Modelling studies on neurodegenerative disease-causing triplet repeat sequences d(GGC/GCC)$_n$ and d(CAG/CTG)$_n$

SHIBASISH CHOWDHURY and MANJU BANSAL*

Molecular Biophysics Unit, Indian Institute of Science, Bangalore 560 012, India

**Corresponding author (Fax, 91-80-3600683; Email, mb@mbu.iisc.ernet.in).*

Model building and molecular mechanics studies have been carried out to examine the potential structures for d(GGC/GCC)$_5$ and d(CAG/CTG)$_5$ that might relate to their biological function and association with triplet repeat expansion diseases. Model building studies suggested that hairpin and quadruplex structures could be formed with these repeat sequences. Molecular mechanics studies have demonstrated that the hairpin and hairpin dimer structures of triplet repeat sequences formed by looping out of the two strands are as favourable as the corresponding B-DNA type hetero duplex structures. Further, at high salt condition, Greek key type quadruplex structures are energetically comparable with hairpin dimer and B-DNA type duplex structures. All tetrads in the quadruplex structures are well stacked and provide favourable stacking energy values. Interestingly, in the energy minimized hairpin dimer and Greek key type quadruplex structures, all the bases even in the non-G tetrads are cyclically hydrogen bonded, even though the A, C and T-tetrads were not hydrogen bonded in the starting structures.

1. Introduction

In recent years, it has been shown that many neuro-degenerative diseases are caused by expansion of triplet repeats. To date, more than twelve human genetic diseases, including myotonic dystrophy (dystrophia myotonica, DM), fragile X syndrome (FraX), Huntington disease (HD), several spinocerebellar ataxias and Friedreich ataxia have been associated with the expansion of GGC, CAG or GAA repeats (Pearson and Sinden 1998). Actually, GGC, CAG and GAA repeats indicate duplex repeat sequences (GGC)$_n$–(GCC)$_n$ [also indicated as (CGG)$_n$–(GCC)$_n$, Han *et al* 1994], (CAG)$_n$–(CTG)$_n$ and (GAA)$_n$–(TTC)$_n$. Various human genetic diseases associated with different triplet repeat sequences are listed in table 1 and it is seen that nine different disease loci have (CAG)$_n$–(CTG)$_n$ repeats and four fragile-X chromosomal sites contain (GGC)$_n$–(GCC)$_n$ repeats. The locations of these triplet repeats are different in various diseased genes. For example, FraX triplets, (GGC)$_n$/(GCC)$_n$ are located in the 5′-untranslated region (5′UTR) of the FMR-1 gene (Yu *et al* 1991). DM triplets, (CTG)$_n$ are located in the 3′-untranslated region (3′UTR) of the myotonin protein kinase (Mt-PK) gene (Fu *et al* 1992; Brook *et al* 1992; Mahadevan *et al* 1992) while HD triplets, (CAG)$_n$ are located inside the first codon of the Huntington's gene (Huntington's disease collaborative research group 1993). However, the following properties are common in all these triplet repeats. In normal individuals, these triplet repeats exhibit mild polymorphism in repeat length whereas in disease phenotypes, they are considerably expanded and are highly unstable (Caskey *et al* 1992). It has been demonstrated that in normal individuals, the number of d(GGC) repeats range between ~ 6 and ~ 56. In phenotypically normal "premutation" carriers, the number of repeats increases to ~ 50 to ~ 200 and in individuals affected with FraX, the number of repeats further increases to ~ 200 to ~ 2000 (Fu *et al* 1991; reviewed in Mandel 1993). Similarly in case of HD, normal individuals have ~ 11 to ~ 34 copies of CAG repeats, whereas affected

Keywords. d(GGC/GCC)$_5$; d(CAG/CTG)$_5$; hairpin structure; model building; molecular mechanics; quadruplex structure

Abbreviations used: DM, dystrophia myotonica; FraX, fragile X syndrome; HD, Huntington disease; rms, root mean square.

individuals have between ~ 37 to ~ 90 copies (Wells 1996) and in DM, normal individuals have ~ 5 to ~ 37 copies of CTG repeats while > 700 CTG repeats are observed in DM affected individuals. It has been also demonstrated that an inverse relationship exists between the triplet repeat length and the age of onset of the disease and a direct relationship exists between repeats length and the severity of the disease (Duyao *et al* 1993) which is clinically referred to as "anticipation".

Molecular mechanisms of triplet repeat instabilities (both expansion and deletions) have been investigated by several groups and it has been proposed that DNA structures may play critical roles in the expansion of these triplets repeats (reviewed in Sinden 1999). It has been observed that linear DNA molecule containing triplet repeats moves anomalously during gel electrophoresis studies, indicating that triplet repeats containing DNA sequences posses greater inherent flexibility to adopt unusual structures (Chastain *et al* 1995; Bacolla *et al* 1997; Chastain and Sinden 1998). Several physico-chemical and NMR studies have demonstrated that these repeats containing DNA sequences can adopt various unusual structures (figure 1) including slipped-strand DNA structures (Wells 1996; Pearson and Sinden 1998), hairpins (Mitas 1997; Darlow and Leach 1998a; Mariappan *et al* 1996a, b, 1998; Chen *et al* 1998) as well as quadruplex structures (Fry and Loeb 1994; Chen 1995; Kettani *et al* 1995; Usdin 1998). On the basis of these unusual DNA structures, numerous models have been proposed to explain the expansion (or deletion) of triplet repeats leading to human diseases (Wells and Sinden 1993; Fry and Loeb 1994; Kang *et al* 1995; Usdin and Woodford 1995; Mariappan *et al* 1996a; Wells 1996; Samadashwily *et al* 1997; Chen *et al* 1998).

In case of d(GGC)·d(GCC) repeat, G-rich strand can align (in term of frame) as GGC, CGG and GCG while possible alignments of C-rich strand are GCC, CCG, CGC. Frames 1 and 2 are stabilized by Watson and Crick G : C base pairs and mismatch C : C or G : G pairs. In frame 3, even though there is no WC pairing, investigators have considered this alignment because of the possibility that at least the G-rich strand might form a quadruplex structure held together by G-quartets. However, d(CAG)·d(CTG) repeat can form stable hairpin structures only with CAG or CTG frame because in other frames (AGC and GCA as well as TGC and GCT) there is no possibility of favourable hydrogen bonds between the strands. As discussed above, secondary structure formation by the single strand of aforementioned triplet repeat sequences have been addressed by several biophysical studies including electrophoretic mobility, NMR studies, chemical and enzymatic probing, DNA base modifications and cleavage as well as by energy considerations [Kohwi *et al* 1993; Mitas *et al* 1995b; Gacy *et al* 1995; Yu *et al* 1995a, b; Smith *et al* 1995; Mariappan *et al* 1996a, b; Petruska *et al* 1996; Zheng *et al* 1996 for d(CAG)·d(CTG), Darlow and Leach 1998a; for d(GGC)·d(GCC) repeats]. From the above studies, it has been concluded that (reviewed by Darlow and Leach 1998a) G-rich strand takes up hairpin structure in frame 1 but can also form parallel stranded quadruplex structures. However, formation of antiparallel quadruplex structures is also possible with hydrogen bonding patterns similar to frame 3 but the formation is very slow at neutral pH. It has been also suggested (Mariappan *et al* 1996b; Darlow and Leach 1998b) that even-membered loop in frame 2 is energetically more favourable while odd-membered loop is favoured in frame 1.

Thus, hairpins and quadruplex structures are the most common forms of higher order structures which could be adopted by triplet repeats containing DNA sequences. In order to understand the relative stability of various higher order DNA helical structures, model building and molecular mechanics studies have been carried out on hairpin,

Table 1. Human genetic diseases caused by the expansion of $d(GGC/GCC)_n$ and $d(CAG/CTG)_n$ triplet repeat expansion (Richards and Sutherland 1997; Sinden 1999).

Disease	Repeat	Location	Normal length	Full disease length	Possible biological effect of expansion
Fragile XA (FRAXA)	(GGC)	5′-UTR	6–52	230–2000	Promoter methylation, gene silencing
Fragile XE (FRAXE)	(GCC)	5′-UTR	4–39	200–900	Promoter methylation, gene silencing
FrgileXF (FRAXF)	(GGC)	5′-UTR	7–40	306–1008	Promoter methylation, gene silencing
Jacobsen syndrome (FRA11B)	(GGC)	5′-UTR	11	100–1000	Promoter methylation, gene silencing
Kennedy syndrome (SMBA)	(CAG)	Coding	14–32	40–55	Polyglutamine tract expansion
Huntington disease (HD)	(CAG)	Coding	10–34	40–121	Polyglutamine expansion
Myotonic dystrophy (DM)	(CTG)	3′-UTR	5–37	80–1000	Altered mRNA processing, altered gene expression
Spinocerebellar ataxia 1* (SCA1)	(CAG)	Coding	6–39	40–81	Polyglutamine expansion
Machado Joseph disease (MJD/SCA3)	(CAG)	Coding	13–44	60–84	Polyglutamine expansion
Haw River syndrome (HRS/DRPLA)	(CAG)	Coding	7–25	49–75	Polyglutamine expansion

*Spinocerebellar ataxia 2, 6 and 7 are also caused by expansion of CAG repeats in coding region.

hairpin dimer and parallel as well as antiparallel quadruplex structures, containing $(GGC)_5$, $(GCC)_5$, $(CAG)_5$ and $(CTG)_5$ repeats. Model building studies explore the conformational features of various hairpin and quadruplex structures while molecular mechanics studies show relative stability and the energetic features of high order DNA structures. Relative energy (enthalpy) of hairpin and quadruplex structures are also compared with corresponding B-DNA type hetero-duplex structure.

2. Models and methods

2.1 *Model building*

Parallel as well as antiparallel quadruplex structures and hairpin models have been built manually for $d(GGC)_5$, $d(GCC)_5$, $d(CAG)_5$ and $d(CTG)_5$ sequence motifs. Hairpin dimer structures are generated with these hairpin structures. All quadruplexes, hairpins and hairpin dimer structures have been energy minimized using normal as well as reduced phosphate oxygens charge. The energy-minimized structures have been analysed for conformational and energetic features.

2.2 *Model building of quadruplex structures*

Earlier molecular mechanics studies (Mohanty and Bansal 1993) and melting experiments (Lu *et al* 1993) have shown that for an all G stretch a parallel all-*anti* alignment of strands is more favourable compared to an antiparallel arrangement of strands. However, the sequences, which consist of runs of guanine residues, interspersed with short stretches of thymines or adenines favour folded back structures (Mohanty and Bansal 1994). The parallel quadruplex structures were built with $(GGC)_5$, $(GCC)_5$, $(CAG)_5$ and $(CTG)_5$ sequences, starting from Arnott's fibre model of parallel poly r(G)-quadruplex structure (Arnott *et al* 1974). In each case, depending on the sequence, some of the guanine bases are replaced by other bases (cytosine, adenine or thymine). The cytosine, adenine and thymine bases are kept in exactly the same conformation as guanines in the fiber model. The parallel quadruplex structures with the GGC, GCC, CAG and CTG repeats (for a single repeat unit) are schematically shown in figure 2. Thus, in case of $(GGC)_5$ and $(GCC)_5$, stacked G-tetrads and C-tetrads occur in the four stranded parallel quadruplex structures. Similarly,

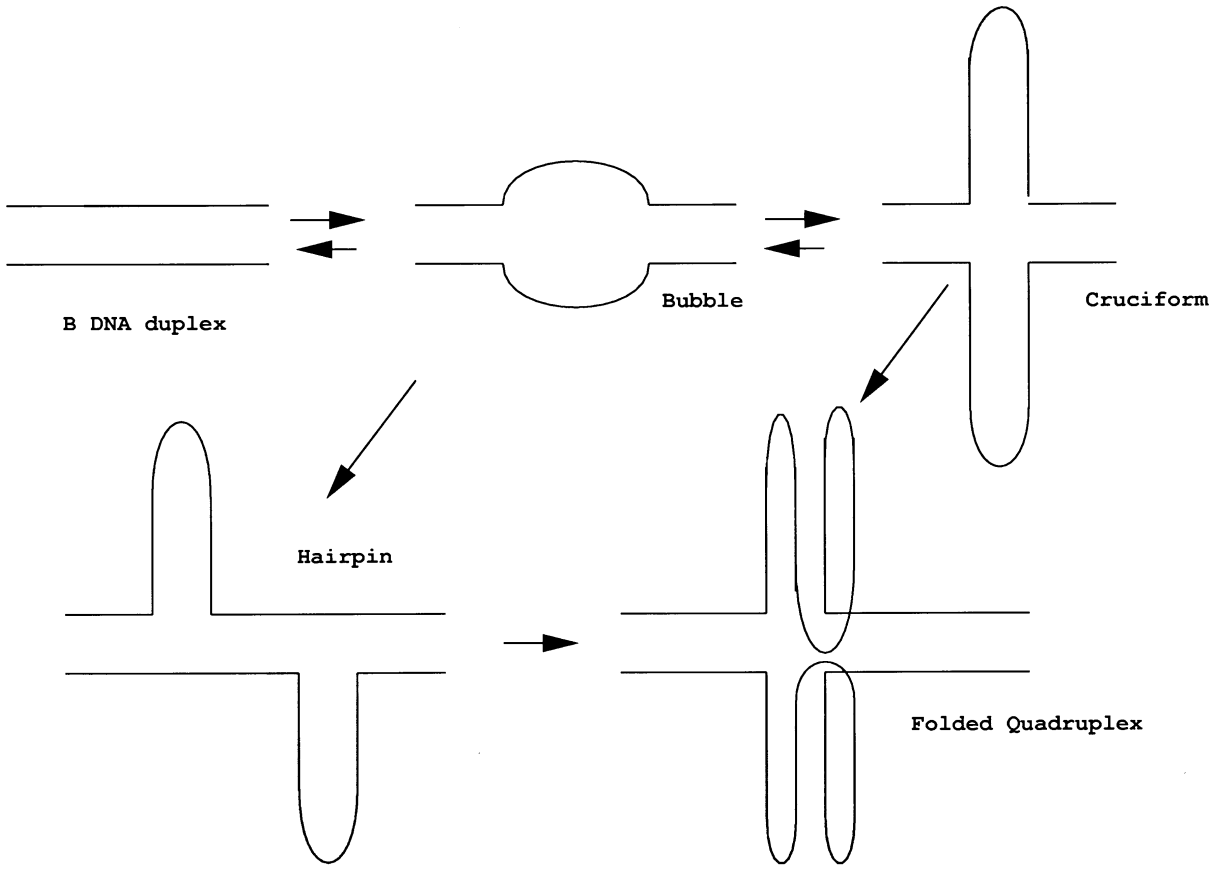

Figure 1. Formation of probable higher order structures from B-DNA duplex.

C, A and G-tetrads are formed in $(CAG)_5$ quadruplex structure while C, T and G tetrads form in the $(CTG)_5$ parallel quadruplex structure. It is worth mentioning that so far parallel quadruplex structures have not been observed experimentally for GCC, CAG or CTG repeats.

In an antiparallel arrangement of quadruplex structure, two antiparallel strands could be adjacent (Greek key type) or diagonal (Indian key type). The formation of G-C-G-C tetrad is common to all the Greek key type quadruplex structures with GGC, GCC, CAG and CTG repeat sequences (figure 3). Similarly, G-G-C-C tetrads

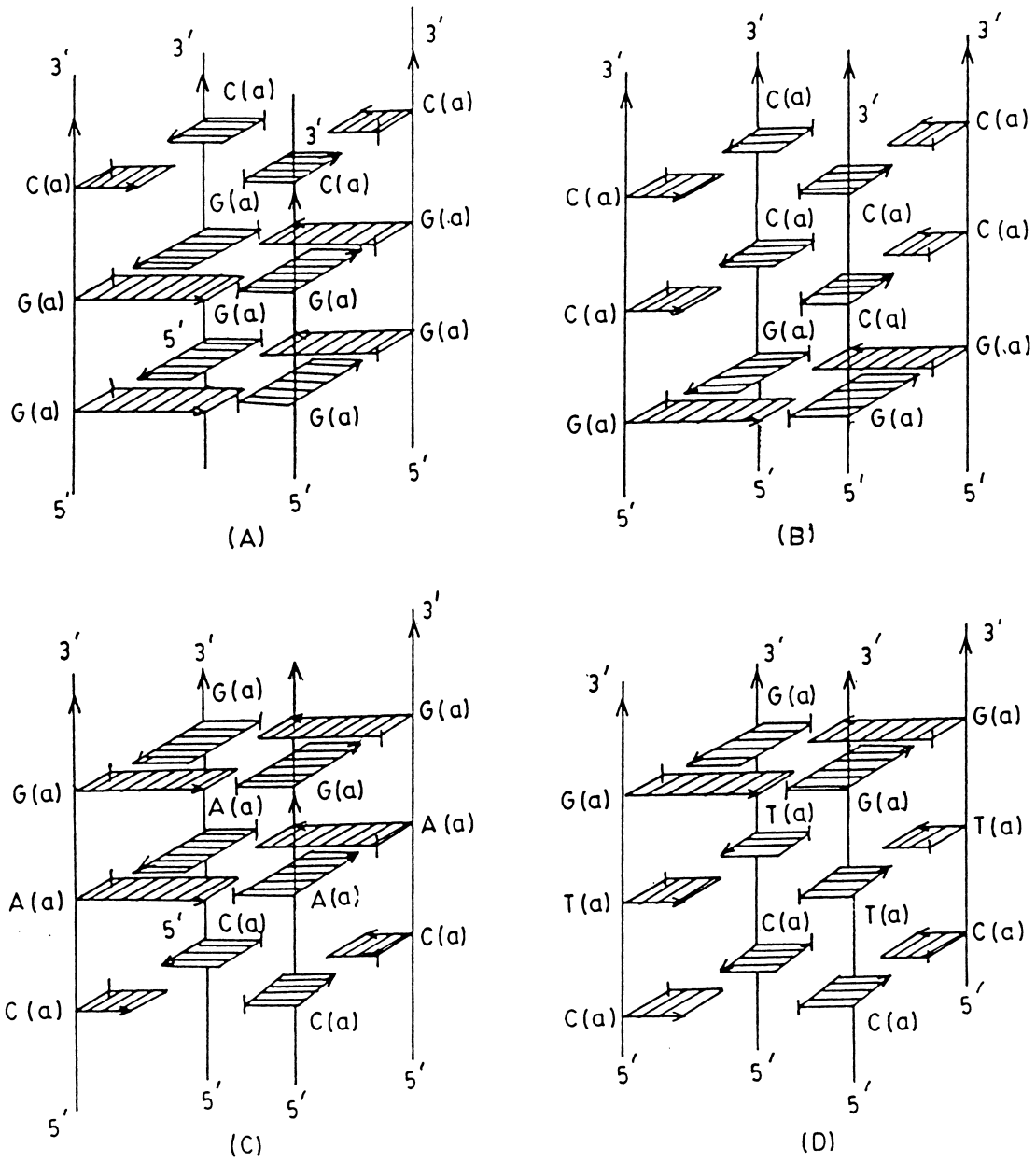

Figure 2. Four-stranded parallel quadruplex structure formed by the (**A**) GGC, (**B**) GCC, (**C**) CAG and (**D**) CTG repeat. Purine bases (guanines and adenines) have been represented as larger rectangular planes while smaller rectangular planes represents pyrimidine bases (cytosines and thymines). The WC faces of all the bases are indicated by → while Hoogsteen faces of purines are shown by —|. Similarly, —| represents C5 and C6 atoms containing side of cytosines and methyl group containing side of thymines. The *a*-face (the faces corresponding to clockwise numbering of atoms visible from top) of guanines and adenines, and *b*-face (anticlockwise numbering of atoms visible from top) of cytosines and thymines, when visible from the top, have been shaded. The glycosidic conformations of all the bases are *anti* (a). The chain polarity is shown by arrow direction.

stabilize Indian key type quadruplex structures with all these triplet repeat (GGC, GCC, CAG and CTG) sequences. Representative Indian key type antiparallel structure with GGC repeat is shown in figure 4. In addition to that, both Greek and Indian key type quadruplex structures with GGC repeats are stabilized by G-G-G-G tetrads while C-C-C-C, A-A-A-A and T-T-T-T tetrads stabilize quadruplex structures with GCC, CAG and CTG repeat sequences respectively. As observed in figure 4, Hoogsteen sides of both the guanines in G-G-C-C tetrad are facing each other while C5 and C6 atoms containing sides in both the cytosines are facing each

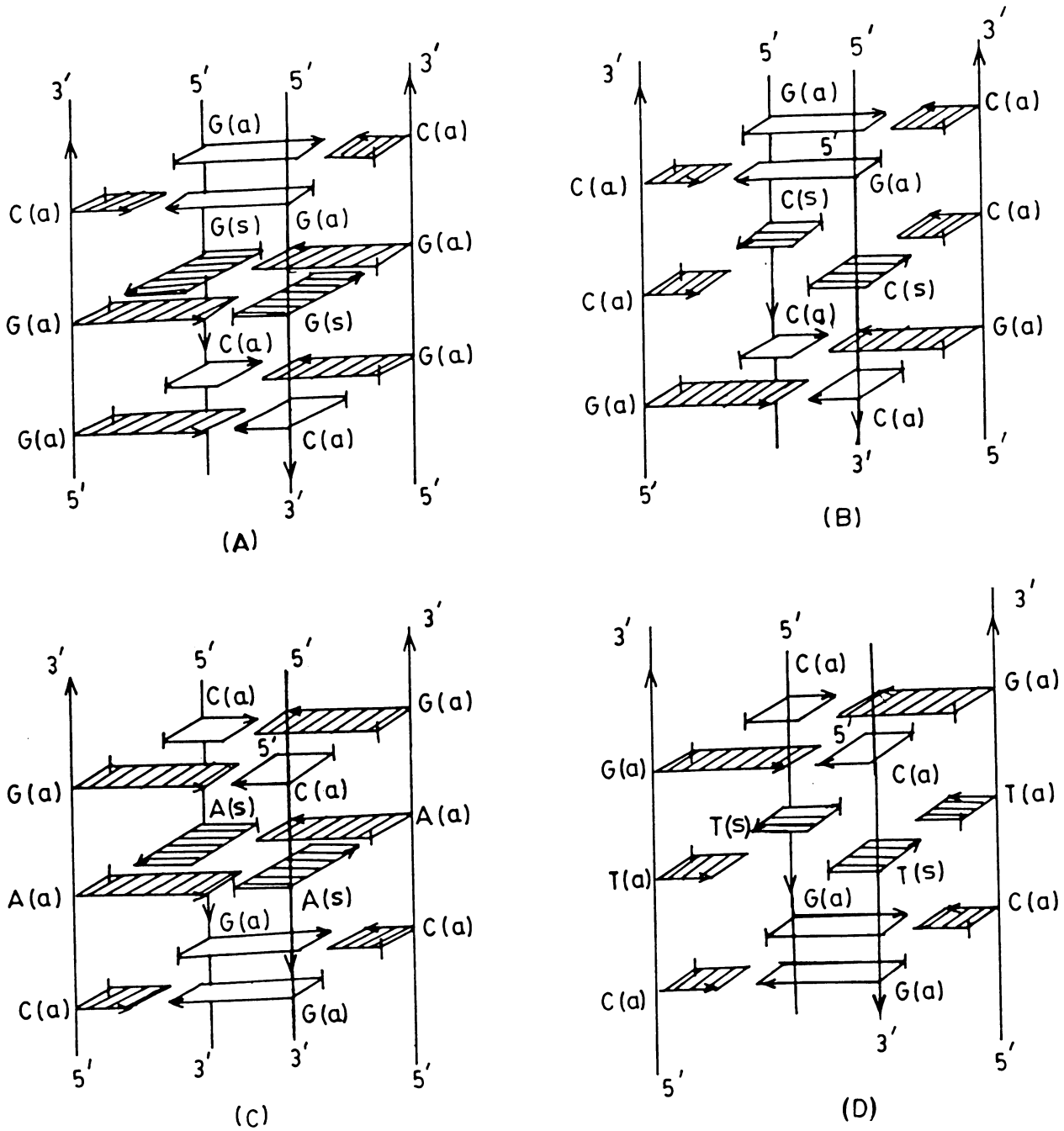

Figure 3. Four-stranded antiparallel Greek key type quadruplex structures formed by the (**A**) GGC, (**B**) GCC, (**C**) CAG and (**D**) CTG repeat. The representations of bases are similar to that described in figure 1. All *anti* glycosidic conformation of bases is observed within the GCGC tetrads while alternating *anti-syn* conformation is observed in G-, C-, A- and T-tetrads.

other. Thus, the hydrogen bond formation within guanines as well as within cytosines is not possible in G-G-C-C tetrad. Hence G-G-C-C tetrads in Indian key type quadruplex structures contain fewer number of hydrogen bonds than the G-C-G-C tetrads in Greek key type quadruplex structures and are considered to be thermodynamically less favourable. Therefore, only Greek key type antiparallel structures with these repeat sequences are considered for model building studies.

Model building of antiparallel Greek key type quadruplex structure was carried out in three stages. Initially, all the base tetrads are generated by a combination of rotations and translation of individual bases or base pairs, using INSIGHT II package. For example, G-C-G-C tetrad is constructed from two Watson-Crick G-C base pairs (as observed in the NMR structure of Kettani *et al* 1995) whereas guanine base tetrad is obtained from the fiber model of G-quadruplex structure (Arnott *et al* 1974) with cyclic Hoogsteen type hydrogen bonds within the tetrad. A, C and T tetrads are obtained from the G-tetrad by replacing guanine bases with adenine, cytosine and thymine bases respectively. The stacked arrangement of base tetrads were generated by fixing the successive tetrads at 30° twist and 3·4 Å rise (as observed in fiber model of parallel G-quadruplex structure). This was followed by suitable backbone generation for all the four strands. The adjacent strands were built in an antiparallel manner. Figure 5 shows the relative orientation of three successive nucleoside units in the single strand of GGC antiparallel quadruplex structure. Then, the adjacent sugar units were connected by phosphate groups. In the course of fitting the backbone, it was attempted to retain the backbone torsion angles in the canonical B-DNA range. The final structure selected had less number of short contacts and maximum number of near canonical torsion angles. In order to get acceptable connectivity between the successive tetrads in the antiparallel Greek key type quadruplex structures, the glycosidic torsion angles within the G-C-G-C tetrads, which is a dimer of Watson-Crick base pairs is necessarily all-*anti* while the glycosidic torsion angles within the G/A/C/T-tetrads should be alternating *syn-anti*. Since the C- and T-tetrads in both the parallel and antiparallel quadruplexes are obtained by replacing larger guanine bases in the G-tetrad with smaller cytosine/thymine bases, the adjacent bases in both C- and T-tetrads are not hydrogen bonded. However, the potential hydrogen bonding groups in both C- and T-tetrads are facing each other. For the starting parallel and antiparallel

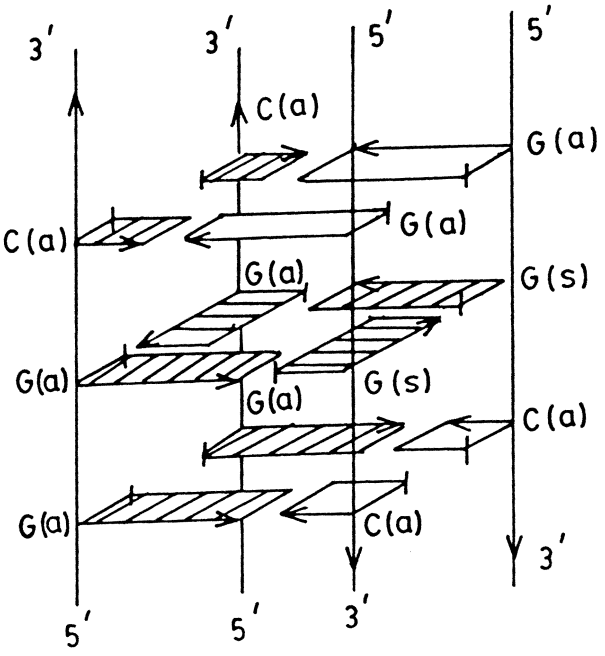

Figure 4. Indian key type strand arrangement is shown for GGC repeat with adjacent strands being alternately parallel and antiparallel. In the GGCC tetrads, Hoogsteen faces of guanines are facing each other while C5 and C6 atoms containing sides of cytosines are pointing each other. The representation of each base is similar to that described in figure 1.

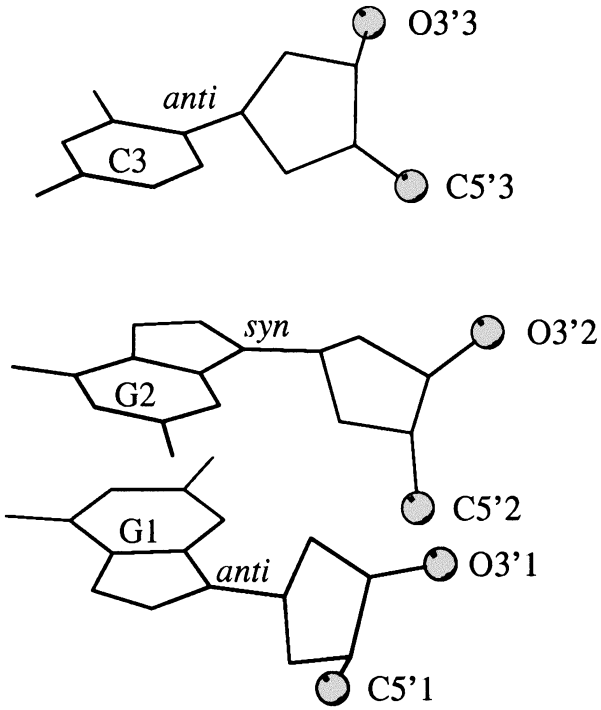

Figure 5. The orientation of C5′ and O3′ atoms of three successive GGC nucleosides in a representative strand of antiparallel quadruplex structure is shown along with the glycosidic conformation of bases.

quadruplexes, the distance between H4 and N3 atoms of adjacent cytosine bases in C-tetrads is ~ 2·9 Å, while the distance between H3 and O4 atoms of adjacent thymine bases in T-tetrads is ~ 2·8 Å.

2.3 *Model building of hairpin structures*

Hairpin structures are frequently proposed higher order structures, which could be adopted by triplet repeat sequences. Model building has been carried out to generate 15-mer hairpin dimer structures with various triplet repeat sequences, which contain hexameric quadruplex stem part and trimeric loop structures. Since these structures are expected to form by single chain folding back on itself, hence both the loops will be formed at one end of the stem. Therefore, in all the hairpin dimer structures, two loops are built on the same side of the stem and can potentially form inter loop hydrogen bonds (representative hairpin dimer structure with GGC repeat is shown in figure 6). Antiparallel Greek key type quadruplex structures with triplet repeat sequences are taken as a stem part of various hairpin dimer structures, and loops are generated to connect the two quadruplex strands which are attached by Watson-Crick type G : C hydrogen bonds. Thus, in case of $(GGC)_5$ and $(GCC)_5$ hairpin dimer structures, GGC and GCC loops connect a cytosine to a guanine, while in case of $(CAG)_5$ and $(CTG)_5$ hairpin structures, CAG and CTG loops connect a guanine to a cytosine. The first guanine/cytosine and third cytosine/guanine base, in the loop could in principle form intra loop G:C/C:G pairing involving Watson-Crick hydrogen bonding. Thus, in the hairpin dimer structures such G:C/C:G base pairs in both the loops can form a G-C-G-C tetrad.

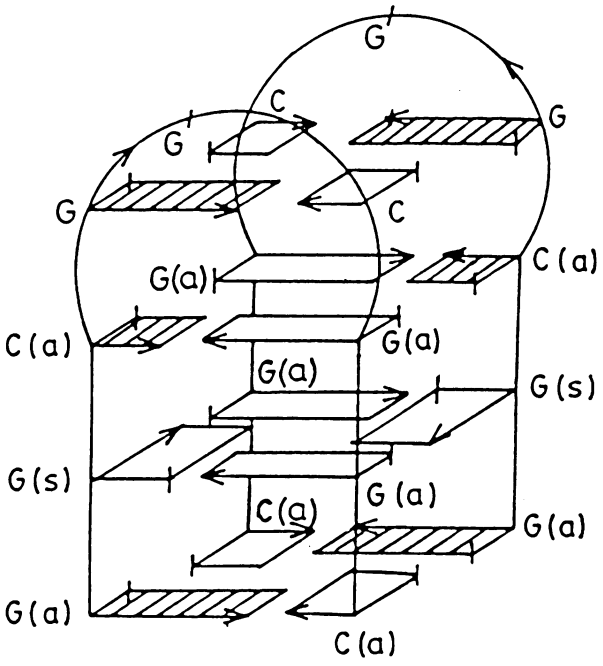

Figure 6. Hairpin dimer structure formed by three GGC repeats. Loops are connecting the WC paired strands. Loops are over the two larger grooves and on same side of hairpins. The GCGC tetrad formed by the first (guanine) and third (cytosine) base in the loop is also shown.

In case of $(GGC)_5$ hairpin dimer structure, the G-C-G-C tetrad was fixed at a z-height of 3·4 Å above the terminal C-G-C-G tetrad of the stem while the middle guanine base in each loop stacks over the first guanine and third cytosine base in each loop (which forms G : C base pair). This is followed by generation of the backbones of these two loops. The helical twist of the G-C-G-C tetrad base pair and the glycosidic torsion angles of guanines and cytosines were varied in the range 0° to 360° to find the values of glycosidic torsion angles which could give rise to a stereochemically satisfactory backbone link for both the GGC loops (Bansal and Sasisekharan 1986). From the grid search of the backbone torsion angles, it is observed that the glycosidic torsion angles of the first (G) and second (G) base in both the loops can remain in *anti* conformation while the glycosidic conformation of the third base (C) should be in *syn* conformation for stereochemically satisfactory connectivity. Similarly GCC, CAG and CTG loops are generated to obtain $(GCC)_5$, $(CAG)_5$ and $(CTG)_5$ hairpin dimer structures. The starting conformations of all the loops are almost identical for these repeat sequences.

2.4 *Energy minimization*

The various quadruplex and hairpin models are outlined in table 2. All quadruplex models are energy minimized to an rms gradient of 0·1 kcal/mol Å using AMBER 4·0 (Pearlman *et al* 1991) all atom forcefield. Solvent effects were mimicked implicitly by using a distance dependent dielectric function $e_{ij} = R_{ij}$ (Mohanty and Bansal 1993; Radhakrishnan and Patel 1993). In addition to the calculation carried out with normal AMBER electrostatic charges, all the structures were also minimized with a reduced charge (by an amount 0·7 of electronic charge, so that total charge on the molecule becomes – 0·3 per phosphate group) on anionic phosphate oxygens, to mimic the effect of counter ions (Mohanty and Bansal 1995; Ravikiran and Bansal 1997).

3. Results and discussion

3.1 *Hydrogen bonding patterns*

All energy minimized quadruplex structures are stabilized by stacked tetrads formed by four cyclically hydrogen

bonded bases. For example, GGC-P and GCC-P quadruplex structures are stabilized by stacked guanine (G) and cytosine (C) tetrads. The hydrogen bonding scheme of representative G and C-tetrads are shown in figure 7A, B. As observed in figure 7A, guanine tetrads are formed by cyclic N1–H1· · · O6 and N2–H2· · · N7 Hoogsteen type hydrogen bonds while C-tetrads are stabilized by cyclic N4–H4· · · N3 hydrogen bond. Energy minimized CAG-P parallel quadruplex structure is stabilized by cytosine (C), adenine (A) and guanine (G) tetrads while CTG-P structure is stabilized by cytosine (C), thymine (T) and guanine (G) tetrads. Hydrogen bonding scheme in the C and G-tetrads are similar to those observed in the GGC-P structure while A-tetrads in the CAG-P structure are stabilized by cyclic N6–H6· · · N1 hydrogen bonds (shown in figure 7C) and T-tetrads in the CTG-P structure are stabilized by cyclic N3–H3· · · O4 hydrogen bonds (shown in figure 7D). In case of T-tetrads, the methyl carbon atom of each thymine is at a distance of < 3·0 Å from O2 of the neighbouring thymine base in (H to O2 distance is about 2·4 Å) the adjacent strand, giving rise to the formation of C–H· · · O hydrogen bonds (Ghosh and Bansal 1999). The Greek key type antiparallel quadruplex structures GGC-G and CAG-G as well as the stems of hairpin dimer structures GGC-HD and CAG-HD are all stabilized by G-C-G-C tetrads. While the GGC triplet repeat has additional G-tetrads, the CAG repeat has A-tetrads in the same positions. The hydrogen bonding pattern of G-C-G-C tetrad is shown in figure 7E while the hydrogen bonding patterns within the G-tetrad are similar to that observed in figure 7A. In case of G-C-G-C tetrad, two Watson-Crick (WC) G : C base pairs align their major groove edges with the non-WC paired cytosine amino protons donating hydrogen bond to N7 of neighbouring guanine bases. Similarly, G-C-G-C tetrads and C-tetrads stabilize GCC-G and stem part of GCC-HD. The hydrogen bonding patterns of C-tetrads and G-C-G-C tetrads are similar to those shown in figure 7B and 7E respectively. In the case of CAG-G and CAG-HD structures, stable A-tetrads are formed by cyclic N6–H6· · · N7 hydrogen bonds involving the Hoogsteen face of adenine bases (shown in figure 7F). Similarly, CTG-G and stem part of CTG-HD structures are stabilized by stacked G-C-G-C (figure 7E) and T-tetrads (figure 7D). It is worth noting that cytosine, adenine and thymine bases were not hydrogen bonded in the starting C, A and T-tetrads. However, in case of GGC-H hairpin structures, WC G : C base pairs and G : G Hoogsteen pairs stabilize the structure. Similarly, WC G:C base pairs and C : C mismatch pairs stabilize GCC-H structure while WC G : C base pairs and A : A mismatch pairs and WC G : C base pairs and T : T mismatch pairs stabilize CAG-H and CTG-H structures respectively. The hydrogen bonding patterns in all the hairpin structures are similar to those observed between the WC paired strands of corresponding quadruplex structures.

In all the energy minimized hairpin dimer structures, first and third bases in the two loops retain G-C-G-C/C-G-C-G (as shown in figure 7E) tetrads. Interestingly, in case of CAG-HD structure, middle bases (adenines) in the two CAG hairpins form an inter loop A : A pair with N6–H6· · · N7 hydrogen bonds, as shown in figure 8.

Table 2. Structural summary of various models with disease-causing $d(GGC/GCC)_5$ and $d(CAG/CTG)_5$ triplet repeats.

Model	Repeat sequence	Structure	Total number of bases in the structure
GGC-P	$(GGC)_5$	Parallel quadruplex	60
GCC-P	$(GCC)_5$	Parallel quadruplex	60
CAG-P	$(CAG)_5$	Parallel quadruplex	60
CTG-P	$(CTG)_5$	Parallel quadruplex	60
GGC-G	$(GGC)_5$	Antiparallel quadruplex	60
GCC-G	$(GCC)_5$	Antiparallel quadruplex	60
CAG-G	$(CAG)_5$	Antiparallel quadruplex	60
CTG-G	$(CTG)_5$	Antiparallel quadruplex	60
GGC-H	$(GGC)_5$	Hairpin	15
GCC-H	$(GCC)_5$	Hairpin	15
CAG-H	$(CAG)_5$	Hairpin	15
CTG-H	$(CTG)_5$	Hairpin	15
GGC-HD	$(GGC)_5$	Hairpin dimer	30
GCC-HD	$(GCC)_5$	Hairpin dimer	30
CAG-HD	$(CAG)_5$	Hairpin dimer	30
CTG-HD	$(CTG)_5$	Hairpin dimer	30

3.2 *Structural parameters*

The backbone torsion angles for all the energy minimized parallel structures are close to the starting model (within 10°) as well as within the canonical B-DNA range i.e. *a*, *b*, *g*, *d*, *e* and *z* torsion angles are in *gauche*$^-$ (g^-), *trans* (*t*), *gauche*$^+$ (g^+), *trans* (*t*), *trans* (*t*), *gauche*$^-$ (g^-) region respectively. Most of the torsion angles show very small standard deviation values. Since non-WC hydrogen bonded G/C/A/T-tetrads are incorporated between the WC hydrogen bonded G-C-G-C tetrads, the backbone of antiparallel Greek key type quadruplex structures are much more non-uniform and quite a few backbone torsion angles stay outside the canonical B-DNA range. It is observed that all the glycosidic torsion angles within the G-C-G-C tetrads remain in the *anti* range while alternating *syn-anti* conformations are observed within the G/C/A/T-tetrads.

Since the stem part of various hairpin dimer structures are obtained from the corresponding Greek key type

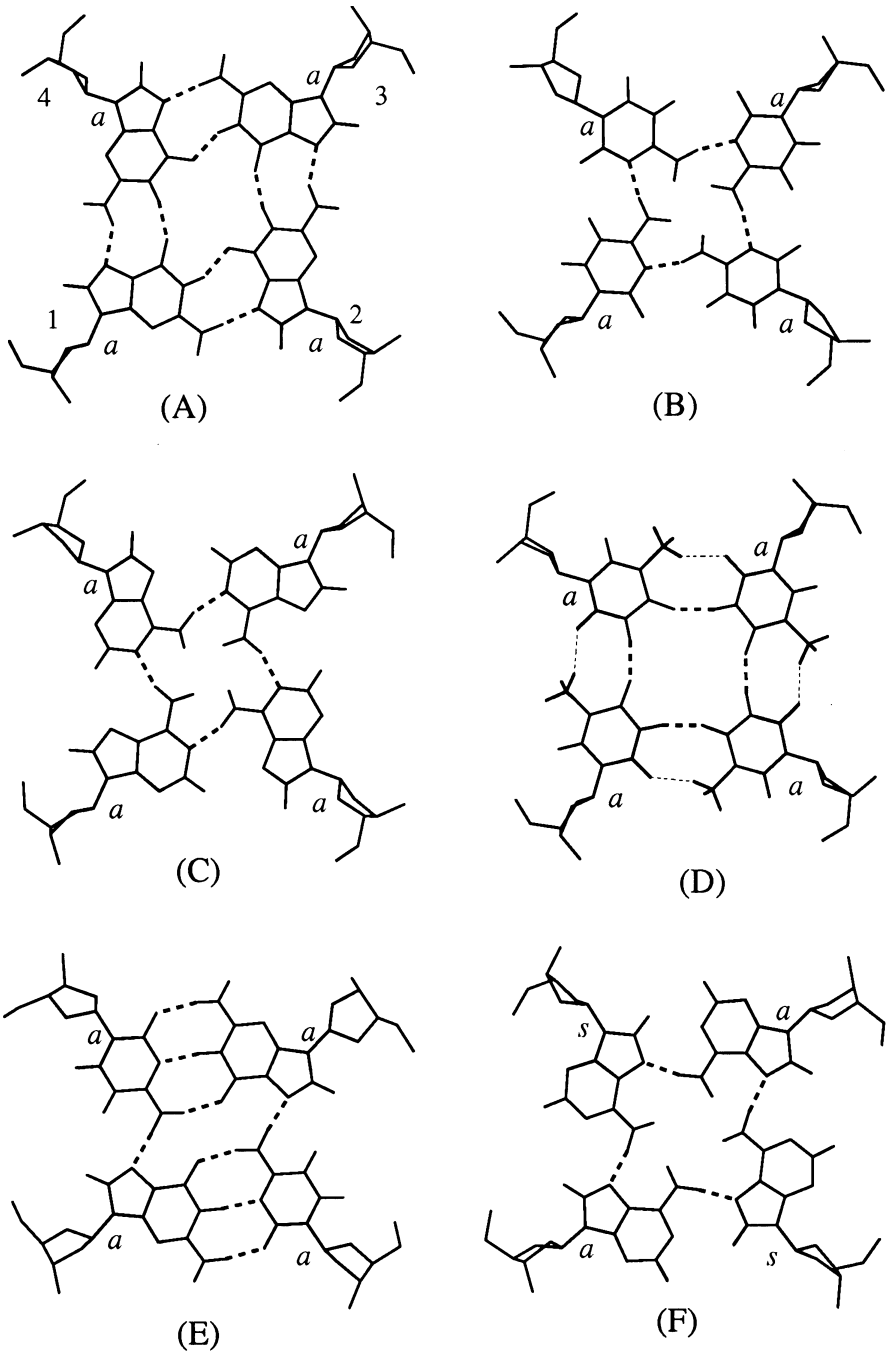

Figure 7. Representative hydrogen bond patterns in the various quadruplex structures minimized using reduced charge on phosphate oxygens. Dotted lines in each case indicate hydrogen bonds with hydrogen to acceptor distances < 2·2 Å. (**A**) Guanine tetrad in the GGC–P structure stabilized by cyclic N1–H1 · · · O6 and N2–H2 · · · N7 hydrogen bonds. The strand numbering scheme used in all subsequent figures is also shown. (**B**) C-tetrad in the GGC-P structure with cyclic N4–H4 · · · N3 hydrogen bonds. (**C**) Adenine tetrad in the CAG-P structure, connected by cyclic N6–H6 · · · N1 hydrogen bonds. (**D**) Thymine tetrad in the CTG-P structure, stabilized by cyclic N3–H3 · · · O4 hydrogen bonds as well as C–H · · · O hydrogen bonds between the O2 atom and methyl hydrogen of neighbouring thymine bases (shown by lighter dotted lines). (**E**) G•C•G•C tetrad in the GGC-G structure wherein a pair of Watson–Crick GC base pairs are also hydrogen bonded to each other through N4–H4 · · · N7 hydrogen bonds. (**F**) Adenine tetrad in the antiparallel CAG-G structure connected by cyclic N6–H6 · · · N7 hydrogen bonds. Adenine bases in this case are in alternating *anti*(*a*)-*syn*(*s*) conformation while all bases in all other tetrads are in *anti*(*a*) conformation. Four symmetrical grooves are seen in all the parallel structures while alternating large and small grooves are observed in Greek key type antiparallel structures (**E** and **F**).

antiparallel structures, the backbone and glycosidic torsion angles of energy minimized hairpin structures are close to the corresponding angles in the antiparallel structures. The backbone and glycosidic torsion angles in the loop region for all the hairpin dimer structures (listed in table 3) can take up a range of values in various models, but the glycosidic torsion angles which are directly correlated to the chain folding can take up values ranging from *anti* to *syn*. In all the hairpin dimer structures, which have loop across the large groove connecting two bases with *anti* conformation and intra loop WC pairing, the first and middle bases in the loops are in *anti* conformation, the third base is in *syn* or low *anti* conformation. Thus, in case of GGC-HD and GCC-HD structures, cytosine (third base) in the loop takes up *syn* conformation, which is expected to be stereochemically less favourable than the loops in CAG-HD and CTG-HD structures with guanines (third base) in the loop taking up *syn* conformation.

Most of the base pair doublet parameters in all the parallel structures show small variation from the starting structures. Figure 9 shows the energy minimized (reduced charge minimization) models of various 15-mer parallel quadruplex structures. As shown in figure 9 various parallel quadruplex structures are quite uniform. Only in the case of CAG-P structure, the four grooves in the energy minimized quadruplex structure are not uniform. During the minimization, alternating grooves have become wider or narrower (clearly observed in the figure 9C), even though the starting quadruplex structure has four symmetrical grooves.

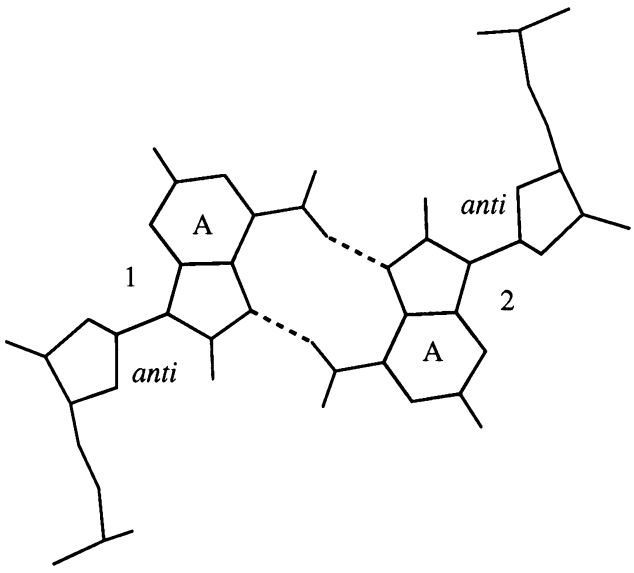

Figure 8. Hydrogen bonding pattern for the adenine–adenine base pair in the loop region of CAG-HD structure minimized using reduced charge on the phosphate oxygens. The middle adenine nucleotide, in the loop region of first hairpin is marked as 1 while same adenine nucleotide of second hairpin is shown as 2. Dotted lines indicate hydrogen bonds with hydrogen to acceptor distances < 2·2 Å

The base pair doublet parameters in the various antiparallel structures show large variation. The twist values for different steps in all the energy minimized antiparallel structures are around $30^{o} \pm 6$. This twist variation could be due to mismatch base pairs, which are placed between the WC base pairs. The 'rise' values in various steps vary between 2·9 to 4·1 Å. However, average rise values are close to 3·4 Å while all steps in the various structures show negative 'slide' values. Overall, as expected, the parallel structures are much more uniform than antiparallel Greek key type structures, even though the former structures also contain non-equivalent all-purine and all-pyrimidine tetrads. The base pair doublet parameters (only for the stem part) for the hairpin dimer structures are close to those observed for the Greek key structures. However, the base tetrads in the hairpin dimer structures remain planar as can be seen in the energy minimized (using reduced phosphate charge) structures of various hairpin dimer structures (figure 10).

3.3 *Stacking interactions*

Stacking between successive tetrads in the various quadruplex structures provides favourable interaction energy. However, the extent of stabilization energy depends on the nature of successive tetrads. The average stacking energy values within the trinucleotide repeat for all the parallel and antiparallel Greek key models minimized with normal as well as reduced phosphate charge, are listed in table 4, considering bases alone and also bases with their deoxyribose sugar units. It has been observed that the stacking energies within the GGC-P and CAG-P structures are marginally more favourable than any other structure while CTG-P and CTG-G structures have less favourable stacking energy values. Both van der Waals and electrostatic components contribute towards these differences. In case of GGC-P structure, both G4 : G4 and C4 : G4 (5′ tetrad is indicated before the 3′-end tetrad) steps have very favourable stacking energy values. Due to better electrostatic interaction between the O6 atom of guanine base and amino group of adenine base, the A4 : G4 steps in the CAG-P structure have more favourable stacking energy. All steps in the CTG-P and CTG-G structures have significantly less favourable stacking energy. If we include the sugar rings in the stacking energy calculation, the stacking energy values in all the steps improve marginally, although the trend in the stacking energy values remain unchanged. All steps in the CTG-P and CTG-G structure have less favourable stack-

ing energy. The stacking energy trend is also more or less same for the stem part of various hairpin dimer structures.

3.4 *Energy component analysis*

Earlier molecular mechanics (Mohanty and Bansal 1993) and experimental (Lu *et al* 1993) studies suggested that parallel orientation of guanine strands in a quadruplex structure is more favourable than the quadruplex structure with antiparallel orientation of guanine strands. However, when the guanine stretches are interspersed with thymine residues, folded back quadruplex structures with antiparallel strands are preferred. Similarly, in order to understand the relative enthalpic energy of GC rich sequences in parallel and antiparallel quadruplex structures, total energy as well as breakup of total energy for all the energy minimized parallel and antiparallel models (both normal as well as reduced charge minimization) are tabulated in table 5. Due to the presence of *syn* bases within the antiparallel structures, the distortion and van der Waals energy components of the antiparallel quadruplex structures are marginally less favourable than the distortion and van der Waals energy of corresponding parallel quadruplex structures. However, favourable electrostatic component of total energy makes the antiparallel structures (GGC-G, CAG-G and CTG-G) comparable with the corresponding parallel structure. Interestingly, in case of normal charge minimization, antiparallel GCC-G structure is noticeably more favourable than the parallel GCC-P structure. In case of reduced charge minimization, all the parallel and antiparallel quadruplex structures are comparable in energy. It has been also observed that both the van der Waals and electrostatic components of intra strand interaction energy favour parallel orientation of strands in the quadruplex structure while the inter strand interaction energy, particularly electrostatic component of inter strand interaction favours antiparallel orientation of strands (data not shown).

The total energy as well as its breakup into various components, for the heteroduplex, the two individual hairpins, one from each strand, as well as hairpin dimer and Greek key type quadruplex structures, are listed in the table 6. The energy values for the hairpin dimers and Greek key type quadruplex structures have been

Table 3. The backbone and glycosidic torsion angles (in degrees) are listed for the three nucleotides in the loop region (moving from 5′-end to 3′-end) of various hairpin dimer structures, minimized with reduced phosphate charge on the phosphate oxygens.

		a	*b*	*g*	*d*	*e*	*z*	*c*
GGC-HD	G	266	64	163	135	176	270	220
		256	63	172	109	182	270	204
	G	173	169	190	127	197	222	164
		165	180	180	127	171	260	175
	C	103	218	201	126	182	270	300
		177	179	179	119	175	270	297
GCC-HD	G	244	64	159	135	180	266	222
		262	64	172	89	183	276	185
	C	166	170	181	143	196	213	194
		168	177	182	149	166	263	187
	C	99	214	211	138	178	271	298
		161	193	181	132	170	262	296
CAG-HD	C	165	188	163	155	176	270	242
		289	165	76	150	170	263	263
	A	169	171	184	134	185	236	167
		181	170	182	128	176	255	167
	G	123	192	186	121	71	306	317
		143	183	180	102	86	306	312
CTG-HD	C	158	186	190	150	170	270	236
		290	165	76	152	165	263	260
	T	176	172	184	136	196	216	189
		182	170	178	147	190	227	190
	G	105	192	193	129	66	311	328
		124	191	186	119	72	299	324

The values in the first row correspond to the nucleotides in the first loop while values in the second row correspond to the nucleotides in the second loop.

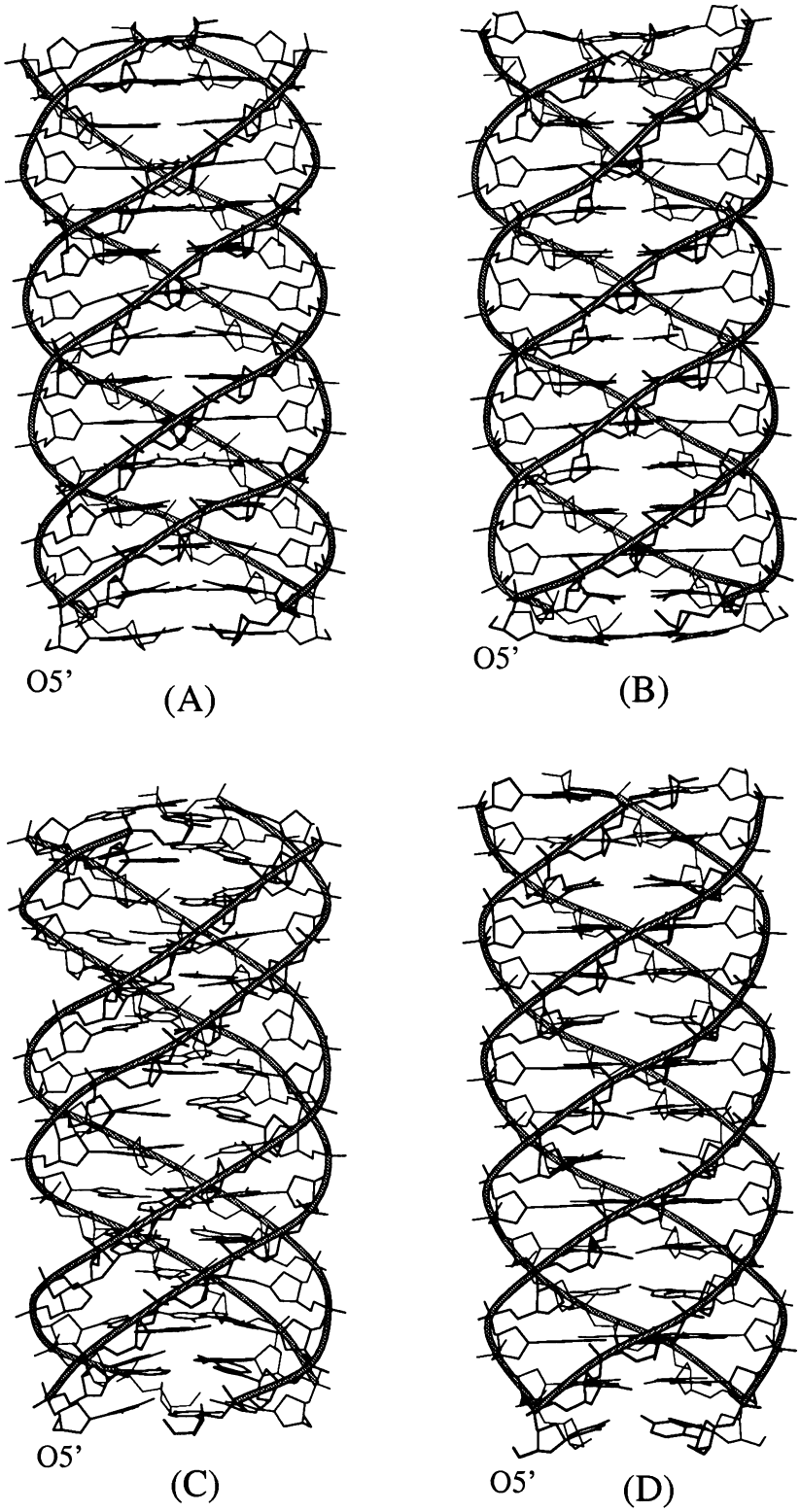

Figure 9. Energy minimized models (using reduced charge on phosphate oxygen) for various 15-mer parallel quadruplex structures (**A**) GGC-P, (**B**) GCC-P, (**C**) CAG-P and (**D**) CTG-P. The 5′-end of each structure is indicated by O5′ and a ribbon has been drawn through the backbone of each structure.

appropriately scaled (as indicated in columns IX and X of table 6) so that energies are compared for the same 30 nucleotide fragments, consisting of two 15-mers (or 5 repeats of each trinucleotide), with the base sequences corresponding to the original Watson-Crick duplex.

Analysis of the results, when the structures are minimized using normal phosphate oxygen charges, reveals that the sum of both individual hairpins as well as the average of two self associated hairpin dimer structures, with GGC and GCC repeats, are energetically comparable with the standard B-DNA type heteroduplex structure with self complementary sequence (as shown in the columns VIII, IX and I respectively, of the top half of table 6A). All the loops in the various hairpin structures provide favourable interaction energy with the stem part. However, due to less favourable electrostatic energy, the scaled energy for the longer antiparallel Greek key type quadruplex structures for $(GGC)_5$ and $(GCC)_5$ sequences

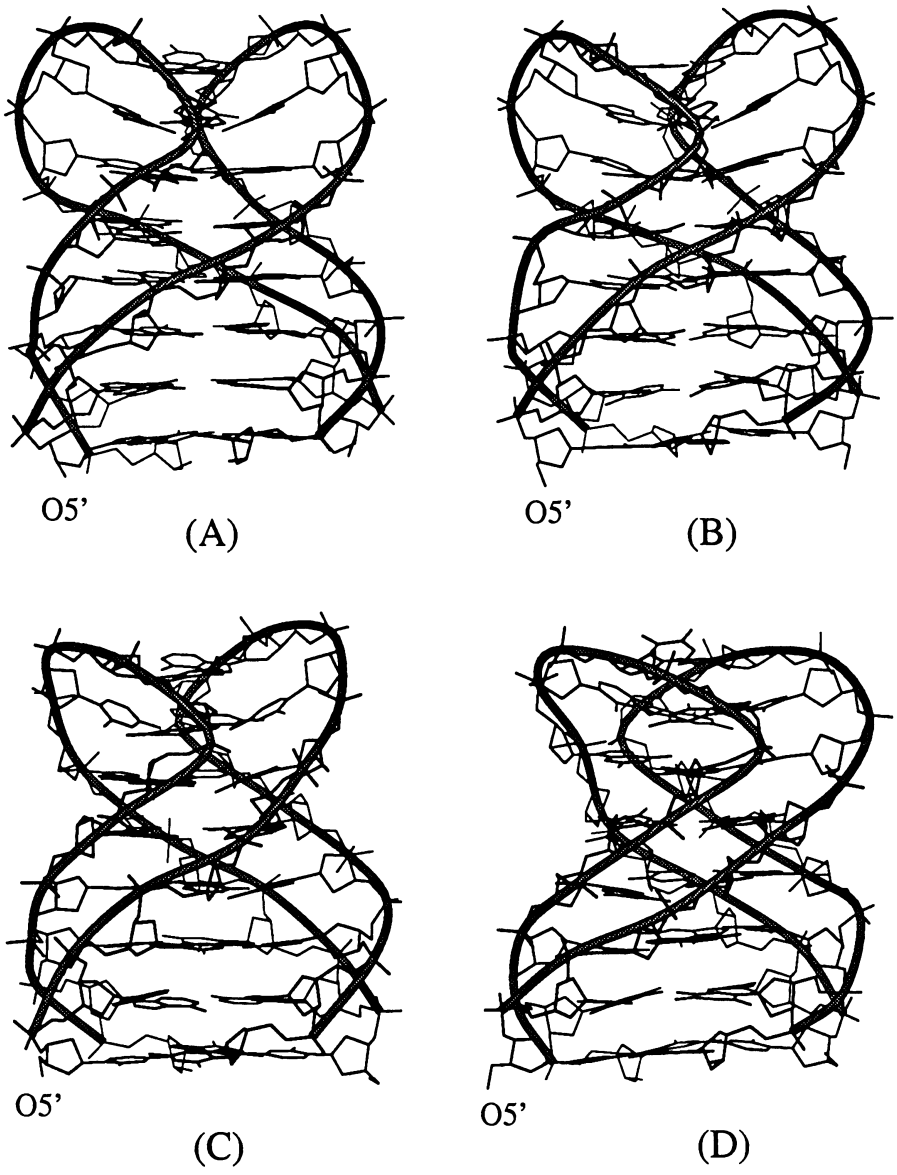

Figure 10. Energy minimized (using reduced charge on phosphate oxygen) hairpin dimer structures with Greek key geometry, for 15-mer sequences (**A**) GGC-HD, (**B**) GCC-HD, (**C**) CAG-HD and (**D**) CTG-HD. In each case, the 5′ and the 3′-ends of the first hairpin are indicated by O5′ and O3′. Backbone of each structure is highlighted by a ribbon representation.

is less favourable (column X) than the B-DNA type duplex structure, as well as the hairpin dimer structures (values given in columns I and VIII respectively). If we compare the structures which are minimized using reduced phosphate oxygen charge, wherein the interstrand phosphate group repulsion is much less, the Greek key type antiparallel quadruplex as well as hairpin dimer structures are energetically comparable with B-DNA type duplex structures, while the duplex hairpin structures are marginally less favourable (as shown in the bottom half of table 6A). Very similar trends are observed for the structures with CAG and CTG triplet repeats (table 6B). Hence, the single hairpins as well as quadruplexes formed by self association of hairpin dimer structures are as favourable as B-DNA type structures under normal phosphate oxygen charge while the antiparallel Greek key type structures are less favourable. However, when the structures are minimized using reduced phosphate oxygen charge, hairpin dimer and Greek key type quadruplex structures are comparable with corresponding B-DNA type structures while single hairpin structures are marginally less favourable.

4. Conclusions

Several human genetic diseases are associated with expansion of triplet repeats. A mechanism of triplet repeat expansion was involved in blockage of DNA replication by unknown higher order structures. Several

Table 4. Average stacking energy values (in kcal/mol) within a trinucleotide repeat for various parallel as well as Greek key type antiparallel quadruplex structures, considering base alone and base with sugar rings (without O3′ and O5′ atoms). Total stacking energy is divided into van der Waals (including 10–12 energy components) and electrostatic components. (A) Structures minimized with normal phosphate charge. (B) Minimized with reduced phosphate charge.

	(A)						(B)					
	van der Waals		Electrostatic		Total		van der Waals		Electrostatic		Total	
	Base	Base with sugar	Base	Base with sugar	Base	Base with sugar	Base	Base with sugar	Base	Base with sugar	Base	Base with sugar
GGC-P	– 42·2	– 52·8	– 2·3	5·1	– 44·5	– 47·7	– 42·8	– 54·2	– 3·6	3·7	– 46·4	– 50·5
GGC-G	– 41·2	– 52·2	2·3	8·2	– 38·9	– 44·0	– 43·4	– 55·3	2·1	8·5	– 41·3	– 46·8
GCC-P	– 35·2	– 45·8	– 4·4	3·1	– 39·6	– 42·7	– 37·4	– 48·6	– 4·8	3·1	– 42·2	– 45·5
GCC-G	– 36·6	– 47·8	– 2·7	2·3	– 39·3	– 45·5	– 37·9	– 48·9	– 1·7	3·9	– 39·6	– 45·0
CAG-P	– 40·0	– 50·0	– 5·3	1·3	– 45·3	– 48·7	– 40·5	– 51·3	– 6·3	0·5	– 46·8	– 50·8
CAG-G	– 37·2	– 48·7	– 0·8	4·9	– 38·0	– 43·8	– 40·3	– 51·5	– 2·6	3·3	– 42·9	– 48·2
CTG-P	– 34·2	– 45·2	– 1·3	5·9	– 35·5	– 39·3	– 35·6	– 46·9	– 1·4	6·4	– 37·0	– 40·5
CTG-G	– 32·1	– 42·4	1·0	4·6	– 31·0	– 37·8	– 33·8	– 44·1	0·9	7·4	– 32·9	– 36·7

Table 5. Various energy components (in kcal/mol) for the all-*anti* parallel and Greek key type antiparallel quadruplex structures for 15-mers with GGC, GCC, CAG and CTG repeats.

	Normal charge on phosphate oxygens					Reduced charge on phosphate oxygens				
	Distortion bond angle dihedral	van der Waal (6–12)	Electro-static	H-bond (10–12)	Total	Distortion bond angle dihedral	van der Waal (6–12)	Electro-static	H-bond	Total
GGC-P	876	– 881	– 1794	– 35	– 1834	851	– 935	– 2922	– 33	– 3039
GGC-G	925	– 871	– 1942	– 38	– 1926	898	– 974	– 2961	– 34	– 3071
GCC-P	886	– 727	– 1714	– 23	– 1579	867	– 792	– 2870	– 22	– 2816
GCC-G	936	– 693	– 1948	– 35	– 1741	936	– 834	– 2932	– 33	– 2863
CAG-P	889	– 815	– 1614	– 26	– 1567	877	– 886	– 2756	– 28	– 2793
CAG-G	904	– 761	– 1746	– 38	– 1641	892	– 840	– 2802	– 32	– 2782
CTG-P	858	– 714	– 1439	– 27	– 1321	825	– 751	– 2543	– 20	– 2489
CTG-G	903	– 660	– 1541	– 32	– 1330	840	– 742	– 2528	– 27	– 2457

The energy values for the structures minimized with normal as well as reduced charge on phosphate oxygens are shown.

gel electrophoresis studies indicate that triplet repeats containing DNA oligonucleotides readily and specifically adopt unusual compact structures that migrate abnormally fast on native, but not on denaturing gels (Chastain *et al* 1995; Mitchell *et al* 1995). Present model building and molecular mechanics studies also indicate that hairpin as well as hairpin dimer structures formed by individual strands, with triplet repeat sequences, are nearly as favourable as the corresponding B-DNA type heteroduplex structures. Furthermore, Greek key type antiparallel structures could also be formed at high salt condition. Thus, our molecular mechanics studies support

Table 6A. Various energy components (in kcal/mol) for the energy minimized (with normal as well as reduced charge minimization) structures with GGC and GCC triplet repeat sequence.

Structure	Hetero duplex I	Hairpin II	Hairpin III	Hairpin dimer IV	Hairpin dimer V	Quadruplex (Greek key) VI	Quadruplex (Greek key) VII	II + III VIII	IV + V / 2 IX	VI + VII / 4 X
Sequence	$d(GGC)_5$• $(GCC)_5$	$d(GGC)_5$	$d(GCC)_5$	$[d(GGC)_5]2$	$[d(GCC)_5]2$	$[d(GGC)_5]4$	$[d(GCC)_5]4$			
Number of residues	15 × 2	15 × 1	15 × 1	15 × 2	15 × 2	15 × 4	15 × 4	30	30	30
Normal charge minimization										
Distortion	410	214	218	456	476	925	936	432	466	465
vdW (6–12)	– 353	– 170	– 135	– 421	– 353	– 871	– 693	– 305	– 387	– 381
Electrostatic	– 1184	– 623	– 619	– 1134	– 1117	– 1942	– 1948	– 1232	– 1126	– 973
H-bond	– 7	– 4	– 4	– 19	– 18	– 38	– 35	– 8	– 19	– 18
Total	– 1134	– 583	– 540	– 1118	– 1012	– 1926	– 1740	– 1123	– 1065	– 817
Reduced charge minimization										
Distortion	403	214	217	449	470	898	936	431	460	459
vdW (6–12)	– 361	– 170	– 140	– 453	– 391	– 974	– 834	– 310	– 422	– 452
Electrostatic	– 1439	– 699	– 699	– 1469	– 1469	– 2961	– 2832	– 1398	– 1469	– 1448
H-bond	– 6	– 4	– 4	– 17	– 16	– 34	– 33	– 8	– 17	– 17
Total	– 1403	– 659	– 626	– 1490	– 1406	– 3071	– 2863	– 1285	– 1448	– 1484

Total energy is divided into bond, angle and dihedral distortion energy, 6–12 van der Waals (vdW) component, electrostatic and 10–12 hydrogen bond energy component. The energies for the two hairpin dimers (columns IV and V) and the Greek key quadruplex structures (columns VI and VII), corresponding to the two complementary sequences, have been added and then scaled down (as indicated in columns IX and X). This allows comparison of energy values for the same number (viz. 30) and type of residues, as in the heteroduplex (column I) and in the two hairpins taken together (column VIII).

Table 6B. Various energy components (in kcal/mol) for the energy minimized (both normal as well as reduced charge minimization) structures with CAG and CTG triplet repeat sequences.

Structure	Hetero duplex I	Hairpin II	Hairpin III	Hairpin dimer IV	Hairpin dimer V	Quadruplex (Greek key) VI	Quadruplex (Greek key) VII	II + III VIII	IV + V / 2 IX	VI + VII / 4 X
Sequence	$d(CAG)_5$• $(CTG)_5$	$d(CAG)_5$	$d(CTG)_5$	$[d(CAG)_5]2$	$[d(CTG)_5]2$	$[d(CAG)_5]4$	$[d(CTG)_5]4$			
Number of residues	15 × 2	15 × 1	15 × 1	15 × 2	15 × 2	15 × 4	15 × 4	30	30	30
Normal charge minimization										
Distortion	410	223	217	457	454	904	903	440	456	452
vdW (6–12)	– 339	– 147	– 137	– 380	– 350	– 761	– 660	– 284	– 365	– 355
Electrostatic	– 1096	– 605	– 549	– 1064	– 983	– 1746	– 1541	– 1154	– 1024	– 822
H-bond	– 4	– 5	– 3	– 17	– 16	– 38	– 32	– 8	– 17	– 18
Total	– 1029	– 534	– 472	– 1004	– 895	– 1641	– 1330	– 1006	– 950	– 743
Reduced charge minimization										
Distortion	401	219	216	444	445	892	840	435	445	433
vdW (6–12)	– 349	– 150	– 142	– 418	– 395	– 840	– 742	– 292	– 407	– 396
Electrostatic	– 1345	– 670	– 622	– 1407	– 1295	– 2802	– 2528	– 1292	– 1351	– 1333
H-bond	– 5	– 4	– 3	– 15	– 15	– 32	– 27	– 7	– 15	– 8
Total	– 1298	– 605	– 551	– 1396	– 1269	– 2782	– 2457	– 1156	– 1333	– 1310

All other details as in table 6A.

the suggestions from various spectroscopic and NMR studies that hairpin and quadruplex structures are the most common structures formed by these sequences. It is also observed that non-guanine tetrads (A, C and T-tetrads) can be nicely accommodated in the quadruplex structures and provide favourable stacking energy. Interestingly, in the minimized hairpin dimer and quadruplex structures, adenine (A), cytosine (C) and thymine (T) bases form hydrogen bonded A, C and T-tetrads, even though they were not hydrogen bonded in the starting structures. Recently, several NMR studies report the formation of A, C and T-tetrads (Patel *et al* 1999, 2000; Patel and Hosur 1999). The hydrogen bonding scheme within these tetrads are similar to that suggested by the present model building and molecular mechanics studies.

Triplet repeat sequences can take up energetically favourable single hairpin, as well as hairpin dimer structures. Under low salt condition, corresponding to the physiological environment (i.e. normal charge minimization) hairpin structures with these triplet repeats are energetically comparable with corresponding B-DNA duplex whereas hairpin dimer structures are marginally less favourable and the longer quadruplex structures are clearly less favourable, due to unfavourable electrostatic interaction between the four chains of the quadruplex. However, at high salt condition (reduced charge minimization) both hairpin dimer and quadruplex structures are energetically more favourable than the corresponding double helical hairpin structures. Thus, these sequences are highly polymorphic and depending on the environmental condition, the Watson-Crick, self-complementary duplex could dissociate and each strand can take up hairpin or quadruplex structures, leading to deleterious effects on the normal functioning of DNA.

Acknowledgements

This work is partially supported by a grant from the Council of Scientific and Industrial Research, New Delhi. The authors are grateful to SERC, Indian Institute of Science, for computational facilities.

References

Arnott S, Chandrasekaran R and Marttila C M 1974 Structure for polyinosinic acid and polyguanylic acid; *Biochem. J.* **141** 537–543

Bacolla A, Gellibolian R, Shimizu M, Amirhaeri S, Kang S, Ohshima K, Larson J E, Harvey C, David Stollar B and Wells R D 1997 Flexible DNA: genetically unstable CTG·CAG and CGG·CCG from human hereditary neuromuscular disease genes; *J. Biol. Chem.* **272** 16783–16792

Bansal M and Sasisekharan V 1986 Molecular Model-building of DNA: Constraints and Restraints; in *Theoretical chemistry of biological systems* (ed.) G N Szabo (Amsterdam: Elsevier) pp. 127

Brook J D, McCurrach M E, Harley H G, Buckler A J, Church D, Aburatani H, Hunter K, Stanton V P, Thirion J P and Hudson T 1992 Molecular Basis of Myotonic Dystrophy: Expansion of a Trinucleotide (CTG) Repeat at the 3′ End of a Transcript Encoding a Protein Kinase Family Member; *Cell* **68** 799–802

Caskey C T, Pizzuti Y H, Fu R, Fenwick R G and Nelson D L 1992 Triplet repeat mutations in human disease; *Science* **256** 784–789

Chastain II P D, Eichler E E, Kang S, Nelson D L, Levene S D and Sinden R R 1995 Anomalous rapid electrophoretic mobility of DNA containing triplet repeats associated with human disease genes; *Biochemistry* **34** 16125–16131

Chastain P D and Sinden R R 1998 CTG repeats associated with human genetic disease are inherently flexible; *J. Mol. Biol.* **275** 405–411

Chen X, Mariappan S V S, Catasti P, Ratliff R, Moyzis R K, Laayoun A, Smith S S, Bradbury E M and Gupta G 1995 Hairpins are formed by the single DNA strands of the fragile X triplet repeats: structure and biological implications; *Proc. Natl. Acad. Sci. USA* **92** 5199–5203

Chen X, Mariappan S V S, Moyzis R K, Bradbury E M and Gupta G 1998 Hairpin induced slippage and hypermethylation of the fragile X DNA triplets; *J. Biomol. Struct. Dyn.* **15** 745–756

Darlow J M and Leach D R F 1998a Secondary structures in d(CGG) and d(CCG) repeat tracts; *J. Mol. Biol.* **275** 3–16

Darlow J M and Leach D R F 1998b Evidence for two preferred hairpin folding patterns in d(CGG)·d(CCG) repeat tracts in vivo; *J. Mol. Biol.* **275** 17–23

Duyao M, Ambrose A, Myers R, Novelletlo A, Persichetti F, Frontali M, Folstein S, Ross C, Franz M, Abbott M, Gray J, Conneally P, Young and Penney J 1993 Trinucleotide repeat length instability and age of onset in Huntington's disease; *Nature Gent.* **4** 387–392

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

Fu Y H, Pizzuti A, Fenwick R G, King Jr J, Rajnarayan S, Dunne P W, Dubel J, Nasser G A, Ashizawa T, Jong P D, Wieringa B, Korneluk R G, Peryman M B, Epstein H F and Caskey C M 1992 An unstable triplet repeat in a gene related to myotonic muscular dystrophy; *Science* **255** 1256–1258

Gacy A M, Goellner G, Juranic N, Macura S and McMurray C T 1995 Trinucleotide repeats that expand in human disease form hairpin structures *in vitro*; *Cell* **81** 533–540

Ghosh A and Bansal M 1999 C–H**...**O hydrogen bonds in minor groove of A-tracts in DNA double helices; *J. Mol. Biol.* **294** 1149–1158

Han J, Hsu C, Zhu Z, Longshore J W and Finley W H 1994 Over representation of the disease associated (CAG) and (CGG) repeats in the human genome; *Nucleic Acids Res.* **22** 1735–1740

Huntington's Disease Collaborative Research Group 1993 A novel gene containing a trinucleotide repeat that is expanded and unstable on Huntington's disease chromosomes; *Cell* **72** 971–983

Kang S, Ohshima K, Shimizu M, Amirhaeri S and Wells R D 1995 Pausing of DNA synthesis *in vitro* at specific loci in CTG and CGG triplet repeats from human hereditary disease genes; *J. Biol. Chem.* **270** 27014–27021

Kohwi Y, Wang H and Kohwi-Shigematsu T 1993 A single trinucleotide, 5′AGC3′/5′GCT3′, of the triplet-repeat disease genes confers metal ion-induced non-B DNA structure; *Nucleic Acids Res*. **21** 5651–5655

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

Lu M, Guo Q and Kallenbach N R 1993 Thermodynamics of G-tetraplex formation by telomeric DNAs; *Biochemistry* **32** 598–601

Mahadevan M, Tsilfidis C, Sabourin L, Shutler G, Amemiya C, Jansen G, Neville C, Narang M, Barcelo J, O'Hoy K, Leblond S, Earle-MacDonald J, Jong P J D, Wieringa B and Korneluk R G 1992 Myotonic dystrophy mutation: an unstable CTG repeat in the 3′ untranslated region of the gene; *Science* **255** 1253–1255

Mandel J L 1993 Questions of expansion; *Nature Genet*. **4** 8–9

Mariappan S V S, Garcia A E and Gupta G 1996a Structure and dynamics of the DNA hairpins formed by tandemly repeated CTG triplets associated with myotonic dystrophy; *Nucleic Acids Res*. **24** 775–783

Mariappan S V S, Catasti P, Chen X, Ratliff R, Moyzis R K, Bradbury E M and Gupta G 1996b Solution structures of the individual single strands of the fragile X DNA triplets $(GCC)_n$–$(GGC)_n$; *Nucleic Acids Res*. **24** 784–792

Mariappan S V S, Silks III L A, Chen X, Springer P A, Wu R, Moyzis R K, Bradbury E M Garcia A E and Gupta G 1998 Solution structures of the Huntington's disease DNA triplets, $(CAG)_n$; *J. Biomol. Struct. Dyn*. **15** 723–744

Mitas M 1997 Trinucleotide repeats associated with human disease; *Nucleic Acids Res*. **25** 2245–2254

Mitas M, Yu A, Dill J and Haworth I S 1995a The trinucleotide repeat sequence $d(CGG)_{15}$ forms a heat-stable hairpin containing G^{syn}-G^{anti} base pairs; *Biochemistry* **34** 12803–12811

Mitas M, Yu A, Dill J, Kamp T J, Chambers E J and Haworth I S 1995b Hairpin properties of single-stranded DNA containing a GC-rich triplet repeat: $(CTG)_{15}$; *Nucleic Acids Res*. **23** 1050–1059

Mitchell J E, Newbury S F and McClellan J A 1995 Compact structures of $(CNG)_n$ oligonucleotides in solution and their possible relevance to fragile X and related human genetic diseases; *Nucleic Acids Res*. **23** 1876–1881

Mohanty D and Bansal M 1993 Conformational polymorphism in G-tetraplex structures: strand reversal by base flipover or sugar flipover; *Nucleic Acids Res*. **21** 1767–1774

Mohanty D and Bansal M 1994 Conformational polymorphism in telomeric structures: loop orientation and interloop pairing in $d(G_nT_nG_n)$; *Biopolymers* **34** 1187–1211

Mohanty D and Bansal M 1995 Chain folding and A : T pairing in human telomeric DNA: A model-building and molecular dynamics study; *Biophys. J*. **69** 1046–1067

Patel P K and Hosur R V 1999 NMR observation of T-tetrads in a parallel stranded DNA quadruplex formed by *Saccharomyces cerevisiae* telomere repeats; *Nucleic Acids Res*. **27** 2457–2464

Patel P K, Koti A S R and Hosur R V 1999 NMR studies on truncated sequences of human telomeric DNA: observation of a novel A-tetrad; *Nucleic Acids Res*. **27** 3836–3843

Patel P K, Bhavesh N S and Hosur R V 2000 NMR observation of a novel C-tetrad in the structure of the SV40 repeat sequence GGGCGG; *Biochem. Biophys. Res. Commun*. **270** 967–971

Pearlman D A, Case D A, Caldwell J C, Seibel G L, Singh U C, Weiner P K and Kollman P A 1991 AMBER 4·0, UCSF

Pearson C E and Sinden R R 1998 *Genetic instabilities and hereditary neurological disorders* (San Diego: Academic Press) pp 585

Petruska J, Arnheim N and Goodman M F 1996 Stability of intrastrand hairpin structures formed by the CAG/CTG class of DNA triplet repeats associated with neurological diseases; *Nucleic Acids Res*. **24** 1992–1998

Radhakrishnan I and Patel D J 1993 Solution structure of a purine·purine·pyrimidine DNA triplex containing G·GC and T·AT triples; *Structure* **1** 135–152

Ravikiran M and Bansal M 1997 Sequence dependent recombination triple helices: A molecular dynamics study; *J. Biomol. Struct. Dyn*. **15** 333–345

Richards R I and Sutherland G R 1997 Dynamic mutation: possible mechanisms and significance in human disease; *Trends Biochem. Sci*. **22** 432–436

Samadashwily G M, Raca G and Mirkin S M 1997 Trinucleotide repeats affect DNA replication *in vivo*; *Nature Genet*. **17** 298–304

Sinden R R 1999 Biological implications of the DNA structures associated with disease-causing triplet repeats; *Am. J. Hum. Genet*. **64** 346–353

Smith G K, Jie J and Fox G E 1995 DNA CTG triplet repeats involved in dynamic mutations of neurologically related gene sequences form stable duplexes; *Nucleic Acids Res*. **23** 4303–4311

Usdin K 1998 NGG-triplet repeats form similar intrastrand structures: implications for the triplet expansion diseases; *Nucleic Acids Res*. **26** 4078–4085

Usdin K and Woodford K J 1995 CGG repeats associated with DNA instability and chromosome fragility form structures that block DNA synthesis in vitro; *Nucleic Acids Res*. **23** 4202–4209

Wells R D 1996 Molecular basis of genetic instability of triplet repeats; *J. Biol. Chem*. **271** 2875–2878

Wells R D and Sinden R R 1993 *Genome Analysis* Vol. 7 (New York: Cold Spring Harbor) pp 107

Yu A, Dill J and Mitas M 1995a The purine-rich trinucleotide repeat sequences $d(CAG)_{15}$ and $d(GAC)_{15}$ form hairpins; *Nucleic Acids Res*. **23** 4055–4057

Yu A, Dill J, Wirth S S, Huang G, Lee V H, Haworth I S and Mitas M 1995b The trinucleotide repeat sequence $d(GTC)_{15}$ adopts a hairpin conformation; *Nucleic Acids Res*. **23** 2706–2714

Yu S, Pritchard M, Kremer E, Lynch M, Nancarrow J, Backer E, Homan K, Mulley J C, Warren S T, Schlessinger D, Sutherlan G R and Richards R I 1991 Fragile X genotype characterized by an unstable region of DNA; *Science* **252** 1179–1181

Zheng M, Huang X, Smith G K, Yang X and Gao X 1996 Genetically unstable CXG repeats are structurally dynamic and have a high propensity for folding: An NMR and UV spectroscopic study; *J. Mol. Biol*. **264** 323–336

MS received 26 June 2001; accepted 2 November 2001

Corresponding editor: XIANGYIN KONG