L., Liénard,H., Labourier,E., Djavaheri-Mergny,M., Lacoste,J., Leffers,H., Tazi,J., Hélène,C. and Mergny,J.L. Identification of two human nuclear proteins that recognise the cytosine-rich strand of human telomeres in vitro. *Nucleic Acids Res.*, 28, 1564–1575.
64. Marky,L.A. and Breslauer,K.J. (1987) Calculating thermodynamic data for transitions of any molecularity from equilibrium melting curves. *Biopolymers*, **26**, 1601–1620.