Nucleic Acids Research, 1997, Vol. 25, No. 21 **4389–4392**

The contribution of thymine–thymine interactions to the stability of folded dimeric quadruplexes

Max A. Keniry*, Elisabeth A. Owen and Richard H. Shafer[1]

Research School of Chemistry, The Australian National University, GPO Box 414, Canberra, ACT 2601, Australia and [1]Department of Pharmaceutical Chemistry, School of Pharmacy, University of California, San Francisco, CA 94143-0446, USA

Received June 3, 1997; Revised and Accepted September 8, 1997

ABSTRACT

The loop of four thymines in the sodium form of the dimeric folded quadruplex [d($G_3T_4G_3$)]$_2$ assumes a well-defined structure in which hydrogen bonding between the thymine bases appears to contribute to the stability and final conformation of the quadruplex. We have investigated the importance of the loop interactions by systematically replacing each thymine in the loop with a cytosine. The quadruplexes formed by d($G_3CT_3G_3$), d($G_3TCT_2G_3$), d($G_3T_2CTG_3$) and d($G_3T_3CG_3$) in the presence of 150 mM Na^+ were studied by gel mobility, circular dichroism and ^{1}H NMR spectroscopy. The major species formed by d($G_3CT_3G_3$), d($G_3TCT_2G_3$) and d($G_3T_3CG_3$) at 1 mM strand concentration at neutral pH is a dimeric folded quadruplex. d($G_3T_2CTG_3$) has anomalous behaviour and associates into a greater percentage of linear four-stranded quadruplex than the other three oligonucleotides at neutral pH and at the same concentration. The linear four-stranded quadruplex has a greater tendency to oligomerize to larger ill-defined structures, as demonstrated by broad ^{1}H NMR resonances. At pH 4, when the cytosine is protonated, there is a greater tendency for each of the oligonucleotides to form some four-stranded linear quadruplex, except for d($G_3T_2CTG_3$), which has the reverse tendency. The experimental results are discussed in terms of hydrogen bonding within the thymine loop.

INTRODUCTION

The structural polymorphism of guanine-rich sequences, principally found in telomere repeats at the end of chromosomes, has generated considerable interest during the past few years (1,2). Substantial effort has been expended toward the complete structure determination of the various forms of G quadruplex with the goal of understanding the biological significance of these unusual structures (3–9). In general, a single tract of two or more guanines can form only a four-stranded structure, an oligonucleotide with two tracts of two or more guanines separated by at least three non-guanine residues can also form a bimolecular fold-back structure and an oligonucleotide with four tracts of two or more guanines each separated by at least two non-guanine residues can also form a unimolecular fold-back structure. In the latter two cases there is frequently an equilibrium involving folded forms and four-stranded structures of ill-defined size, under kinetic control, that may complicate a detailed structure determination.

Although there is little doubt that the guanine tracts primarily determine the quadruplex structure, the base sequence in the non-guanine linker regions may have a secondary role. Surprisingly, little work has been published on the effect of the base sequence in these linker regions on quadruplex structure and stability, although the sequence and content is highly conserved. The non-guanine bases in guanine-rich telomeric sequences are usually thymine residues and less frequently adenines and cytosines.

In a previous publication we reported the detailed structure of the Na^+ form of the dimeric folded quadruplex d($G_3T_4G_3$)$_2$ (10). The T_4 linker sequence between the two G_3 guanine tracts formed a highly structured loop which appeared to be stabilized by hydrogen bonds between the thymine residues. In order to probe the importance of the hydrogen bond interactions within the loop we have individually replaced each thymine base within the loop with a cytosine base and studied the effect by non-denaturing gel mobility, circular dichroism and ^{1}H nuclear magnetic resonance.

MATERIALS AND METHODS

Oligonucleotide synthesis and purification

The decanucleotides d($G_3T_4G_3$), d($G_3CT_3G_3$), d($G_3TCT_2G_3$), d($G_3T_2CTG_3$) and d($G_3T_3CG_3$) were synthesized by a cyanoethylphosporamidite method on an Applied Biosystems 394 DNA synthesizer and purified by anion exchange HPLC, desalted by HPLC and finally dialysed against water over 24 h with one change of dialysate.

Qaudruplex formation

DNA quadruplexes were formed by dissolving solid lyophilized oligonucleotide in 10 mM phosphate buffer containing 150 mM NaCl at pH 7. For some experiments the pH was adjusted to 4 by

*To whom correspondence should be addressed. Tel: +61 6 249 2863; Fax: +61 6 249 0750; Email: max@rsc.anu.edu.au

Figure 1. Non-denaturing 18% polyacrylamide gel in Tris–borate buffer containing 150 mM NaCl of: I, $[d(G_3CT_3G_3)]_2$ (C4); II, $[d(G_3TCT_2G_3)]_2$ (C5); III, $[d(G_3T_2CTG_3)]_2$ (C6); IV, $[d(G_3T_3CG_3)]_2$ (C7); P, $[d(G_3T_4G_3)]_2$.

adding aliquots of 0.1 M HCl. The quadruplexes formed by $d(G_3CT_3G_3)$, $d(G_3TCT_2G_3)$, $d(G_3T_2CTG_3)$ and $d(G_3T_3CG_3)$ in this buffer are denoted C4, C5, C6 and C7 respectively.

Non-denaturing gel electrophoresis

The oligonucleotides were analysed by electrophoresis on 18% polyacrylamide gels containing 90 mM Tris–borate buffer, 0.2 mM EDTA and 150 mM NaCl, pH 8. All DNA samples were preheated to 85°C for 5 min, slowly cooled and loaded onto the gel in 10 mM phosphate buffer, pH 7, containing 150 mM NaCl in 40% Ficoll. Electrophoresis was carried out at 4°C at 1.5 V/cm for 48 h. The gels were visualized by UV shadowing and then photographed under UV light at 254 nm.

Circular dichroism

CD data were recorded on a Jobin Yvon (Longjumeau, France) Dichrograph CD6 spectropolarimeter interfaced to a Haake (Karlsruhe, FRG) F3 circulating water bath. All samples (1 mM in 10 mM phosphate buffer containing 150 mM NaCl at pH 7 or pH 4) were preheated to 80°C for 5 min and slowly cooled for 40 min prior to measurement. A 0.1 mm path length cell was used for all measurements and each spectrum was an average of five scans.

Nuclear magnetic resonance

One-dimensional ^1H NMR spectra were recorded at 15°C on a Varian INOVA 500 NMR spectrometer. All samples were 1 mM in strand concentration in 10 mM phosphate, 150 mM NaCl buffer in 90% H_2O/10% D_2O at either pH 7 or pH 4. All samples were preheated to 80°C for 5 min and cooled slowly for 40 min prior to measurement. All spectra were recorded using a $1\bar{1}$ echo (11) over a sweep width of 8000 Hz, 8128 complex points, 512 scans and a recycle time of 2 s. All spectra were apodized with an exponential line broadening factor of 2 Hz.

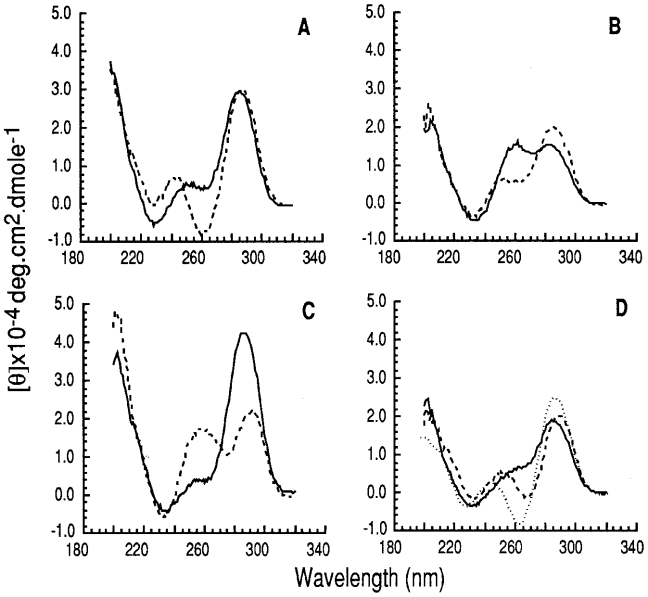

Figure 2. CD spectra at ~1 mM (in strand) in 10 mM phosphate buffer containing 150 mM NaCl at pH 7 (- - - -) or pH 4 (——) at 20°C of: (**A**) $[d(G_3CT_3G_3)]_2$ (C4); (**B**) $[d(G_3TCT_2G_3)]_2$ (C5); (**C**) $[d(G_3T_2CTG_3)]_2$ (C6); (**D**) $[d(G_3T_3CG_3)]_2$ (C7). The spectrum of the parent quadruplex, $[d(G_3T_4G_3)]_2$ under the same conditions at pH 7 is displayed as a dotted line (·····) in D.

RESULTS

Gel electrophoresis

Gel electrophoresis is a convenient method for characterizing macromolecular sizes and distribution. In previous work we have shown that the folded dimeric quadruplex $d(G_3T_4G_3)_2$ migrates as a single species at a rate faster than the unstructured monomer in the presence of moderate concentrations of monovalent cations (12). $d(G_3T_4G_3)_2$ and $d(G_4T_4G_4)_2$ can also form linear four-stranded qudruplexes in the presence of K^+ ions, which always migrate slower than the more compact dimeric folded form (13).

We have studied the electrophoretic mobility of C4, C5, C6 and C7 at pH 7 in a non-denaturing 18% polyacrylamide gel containing 150 mM NaCl. We observe that each of the oligonucleotides C4, C5 and C7 migrates as a single band (Fig. 1, columns I, II and IV respectively) each slightly slower than the band of the parent quadruplex (Fig. 1, column P), but C6 migrates as two bands (Fig. 1, column III), one of which migrates at a rate slightly slower than the parent quadruplex and the other at a rate considerably slower than the parent quadruplex. Under the electrophoretic conditions the parent quadruplex, C4, C5 and C7 are mainly dimeric folded quadruplexes, but C6 is a mixture of dimeric folded quadruplex and four-stranded linear quadruplex.

Circular dichroism

CD spectroscopy has been used to distinguish between parallel and antiparallel G quadruplexes. The former quadruplexes exhibit a positive CD band around 260 nm and a negative CD band at 240 nm (13,14). The latter quadruplexes exhibit a positive CD band around 290 nm and a weaker negative band at 265 nm (13). In previous work we demonstrated that the K^+ form of

d($G_3T_4G_3$)$_2$, which we subsequently established by NMR techniques to be an antiparallel quadruplex (12), displays an intense CD band at 290 nm and a weaker negative band at 267 nm. The spectra of each of the oligonucleotides C4, C5, C6 and C7 at pH 7 have a strong positive band at 290 nm, another weaker positive band at 240–255 nm and a weak negative band at 235 nm (Fig. 2A–D). The spectrum of the parent quadruplex (the dotted curve in Fig. 2D), which has a strong positive band at 290 nm, is displayed for comparison. The spectrum of C6 is distinctly different from the spectra of C4, C5 and C7 because the band at 255 nm is the same intensity as the band at 290 nm. At pH 4 the band at 255 nm has a higher relative intensity in the spectra of C4, C5 and C7 and shifts to a longer wavelength when compared with the equivalent spectra at pH 7. In contrast, the positive band at 290 nm is much more intense than the positive band at 255 nm in the spectrum of C6 at pH 4. The spectrum of the parent quadruplex shows no pH dependence in the range pH 4–7. CD spectra recorded at a concentration of 0.1 mM in strand in a 1 mm cell (data not shown) have almost identical spectra at pH 7, but at pH 4 the intensity of the positive band at 255 nm is substantially less for each of C4, C5, C6 and C7. This result would be anticipated if the parallel quadruplex is a linear four-stranded conformation and the antiparallel quadruplex is a two-stranded folded conformation. A linear four-stranded quadruplex is considerably more sensitive to mass action effects than a two-stranded quadruplex and hence lower oligonucleotide concentrations are expected to favour the two-stranded folded quadruplex.

Nuclear magnetic resonance

The imino proton region of the ^{1}H NMR spectrum is an indicator of symmetry and stoichiometry. The ^{1}H NMR spectrum of the parent quadruplex d($G_3T_4G_3$)$_2$ displayed 12 distinct imino proton resonances corresponding to the 12 distinct guanine residues in the three quartets (12,15,16). The spectra of C4, C5, C6 and C7 at pH 7 demonstrate at least two components (Fig. 3A–D), one an underlying broad component and the other a series of narrow imino proton peaks. Apart from this observation, each spectrum is distinctly different, indicating either a different loop structure or, at the very least, altered hydrogen exchange rates. The two most closely related spectra are those of C4 and C7, which are also similar to the imino proton spectrum of the parent quadruplex, d($G_3T_4G_3$)$_2$ (15,16), except that there are more than the expected 12 imino proton peaks in the spectrum of C7. We have also observed that the spectrum of C5 can be converted to a spectrum similar to C4 and C7 after cycling the pH from high to low to high pH again. The spectrum of C6 is always observed to have a much larger broad component than any of the other oligonucleotides at pH 7.

At pH 4 the relative ratios of the broad and narrow components change compared with the spectra at pH 7 (Fig. 3E–H). There is a substantial increase in the broad component for C4, C5 and C7 (Fig. 3E, F and H), whereas there is no significant change in the amount of broad component relative to the narrow component for C6 (Fig. 3G). No broad component is ever observed in the spectrum of the sodium form of the parent quadruplex, d($G_3T_4G_3$)$_2$, in the pH range covered by this study (data not shown).

DISCUSSION

The overwhelming majority of recent studies of DNA quadruplex structure and stability have concentrated on the interactions

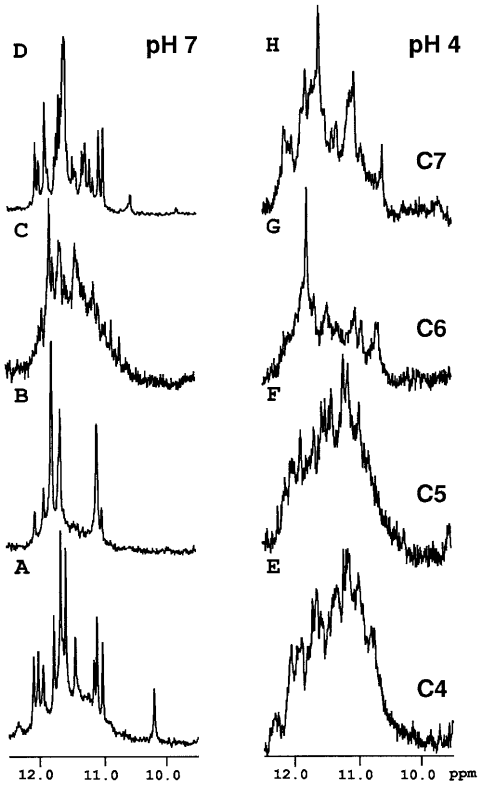

Figure 3. ^{1}H NMR spectra at 500 MHz at pH 7 of: (**A**) [d($G_3CT_3G_3$)]$_2$ (C4); (**B**) [d($G_3TCT_2G_3$)]$_2$ (C5); (**C**) [d($G_3T_2CTG_3$)]$_2$ (C6); (**D**) [d($G_3T_3CG_3$)]$_2$ (C7). ^{1}H NMR spectra at 500 MHz at pH 4 of: (**E**) [d($G_3CT_3G_3$)]$_2$ (C4); (**F**) [d($G_3TCT_2G_3$)]$_2$ (C5); (**G**) [d($G_3T_2CTG_3$)]$_2$ (C6); (**H**) [d($G_3T_3CG_3$)]$_2$ (C7). Each solution was ~1 mM (in strand) in 10 mM phosphate, 150 mM NaCl buffer in 90% H_2O/10% D_2O at 15°C.

within the guanine quartet. There is little argument that the major forces driving the formation of G quadruplexes are ionic interactions between the phosphate backbone and the monovalent cation, hydrogen bonding between the guanine bases and stacking forces between the individual G quartets. However, there is mounting evidence that interactions between the non-guanine bases contribute to the stability and final form of the quadruplex (10). Although the number and sequence of non-guanine bases is highly conserved in telomeric sequences, only Balagurumoorthy *et al*. (13) have addressed the contribution of these non-guanine bases in a systematic study of the effect of the length of the thymine region on the stability of the dimeric folded-back quadruplex structure.

We have determined the detailed three-dimensional structure of the dimeric folded-back quadruplex d($G^1G^2G^3T^4$-$T^5T^6T^7G^8G^9G^{10}$)$_2$. In this quadruplex the thymine loop assumes a highly ordered conformation in which the hydrogen acceptor and donor groups of the thymine bases assume positions that suggest that hydrogen bonding between the thymine residues in the loop may contribute to the overall structure and stability of the quadruplex. In particular, the imino proton of T^6, the third thymine from the 5′-end of the loop, was sufficiently close to form a hydrogen bond with the O2 of T^4, the thymine at the 5′-end of the loop. The hydrogen exchange rate of this thymine imino proton was sufficiently slow that it produced a narrow resonance in the ^{1}H NMR spectrum (10). Figure 4 illustrates the structure of the loop,

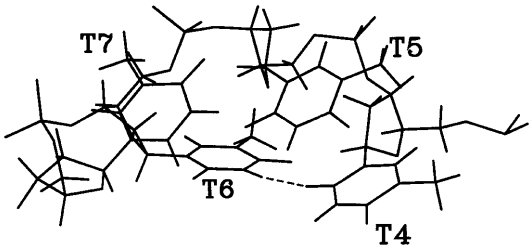

Figure 4. Monoview of the loop structure of $[d(G_3T_4G_3)]_2$ emphasizing the close proximity of T^6 and T^4.

in particular the unusual conformation of T^6 that permits this base to stack over the quadruplex quartet proximate to T^4.

If this hydrogen bond is important for the structure of the loop and consequently for the conformation that the quadruplex assumes, then by replacing a thymine base with a cytosine base, which at neutral pH does not have an imino proton, the dimeric folded form of this quadruplex would be destabilized. Furthermore, this destabilization may be partially reversed when the N3 of cytosine is protonated at lower pH. Of course, in the latter experiment protonation of the cytosine adds a positive charge to the quadruplex. Hardin and co-workers have observed stabilization of the linear quadruplex at low pH when there is a mixture of quadruplex and Watson–Crick duplex (17). This effect was attributed to hemiprotonated $C.C^+$ base pairing, which is possible in the linear parallel quadruplex but not in a Watson–Crick duplex or in a folded two-stranded quadruplex when the loops are at opposite ends of the stem. Since we observe no change in the CD spectrum or broadening of the NMR spectrum of the parent quadruplex over the pH range 4–7, then it is reasonable to attribute stabilization of the linear parallel quadruplex at pH 4 in C4, C5 and C7 to formation of hemiprotonated $C.C^+$ base pairs.

The gel electrophoresis, CD and ^{1}H NMR data establish that the difference in free energy between the dimeric fold-back and linear four-stranded quadruplexes formed by C4, C5, C6 and C7 is smaller than the free energy difference between the equivalent quadruplexes formed by $d(G_3T_4G_3)$. In the pH range 4–7, $d(G_3T_4G_3)$ spontaneously associates into a dimeric folded quadruplex in the presence of as little as 65 mM Na^+ and under these conditions no linear four-stranded form is observed. The sequence-specific spectral and structural variations of individual thymine to cytosine substitutions is surprising. Although theoretically the results could be interpreted in terms of stabilization of the linear quadruplex, the sequence-specific effects, particularly the important $T^6 \rightarrow C^6$ substitution, suggest, from symmetry arguments, that destabilization of the folded form is more likely. The gel mobility, CD spectra and ^{1}H NMR spectra, therefore, indicate that the dimeric folded quadruplex of C6 is destabilized at pH 7 and some other ill-defined species, most likely oligomers of linear four-stranded quadruplex, is favoured. Protonation of the cytosine at pH 4 does not induce further destabilization of the C6 folded quadruplex and the CD spectra indicate an increase in the amount of folded quadruplex relative to linear four-stranded quadruplex at the lower pH. These findings support the role of T^6 as the hydrogen donor in the hydrogen bond with T^4.

Thymine is the base most frequently found in telomeric sequences apart from guanine. There are two possible reasons for this. Firstly because of steric considerations, thymine can fit into a compact structure, especially hairpin loops. Secondly, the imino proton is a potential hydrogen bond donor that may stabilize a loop structure. Individual adenines are observed in telomeric sequences but, because of the bulky purine ring, multiple adenines are rarely found. Cytosines are rarely found in telomeres, probably because of a tendency to Watson–Crick pair to nearby guanines, but also because at neutral pH cytosine lacks a hydrogen bond donor at the N3 position. The significance of hydrogen bonding in the two-stranded versus four-stranded quadruplex equilibrium is demonstrated by stabilization of the dimeric fold-back quadruplex of C6 at low pH, even though an additional positive charge at the protonated cytosine should be a destabilizing factor in the formation of dimeric fold-back quadruplexes. This work suggests that any alteration in the loop conformation may substantially change the thermodynamic stability of the quadruplex by changing the hydrogen bond interactions. Of course, there are other factors that may also affect the loop structure, including base stacking interactions and differences in the charge distribution in the thymine and cytosine rings. Recent studies on the stability of intrastrand loops in quadruplexes using a DNA polymerase arrest assay provide further evidence that thymine-rich tracts stabilize these loops (18).

ACKNOWLEDGEMENTS

We thank Dr Ivan Smirnov for his expert assistance with the gel mobility studies. This research was supported by NIH grant GM 51650, DHHS and benefited from the use of NMR facilities at the Australian National University. We also acknowledge the Computer Graphics Laboratory, UCSF, supported by grant RR01081 from the Division of Research Resources, NIH, DHHS.

REFERENCES

1 Williamson,J.R. (1994) *Annu. Rev. Biophys. Biomol. Struct.*, **23**, 703–730.
2 Rhodes,D. and Giraldo,R. (1995) *Curr. Opin. Struct. Biol.*, **5**, 311–322.
3 Wang,K.Y., Swaminathan,S. and Bolton,P.H. (1994) *Biochemistry*, **33**, 7517–7527.
4 Schultze,P., Macaya,R.F. and Feigon,J. (1994) *J. Mol. Biol.*, **235**, 1532–1547.
5 Wang,Y. and Patel,D.J. (1993) *Structure*, **1**, 263–282.
6 Schultze,P., Smith,F.W. and Feigon,J. (1994) *Structure*, **2**, 221–233.
7 Laughlan,G., Murchie,A.I.H., Norman,D.G., Moore,M.H., Moody,P.C.E., Lilley,D.M.J. and Luisi,B. (1994) *Science*, **265**, 520–524.
8 Kang,C., Zhang,X., Ratliff,R., Moyzis,R. and Rich,A. (1992) *Nature*, **356**, 126–131.
9 Wang,K.Y., Krawczyk,S.H., Bischofberger,N., Swaminathan,S. and Bolton,P.H. (1993) *Biochemistry*, **32**, 11285–11292.
10 Keniry,M.A., Strahan,G.D., Owen,E.A. and Shafer,R.H. (1995) *Eur. J. Biochem.*, **233**, 631–643.
11 Sklenár,V. and Bax,A. (1987) *J. Magn. Resonance*, **74**, 469–479.
12 Scaria,P.V., Shire,S.J. and Shafer,R.H. (1992) *Proc. Natl. Acad. Sci. USA*, **89**, 10336–10340.
13 Balagurumoorthy,P., Brahmachari,S.K., Mohanty,D., Bansal,M. and Sasisekharan,V. (1992) *Nucleic Acids Res.*, **20**, 4061–4067.
14 Lu,M., Guo,Q. and Kallenbach,N.R. (1992) *Biochemistry*, **31**, 2455–2459.
15 Smith,F.W., Lau,F.W. and Feigon,J. (1994) *Proc. Natl. Acad. Sci. USA*, **91**, 10546–10550.
16 Strahan,G.D., Shafer,R.H. and Keniry,M.A. (1994) *Nucleic Acids Res.*, **22**, 5447–5455.
17 Hardin,C.C., Corregan,M., Brown,B.A. and Frederick,L.N. (1993) *Biochemistry*, **32**, 5870–5880.
18 Weitzmann,M.N., Woodford,K.J. and Usdin,K. (1996) *J. Biol. Chem.*, **271**, 20958–20964